Université de Montréal

# Intentionality and Concept Attribution: The Search for Mental States in the Animal Kingdom

Joanne Downs

Département de philosophie
Faculté des arts et des sciences

Thèse presentée à la Faculté des études supérieures
en vue de l'obtention du grade de Doctorat
en philosophie

Février 2004

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée
Intentionality and Concept Attribution:
The Search for Mental States in the Animal Kingdom

présentée par
Joanne Downs

A été évaluée par un jury composé des personnes suivants:

Jean-Pierre Marquis
président-rapporteur

Daniel Laurier
directeur de recherche

Michel Seymour
membre du jury

M. Luc Faucher (UQAM)
examinateur externe

Daniel Perusse (Anthropologie)
représentant du doyen de la FES

# Acknowledgments

A great many of my friends and family supported my writing of this dissertation. I would first like to thank Paul Bernier, my supervisor at the Master's level, for encouraging me to pursue a PhD in the first place, as well as two ex-professors of mine, Bill Massicotte and Brian MacPherson, for always believing that I had something worthwhile to contribute to the field of philosophy.

My supervisor, Daniel Laurier, played the biggest role in my writing of this dissertation. I thank him for his support, his encouragement, his editing prowess, but most of all I admire and thank him for putting up with me.

Many thanks go to my partner in life and in crime, Willie 'Devilman' Morelli, in particular for not believing in philosophy and for playing Country music (which I can't stand) extremely loudly and repetitiously while I was trying to write.

I would also like to thank my parents, who bore the brunt of my writing angst in the form of endless, lengthy conversations about everything from politics to topics they had not a clue about. I appreciate in particular their intellectual stamina in keeping up with me when I was at my worst and most intense. I also thank my sister Sara for living in a parallel universe and therefore knowing exactly what to say, and my brother Jolyon for his knack at being the devil's advocate.

I much appreciated and continue to enjoy lengthy conversations with members of our unofficial metaphysical club, The Überwenches, composed of Marianne Jeadah, Sharon Burgher, Chantal Giroux, Heidi O'Brien, Shelley Rohar and myself. Each of these women provided enormous support in their own ways: Marianne and Sharon for help in overcoming philosophical obstacles, Heidi for her expertise as my second editor and thesis layout designer, Chantal for forcing me to think about things other than my dissertation, and Shelley for having that uncanny ability to peel me off the ceiling, for bringing me back to earth from many an existentialist crisis.

I thank the various pets I have had and in particular my dogs, Rushton and Lobo and my beloved Buddy (rest in peace) for providing the inspiration to write about animals and how smart they are.

I would also like to thank a few musicians for putting out fantastic tunes that served as a vital and welcome retreat from writing. They are: Kid Rock, Everlast, Limp Bizkit, The Chemical brothers, The Crystal Method, and Rob Zombie.

I dedicate this dissertation to a personal source of continued inspiration and hope, my nieces and nephews: Oscar, Megan, Shannon, Cory and Stacy.

## Résumé

La discipline de l'éthologie cognitive étudie la nature et l'évolution des capacités cognitives chez les animaux. Selon deux éthologues éminents, C. Allen et M. Bekoff, le statut scientifique de cette discipline fait l'objet de vives contestations et se heurte à plusieurs objections. Cette thèse a deux objectifs. Dans la première partie, j'examine les quatre objections les plus importantes qui ont été avancées contre le projet de l'éthologie cognitive. Il s'agit d'objections qui visent à montrer, en particulier, qu'il n'est pas légitime d'attribuer des états mentaux intentionnels aux animaux. Les objections sont: i) que les animaux n'ont pas de langage, et donc pas de pensées, ii) que l'attribution d'état mentaux aux animaux n'est qu'une forme d'anthropomorphisme, iii) que, même si les animaux possédaient des état mentaux, on ne pourrait pas les connaître, car les esprits des animaux sont inacccessibles, et iv) que sur le plan expérimental, il est pratiquement impossible d'élaborer des expériences qui démontreraient que les animaux ont des états mentaux. J'essaierai de montrer que ces objections n'ont qu'une force 'prima facie' et qu'elles ne réussissent pas à faire avorter le projet. Après cette discussion, il restera un problème méthodologique qui ne sera résolu que dans la deuxième partie de la thèse, à savoir le fait que les théories de type behavioriste expliquent aussi bien les comportements observés chez les animaux que les théories de type mentaliste. Par conséquent, il n'y aurait aucune raisons valable de postuler des états mentaux.

Dans la deuxieme partie, j'examine différentes théories de l'intentionnalité et de l'attribution de concepts qui ont été avancées par des éthologues cognitifs, dans le but de dégager les hypothèses les plus fructueuses en ce qui concerne le potentiel intentionnel et conceptuel des animaux. Les théories de l'intentionnalité que je présente visent toutes à identifier ou à attribuer des états mentaux intentionnels aux animaux. J'éxamine des théories de quatre types: une théorie behavioriste, une théorie normative, la théorie de la stratégie intentionnelle de D. Dennett, et la théorie téléologique de J. Bennett. Il s'avère que la théorie de Bennett permet de résoudre le problème méthodologique laissé en suspens à la fin de la première partie. Dans le dernier chapitre j'examine le projet de N. Chater et C. Heyes de trouver une théorie des concepts qui s'appliquent aux animaux. Ils prétendent que ce projet est voué à l'échec, puisqu'ils sont incapables de concevoir que les concepts puissent être indépendants du langage. Je ne suis pas d'accord avec eux, et je montre que ce sont plutôt leurs critères qui sont suspects. Je finis par esquisser une théorie de l'attribution de concepts qui est indépendante du langage et applicable aux animaux.



Mots clé: éthologie cognitive; concepts; anthropomorphisme; contenu propositionel; expérience subjective.

**Abstract**

The discipline of cognitive ethology is concerned primarily with an investigation into the nature and evolution of cognitive capacities in non-human animals (hereafter animals). According to two eminent cognitive ethologists, the discipline of cognitive ethology faces challenges to its scientific status. My aim in this thesis is two-fold. In Part One, I will examine the four most important of the objections made to the discipline of cognitive ethology, in particular as they relate to the search for mental states in animals, and show these objections as providing no real obstacles to the search for mental states in animals. The objections are that animals have no language therefore they cannot have mental states and other types of thought; that all mental state attribution to animals is anthropomorphism; that even if animals have mental states, we will not be able to gain access to them because animals are 'other minds', and that the search for mental states in animals is rendered nearly impossible from an experimental point of view, since all explanations of the animals' behavior are also accounted for by behaviorist explanations. This methodological problem will not be completely solved in the first half of the thesis.

Once I have demonstrated that there is no clear prima facie reason not to examine the potential for mental states in animals, I examine in Part Two various theories of intentionality and concept attribution that have been advanced by cognitive ethologists, with the aim of pointing to the most fruitful advances made by the discipline with regard to exploring the potential for intentional mental states and concepts in animals. The types of intentional theories that I examine are all concerned with the identification or attribution of intentional (purposeful) mental states in animals. I examine four different types, one that is behaviorist in nature, one that is normative in nature, Daniel Dennett's Intentional Stance theory, and Jonathan Bennett's teleological theory. As it turns out, Bennett's theory ends up solving the methodological problem left over from chapter four. In the last chapter, I examine Nick Chater and Celia Heyes' attempt to search for a theory of concepts that applies to animals. They claim that their search is unsuccessful because they cannot find a sense of the term 'concept' that is independent of language. I do not agree with their view and instead argue that it is their set of criteria that is at fault. I end up finding a theory of concept attribution that is both independent of language and applicable to animals.

Keywords: cognitive ethology; concepts; propositional content; anthropomorphism; subjective experience.

## Contents

## Introduction

This thesis is concerned with an examination of the field of cognitive ethology particularly as it bears on the search for mental states in animals. The field of cognitive ethology's main aim is the study of the cognitive processes of animals.

Donald Griffin is credited with starting the discipline of cognitive ethology. He had been working on the echolocation capacities of bats since the early 1970's and was giving a talk on his findings at Rockefeller University. Tomas Nagel was in the audience, and witnessed Griffin's post-colloquium treat of releasing bats in the auditorium to demonstrate the process of echolocation. Nagel was inspired to write an essay in 1974 entitled "What is it Like to be a Bat?" (1974) which was about the elusive phenomenon of the subjective point of view and the apparent failure of objective scientific theories in capturing it. This essay in turn inspired Griffin to write a book published in 1976 entitled "The Question of Animal Awareness" (1976), thus introducing the question of the possibility of conscious awareness and other cognitive capacities in animals. Hence was born the discipline of cognitive ethology.

Characteristic to the discipline of cognitive ethology is its interdisciplinary nature. Input to the field includes philosophy, biology, and psychology as well as evolutionary psychology. This interdisciplinary nature makes for rich and varied discussion amongst its participants. However, one drawback stemming from the lack of a common background in the various participants in discussions is the absence of any standardized agreement over what should be the proper objects of study and what methods should be used to study them. Central to the study of cognitive ethology is a new emphasis on discovering the mental processes of animals. This new emphasis was inspired in part by Donald Griffin's hypothesis that some animals might have conscious awareness. These two issues, particularly that of whether or not and to what degree animals might be the locus of cognitive processes, have sparked off much discussion.

Many thinkers have voiced their doubts about the viability of studying animal cognition, offering up seemingly powerful arguments as to why animals cannot share the capacity for possessing intentional mental states, concepts, thoughts and other qualities

that we humans possess. As the eminent ethologists Colin Allen and Marc Bekoff note, the discipline of cognitive ethology has recently faced a challenge to its scientific status (1997:314). These arguments or objections to the project of cognitive ethology are pervasive in the literature, and presented as obstacles to the project by those who question its viability.

The inspiration for this thesis comes from an article titled "Slayers, Skeptics and Proponents" (1997) written by Allen and Bekoff. In this article they categorize the views of commentators on reviews of the work of Donald Griffin into three possible points of view: that of a detractor (Slayer), a hopeful fence-sitter (Skeptic) and an advocate (Proponent). From there, they were able to distill the objections against and arguments for and against the discipline of cognitive ethology based on book review articles commenting on the work of Donald Griffin. The objections treated by Allen and Bekoff include that anthropomorphism is unscientific; that anecdotes are illegitimate forms of data; that attribution mental states to animals is impossible and that cognitive ethology is a soft science (1996:315).

Finding Allen and Bekoff's treatment of the main objections to be adequate but somewhat too superficial and limited to the work of a single author, I chose to limit my focus to the most damaging point of view vis-à-vis the status of cognitive ethology, that of the detractor, and also to distill and compile a set of objections that were most recurrent in my reading of the literature. I have devoted the first part of this dissertation to an examination of some of the main objections from the detractor point of view made to the project of investigating mental states in animals. My aim is to demonstrate that although these objections might have *prima facie* force to them, upon further scrutiny they end up being baseless. For each of the four objections, I have chosen to discuss an author whom I believe is most representative of the objection. This allows me to discuss the objection in some depth without compromising details.

The first objection has to do with language. It is argued that language is necessary to thought, animals do not possess a reasonably humanlike form of language, and so animals cannot be said to have thoughts. Since intentional mental states fall into the category of thoughts, it is argued that language is necessary for the possession of intentional mental states. Donald Davidson is the most thorough and well-articulated

proponent of this view. This objection hinges on the definition of language, and there are two argument strategies one could take with regard to this claim. If language is more narrowly defined, in other words construed as reasonably humanlike in nature, then it is true that animals do not possess language, construed as such. The strategy in this case is to then argue that language is not necessary for the having of thought. If language is taken in its broadest sense to mean a system of communication, the strategy is to argue that some animals do indeed possess a language, and the option is then open to investigate whether animals possess mental states or not. In chapter one I examine both of these strategies.

The second objection has to do with anthropomorphism. The term refers to the tendency to attribute humanlike qualities to non-humans. The objection is that it is anthropomorphic, and thus a category mistake, to attribute humanlike qualities to animals. Anthropomorphism is considered a category error according to this objection because of an underlying assumption that humans and animals belong to two separate categories, and the attribution of traits across categories is an error of misattribution. Mental states fall into the set of traits that are presumed to be restricted to humans, and so it is anthropomorphic and thus an error to attribute mental states to animals. The claim that animals and humans are two separate categories has not yet been borne out, and so this objection turns out to be a case of begging the question. However, there is also the more general objection regarding the issue of anthropomorphism, based on the claim by detractors that it occurs rampantly, regularly and in an unchecked manner in research in cognitive ethology. This pervasive phenomenon supposedly stems from the fact that it is an innate natural human tendency to anthropomorphize. It is argued that researchers should not use the same terminology that they use for humans in their descriptions of animal behavior. Chapter two is thus concerned with an examination of the claim that the charge of anthropomorphism is based on an error of categories as well as the rejoinder claim made by ethologists that it is useful when employed as a heuristic tool in research, in the context of hypothesis-testing.

The third objection is not an objection in and of itself, but rather underlies many of the others discussed. It is the problem of 'other minds', also known as the problem of 'other species of mind' in Cognitive Ethology. The fact of the matter that is often

referred to is that we can never have direct access to the contents of the mind of another human. We have even less direct access to the minds of animals, since they cannot verbalize their mental states in a manner understood by us. Since we have twice removed access to the mental states of animals, it is futile to search for them.

The problem of other minds stems in part from the mutual influence of the philosopher Thomas Nagel and the ethologist Donald Griffin mentioned above. As mentioned, the discipline of cognitive ethology was in part born as a result of Griffin's suggestion that animals may have subjective awareness. This claim led Nagel to ask of the nature of the phenomenon of subjective experience and whether we could have access to it. If anything is not knowable, it is certainly the subjective point of view of another human being, and even more so that of an animal. Nonetheless the attempt is being made by cognitive ethologists to research the subjective experience of an animal, due in part to the realization that much of the research done to date is from an anthropocentric (human centered) perspective and potentially masks whatever cognitive capacities the animal might truly have. Chapter three is thus concerned with an analysis of some of the various reactions to the tension created by Nagel's consideration of the subjective point of view, as well as an evaluation of this new research strategy. Notwithstanding questions of tractability, research into the animal's subjective world is certainly a step in the right direction simply because it draws us further away from the human centered or anthropocentric perspective that is characteristic of much of cognitive ethology, and more toward the area in which subjective awareness might be found, if it exists.

The final objection has to do with methodology. It is argued that the study of animals is empirically intractable, in part because they lack language. In most experiments involving human subjects, language is the medium by which subjects are briefed and de-briefed as to the aim of the experiment. Animals cannot be briefed nor can they be asked or answer questions.

Experimental design must become a lot more intricate and sophisticated to get around the lack of a common information-sharing medium between humans and animals. This new level of sophistication, along with a consideration of the type of phenomenon being studied (mental states), invites the question of interpretation, most often whether the experimental results support the hypothesis advanced. Chapter four looks at a

snapshot of this problem occurring in a subset discussion area within the discipline of cognitive ethology, where the existence of a theory of mind is being investigated in primates. Theory of mind theories are concerned to explain the possible mechanism underlying the human ability to explain and predict each other's behavior. It is hypothesized that humans interpret each other's behavior by attributing mental states to themselves and others. The view of Celia Heyes, ardent opponent of the theory of mind theory, will be examined in this chapter. Her argument is that since current experiments cannot demonstrate univocally that primates use a theory of mind, we should halt research into this branch of cognitive ethology until a decisive method can be found. An important methodological problem remains from the discussion regarding the apparent ambiguity in interpretation in current experiments, the idea that theory of mind theory cannot eliminate other alternative explanations. This problem will be resolved in the second half of the thesis.

It should be clear to the reader of the first four chapters that philosophy informs a large portion of the theoretical underpinnings of cognitive ethology. The second half of the thesis is thus devoted to evaluating two of the most important and fruitful outcomes of the marriage between philosophical theory and empirical research in cognitive ethology: the attribution of intentional mental states and the attribution of concepts.

One of the aims of research in cognitive ethology and the topic of this thesis is the investigation of mental states in animals. Theories of intentionality, on my interpretation and in the context of this thesis, are the theoretical 'spelling out' of both the constraints necessary for the attribution of mental states as well as the content of these mental states. Chapters five and six will thus be concerned with an evaluation of four theories of intentionality. These theories will be subject to four conditions that I have retained from the discussion on objections entertained in the first half of the dissertation. Stealing a trick from Dennett in his 1969 book "Content and Consciousness", my aim is to elucidate the constraints from within which any satisfactory theory of intentionality must evolve, in order to be applicable to animals. The four conditions are the degree of empirical applicability of the theory, the ability of the theory to account for error, whether or not the theory can specify content of mental states to a reasonable degree and most importantly, whether or not the theory can vindicate the attribution of mental states in the

animal. This last condition, if fulfilled by the theory, will give us a solution to the methodological dilemma encountered in chapter four in the discussion on eliminating other explanations.

A second philosophical evaluatory tool that enters into the search for intentional mental states of animals is the notion of a concept. On a philosophical construal, concepts are the constituents out of which thoughts are built. The contentious claim with regard to the link between mental states and concepts is the following. If a creature is to be attributed propositional attitude mental states such as hopes, desires and fears, then that creature must also possess the concepts of these propositional attitudes. I will not concern myself with this dense claim in the last chapter, preferring instead to tackle the preliminary task of divorcing concepts from language and offering one example of a theory of concept attribution that does not depend on language and that is empirically tractable.

In the last chapter I will thus evaluate Cecelia Heyes and Nick Chater's self-fulfilling prophecy of a failed search for a theory of concepts that applies to animals. The reason that no theory of concepts is applicable to animals is due, in their opinion, to the tight link between concepts and language. The search will be doomed from the start because there is good reason to think that a theory of concepts that applies to humans will not apply to animals precisely because human theories are often linked to human language. I will argue that there is no reason to accept this conclusion by pointing to one of the most fruitful theories of concepts in my opinion, employing behavioral criteria, that applies to animals and that does not depend on language, developed by Colin Allen.

## Chapter One

## No Thought Without Language

### 1. Introduction: Two Strategies

The idea that animals lack a reasonably discernable human-like language is often cited as a reason not to attribute thoughts to them. In schematic form, the argument is as follows: Language is necessary for having thought, animals lack language, thus animals cannot have thoughts. The language argument is one of four major objections raised against the idea that animals might have intentional mental states.

Donald Davidson has written three articles on the topic of thought, language and their relation to animals. Three separate but related arguments can be discerned from these three articles. The first article "Thought and Talk" (1975) discusses the issue of the interdependence of thought and language, in it Davidson argues that thought depends on speech. The second article "Rational Animals" (1982) makes the case that the having of propositional attitudes requires rationality, thereby linking rationality with thought. The third article "The Emergence of Thought" (1999) is an argument for the holism of thought, or the interdependence of various aspects of mentality, which makes difficult the tracking of the exact emergence of each of these aspects. These three articles taken together constitute Davidson's overall view that there is no thought without language and lacking language, animals cannot thereby reasonably be claimed to have thought.

There are three aspects of mentality that Davidson will try to link together in the three articles: language, thought, and rationality. Due to the holistic nature of his arguments and the interdependence between the various aspects he wants to evidence, his arguments seem to hinge on one another, i.e., one cannot discuss one without discussing all of them. Some points are revisited in different articles, but each time with a different aspect that is amplified. It is very difficult to convey his arguments in a sympathetic manner without first exposing them in their entirety, which is what must be done in order to get a comprehensive view of them.

We will see in the next chapter that Davidson has been labeled a "hard 'centrist" by John Fisher. Hard anthropocentrists, generally speaking, are committed to a sharp

divide between humans and animals. The position of the hard anthropocentrist is that any attribution of any mental predicate to any non-human animal is a form of categorical anthropomorphism (Fisher, 1996:7). Although Davidson does not actually claim that language is an exclusively human trait aside from in a footnote in the "Rational Animals" article (1982:319, note 1), the fact that he explicitly claims that only creatures with language can think may be taken as an anthropocentric view. At any rate Davidson himself raises this issue in this second article.

As mentioned, one of the issues that Davidson discusses in the first article is the relation between thought and language. Does language depend on thought, does thought depend on language, or does neither have conceptual priority? There are two possible ways to interpret the claim that neither thought nor language has conceptual priority. One is that they are interdependent, which is Davidson's view, and the other is that they are independent of each other. Davidson doesn't believe that an adequate argument has been given for the view that thought depends on speech, and so this is one of his aims in the article. It should be noted here that his focus is on the interpretation of speech rather than speech *per se*. That is, he is interested in highlighting the role of the interpreter of speech rather than the speaker. He will not try to demonstrate that an interpreter must be a speaker, although there are good reasons, on his view, to think this. He is ultimately interested in demonstrating that thought depends on the ability for the interpretation of speech, the ability to understand the utterances of another. The revised claim is thus that thought depends on the interpretation of speech. The ultimate conclusion he wishes to draw from this argument is that a creature cannot have thoughts unless it is an interpreter of the speech of another.

There are two possible strategies to take for those who are not in agreement with Davidson's claim that language and particularly the interpretation of speech is necessary for thought. The first is to disagree that language is necessary for thought. In taking this strategy, one must examine Davidson's entire justification for the claim and try to find flaws in it. This second strategy one can take with regard to Davidson's claim is to agree that indeed, language is necessary for thought and then argue that animals have language, therefore they have thought. I believe that this second strategy is the one pursued by most cognitive ethologists. In taking this strategy the onerous task is that of convincing

the audience that the system of communication in animals does constitute a language. At issue here is the definition of language, among other things. It is obvious that, if animals had human language, the objection would never have been raised in the first place and cognitive ethology would not have to defend itself against this objection. The issue is thus how much the system of communication in animals must resemble human language in order for it to be construed as a language, i.e., the kind of language that is supposed to be necessary for thought. Davidson has a few conditions regarding this issue. A second issue concerns whether we should accept Davidson's elements that are necessary for language and conditions for thought as applying to language in general or just applying to human language.

Both of the above strategies will be examined in this chapter. I choose to examine both because both are viable strategies to take. In particular I want to examine the strategy of claiming that animals have language because I don't think that Davidson has considered the evidence. To say this brings an immediate objection that this is not an empirical question, decidable exclusively by citing evidence or naming names of particular species. However, I think it is, to a certain extent, necessary to look at empirical demonstrations of animal language, or at least considerations on the matter, since the issue hinges on whether a language can be found in species of animals, and whether this so-called language can be demonstrated to resemble the type of language that humans have.

## 2. Thought and Talk

Davidson's point in this first article (1975) is to outline the relation between thought and speech. He claims he will demonstrate, through a series of interrelated arguments, that thought depends on speech. He believes that the relation has never been entertained for its own sake, that the assumption is usually made that one is more complex a concept than the other, and that the more complex term can be explained in terms of the simpler term (1975:156). Neither of the two concepts can be fully explained in terms of the other, in his view.

The term 'thought' must be defined since it has such a central role in Davidson's views. Davidson's use of the term should not be conflated with the ordinary-use sense of

the term. In ordinary usage, thoughts encompass all mental states that have content. On Davidson's construal, thoughts are the contents of propositional attitudes such as belief and desire. Propositional attitude reports, such as 'John hopes that it will rain today', are characterized by the fact that they exhibit semantic intensionality. That is, substitutions of co-extensive terms in the sentences can alter the truth-value of the sentence (1975:156).

The first thing to notice about Davidson's argument is that he is interested in demonstrating the relation between thought and speech as opposed to merely thought and language, for he will claim that without the ability to interpret the speech of another, thoughts and beliefs and desires cannot be attributed. The relation of dependence he wants to emphasize is that thought depends on the ability to interpret speech. This might appear to really obstruct the cognitive ethologist's project, even more so than if Davidson were able to prove that thought depends on language, since there is the implicit additional assumption of verbal utterances or speech. To prove that animals need language in order to think is one thing, and could perhaps be circumscribed by enlarging the definition of language to include the connotation of a system of communication. It could then be argued that it is possible to attribute concepts to an animal, since animals pass the criterion of having a language, if language is construed as a system of communication. To make the case that thought depends on the interpretation of speech presents a seemingly insurmountable challenge to the cognitive ethologist. First, it must be shown that some species of animals speak. It could be objected that this requirement is unnecessary, for as stated above, Davidson insists that the focus is not on speaking as such. However, the idea of an animal that cannot speak but can nonetheless interpret the speech of others, in other words a mute interpreter, makes no sense unless we identify whose speech the mute is interpreting. Demonstrating that some animals speak might be impossible since animals lack the developed vocal cords that humans have. Second, one must show that animals must be able to interpret each other's utterings in a meaningful way in order to be said to have thought. Since the second challenge requires the first in order to be fulfilled, and the first has so far been found physically impossible to fulfill, there does not appear to be much hope for the cognitive ethologist to demonstrate that animals satisfy Davidson's condition for having thoughts.

The argument in the article can be divided into three sub-arguments. The first sub-argument makes the case for the endless interlocking of belief, or holism with regard to belief. One of the claims is that the propositional attitudes cannot be reduced to each other. For instance, desires cannot be reduced to hopes, and there are no basic propositional attitudes. Belief, however, is central to all types of thought, and often underlies other attitudes. A desire for an object, for instance, is often accompanied by a belief that the object exists. The next claim is that having a thought requires that there be a background of beliefs, but a thought does not depend on a particular belief. So although a list of potential beliefs can be attributed to an individual in a particular scenario, the thought of the individual cannot be fixed to particular beliefs. The last claim makes the case for holism of belief. Here follows the first set of claims in point form.

1. The various sorts of thought cannot be reduced to one another.
2. Belief is central to all kinds of thought.
3. Having a thought requires that there be a background of beliefs.
4. It is necessary that there be endless interlocked beliefs (1975:156-7).

Davidson then suggests looking at the relation between thought and language from another angle, namely by inspecting the theory implicit in the explanation of behavior, the teleological explanation of action. He gives a mundane example of an action, that of a man raising his arm, that is explained by a series of beliefs and desires. For instance, the person raises his arm because he desires to attract the attention of his friend. This person must also have the belief that raising his arm will indeed attract the attention of his friend. The fact that behavior can be explained by patterns of beliefs and desires leads Davidson to claim that attributions of belief and desire are supervenient on behavior. Supervenience in this context means that there is a relation of dependence between beliefs and desires and of behavior of the following type. There cannot be a difference in beliefs and desires without their being a difference in behavior, but there can be a difference in behavior without an ensuing difference in the beliefs and desires (Honderich, 1995:860). There is a further implicit assumption made by Davidson here, that teleological explanation is a form of rational explanation. Davidson appeals to two factors to make the case for the cogency of teleological explanation. The first is that the

action to be explained must be reasonable in the light of the assigned beliefs and desires, and the second is that beliefs and desires must fit with one another. Here follows the second group of claims in point form.

5. Attributions of beliefs and desires are supervenient on behavior.
6. The cogency of belief-desire (telelogical) explanation rests on the ability to discover a coherent pattern (1975:158-9).

This assumption of rationality, according to Davidson, constrains the range of beliefs that are potentially attributable to an individual in a particular scenario in the sense that it is still possible to attribute irrational beliefs to someone, but the possibility is less likely given the rationality constraint. The fact remains, however, that it is possible to attribute to a thinker an explanation of behavior that is made up of irrational beliefs as well as numerous different sets of rational beliefs and this creates a problem of under-determination. That is, that many equivalent sets of beliefs and desires can be attributed to any given behavior, and there is no way to choose which set is the one. The problem is further exacerbated by the fact that behavior, which is the main evidential basis for attributions of belief and desire, is observable, while beliefs and desires are not. In order to narrow down the possible set of beliefs and desires that can be attributed to a thinker in order to then begin to identify a particular belief or desire, Davidson claims that the attributer must be an interpreter of speech.

The next half of Davidson's article is thus concerned with making the case for the main thesis, the claim that a creature cannot have thoughts unless it is an interpreter of the speech of another. Central to this claim is the idea of an interpreter, one who can understand the utterances of another. Davidson insists that the idea of a language is not necessary for making his point. Two speakers could interpret each others utterances without there being, in any ordinary sense, a common language (1975:157). While this claim might be true, I wonder if two interpreters could understand each others utterances without there being a common language? I shall come back to this question.

So far, Davidson has shown that the attribution of belief and desire must go hand in hand with the interpretation of speech, but has said nothing about why the attribution of thought depends on the interpretation of speech. He first offers an uninformative reason: that without speech we cannot make the fine distinctions between thoughts

essential to explanations of behavior (1975:163). He gives the example of a dog that believes that his master is home. He asks, does the dog also believe that Mr. Smith is home, or that the manager of the local bank is home? All three beliefs are equivalently attributable, and there seems no way to decide between them, especially in the absence of speech (1975:164). The above does not constitute an informative reason however, according to Davidson, all he has shown is that unless there is behavior that can be interpreted as speech, the evidence will not be adequate to justify the fine distinctions we make in attributions of thought.

An argument is needed that will show that only creatures with speech have thoughts. To develop the argument, Davidson appeals to the notion of interpretation. A central aspect of interpretation is to give knowledge to the interpreter of the circumstances under which someone holds sentences true (1975:162). To make this point, Davidson draws an analogy with belief. Just as it is the pattern of beliefs that allows us to identify a particular thought, it is the pattern of sentences held true that gives sentences their meaning. In drawing this analogy, nothing has been said about how interpretation is able to serve this function of giving to the interpreter knowledge of sentences held true thus giving meaning to sentences. The difficulty in saying how interpretation is able to carry out this function is due to the fact that two factors enter into the situation: what the thinker takes the sentence to mean, and what the thinker believes. A method is needed to hold one of these factors steady while the other is studied (1975:167).

The assumption that most beliefs held by the thinker are true enables one to hold steady the factor of what the thinker believes. This assumption is too strong, however, for it assumes that the thinker has no false beliefs at all and can therefore never err. Davidson thus claims that the intelligibility of the identification of false beliefs must depend on a background of largely unmentioned and unquestioned true beliefs (1975:168). What makes interpretation possible is that we can dismiss the chance of massive error. A good interpretation counts a sentence true just when a speaker holds it to be true, and given that both the speaker and the interpreter may be wrong in some cases, Davidson modifies the original claim to the idea that a good theory of interpretation maximizes agreement between the interpreter and the speaker (1975:169).

Given the account of interpretation above, Davidson claims that the concepts of objective truth and of error are central in the context of interpretation. The distinction between a sentence held true and being in fact true is essential to the existence of an interpersonal system of communication. When there is a gap between that which is objectively true and that which is held true by the speaker, this gap must be called error. Since the attitude of holding true by the speaker is the same whether the sentence is true or not, it corresponds directly to the concept of belief. The concept of belief is what takes up the slack between objective truth and that which is held true by the individual (1975:170).

Davidson then makes a rather bold set of claims. The first is that we have the idea of belief only through its role in the interpretation of language. As a private attitude it is unintelligible except as an adjustment to the public norm provided by language. Thus, he claims, a creature must be a member of a speech community if it is to have the concept of belief. Given the dependence of the other attitudes on belief, only a creature that can interpret speech can have the concept of a thought. Below is the last part of the argument in point form.

7. We have the idea of belief only through the interpretation of language.
8. To have the concept of a belief one must be a member of a speech community.
9. Given the dependence of the other attitudes on belief, only a creature that can interpret speech can have the concept of a thought (1975:170).

Davidson then asks, at the very end of the article, if a creature can have a belief if it does not have the concept of belief. He thinks not, because in order to have a belief, a creature must understand the possibility of being mistaken, this requires grasping the contrast between truth and error- true belief and false belief. (1975:170). This contrast, he argues, only arises in the context of interpretation, which alone forces us to the idea of an objective public truth. The stipulation here is that in order to be able to entertain beliefs, a creature must understand the contrast between true belief and false belief, which necessarily implicates the concept of belief.

There are numerous points of disagreement to be found with Davidson's view even with an examination of just this first article, since it represents a schematic for his general viewpoint. The other two articles deal with amplifying two other aspects of

thought: its relation to rationality and the holistic character of thought. It should be noted that I am only interested in disagreements with Davidson's views in these three articles as they directly bear upon the possibility of attributing thought to animals. The first obvious point of disagreement concerns his definition and construal of 'thought'.

Hans-Johann Glock takes issue with Davidson's use of the term 'thought', in particular his inclusion of 'concept' and 'propositional attitude' within the set of 'thoughts' (Glock, 2000:42). To take issue is legitimate, since Davidson explicitly mentions his intention to interchange certain terms at various points in the articles. Concerning the inclusion of propositional attitudes into the realm of thoughts, Davidson states in the "Rational Animals" article "Let me speak of all the propositional attitudes as thoughts." (Davidson, 1982:321). Concerning his inclusion of concepts as well as propositional attitudes within the realm of thoughts, he states in an article called "Seeing Through Language": "Thus there is in fact no distinction between having a concept and having thoughts with propositional content." (Davidson, 1997:25). Glock notes that Davidson has included concepts within the realm of thoughts and that this causes a problem. Glock writes: " He (Davidson) insists, firstly, that concept possession and the ability to have thoughts amount to one and the same thing, and, secondly, that both are confined to language users." (Glock, 2000:42). The first claim, according to Glock, provides the rationale for the second, in that to attribute thoughts to animals on the basis of non-linguistic behavior is misguided, since these thoughts involve concepts which cannot be attributed on such a basis. Glock has a point here, the first claim does provide a rationale for the second claim. This is merely a symptom, however, what does the source of this problem stem from? I believe it stems from the inclusion of both concepts and propositional attitudes under the heading of thoughts. Putting all three elements into one set works well for ease of discussion, and it is true that concepts and propositional attitudes are thoughts. The problem with conflating all three is that distinctions that do in fact exist between the three terms are masked. One can therefore not ask questions that draw on the distinctions between the three elements. For instance, can one have thoughts without possessing concepts? Can one have propositional attitudes without having the concept of one, such as belief? It would be interesting to ask if it is possible to have

propositional attitudes without having the concept of them, particularly in relation to animals.

This conflation also raises the question of whether or not concept possession really precludes animals from having thoughts or beliefs. As Glock notes, not making the distinction between thoughts and concepts does not even permit the question to be asked. This issue will be taken up again later, because as will be seen, Davidson will add two more items, membership in a speech community and language, to what is necessary in order to be said to have a belief.

Another point of disagreement to be found is with Davidson's holistic characterization of belief outlined in the first four premises. One can disagree with his premise 4, that it is necessary that there be endless interlocked beliefs. One can also disagree in the same vein with premise 3, that having a thought requires that there be a background of beliefs, but a thought does not depend on a particular belief. There exists a contrasting point of view to this holistic view of belief and thought, that of atomism. Atomism about concepts holds that instead of concepts being individuated by their relations to one another as a holistic view would dictate, concepts are instead individuated by their relations to the world (Margolis, 1999:551). Applied to beliefs, atomism entails that beliefs are individuated by their relation to the world. In the case of animals, this view would allow animals to be attributed single beliefs without having to assume a whole background of other beliefs that the animal may or may not have. It would also allow for the identification of a particular belief, for it does not assume that a thought does not depend on a particular belief, i.e., the indeterminacy of belief in the case of Malcolm's dog. On Davidson's view, a thought does not depend on a particular belief, i.e., there is no one-to-one correlation between beliefs and thoughts.

It is also possible to counter-argue Davidson's claim that the under-determination problem is solved only by language. As will be recalled, the problem of under-determination is caused by the rationality constraint, and exacerbated by the holistic nature of belief. It is the insistence that many of the creature's beliefs must be rational along with the insistence that there be endless interlocked beliefs that gets Davidson into the under-determination situation where individual beliefs cannot be identified. As will be seen in the second half of the thesis, the issue of identification of beliefs also arises in

relation to theories of intentionality as applied to animals, in the search for mental states. The problem must be solved in various ways other than by recourse to language, since animals do not possess a language that is strictly humanlike in nature.

The case could be made that there is an inconsistency with regard to the issue of language in Davidson's arguments. At the beginning of the article, he insists that the idea of a shared language is not necessary for making his point. He claims that two speakers could interpret each others' utterances without there being a common language. We may agree that interpretation is still possible between two individuals whose mother tongue is different, for instance. He then insists that a speaker must be a member of a language community in order to have the concept of a belief. It is not clear whether speakers of this community speak the same language. Is the notion of a shared language necessary or not to Davidson's arguments? I don't think his argument can get off the ground without the preliminary assumption of a shared language. On the other hand, if we take his view at face value, and accept that a shared language is not necessary, then just as humans who do not have a shared language may nonetheless be able to interpret each other's foreign language, the door is open to argue that humans may also eventually be able to decipher the 'language' of animals. Humans and animals then may eventually be able to interpret each other's 'language' even though it is not a shared language.

**3. Rationality**

Davidson's second article (1982) on the question of thought and language is interesting, among other reasons because he notes at the beginning of the article that he is not interested in the empirical question of whether animals have propositional attitudes, contrary to most of the other commentaries on the issue. The question he is interested in is rather what sort of empirical evidence is relevant to the question of whether an animal has propositional attitudes (1982:318). What subtle difference is he trying to emphasize here by his turn of phrase? Perhaps the distinction can be stated as follows: 'What animals are rational?' constitutes the empirical question, as opposed to the question Davidson is interested in, which is 'what makes an animal rational?' With regard to the first empirical question, one presumes he means that whether or not animals have concepts is a question that gets a yes or no answer and is decided purely on empirical

evidence for or against, and not on theoretical considerations. Davidson is interested in theoretical considerations. I have to wonder about Davidson's underlying motive, in not looking at the question empirically and not engaging in, as he calls it "naming names or names of species" (1982:318), is he trying to eliminate the possibility that looking at empirical work would constitute evidence for or against the issue?

The answer to the question of what constitutes rationality in an animal largely hinges on the same arguments already discussed in the first article. Davidson cites the first criterion, that of having propositional attitudes. The having of propositional attitudes is thus a criterion for rationality. His argument for holism follows closely behind. Talk about propositional attitudes naturally leads him to claim that "to have one is to have a full complement." (Davidson,1982:318). He goes on to list a second criterion, that of language, although he gives no justification for it yet. He just states that, according to holism, one belief needs other beliefs, and other propositional attitudes such as desire, intention, and perhaps even the gift of tongues.

These two criteria, that one either has none or many propositional attitudes, and that one must have language in order to be rational might lead one to, as he anticipates, accuse him of being anthropocentric. Anthropocentrism construed generally is the view of regarding man as the center of existence. An anthropocentric view is a human centered view. With regard to language it is the view that only humans have the cognitive capacity for it (Mitchell et al 1997:11). Language is an exclusive property of the human species (Kiriazis & Slobodchikoff, 1997:365). Davidson believes that the charge is fair but ought not be levied against him since, by his lights, he is only describing a feature of certain concepts. In other words, it is a feature of propositional attitudes that 1) to have one is to have many and 2) to have propositional attitudes is to also have language. His reason why he should not be charged with anthropocentrism is that he is merely pointing out two special features of language. He then gives two examples of fine distinctions that exist in language. The first is the fact that our language is rich enough to describe the differences between humans and other creatures. The second is the fact that the Inuit language is rich enough to contain 16 different words for snow. These two aspects of language go to show that we strive to make our language and us seem special (1982:319). Nowhere, he insists, is he claiming that language is unique

to humans. I am of the opinion he doesn't have to explicitly claim that language is unique to humans, that it follows from what he says.

Using Norman Malcolm's story in 'Thoughtless Brutes' (1972) as a point of departure, Davidson announces that he has an argument that will throw doubt on Malcolm's conclusion that the dog has a particular belief: that the cat went up the tree. A dog is chasing a cat in a backyard. The cat is heading toward an oak tree but at the last second, unbeknownst to the dog, veers off and climbs a maple tree instead. The dog, thinking that the cat went up the oak, stands under it and barks up at the branches. Someone observing the scene would say ' the dog thinks the cat went up that oak tree.' Malcolm claims that the observer would be justified in attributing this belief to the dog under the circumstances. Davidson's challenge is that we cannot attribute a definite belief or set of beliefs to the dog, and there are many to choose from: the fact that the tree is oak, the fact that it is the oldest tree in the park, the fact that it is the same tree as the last one the cat went up, etc. In order to be able to attribute to the dog the belief that the cat went up the tree, we would have to assume that the dog had many other beliefs as well. As he puts it " There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs, there would be no reason to identify a belief as a belief about a tree, much less an oak tree." (1982:320). So many or at least more than one belief is necessary in order for a single belief to be attributable to a creature. He claims that one runs into trouble quite quickly as soon as one wonders how one would decide if the animal had the peripheral set of beliefs necessary to make the initial one make sense. One cannot distinguish between the various beliefs that the dog might have, one is not able to tell if the dog has them or not. Each belief requires a world of beliefs in order to give it content and identity, and every other propositional attitude depends for its particularity on a similar world of beliefs (1982:321). In brief terms, the holistic nature of belief is such that it brings about the situation of underdetermination of content with the consequence that we cannot attribute a single belief to the dog without attributing many, and it is impossible to identify any single belief in the dog.

Davidson then gives a reason why to have propositional attitudes is to be a rational creature. He starts with the propositional attitude of belief, saying that although there need not be a fixed set of beliefs attributable to the dog, many true beliefs are

necessary. Within this set of beliefs, some may be particular, general or logical. He then claims that since belief is so central to the propositional attitudes, that he is going to hereafter refer to all the propositional attitudes as thoughts. This allows him to claim that thoughts have logical relations. The identity of a thought cannot be displaced from its place in the logical network of other thoughts, it also cannot be relocated in the network without becoming a different thought. Radical incoherence in belief is therefore impossible. He is thus able to conclude that to have a single propositional attitude is to have a largely correct logic, in the sense of having a pattern of beliefs that logically cohere (1982:321). This is one reason why to have propositional attitudes is to be a rational creature.

Davidson then goes on to argue for language as a necessary condition for thought. He starts with the claim that it is justifiable to attribute attitudes to a creature given the observance of a reasonably complex pattern of behavior, because there is enough of a conceptual tie between behavior and the attitudes. Then there is a stipulation that the pattern of behavior being observed must be quite complex to warrant the attribution of a single thought. There is such a complex pattern of behavior only if the agent has language. The implication here is that Malcolm is only justified in attributing the belief 'that the cat is up the oak tree' to the dog if the dog has language. In order to be a thinking rational creature the dog must be able to express many thoughts, and above all, be able to interpret the speech and thoughts of others (1982:323).

Against this it has been argued that given the success of explaining and sometimes predicting behavior by attributing thought to languageless creatures, why postulate the additional stipulation of language? Davidson admits that although we do predict and explain the behavior of animals by attributing beliefs, desires and intentions to them, there is a sense in which it is wrong to claim that non-verbal animals have propositional attitudes. He compares animals to missiles, whose behavior can also be explained by attributing propositional attitudes to them, although it is clearly unwarranted. In the case of the missile, it is the designer of the missile who must have propositional attitudes attributed to he or she, such as believing and desiring that the missile should destroy an enemy airplane, rather than the missile itself. Describing the missile as having propositional attitudes is a manner of speaking, it is not the case that the

missile really has propositional attitudes. Animals are different from missiles in two ways. One, they are far more like humans in the range of their behavior than missiles and two, we do not know of any better way to explain their behavior than to ascribe propositional attitudes. If we had a solid condition for the necessity of language for thought, we could continue to attribute propositional attitudes to dogs, even though we know that they do not really have them (1982:324).

So far Davidson has not really provided what he had set out to do, which is a necessary condition for thought. As he sees it, all he has really thus far shown is that there can't be much thought without language. The condition for thought that only language can supply comes in two premises:

1. In order to have a belief, one must have the concept of a belief.
2. In order to have the concept of belief, one must have language (1982:324).

Davidson begins by contrasting his construal of 'belief' with Malcolm's construal. Malcolm, unlike Davidson, restricts the term 'thought' to cover only the higher level of thinking, i.e., reflexive thinking. Thus he makes a distinction between simply having a belief or believing something, and knowing that one believes something, or being aware that one has a belief. Malcolm considers only the second type of higher order belief as thought and only it requires language. The dog can thus believe that the cat went up the tree but it cannot have the thought that the cat went up the tree. Davidson makes no distinction between beliefs and thoughts, and so both types require the concept of belief. Even the lower form requires it: to have a belief, one must have the concept of belief, which requires language (1982:324).

One of the criteria for having the concept of belief is the phenomenon of surprise. Surprise is an indication of the contrast between what the agent did believe and what the agent now comes to believe. Such awareness amounts to a belief about a belief. The phenomena of surprise points to the difference between the subjective way things are according to the thinker, and the objective way things really are, according to the world. Another way of saying this is to say that surprise involves a belief that a prior belief was wrong. This distinction implies the idea of an objective reality that is independent of my prior belief. A creature may react to the world, be able to discriminate colours, learn new reactions, and generalize its behavior to new categories of stimuli without entertaining

propositions. None of these things, according to Davidson show that the creature commands the subjective-objective contrast, as required by belief. The only thing that does demonstrate command of the subjective-objective contrast is linguistic communication. Communication depends on each communicant having and correctly thinking that the other has the concept of shared world, an inter-subjective world. The concept of an inter-subjective world is the concept of an objective world (1982:325-6).

To complete the argument, Davidson needs to show that the only way one could come to have the subjective-objective contrast is through having the concept of inter-subjective truth. In place of an argument, he offers an analogy where he introduces the notion of triangulation. He asks us to imagine what it would be like to be bolted to the earth. One implication would be not knowing where objects were located relative to oneself. The reality of our situation here on earth at the present time is that in not being bolted to earth, we are free to triangulate with objects. He asks us to imagine a sense of triangulation, involving two creatures, one that brings about the consequence of objectivity. The fact that the two creatures share language and therefore the concept of truth means that rationality is a social trait and only communicators have it (1982:327). This notion of triangulation is elaborated on further in the last article.

Davidson is concerned with the question of what constitutes rationality in this article. He specifically asks what makes an animal rational. He gives criteria for the attribution of rationality to an animal, one is the having of propositional attitudes, and above all, as he argued for in the first article, the ability to interpret the speech and thoughts of others, for these two things occur as a result of triangulation. He is able to conclude from this that rationality is a social trait and that only communicators have it.

One point of potential disagreement occurring in Davidson's second article is his argument concerning the criteria for the concept of belief. With regard to belief, it will be remembered from the first article that he does not think that a creature can have a belief if it does not have the concept of a belief. In that same article, he claimed that to have the concept of belief one must be a member of a speech community. In this article, he additionally claims that in order to have a belief one must have the concept of a belief and in order to have the concept of belief one must have language.

Concerning the two premise argument that links belief with language, Johann Glock has found an inconsistency within it. Davidson's argument is as follows.

1. To have a belief, one must have the concept of belief.
2. To have the concept of belief, one must have language (1982:324).

Glock believes that premise one is mistaken and that premise two, while true, cannot be argued for in the way that Davidson does (Glock, 2000:54-6). As a starting point, Glock takes Davidson's answer to the question treated at the end of his article: can a creature have a belief if it does not have the concept of belief? It will be recalled that Davidson thinks not, because a creature must understand the possibility of being mistaken, and this requires grasping the contrast between truth and error, true belief and false belief. Glock's answer to this is to advance the claim that it is possible to switch from belief A to belief B without realizing that one's prior belief was mistaken. Realizing that one's prior belief was mistaken is akin to having a belief about a belief, and Glock claims that this middle step is not necessary in the switch from a mistaken belief to a new belief (2000:46). Moreover, Davidson cannot rule out this possibility.

I think that Malcolm's distinction between beliefs and higher order beliefs that I made reference to above is a sound one. I believe that Glock agrees with this construal and it is the one that Glock is trying to point out in the argument of the previous paragraph. This construal allows for attributing the thought to Malcolm's dog that the cat went up the tree, and saves the higher order reflexive thoughts, beliefs about beliefs, for reflexive creatures like humans.

Against Davidson's criteria for attribution for a concept, Glock offers an alternate construal of concept that is based on a behavioral criterion, and constitutes the type of construal that Davidson has already discussed and argued is not sufficient for concept attribution: "Concepts are principles of discrimination, and to possess a concept is to have the ability to recognize or discriminate different types of things." (Glock, 2000:45). Davidson insists that mere discriminatory capacities are not enough, that the ability to discriminate an object from others does not mean that a creature has the concept of that object (Davidson, 1999:8).

Davidson has two arguments against this construal of concepts and concept possession, found in another of his articles titled 'Seeing through Language' (1997). The

first is *reductio ad absurdum*, and overstates the case a bit, in my opinion. The quote from Davidson is: " Unless we want to attribute concepts to butterflies and olive trees, we should not count mere ability to discriminate between red and green or moist and dry as having a concept, not even if such selective behavior is learned."(Davidson, 1997:25). Davidson overstates the case here, no-one wants to attribute concepts to an olive tree based on the fact that it withers in dry soil and flourishes in moist soil, even if it were to turn itself toward the sun like a plant does.

One might be tempted to make sentience the distinguishing factor between plants and animals and humans, and follow Glock's suggestion which is to attribute concepts only to creatures that are sentient (2000:45). Sentience, on his view, is the dividing line between differential reactions to causal inputs, in the case of the tree, and real discrimination, which is tied to creatures with perceptual capacities. In my opinion, making the distinction hinge on sentience is a wrong way to go, for it is an *ad-hoc* distinction. As it turns out, a more appropriate distinguishing feature is contained within the claim itself; it is the ability to learn. Learned selective behavior should be considered as real discrimination, contrary to Davidson's dismissal of it, since the ability to learn is an ability, not a mere disposition, and furthermore it is the ability to modify one's behavior in the face of changed circumstances or circumstances that do not lead to the desired goal. It might even involve recognizing a mistake, it at least involves some kind of recognition that causes the behavior of the creature to be modified. As will be seen in chapter 7, learning from one's mistakes is one criterion for the attribution of a concept according to the eminent cognitive ethologist, Colin Allen.

The other argument of Davidson's is that there is a difference between discrimination and classification. Discrimination is a mere disposition and has no normative force, on Davidson's view. Classification is required on Davidson's construal of concept possession. Classification requires the ability to recognize a mistake, and is not among the abilities of non-linguistic creatures, according to Davidson (1997:25). As mentioned above, the case can be made that the ability of learned discrimination also requires the recognition that a mistake has been made, proof of this is that the behavior is modified on the basis of the mistake. Thus learned discrimination passes the criterion and should be accepted as an indicator of concept possession. As to why language is

required for the recognition of a mistake, this hearkens back to arguments visited earlier in the chapter. As Glock sees the issue, animals display non-linguistic or behavioral indicators that a mistake made has been recognized, and so the linguistic criteria should be dropped. This phenomenon of mistake recognition has not yet been explicitly studied in animals, but there is one case that has been cited by Colin Allen (1999:38). In the experiment, pigs were rewarded for making same/different choices with regard to pictures of faces and other body parts. The pigs performed at about 90% accuracy, and when the mistakes were analyzed, it was found that pigs physically backed away from 22 out 23 of their wrong choices made. Obviously language is a convenient indicator for communicating that a mistake has been made, but is not the only indicator. Moreover, humans often don't indicate by language but rather through body language that a mistake has been made. The same reasoning could be applied to animals.

**4. Emergence of Thought**

The third article by Davidson is titled 'The Emergence of Thought' (1999). It is in this article that Davidson's views on the holistic character of thought are detailed. Davidson claims that emergence is relative to a set of concepts, since when a phenomenon emerges for the first time a concept is instantiated. He cites holism of the mental as a reason for the difficulty in saying anything about the emergence of various aspects of mentality. We have seen him appeal to the holism argument with regard to the phenomenon of belief earlier. Holism of the mental is the interdependence of various aspects of mentality (1999:7). The fact that various aspects of mentality are interdependent means that it will be difficult to plot the emergence of any single one. Holism about belief entails that one cannot have just one belief. About this he states " Beliefs do not come one at a time; what identifies a belief and makes it the belief it is is the relationship (among other factors) to other beliefs" (1999:8).

The argument against the idea that a dog can have a single belief has already been seen in the second article. In this version Davidson looks at the issue from a slightly different angle. The argument from this angle contains two sub-claims. The first is that a belief is identified by its propositional content. The second is that one must have mastery of the concepts involved in the propositional content (1999:8). The ability to

discriminate an object from other objects, for instance, which is a capacity that animals are often claimed as having, is not the same as possessing a concept. Animals presumably can discriminate some objects from others and this is what leads people to attribute concepts to them. This is a mistaken line of reasoning for Davidson, for to have a belief or a concept is to be able to make sense of the idea of misapplying a concept. Cats and dogs cannot, on his account, make sense of the idea of misapplying a concept, that is, of believing or judging that something is a cat which is not in fact a cat. Furthermore, concepts, due to their holistic nature, also have logical relations to each other. One cannot identify the content of one's belief unless beliefs are mostly consistent with each another. Here Davidson equates consistency with rationality and concludes that a degree of rationality is also a condition for having beliefs (1999:8).

After much preamble about how difficult it is to say something about the emergence of thought, Davidson describes a pre-linguistic pre-cognitive situation which constitutes a necessary condition for thought and language, called triangulation. This notion of triangulation was introduced at the end of the rationality article, as the only way that a thinker could come to have the concept of the subjective-objective contrast. Triangulation is defined similarly in this article as a relationship between two agents, each who also have a relation to the world. Each agent tracks changes in the other agent based on the other agent's interaction with the world. Triangulation is so named because it is a threefold interaction, of two creatures and the world, but an interaction which is twofold from the point of each of the interacting agents. Davidson admits that triangulation can be observed to obtain in the preverbal child and the animal, because it can exist independently of thought and therefore preclude it (1999:12).

Davidson gives two examples of triangulation occurring in its simplest form (1999:12). He first gives an example of what he thinks is a triangulation situation that is at its source a wired-in reaction. This would be a fish reacting to the slightest movement of other fish in its school, tailoring its movements so that the formation of the school is not changed. Another example would be the Canadian geese who migrate to warmer climates every fall. They fly in a V-shaped formation, and often change position within the formation according to the movements of each other without disturbing the V-shape to a great degree. Davidson then gives two examples of a learned triangulation reaction.

The first is that of vervet monkeys in Kenya that have been found to give three significantly different vocalizations depending on whether they see an eagle, a lion or a snake approaching. The other members of the group, regardless of whether they have themselves seen the predator, flee to safety in some manner. The second example is that of a honey-seeking bird found in Africa. This bird knows how and where to locate honey but cannot open the source. The bird thus directs human hunters to the honey source to open it and the hunters then share the honey with the bird. He says about this that we cannot nonetheless conclude that the bird's behavior or the monkey's vocalization, however complex and purposeful it is, is due to propositional beliefs, desires or intentions. The bird's flight and the monkey's call, however instructive they may be, do not constitute a language. In order to constitute a language, the bird's behavior would have to be due to propositional beliefs, desires and intentions (1999:12).

Triangulation is essential to the existence and hence to the emergence of thought. The triangle can account for two aspects of thought that cannot otherwise be accounted for: the objectivity of thought, and the empirical content of thoughts about the external world (1999:12-13). The first aspect, the objectivity of thought, refers to the fact that propositional content is true or false independent of what it is to the thinker. The thinker must be aware of this situation. Wittgenstein has suggested that we could not have the concept of getting things right or wrong if it were not for our interaction with other people. The triangle stands for the simplest interpersonal situation. Two or more creatures each correlate their own reactions to external phenomena with the reactions of the other interacting agent. Language as well as thought is necessarily social.

As for the second aspect, the empirical content of thoughts about the world, Davidson believes that social interaction is the only account of how experience gives content to our thoughts. Without the situation of triangulation, there is no other medium that could tell us what it is in the world we are responding to. This is due to the ambiguous nature of the concept of cause. It is in our interest to resolve the ambiguity because the phenomenon of cause contributes to giving beliefs their content. There are two sources of cause, both of which are provided by social interaction, that of width and distance. The question of width is to determine how much of the content of belief is relevant to cause, and it is the similarity of reactions among participants that brings about

the answer. It is the social sharing of reactions that makes the objectivity of content available. The question of the distance of the relevant stimulus from the participant is again socially determined, it is distal as opposed to proximal because it is intersubjectively shared. The distal stimulus is thus triangulated, it is where causes converge in the world (1999:13).

Davidson is careful to note that triangulation is a necessary condition to thought. It cannot be a sufficient condition, because it exists in animals that he would not credit with judgement. He thus concludes that although triangulation must be present if thought is present, it can also exist independently and should be viewed as preceding thought in the order of things.

As things stand with Davidson's arguments, it is possible to credit animals with triangulation, and since triangulation exists where thought exists, it is not a far greater leap to credit them with thought, a leap that Davidson does not want to make. In order to stop this from occurring Davidson must add something further to prevent animals from being credited with thoughts. That further thing is language. Language is the instrument that enables a creature to communicate propositional contents out into the world, and it is that missing element that enables creatures in the triangle to form judgments about the world (1999:13).

In this article Davidson has claimed holism of the mental as a reason why not much can be said about the emergence of various aspects of mentality. He has also identified triangulation as a necessary condition for thought, and because it can occur in pre-verbal infants and non-verbal animals, he has been forced to claim a sufficient condition for thought, that of language.

A potential point of disagreement to be found in this article concerns this issue of Holism. Holism is central to Davidson's arguments against animals having concepts, thought and propositional attitudes. It can be seen from the way he argues his point for holism of the mental, that it makes difficult the tracking of the exact emergence of various aspects of mentality, that he is going to have trouble accounting for anything to do with the phenomenon of acquisition, be it of language, concepts or propositional attitudes. Because he has this 'all or nothing' attitude toward concept acquisition

language and the propositional attitudes, he is at pains to account for recent findings in child development, for instance.

Simon Evnine also takes issue with the use of the argument for holism by Davidson. His tactic is to find incompatibilities with Davidson's espousal of the general principles of holism and the strange implications that result from their application to the pre-verbal child (1995). The situation as Evnine sees it is that Davidson's denial of thought and language to animals is counter-intuitive to most people. This is not a very strong argument since it is well known that science has a history of overturning intuition and even common sense. However, if one applies this holistic view of concepts and language to infants, as Davidson does, the child cannot have language and so any utterings the child might make have no meaning, at least not to any adult. The intranslatable child somehow grows into the translatable adult, and somewhere during the course of that change from child to adult, incoherence in the child seemingly magically becomes coherence. Two explanations are possible for the result. One is that the child had some but not all of the conceptual and linguistic resources of the adult. This means giving up holism with regard to language. The other possibility is that the child jumped from not having language to having one all at once, and this conclusion is implausible, on Evnine's view. The point is that a holistic view with regard to even the acquisition of language is untenable. This throws doubt on Davidson's view that all aspects of mentality emerge simultaneously, thus opening the door to the possibility that one can have one or some of these aspects without having all. One could then have beliefs, for instance, without having the concept of belief.

The question that Achim Stephan is interested in is whether a creature that lacks the concept of belief can be said to have any beliefs at all. Stephan's tactic is the following (1999: 80-83). He takes one of Davidson's examples of what it means to have the concept of something, in this case the concept of a cat. In order to have the concept of a cat, Davidson stipulates, one must have a lot of beliefs about what a cat is, as well as a lot of other concepts, such as the concept of an animal, the concept of a continuing physical object etc (Davidson 1999:8). Stephan then transfers this stipulation, complete with criteria, to the case of belief, and out of this comes the Munchhausen, as he calls it, conclusion, that without having the concept of a belief one can have neither beliefs nor

concepts. It is Munchhausen because as Munchhausen pulled himself out of the swamp by his forelock, all concepts and beliefs get pulled out of the realm of the pre-mental by the higher-order concept of belief, according to Davidson's holistic account of belief. The problem, according to Stephan, is that this holistic view of concepts has the implication that neither animals nor infants nor demented adults can be said to have beliefs or concepts, because they probably don't have the concept of belief.

Stephan is looking for a category of creatures, that would include infants, animals and demented adults, that falls in between the set of creatures who can only perform rudimentary acts of discrimination and those truly concept-possessing individuals. He thinks that Davidson believes that the set is empty, especially given the quote from Davidson, reproduced below because it seems to precisely sum up the difficulty.

> We have many vocabularies for describing nature when we regard it as mindless, and we have a mentalistic vocabulary for describing thought and intentional action; what we lack is a way of describing what is in between. This is particularly evident when we speak of the 'intentions' and 'desires' of simple animals; we have no better way to explain what they do. (Davidson, 1999:11)

Achim Stephan thinks that there is a set of creatures that have beliefs without having the concept of belief. He thus suggests another construal of concept possession, conceived of by a notorious ethologist, Colin Allen. Creatures do not need to possess the concept of belief in order to have any concepts or beliefs at all. Allen's construal is rather based on a more enriched discriminatory capacity than Davidson's, one that includes recognition and correction of mistakes. It has three criteria:

1. The creature must be able to systematically discriminate between Xs and non-Xs,
2. The creature must be able to recognize its own discrimination errors, and
3. hereby be able to learn to better discriminate between Xs and non-Xs. (Allen, 1999:37).

It should be noted here that Colin Allen's second criterion replies to Davidson's objection on why discrimination is not proper concept possessing behavior, that a creature must be able to recognize when it has made a mistake in order to truly possess a concept. The phenomenon of mistake recognition can be observed behaviorally by the

fact that the animal changes its subsequent behavior. Allen's account also makes reference to learning, in that the creature learns to better discriminate by modifying or correcting its behavioral reaction the next time the situation arises. This set of three criteria for concept attribution will be examined in greater detail in the last chapter on concept attribution. In this context, Allen's criteria are offered as an alternative construal of concept attribution that does not assume language nor possession of the concept of belief, and would be ideally applicable to animals.

As I mentioned in the introduction to my comments on Davidson's views on the issue of the necessity of language for thought, I am only interested in critiquing his views as they bear directly on animals. As it turns out, the holism issue does not appear at first glance to bear on the animal issue. The fact that Davidson's holistic view of mentality is incompatible with aspects of human development of language and mentality in general in the child may not seem to have anything to do with the issue of animals. Appearances to the contrary, however, it does. If we create a set of humans that has as a common factor the lack of language, that would contain pre-verbal infants, humans who are hearing impaired and speech impaired and others, the case could be made that Davidson's views, although they could be said to be advanced to support the thesis that human language is necessary to human thought, that they do not even succeed at this level, since they leave out a portion of the population from the explanation. The door is then open to claim that animals may be included in this set of creatures, in need of a theory of thought that does not have language as its main condition. I thus would like to agree with Davidson on the one hand, that human language might be necessary to human thought. I do not, on the other hand, think that he succeeds even in making the case for this claim. In any case, his conclusion has nothing to do with the animal issue. That is, one cannot move from the claim that 'animals do not have human language' to the claim that 'animals do not have thought'. All that can be said as a result of Davidson's views is, following Searle, 'Humans have language in a sense that animals do not' (Searle, 1994:209).

**5. Conclusion: Taking the Second Strategy**

Remembering the two strategies that could be taken regarding Davidson's claims that I outlined at the beginning of the chapter, one could also take the second tactic, and

agree that language is indeed necessary to thought. This tactic would entail arguing that the term 'language' should be defined in a wider sense as a system of communication. In fact, one could follow Bennett's construal cited in the preface of "Linguistic Behavior" and interpret language as systematic communicative behavior (Bennett, 1976:ix). Taken this way, it could be argued that alarm calls and barks constitute the language of an animal and are interpreted by other animals of the same species. This constitutes the only point of comparison between Davidson's theory and animal theories that I can see, and it requires not only widening the definition of language, a move that Davidson might not agree with, but also looking at empirical evidence. Let us nonetheless look at this second strategy, for it has given up some surprising results and has added some interesting considerations to the issue.

The second strategy also offers two possible routes that can be taken. One is to see if certain species of animals can be taught to use some form of human language. The other entails examining the communication system of animals and comparing it with human language for common elements. Both of these routes are compatible with acknowledging that language is necessary to thought. Instead of taking the second route which entails trying to demonstrate that animals have a language of their own that should be construed as a language, some researchers have taken the first route and decided to instead teach chimps to communicate using human language. The work of Sue Savage-Rumbaugh (1998) is the most thorough example of this attempt. Generally speaking, success at the task of teaching human language to primates would indicate that human language is not a characteristic or capacity that is restricted to humans. The sharp divide between animals and humans that has been thus far claimed could then be thrown into question, since the original differentiating characteristic of human language would be a capacity also possible in animals.

The reason why this strategy has been criticized even by ethologists is because it entails applying an anthropocentric view of language, i.e., human language, and attempting to show it to be present in non-humans. Why should animals be shown to have human language rather than their own species-specific language? Of course, it makes perfect sense to take this strategy if one is intent upon answering Davidson's arguments directly. In this case, the strategy is to claim a victory on Davidson's own

territory by showing non-human animals to be capable of human speech. I think that the work of Savage-Rumbaugh and others, while it constitutes 'feeding into the hand of the enemy', can also be taken in a more foundational way to question the idea that humans and animals are different and the view that language is the defining trait that reinforces this divide, which is basically the view of Davidson. While I think that this is a good thing, this work also takes ethologists one step back because it excludes the path of determining whether species-specific systems of communication exist and then debate whether these systems should also be constituted as languages. Moreover, Davidson's view is still vindicated because he claims that thought ultimately depends on many other things, among them the successful interpretation of another's utterances. Savage-Rumbaugh has only so far shown that primates can communicate using ESL, a set of symbols that are communicated through hand movements, and lexical pictures. She cannot show that primates speak human language because the vocal cords of primates are not bent at the same angle that those of humans are. I am thus led to wonder whether Davidson would even accept that ESL constitutes a legitimate form of language.

One issue that has become central to primates being taught human language is exactly this question, whether communication using lexicon images or sign language constitutes a legitimate form of human language use. Skeptics (and I include Davidson in this camp) argue that such practices do not constitute legitimate forms of language use. They argue that in order to qualify as language, sign language and lexical pictures must contain all the syntactical structure that human language contains, including logical connectives. Some even go so far as to claim that the primates must be able to produce spontaneous novel sentence fragments, as children are observed to do in the language acquisition phase. Arguments of this type always take on the same form, deemed the strategy of 'upping the ante' by Talbot Taylor (1998). Skeptics argue that only if it can be shown that it is justified to attribute more or less all the communicational abilities that this or that theory of language attributes to an adult human can the primate be said to possess human language. We will see in later chapters that this type of argument is also used with regard to mathematical ability and concept attribution.

On the other side of this issue, there is also the question of what exactly is taught to the primate. The leap is not easily made from teaching ESL or lexical images to the

primate to the conclusion that human language has now been acquired by the primate. In looking at what exactly is taught via ESL or lexicon images, Bennett argues that it is a much impoverished version of human language, lacking in particular in the areas of sentence structure and vocabulary. Claiming that ESL and lexicon images are not representative of what is characteristic of human language, Bennett concludes that if one teaches an impoverished version of language to the primates, one can only expect such an impoverished version to be demonstrated by the primates (1988:203-4). This would explain why no primate has been observed to produce spontaneous novel sentences. It is because the primate has not been taught sentence structures to begin with. It can be concluded here that the primate certainly demonstrates what it has been taught, and leads one to wonder what would happen if all aspects of language use were taught to the primate.

Rather than try to teach human language to species of animals, some authors have instead chosen to take the second route and attempted to study and characterize the communicational systems of animals. One possible starting point is to devise a set of essential characteristics to human language and see if any of these elements occur in the communication systems of animals. It could be immediately objected that this is a non-starter since the elements of comparison are from a human language, and if animals possess a language, it is going to be an animal language and not a human language. Perhaps this animal language will not look anything like the human language. Nonetheless some authors have argued that it has been assumed in an almost *a priori* manner, without evidence, that the communication systems of animals share none of the elements of a human language.

Other authors argue that it's the methods of study that are limited, and not the system of animal communication under scrutiny. If we proceed to investigate the communication systems of animals with a human benchmark or anthropocentrically defined notion of language in mind, we will not find human language in animals. We will also forgo the possibility of finding something like language occurring in the animal, if it exists. It is entirely possible that the syntax/semantics of animal language is entirely unrecognizable to us, since it may have evolved along different pathways from an evolutionary point of view (Kiriazis & Slobodchikoff, 1997:367). Testing for language

in animals is thus rendered more difficult because researchers have this implicit anthropocentric bias for lexicons and syntax/semantics which, while these might be characteristic to human language, need not be present in the languages of other species.

Given the anthropocentric objection above, i.e., that animals will definitely not possess a human language if they possess any language at all, a more promising strategy has been to devise a set of what are known as 'equivalence relations' for the essential characteristics found in human languages that can then be looked for in animal languages (Schusterman & Gisiner, 1997). Devising equivalence relations is a form of bottom up processing, a way to circumvent the anthropocentric objection while still having a basis for comparison. The situation is as follows. In devising a set of essential characteristics for human language, the fact remains that these characteristics are of a human language. Assuming that the essential characteristics of an animal language will be animal in nature is a safe assumption to make, but the problem then arises that we might not recognize these characteristics as anything, because the basis for comparison is lost. Setting about the question from a human perspective is a form of top down processing in that it starts out with a set of human characteristics, but quickly runs into an anthropomorphic objection, that the characteristics of an animal language will probably not be human-like. So, in order to circumvent the charge of anthropocentrism, one starts from a set of equivalence relations.

It is in the context of a neutral setting, teaching an artificial language to dolphins, that Schusterman and Gisiner have been able to develop a set of descriptions that represent what an animal does when it learns a language in order to then compare this set with language acquisition by humans. From here, parallels between animals and humans will be matched up and a set of equivalent relations developed. For instance, it is found that dolphins have the ability to classify or categorize signs. This ability could parallel the human ability to recognize items as belonging in a category or even employ concepts. These two species-specific abilities would then comprise an equivalence relation. The abilities implicit in language learning could then be compared between humans and animals.

It appears that in taking the second strategy around Davidson's arguments and arguing that language is necessary to thought, most of the crux of the issue then lies in the

parameters of the definition of language. In defining the term, it is found that the issue of anthropocentrism arises, since the issue often boils down to what a human language would consist in. There is a high likelihood that animals will not possess something that is strictly human-like in nature. The notion of equivalence relations, in my opinion, pulls the issue out of the anthropomorphic-anthropocentric stand-off and has the benefit of allowing for comparisons to be made between humans and animals as well as allowing for the development of something along the lines of a theory of language acquisition in animals.

In light of a consideration of this second strategy which entails taking a look at recent advances made in animal language research (ALR), I think that two charges can be levied at Davidson, both albeit of an ad-hoc nature. One is that his views on language are anthropocentric in nature, and the other is that he can be viewed as falling back on the rhetorical strategy of 'upping the ante', given his anthropocentric view of language. Both these charges stem from the fact that Davidson has attempted to outline the case for human thought and its dependence on human language. Concerning the first charge, why should we think his system would apply to other species of creatures? Concerning the second, if it was found though equivalence relations that animals have a language that they use to communicate information to each other, why would these animals also have all these other human-like capacities such as the concept of a concept and explicit knowledge of truth and falsity?

I have entertained two different strategies with regard to rebutting the claim made by detractors of cognitive ethology and most thoroughly articulated by Davidson. The claim is that thought depends on language and particularly the successful interpretation of speech. Davidson's belief in the holistic nature of all of the various aspects of mentality leads to an implausible view concerning the acquisition in humans of all aspects of mentality, especially language. If he didn't hold such a holistic view of mentality, the door would be open for ethologists to justify their search for mental states in animals by comparing them with preverbal children or mute adults, claiming that these three groups have some but not all or a primitive version of fully-fledged adult human mentality. In not agreeing with the holistic nature of mentality, one can then disagree with Davidson's claim that thought depends on language.

The second strategy entails agreeing with Davidson's view that thought depends on language, but arguing that animals do have language construed in the wider sense of the term. In entertaining this strategy one gains a wider perspective on the whole issue and the possibility arises that perhaps Davidson's arguments hinge on language uniquely humanly construed. Animals perhaps have a language in the sense that they communicate information to each other, but it does not have all the elements or the same elements that human language has. Attempts at forging a comparison between the two systems could be successful to the extent that the elements are based on relations of equivalence. The notion of anthropocentrism as a hindrance to the study of animals will treated in detail again in chapter three on other minds. It turns out to have a major role to play in objections to the study of mental states in animals. At this point, I would conclude that the "no thought without language" argument, upon further examination, does not constitute a viable objection to the study of mental states in animals.

## Chapter Two
## Anthropomorphism

### 1. Introduction: Assumptions

Many critics of the project of cognitive ethology cite anthropomorphism as one of the main reasons why it is a mistake to attribute mental predicates to animals. In its general form, anthropomorphism is a tendency to attribute specifically human characteristics to non-humans. It is thus anthropomorphic, according to these critics, to attribute mental states to non-human animals. Historically, the term was used to characterize the attribution of human characteristics to gods. Anthropomorphism is an interesting phenomenon, partly because of its nature and the way it occurs. Anthropomorphizing is not in itself an error; it is merely a tendency or a practice. In order for anthropomorphism to count as an error a person must attempt to attribute human characteristics to non-humans, most often animals, that are in fact not applicable to that species. The phrase "are in fact not applicable to the species" is of utmost importance: much of the dispute hinges on this phrase. Two things must be noted here. One is that the charge of anthropomorphism must be levied in order to distinguish it from the mere practice or tendency. The attributer is then 'guilty' of the error of anthropomorphism. The second thing is that there must be a question as to whether or not these characteristics can in fact be attributed to the non-human. The charge of anthropomorphism is thus based on an assumption, the assumption that to attribute mental traits to animals is erroneous.

In this chapter I intend to examine the phenomenon of anthropomorphism in order to ultimately show that as a charge it holds no weight mainly because it begs the question, and that the practice of 'critical' anthropomorphism, as it is construed by Gordon Burghardt (1991), is legitimate and even necessary in formulating hypotheses for research in cognitive ethology. In between these two ideas there exists quite a bit of ground, having to do with proposed sources of the problem, solutions to the problem and explanations for the supposed inevitability of the practice of anthropomorphism. I will thus first examine the idea that various different strands of anthropomorphism exist, as

advanced by John Fisher, who offers a framework for distinguishing them (1996). I will also examine one of the various sources for the tendency that is cited by Hank Davis (1997). There have additionally been numerous solutions advanced to solve the problem as indicated by John Kennedy (1992). I will examine Pamela Asquith's arguments against adopting Kennedy's solutions, and her argument for the inevitability of anthropomorphism (1984, 1997).

In the literature, anthropomorphism is usually submitted by authors as a charge based on the error of categories, in a dismissive manner that is meant to put the issue to rest, and almost never backed up by supporting arguments. It often has the effect of leading the reader to dismiss the whole issue of mental states in animals and thus the whole project of cognitive ethology prematurely. As a prima facie argument anthropomorphism is taken seriously and usually prevents further discussion in the form of rebuttal by cognitive ethologists being charged with it. How does it have this effect?

**2. Category Error**

John Fisher thinks that anthropomorphism is not clearly defined and is the topic of so much confusion that it fails to make its point. It also fails as a viable critique of the project of cognitive ethology, at least when it is based on the underlying charge of category error. In citing the history of its usage, he characterizes it as a vacuous rhetorical weapon. It is vacuous because the charge of category error doesn't go through, as will be seen in a moment, and it is rhetorical because of the way it has been employed throughout history. He cites an example that serves to give the term anthropomorphism its rhetorical connotations. Ernst Cassirer used the term to describe outmoded forms of explanation that have been replaced by more modern ones, i.e., classical physics versus quantum physics (1996:3). The term has thus come to be related to forms of thought that are arcane, outmoded and quaint but false, and that must be overcome by new discoveries in science.

One of the most often cited sources of anthropomorphism according to Fisher is that it rests on a category error (1996:3). As Fisher notes, Gilbert Ryle originally gave us the notion of category error, which is the practice of treating an entity of one type as if it were an entity of another type. This notion applied to animals would entail that it is a

category error to apply characteristics that are restricted to humans, entities of one type, to animals, entities of another type. Here the underlying assumption is that humans are in fact entities of a different type than animals, otherwise the charge does not go through. Fisher claims that the underlying assumption has not been demonstrated and that the charge does not go through.

I am in agreement with Fisher on both the points made above. I think that both notions taken together are what give the charge of anthropomorphism its weight. Upon closer examination, we find out that the phenomenon upon which rests the charge has not been demonstrated, and that the charge itself rests more on reputation than actual factual error. Add to this the idea that anthropomorphism was initially associated with the attribution of human characteristics to god-like entities, which was considered a form of sin, and we now get that it is a sin by association to attribute human characteristics to animals, even though it has not been demonstrated yet that there is not some sharing of characteristics between the two species.

Fisher asks the question "Even if we were to find out, at some point in the distant future, that humans are of a different category than animals, does this mean that it is always a category error to attribute a human characteristic to an animal?" And is it also a category error to attribute animal characteristics to a human? (Fisher, 1996:4) He claims that it depends on what the set of human characteristics in question is, but that these characteristics must be definitively uniquely human in order for the charge of anthropomorphism to stick. For instance, it would be anthropomorphic to attribute human speech, in the form of words, to a dog, for they clearly cannot speak human language. There are not many researchers who would claim that dogs can speak in the way that humans can, however. It is not so clear, however, that some mental attributes such as thinking and reasoning cannot be attributed to animals, that these characteristics truly are uniquely human. Whether or not these predicates can or cannot be attributed to animals is an empirical question, according to Fisher, decidable only by empirical evidence. There is reason to think, in my opinion, that animals partake of some of the mental predicates characteristic of humans just as they partake of some of the physiological apparatus of humans, such as possessing similar senses and internal organs.

### 3. Varieties of Anthropomorphism

Fisher offers a framework for the varieties of anthropomorphism that exist, because part of the problem, according to him, is that there exist different forms of anthropomorphism, and authors either conflate several different forms or fail to take the different forms into account when making their charge. For instance, authors try to associate what I would call warranted forms of anthropomorphism such as hypotheses concerning attributions of mental states to animals with unwarranted forms such as the attribution of human speech to a dog. Another common occurrence is authors who use the term to make the charge and don't back up their claim with any argument, as if there is a single form of anthropomorphism that speaks for itself and nothing more needs to be uttered apart from the word.

There are two broad categories of anthropomorphism according to Fisher's framework, one of which can then be further bifurcated into two other types. The two broad categories are Imaginative and Interpretive. Imaginative anthropomorphism refers to the practice of representing animals in fiction or animated movies as similar to us (1996:6). This practice often shows up in children's movies, for instance, a talking dog. As mentioned above, this is not the type of attribution toward animals that researchers make claims about and attempt to find proof for. It is the conflation of this type of 'Disneyesque' obvious anthropomorphism with the interpretive type that seems to make the interpretive type less viable, through a sort of 'guilt by association'. Interpretive anthropomorphism concerns cases of inference from animal behavior to attributions of mental predicates, where these include descriptions of the animal's physical behavior in terms of intentional actions (1996:6). It should be noted here that this is the form most often represented in the literature on cognitive ethology. Any attribution of any mental predicate to an animal is considered anthropomorphism on this definition. To conflate interpretive anthropomorphism, a much debated topic in the literature, with the imaginative type, encountered in Disney movies, has the effect of diminishing the credibility of the interpretive type.

Fisher then divides interpretive anthropomorphism further into two other types, situational and categorical. Situational anthropomorphism involves the attribution of a

mental predicate to an animal which, while it might not apply to the animal in that particular situation or context, might apply in another situation (1996:6). For instance, attributing the mental state of anger to an ape that bares its teeth while sitting in the lap of Jane Goodall might not be appropriate to the situation, while attributing the state of happiness might. Here, the baring of teeth by an ape is ambiguous until it is contextualized, it occurs in aggressive displays, in the form of teeth baring, as well as in happiness displays, in the form of smiles. We might say, generally speaking that the situational type allows for attribution of mental predicates to animals, and it is the appropriateness of the situation or context that is at issue. Categorical anthropomorphism, on the other hand, is the label given to the attempted attribution of all mental predicates to all animals. This is the type of anthropomorphism that is based on a category error (1996:6). The idea behind it is that humans and animals are two separate categories. It is thus an anthropomorphic error to commit the category error of attempting to apply any or all characteristics that are of the human type to any animal. The problem, as mentioned above, is that it is not in fact an error to attribute the mental predicates to the animals until the empirical evidence has been carried out, because it is not yet an established fact that humans and animals are in fact separate categories.

Here one might wonder why Fisher has chosen to split interpretive anthropomorphism into two other types, rather than just contrast the interpretive type with the situational type. Interpretive and categorical anthropomorphism are identical, both are concerned to deny all proposed attributions of mental predicates to animals. There does not seem to be enough of a difference to warrant two distinct categories.

Categorical anthropomorphism is then further divided into two types, that of species and that of predicate type, to distinguish the conditions under which it is committed. Species type has to do with applying a certain predicate to the wrong species of animal. Predicate type has to do with applying the wrong type of predicate to a certain species of animal.

Fisher gives an example of a charge of anthropomorphism that involves both species and predicate type, that of Peter Carruthers in his article 'Brute Experience' (1989). Carruthers doesn't back up his charge with any supporting arguments or proof in the article. The quote in question is: "For only the most anthropomorphic of us is

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." (Carruthers, 1989:261). Here the so-called anthropomorphic error would be in attempting to attribute higher-order predicates such as beliefs about beliefs, that on some accounts apparently require an individual to possess the concept of belief, to a lower species on the food chain, such as a toad or a mouse. This is predicate type anthropomorphism, according to Fisher, because no one has yet claimed that toads have higher-order beliefs, although other types of lower-order predicates may be applicable to them. It is also species anthropomorphism, since no one has claimed that other lower-on-the-food-chain-animals such as mice have beliefs, while it might be claimed that higher-end animals, such as chimps, have such higher order beliefs.

Carruthers' quote makes two points clear, one is the flippant manner in which charges of anthropomorphism are made, and the other is why species and predicate type anthropomorphism are vacuous categories. First, the article in which this quote occurs is about the ethical treatment of animals, not about what mental predicates we can attribute to animals. His point is that we treat the intellectually-challenged sub-set of our population no less ethically than anyone else, in spite of the fact that they are intellectually challenged. We should thus treat all animals equally ethically, without regard for each species' degree of intellectual sophistication. The article is not about anthropomorphism *per se* and he doesn't qualify or offer any arguments to back up his claim relating to anthropomorphism. This is an example of the dismissive way in which remarks pertaining to anthropomorphism are made.

Second, Fisher's categories of species and predicate types can only be discussed in relation to each other, and not in isolation. Recalling Carruthers' statement as an example, Carruthers hesitates to apply a certain predicate X to two different species of animals. The problem is that one can never have an example of species type anthropomorphism without also having an example of predicate type or vice versa. Applying the mental predicate of belief to a toad is an error of both species and predicate type as much as applying the ability for verb conjugation is to the species of rhesus monkeys. Because one cannot have one type of error without the other, no new information is gained and the splitting of categories in the first place is superfluous.

I think that splitting categorical anthropomorphism into two further types is unnecessary. Categorical anthropomorphism is based on the error of categories, and as mentioned above, the charge does not go through because the empirical evidence on which it is based has not been carried out yet. In addition to this problem, there is the consideration that research in cognitive ethology hasn't got to the point of species specific mental predicate attribution, so Fisher's two further categories based on species and predicate are, in a sense, before the fact. One implication stemming from Fisher's speculative act of splitting a category of anthropomorphism into further types is indulgence in speculation on further types that are also ultimately based on a category error. If cognitive ethology was at the point where species specific attribution was the issue rather than whether any predicates at all can be attributed to animals, all the discussion about anthropomorphism as a category error would already be moot.

On the other hand, if cognitive ethology was at the point in the distant future where it was found that animals and humans shared traits, and it was agreed that mental predicates could be attributed to an animal, the issue of species and predicate attribution could potentially be a subject of debate. One would then be charged with anthropomorphism only if one attempted to attribute a mental predicate that was not attributable to a certain species (but was attributable to another species). This is not a viable path to take with regard to this issue, however. Although the charge of categorical anthropomorphism would no longer apply here, because the attribution of mental predicates would become sanctioned, the division into the two categories of species and predicate is nonetheless untenable, since one cannot involve one without involving the other. One cannot charge someone with attributing the wrong type of predicate to a particular species without also mentioning the other type of error because it acts as a comparison. Without the other category, the charge is made in a vacuum.

Rather than splitting categorical anthropomorphism into species and predicate type, Fisher could instead make a distinction between strong and weak categorical anthropomorphism. Strong anthropomorphism would correspond to the extreme position where it would always be a category error to attribute any human characteristic to any animal. Strong categorical anthropomorphism would imply that humans and animals are two distinct categories and that no properties are shared by the two categories. A less

extreme position, and the more popularly held of the two is weak categorical anthropomorphism, which holds that the attribution of certain, usually higher-order, mental predicates to some animals is a category error. It is futile to further divide the phenomenon of anthropomorphism into distinctions other than strong and weak as long as cognitive ethologists are still debating with opponents the question of whether it is possible to apply any mental predicates at all to any animal.

This distinction between strong and weak anthropomorphism, while it is an improvement on Fisher's complex schematic, is still not satisfactory in my opinion, for it assumes that Gilbert Ryle's notion of a category mistake truly applies in the case of animals and humans. That the charge of anthropomorphism is made in advance of the evidence concerning humans and animals as separate categories is problematic, in my opinion. What could be wrong with the research strategy of attempting to find out just what kind of mental life animals have, if any, and comparing this mental life with that of humans? It is often claimed that humans are animals. This claim has been no more shown to be true than the claim that humans and animals are separate categories, and so the issue is moot. If we give up trying to apply Ryle's notion of categories to this situation, what other arguments are left for the opponent to cognitive ethology, that aren't speculative? We should cease trying to find a distinction between humans and animals in this theoretical manner because it is a futile effort, and start to investigate empirically whether animals and humans share any mental traits.

Fisher also identifies an extreme position most often occupied in the literature for those who believe that any attribution of any mental predicate to animals is the error of anthropomorphism, that of the hard anthropocentrist (1997:7). The label of hard 'centrist, as he calls it, usually occurs in conjunction with anthropomorphism, and is usually charged by those who believe that animals share some of the mental life of humans, such as cognitive ethologists. As mentioned in the previous chapter, the term refers to a human centered perspective. It is thus a charge that is thrown back at the opponent as a sort of invitation to justify why we should hold a human centered view of the world. Fisher has identified Donald Davidson as a hard anthropocentrist. As we have seen in chapter one, Davidson is indeed a hard anthropocentrist who believes that any attempt to attribute any mental predicate to an animal is a form of categorical

anthropomorphism. He feels there is an unbridgeable gap between humans and animals, caused in part by language. Any attempt to attribute mental predicates, and particularly language, to animals is therefore a form of categorical anthropomorphism. In charging that Davidson's views are anthropocentric, the cognitive ethologist is shifting the burden of proof over to Davidson to justify why language and mental state possession should be restricted to humans.

The issue, in my opinion, centers on the applicability of a set of characteristics across species. The charge of anthropomorphism is meant to imply that mental traits are human, and that it is wrong to apply these traits across species. Further splitting anthropomorphism into different types based on traits or situations is a waste of time in my opinion, since these distinctions are based on indulgence of an assumption, that animals and humans are separate categories. What is lacking is a symmetrical term for the animal kingdom, that would mean it is a mistake to apply animal traits to humans. Lacking such a term at this point, I think we should use the term 'anthropocentrism'. The charge of anthropocentrism is meant to question the claim that the traits in question are uniquely human and seems to me to come closest to symmetrically opposing the charge of anthropomorphism, without implicating the assumption of category mistake. In my opinion, there should be two sides to the debate, that of the anthropomorphist versus the anthropocentrist. The issue would be the applicability of traits across species: which human traits can also be found in animals as well as which animal traits can also be found in humans.

**4. Sources: Affirming the Consequent**

The claim has been advanced that the practice of anthropomorphism stems from the result of a classical logical error known as affirming the consequent. Hank Davis is the only proponent of this view, to my knowledge. His characterization of anthropomorphism is that it is a form of intellectual laziness, resulting from a failure to make proper species differentiations (Davis, 1997:336). In Fisher's terminology, he would be said to espouse a form of categorical anthropomorphism as well, since he believes that humans and animals are separate categories. Moreover, according to his view, the boundary separating humans from animals is not defined sharply enough, and

this is a result of intellectual negligence in making species differentiations. He believes that the impact of anthropomorphism in the literature is disastrous, resulting in not only the inference of higher mental properties in animals, but also in inferences about the function of these entities, which he argues are baseless to begin with.

The source of the tendency to anthropomorphize, he argues, is that its nature is rooted in a classical logical error called 'affirming the consequent'. This is where one lists an antecedent and then a consequent in the form of a syllogism, and by affirming the consequent, one thereby fallaciously proves the antecedent. He gives an example for the existence of God:

A) is taken to be 'If there were a god'

B) is taken to be 'The world would be a beautiful place'

C) The world is a beautiful place, therefore there is a god (1997:336).

The affirmation of the consequent, that the world is a beautiful place, thereby proves the antecedent, that there is a God. His point is that this is a fallacy, and can be shown by the fact that there could be many reasons why the world is a beautiful place, the fact that there is a god is only one of many possible causes. There is a problem with the form of the fallacious argument as he has laid it out however, which prevents his argument from being successfully made. According to Hurley's definition of the error of affirming the consequent, it consists of one conditional premise, a second premise that asserts the consequent of the conditional and a conclusion that asserts the antecedent (Hurley, 1982:323). Davis' first premise is not a conditional. In order for it to be a conditional, it would have to be of the following form:

A) If there were a God, then the world would be a beautiful place.

B) The world is a beautiful place.

C) Therefore, there is a God.

This observation of mine is a technical point. It is merely the tip of the iceberg concerning Davis' argument, however. Not only is he off to a bad start in making his argument, but this is not the only problem with his argument.

The example he gives relating to anthropomorphism is the following.

A) If I think

B) Then I scratch my head

I am scratching my head

Therefore I must be thinking (1997:336).

Davis claims that this is the form of fallacious reasoning behind anthropomorphism and the form typical of claims pertaining to mental states in animals made by cognitive ethologists. As the argument stands, it is impossible to make Davis' point. Again in this case it is first necessary to rewrite the premises into the true form of the argument for the fallacy of affirming the consequent:

A) If I think then I scratch my head

B) I am scratching my head.

Therefore I am thinking.

Davis is attempting to draw an analogy between this fallacy argument form and claims made in cognitive ethology. The general argument form is the following. A capacity for a mental state is conditional upon a certain piece of behavior in the antecedent and the piece of behavior is then claimed in the consequent, thereby proving the antecedent, the existence of the mental state. Davis goes on to claim that there is nothing inherently wrong with premise pairs such as the one above, it is the illogical use to which they are put that contaminates our conclusions regarding the mental life of animals (1997:336). For Davis, the problem with these claims is that other causes that are non-mental in nature have not been ruled out as an explanation for the cause of the behavior evidenced by the animal. For instance, various other antecedents, such as dandruff, cannot be ruled out as causes of scratching one's head. We will see in chapter four on methodology that this strategy of arguing that various other antecedents besides mental states cannot be ruled out as causes of a particular behavior, is a tactic most often used to argue that experimental results do not support the hypotheses of the cognitive ethologists.

Davis states in an endnote (p. 347) that affirming the consequent may actually be a useful adjunct to scientific discovery. He sees a difference, however, in the syllogism put to legitimate use, presumably in science, and illegitimate use, presumably in cognitive ethology. He believes that the difference lies in the extent to which other possible antecedents have been ruled out, for the logical syllogism itself is blind as to whether adequate steps have been taken so that the antecedent we are considering stands alone among causal possibilities (1997:327). Applied to the above, the question then is:

is there a causal relation between scratching one's head and thinking? This first has to be established before we can test the argument in a syllogistic form, according to Davis' view.

There are many shortcomings with the argument advanced by Davis. There is first the problem that the form of the syllogism that he has advanced does not even correspond to the fallacious syllogism of affirming the consequent. The problem with this syllogism as Davis has listed it concerns the order of the antecedent and consequent. Normally, in philosophical discussions, the usual practice is to reverse the order in which Davis has listed the antecedent and consequent. In this case the order would then be the following: If I scratch my head, then I think. Usually a behavioral indicator is taken to indicate a higher mental process, and not the other way around. This modification makes the syllogism a correct form of modus ponens, and invalidates the conclusion he originally wished to make about anthropomorphism.

This does not get at the main problem with Davis' whole argument, however. The fact that there is a prima facie problem with the form of the syllogism is a superficial critique. The real problem with the argument is that his claim that this type of reasoning is used in experimentation isn't even accurate. Experimental techniques in the social sciences do not normally employ syllogistic reasoning. That is, researchers normally does not use syllogisms of the form 'modus ponens' to argue for experimental results. In this case, that the behavior of scratching one's head is an indicator of a propositional attitude, such as thinking, in the individual. The usual procedure in experimental disciplines with regard to inferring intervening variables such as mental states is inference to the best explanation, which takes the form of hypothesis testing.

Davis takes the work of Donald Griffin, grandfather of the discipline of cognitive ethology, as an example of the illogical processes implicit in mental state attribution in animals. He quotes Griffin from his 1984 book 'Animal Thinking': "Animals make so many sensible decisions concerning their activities….. that it has become reasonable to infer some degree of conscious thinking"(1984:3-4). Davis misrepresents Griffin's remarks into a syllogistic argument form, neglects to put the first premise into the form of a conditional, and again reverses the order of the premises:

A) If I engage in conscious thought

B) Then I will behave in a sensible manner (1997:338).

He lists no conclusion, so we have to presume what it is and that the argument really looks like the following:

A) If I engage in conscious thought then I will behave in a sensible manner.

B) I am behaving in a sensible manner

Therefore I must be engaging in conscious thought.

Davis states that the problem lies not with the premises themselves, but in the claim that conscious thought is the sole basis for sensible behavior. What happens if one reverses the order of premises to represent the way Griffin has originally stated it? The argument is no longer an example of the logical error of affirming the antecedent. It becomes a valid form of modus ponens. What it comes down to in Davis' opinion, however, is whether a causal link has been established between sensible behavior and conscious thought. This has nothing to do with the correctness of Griffin's claim put into the form of a syllogism.

A few things must be stated here. First, the examples chosen by Davis are ridiculous. No one has attempted to prove a causal relation between scratching one's head and thinking, in humans or in animals. Perhaps Davis has chosen these examples deliberately to amplify his point. The effect it has is to discredit legitimate attempts to find out what mental predicates are applicable to animals and which are not. He chooses far-fetched examples to discuss and hopes that the legitimate examples will become far-fetched by association. A better example to discuss might be the following: a researcher hides an apple in the sand in the courtyard of a windowed compound containing numerous chimpanzees. He is watched by a single female chimp, who sees him 'hide' the apple in the sand. Later the chimps are let out into the courtyard, and the female who witnessed the burying of the apple waits for the other chimps to disperse before digging up the apple and eating it by herself. Here is the argument, with the antecedent in the proper conditional form and the antecedent and consequent within themselves in proper order:

A) If the chimp waits until the other chimps disperse before digging up the apple, then she is selfish.

B) She waits until the other chimps have dispersed before digging up the apple,

Therefore she must be selfish.

Although this anecdote can be made into a valid form of modus ponens, it does not constitute one of the accepted research methods for disciplines such as cognitive ethology or even experimental psychology. Modus ponens syllogisms, although an accepted form of logical argumentation, are not part of the tools of research in either of these two disciplines concerned with animal experimentation.

Second, although Davis mentions in passing the great advances in our understanding of bats' echolocation done by Griffin, he mentions none of the conclusions of Griffin related to this work. I'll venture to bet that none of Griffin's conclusions were advanced using syllogistic reasoning. Instead of examining the conclusions, Davis mentions an excerpt from an introduction to one of Griffin's many books where Griffin is commenting in a general way on the convergence of results that have been obtained recently in cognitive ethology. Griffin did not put the excerpt mentioned above into the form of a syllogism, and the way Davis has laid it out, it looks like the argument for the contents of the book. The premises in the so-called argument have also been reversed. Griffin's conclusion in that paragraph, not included in Davis' extrapolation, is that communicative behavior among various species of animals might offer an opportunity for ethologists to 'listen in on' (obtain access to) and gather data about the nature of animal consciousness. The only claim that Griffin is putting forth in this context is that perhaps observing the communication patterns of various species can be seen as one method among many offering an opportunity to understand what degree of awareness these species possess.

Davis' overall point in the article is twofold. He claims first that the idea that conscious thought is the sole basis for sensible behavior is an unwarranted assumption to make even about human behavior and two, that it only gets worse when attempting to apply it to animals (1997:338). His objection is that we do not know enough about the role of conscious thought in determining human behavior to even attempt to extrapolate to any other species. He concludes that we must first achieve sufficient understanding of the mental processes in humans that we then wish to extrapolate to animals. Thus it may be said that one criterion for determining conscious thought in animals would be to work

out a theory of consciousness in humans, otherwise extrapolation to animals rests on shaky grounds, in Davis' opinion.

Davis questions the role of thought even in determining human behavior, claiming that it is still the subject of considerable debate. Folk psychology, according to Davis, is responsible for offering credibility to anthropomorphism, because it offers a possible mechanism that humans use for predicting and explaining the behavior of each other. However, it is used only by the folk, it has not been experimentally proven, and it should therefore be discredited, according to him (1997:334).

Even if folk psychology were to be proven as the mechanism underlying human behavior prediction and explanation, the practice of anthropomorphism rests on an additional error. The practice of extrapolating specific traits from humans to animals rests on the questionable assumption of continuity between the species. This continuity, according to Davis, has not been demonstrated. We must be convinced that there is a continuum of mental life that includes both humans and animals that we wish to extrapolate from in order to sanction the attempt at extrapolation. Thus a second criterion for determining conscious thought in animals would be to have settled the continuity issue, on Davis' account.

A few things can be said about this two-fold conclusion made by Davis at the end of his article. In a general way, neither of the two criteria he lists should stop cognitive ethologists from continuing their studies with a view toward determining what mental predicates we can attribute to animals. The first criteria that Davis thinks should be achieved is an age-old problem in philosophy. The problem of defining and plotting mental processes in humans has its own branch in the discipline known as 'philosophy of mind' and there are a number of good theories out there, none of which is sufficiently robust to withstand criticism. This should not prevent attempts at constructing theories of animal mentality just because such theories presuppose an answer to the human question. We cannot put off interest in things that presuppose an answer to the issue of mental processes in humans until a viable theory of mental processes is found and accepted, for the search could go on indefinitely. There are entire branches of knowledge devoted to this study, psychology is one of them, and it does not put other research stemming from this notion on hold because a definitive answer has not been found to this question.

Davis' second criterion, that of settling the issue of Darwin's continuity hypothesis, is also an age old question that pits two opposing hypotheses about the relation between humans and animals against each other, neither of which are easily proven. One of these is the continuity hypothesis, expanded on by numerous authors, among others Charles Darwin. From an evolutionary perspective, it is difficult to plot animals that are already extinct along a continuum and show that all species are related, or at least that man is an animal. If this were to be proved, then the issue of categorical anthropomorphism would no longer be considered an objection. The opposing theory, call it the discontinuity theory, that may lend credibility to animals and humans being separate categories, relies largely on the idea that language is innate and is responsible for the break in the continuum between animals and humans. We have not resolved this debate by any means, and are not in danger of solving it anytime soon. It should not prevent ethologists from pursuing their studies. Additionally, the discontinuity theory by itself, even if proven, does not resolve the question adequately either. The question is whether there are any mental properties among the properties of the animal category, not necessarily whether any human category properties can be carried across the category and attributed to animals. The charge of anthropomorphism, when it is based on a category error, assumes a resolution to the question in favor of discontinuity and is thus a case of begging the question.

**5. Solutions: Metaphor and Analogy**

John Kennedy (1992) has been identified as a dissenter of sorts in the anthropomorphism debate. He is an ethologist but thinks that the threat of the anthropomorphic error ought to cause change in the way that ethologists report their data. One solution, according to him, is that ethologists avoid making attributions to animals that could be considered anthropomorphic. This puts him in the middle, conciliatory position, for he is an ethologist, but he wants to avoid the practice entirely and so prescribes avoiding using the language that causes one to commit it. He has been criticized by other ethologists for espousing this avoidance attitude.

Kennedy believes that the tendency to anthropomorphize is built into us. It has been pre-programmed into our hereditary make-up by natural selection (1992:5).

Characteristic to his view is that the tendency to anthropomorphize is largely an unconscious process. This is a convenient claim to make, for it guarantees that we cannot avoid the tendency precisely because it is unconscious, and therefore outside of conscious control. To claim it as an unconscious process also eliminates the need to show that the tendency occurs in the first place, and also how the tendency arises. In my opinion, this claim is conveniently indisputable because of its reference to the unconscious.

An analogy can be drawn between Davidson's view in the previous chapter and Kennedy's view. As we saw at the end of the last chapter, Davidson thinks that there are two vocabularies, one for describing mindless things and a mentalistic or intentionalistic one for describing purposeful or reflective human behavior. He thinks that attempts to apply the mentalistic vocabulary to animals is unwarranted, but that there is no other vocabulary available to describe the creatures that fall 'in between' the two vocabularies. As I interpret him, Kennedy is in agreement with Davidson. Both are seeking to delimit an in-between vocabulary that would be of a higher level than mindless and yet lower level than full-blown mentality and purpose. A middle level vocabulary such as the one they seek does not as yet exist. Davidson is resigned to the fact that we will continue to use the mentalistic vocabulary, while hopefully keeping in mind that we do not really mean such terms in their full sense when we use them to describe animals. One interpretation of this idea is that such vocabulary can be used as long as it is understood that such terms are meant metaphorically and not literally.

Given that such a middle level vocabulary does not exist, Kennedy has three prescriptions on offer to remedy the problem of unconscious anthropomorphism, the first of which resembles Davidson's (1992:162-5). It is that if ethologists insist on using the mentalistic vocabulary as opposed to the mindless one, they should explicitly state that the mentalistic terms being used to describe the animals are being used in a metaphorical sense. The second prescription is an extension of the first, it is to translate the metaphorical terms used back into their literal counterparts. There is the assumption here that the metaphorical terms are equivalent to subjective terms and that literal terms are equivalent to objective terms. The third prescription follows from the idea of avoiding mentalistic terminology altogether and it is to use only objective terms when describing

animals. Each of these prescriptions will be taken up in turn. None of them, in my opinion, constitute good solutions to the problem.

Pamela Asquith has taken issue with the use of metaphor in anthropomorphizing (1997). She thinks it is an inaccurate use of the term 'metaphor' that is being used to justify anthropomorphic attributions. Her conclusion is that the terms used in anthropomorphic attributions to animals are actually meant in their literal sense and not their metaphorical sense. If she can make the case that attributions to animals are based on an inaccurate use of the term metaphor, then Kennedy's first and second suggestions will no longer work. The link between metaphor and anthropomorphism is such that showing the role of metaphor in anthropomorphism to be inaccurate will disqualify Kennedy's first two prescriptions.

Metaphor is defined as the application of a description to a term that is not literally applicable. An examination of the distinction between metaphor and literal meaning makes clear that the term is misapplied in making anthropomorphic attributions. Pamela Asquith offers a pertinent comment on the distinction between the two forms of meaning, arguing that the distinction between literal and metaphorical meaning is fundamental to any discussion of metaphor (Asquith, 1997:31). Here I quote her at length:

> "Two distinctions can be drawn between literal and metaphorical meanings. First, literal meaning is that which is agreed upon by speakers with a common language. The meaning can only be judged right or wrong with reference to common or accepted usage. Literal meaning can, therefore, change. By contrast, a metaphorical word cannot be corrected by referring to proper usage-it can only be criticized as inappropriate or inept. A metaphor will only be appropriate if the meaning of the word used metaphorically can somehow be associated with at least some of the literal meanings of the word used in a metaphorical way, or with other words in the sentence. Second, metaphorical meaning is parasitic on the literal-that is, the force of the metaphor is derived partly from the literal meaning of the word, but no literal meaning is derived from the metaphorical." (1997:31).

The two things to keep in mind are first the idea that metaphor is never correct or incorrect, it can only be criticized as being inept or inappropriate. The second thing to note is that the metaphorical is parasitical on the literal, but that the literal is not parasitical on the metaphorical. We can start with an example of a metaphor applied to

humans in order to see how calling anthropomorphism a form of metaphor is a misapplication of the term. Here is one from Shakespeare's Hamlet: 'When sorrows come, they come not in single spies, but in battalions.' The reader is being asked here to link the notion of sorrow with a contrast between how spies act and how battalions act. The difference here between spies and battalions, and hence in sorrows, is that single spies are stealth-like and strike without warning, whereas battalions strike in an obvious and large manner (Honderich, 1995:555). Applied to sorrow, Shakespeare wishes to draw the reader's attention to the metaphorical image that sorrow often comes in large doses and is often obvious. The point is that this is a metaphor, sorrow does not in fact act like a spy or a battalion, it is an emotion, not a human spy or a group of men. Sorrow has its own characteristics, it is just more interesting to try and find parallels with other terms that one wouldn't normally associate with sorrow, in order to bring out a richer connotation of the term.

Taking a typical example of a metaphor as applied to animals, a cat turns its back on a veterinarian after having received a hypodermic injection and is described by the following statement: 'The cat is indignant'. Here one is invited to apply a description of how a human would react whose feelings were hurt due to a subjective feeling of being treated unjustly and apply this description to the cat. A cat does not really get indignant in the form of hurt feelings, so the argument goes, it is just a metaphorical way to describe the cat's behavior that neatly illustrates what the person offering the metaphor intends. According to the proper use of metaphor, it is a metaphor to describe the cat as being indignant, and not the real case. But what constitutes the real case here? The literal meaning implied here is absent. In the case of sorrow, sorrow has its own set of literal characteristics, one employs a metaphor to bring out other potentially interesting and certainly flowery characteristics. In the case of the cat, what is the literal description that the attribution of indignity is parasitic on?

Asquith gets at this issue by asking the pertinent question " What then is inferred when an author maintains that mentalistic terms to describe animal behavior are being used metaphorically and not literally?" (1997:32). Remembering that the metaphorical is parasitical on the literal, the literal meanings must be those used to identify human characteristics. She asks "are not the metaphorical terms being meant in the same

(human) way when applied in the animal case?" (1997:32). I think that the answer is affirmative, the terms are being meant in the human way when applied to animals. Thus the terms are being used, not in a metaphorical sense, but in a literal sense, when applied to the animals. In the case of the cat, the term 'indignant' is not a flowery substitute for a more concrete, objective, less flowery description, for there is none. Asquith is thus right to say that anthropomorphism cannot be seen as a valid use of a form of metaphor, because the so-called metaphorical terms used are actually intended literally, and the literal counterpart is often missing.

It will be remembered that Kennedy's second suggestion is that we should avoid the practice of anthropomorphism altogether by translating metaphorical or subjective terms, such as 'searching,' back into their objective counterparts, such as 'scanning' (1992:162). Kennedy justifies this switching back of terms by the idea that translating back into the objective description will bring to light many of the features that the subjective term misses or glosses over. This is the exact same argument that cognitive ethologists offer as justification for their use of intentional terms in attributions to animals, only in reverse. Ethologists prefer to use purposeful or intentional vocabularies precisely because the objective purposeless vocabulary masks or leaves out certain necessary aspects of behavior, such as the fact that it is purposeful. This extra ingredient of purpose is exactly what Kennedy and others believe is unwarranted in the description of an animal's behavior. Kennedy's third suggestion is thus to remove the intention from the description and instead add in some long-winded mechanical description for which the intentional term is often a shorthand. Kennedy's suggestion seems to cancel itself out and leave us back where we started.

**6. Inevitability**

Pamela Asquith (1984) offers an argument for the inevitability of the practice of anthropomorphism and then argues for its utility in primate studies. The inevitability arises through the practice of using ordinary language terms to describe animals in scientific settings such as journal articles and textbooks, because there is no other mid-level language available. The utility of anthropomorphism is found in the claim that too much valuable information would be lost if we were to stop using the terms. As we saw

above, it is the intentional or purposive element of the description that is at issue. Its removal constitutes the valuable information that would be lost, according to Asquith.

One can see how anthropomorphism might arise in the scientific literature, if researchers are using ordinary language, usually reserved for humans, to describe the behavior of animals. There is a purposive element to the language, taken for granted or assumed in humans, that according to Kennedy and others, cannot be assumed in animals.

To illustrate one way in which anthropomorphism arises from the use of ordinary language terminology, Asquith borrows Purton's distinction between A-purposive and O-purposive terminology used in descriptions of animal behavior (1984:143-5). In A-purposive, the A stands for agent, who is aware of what behavior he needs to display in order to attain the goal, whereas in O-purposive, the O stands for organic, which refers to the functional explanation for the behavior in question. O-purposive descriptions of behavior are given in non-intentional terms, whereas A-purposive descriptions imply purpose and conscious awareness. Drawing an analogy with Davidson's reference to the two levels of language in chapter 1, O-purposive descriptions correspond to the language we use to describe mindless beings, and A-purposive descriptions correspond to mentalistic vocabulary he would like to reserve for humans. Asquith argues that anthropomorphism arises due to the use of ordinary language terms that imply agency or A-purposiveness that are then attributed to the behavior of animals. Again we seem to find ourselves in Davidson's predicament where we are looking for a mid-level language that would apply to animals but that, for whatever reason, does not yet exist.

Asquith also demonstrates how anthropomorphism occurs in the process of data collection (1984:145-9). Borrowing a distinction used in data recording, between behavioral units and behavioral categories, she illustrates how purposive terminology is consciously removed and how it creeps back in at a later stage in the process. Behavioral units are single bits of data, such as discrete movement patterns and vocalizations, that then get grouped into behavioral categories according to pre-set criteria (1984:145). In a data-recording situation, such as a group of researchers observing a group of primates in a facility, these researchers would describe the behavior of the primates by breaking it up into behavioral units. The behavioral units grouped together to form a grouping for aggression, for instance would include things such as spitting or yelling at someone,

hitting or grabbing someone, menacing facial expression, or even smaller data bits that we have no name for. There is no name for the bits because they are partial descriptions, they have no intention, hence the term 'bits'. These bits will get a name once they get grouped together again into behavioral categories. They also have no name so that the highest degree of objectivity can be achieved in the process of data recording. An example of a behavioral category often used in the observation of humans is that of anger or aggression. The exact point at which anthropomorphism arises is in the grouping of the behavioral units into behavior categories (1984:146). This is also the point at which purpose or intention creeps back into the descriptions.

The process, at least with regard to the gathering and writing up of data, is thus the following: The generic form of anthropomorphism, that attributes a general purposiveness to behavior, arises when researchers employ ordinary language in the discussions of their observations of animal behavior. In order to mitigate the tendency in behavioral observation, the behaviors are broken down into small bits that have to do with physical movement. Anthropomorphism enters in when these bits of movement are collated back into behavioral categories which are in the form of ordinary language in order to be intelligible for discussion.

The problem is the same in both cases, and it stems from the fact that there is only one vocabulary of terms available, the same that we happen to use for humans. This vocabulary contains a purposive or intentional element to it that detractors believe is unwarranted in its application to animals. On the other hand, it is not that ethologists insist on using these terms despite the fact that these terms have proven to be too sophisticated to describe animals, and that there is another vocabulary available to use instead. There is no other vocabulary available. Moreover, if Davidson is right in claiming that there exist only two vocabularies, one for mindless objects and one for mind-full creatures, then ethologists are justified, in a certain sense, in choosing the mind-full vocabulary over the mind-less one to apply in hypotheses regarding mental state attribution in animals. Moreover, it is not as if ethologists assume without testing that animals have all the capacities that humans enjoy. The aim of cognitive ethology is to find out which, if any, of the capacities that humans have can also be applied to animals.

## 7. Conclusion: Utility

Against the argument that ordinary language terms with purposeful connotations should not be used in scientific reports about animal behavior, Asquith argues that this terminology is the best one to use because it gives the clearest, most understandable presentation of animal behavior (1984:165). It allows for connections to be made between behaviors that descriptions in behavioral units do not. It also allows for general theories to be formulated about higher order behavior, i.e., on a social level, or community organizational level (1984:166). This is because both these phenomena, connections between behaviors and social behavior, require the element of purpose in order to be made sense of.

As it turns out, there is one construal of anthropomorphism that constitutes an accepted practice, but only in its restricted form. Kennedy calls it 'mock anthropomorphism' to distinguish from all the other construals. It is taken to refer to pretending, for argument's sake, that an animal can think or feel as we do. It is sanctioned as a legitimate practice because of the heuristic value it has, in particular for hypotheses that can be generated about the function of the animal's behavior.

It is in the mention of hypotheses about function that the constraints of the method of mock anthropomorphism emerge. Ethologists are interested in two aspects of explanation with regard to an animal's behavior, its function, which corresponds to ultimate cause, and its intention, which corresponds to proximate cause. Mock anthropomorphism is the practice of assuming that animals can think and feel as we do for the sake of generating hypotheses related only to the function of an animal's behavior. All other forms of anthropomorphism, because they are concerned to seek the proximate causes of behavior, related to mechanism or intention, are to be avoided. Predictions based on mock anthropomorphism, according to Kennedy, are no more than hypotheses that need to be tested. The problem with mock anthropomorphism is the constraint restricting study to the function of behavior only. Cognitive ethologists are interested in the proximate causes of behavior that relate to purpose or intention. This practice of mock anthropomorphism is unhelpful in that it still does not sanction the seeking of such explanations.

There exist other legitimate forms of anthropomorphism construed as a practice that do not contain the above constraint of restriction to ultimate cause. Other versions include 'critical' anthropomorphism or 'pragmatic' anthropomorphism (Burghardt, 1991, Silverman, 1997). All such construals have the common element of legitimacy, when used for heuristic purposes only, to generate hypotheses about animal behavior. The task remains, according to Kennedy, of how to discriminate between the legitimate form and the illegitimate form, for the onus is now on the author making the claims to state how the mental state or intentional term is meant. As Kennedy notes, unless the author explicitly states that he or she means to anthropomorphize in either a mock or genuine manner, we can't tell the difference. I don't agree with Kennedy on this point, and in my opinion his view leads to an inconsistency on this point. On his view, it will be recalled, the tendency to anthropomorphize is unconscious. It is thus impossible for the author to consciously state whether the attribution is a case of genuine or mock anthropomorphism. Moreover, the practice is sanctioned in the case of generating hypotheses, there is thus no need to state whether it is mock or genuine. In hypothesis generation no conclusions are made, so there is no fear of being guilty of anthropomorphism since it is hypotheses that are being generated and not conclusions that are being advanced.

**8. Conclusion**

Upon closer examination of the term anthropomorphism, it would appear that making a distinction between the practice of anthropomorphism and the charge of anthropomorphism is the crux of the issue. I hope I have demonstrated that construed as a charge, it does not hold much weight, and boils down to a case of question-begging. When the charge is based on a category error, anthropomorphism is actually an empirical question that has not yet been answered. It is one of the goals of cognitive ethology to answer this very empirical question. The various sources of the tendency to anthropomorphize that have been advanced in the chapter range from logical error, to unconscious compulsion, to a problem of missing vocabulary. The claim of logical error misses the mark, and as it turns out logical syllogisms aren't even employed in scientific hypothesis testing. The claim for the phenomenon of unconscious compulsion turns out to be circular and therefore vacuous. We are again left with Davidson's dilemma of a

missing mid-level vocabulary. The fact remains however, that anthropomorphism construed as a practice for hypotheses generation is at the least a legitimate heuristic tool that can be employed while we wait for the mid-level vocabulary to be conceived of.

## Chapter Three
## The Problem of Other Minds

### 1. Introduction: The Vanishing Subjective Point of View

The third challenge I will be looking at in this first half of the thesis is the problem of other minds. It is not advanced as a challenge in and of itself to the project of cognitive ethology. That is, no one has seriously claimed that cognitive ethology cannot pursue its investigation into the possibility of mental states in animals because animals are other minds, and scientific investigation into other minds is impossible. While it is not a challenge in and of itself, it is worthwhile to investigate because it underlies other issues. The case could be made that the other minds problem involves or underlies issues visited in the two previous chapters, that is, the issues of language and anthropomorphism. Although we can infer mental states in other humans based on behavior and verbal reports, the verbal report avenue is blocked in the case of animals. The argument by analogy is made more difficult in the case of animals, because one is basing one's analogy on something that is absent in animals: verbal reports and language. It might thus be said that the other species of mind problem underlies the language problem in animals.

This other species of mind problem is also what underlies the complaint that attribution of mental states to non-humans is anthropomorphism. Allen and Bekoff (1997:52) are of this opinion. They define anthropomorphism as the interpretation of what is not human in terms of human characteristics. As seen in chapter 2, underlying the charge of anthropomorphism is the assumption that humans and animals are two different categories. The fact that animals lack language can also be construed as a reason to believe that animal minds are a different species than human minds, and to attribute the same kinds of mental states to them would thus be anthropomorphic, on this reasoning. The charge applies only to other animals since attributing mental states to other humans cannot be considered anthropomorphism by definition.

In this chapter I will be examining the problem of other minds as it relates to two issues. The first is the idea that we cannot have direct access to the mental states of

animals. The second, slightly different, is that by extension we cannot have direct access to the subjective experience of animals.

The general problem of other minds is whether and if so how one can know or be justified in believing that other individuals have thoughts or feelings (Honderich, 1995:637). Introspection provides one an access point for one's own mind, so there is no apparent problem there. We don't have the same direct introspective access to other humans' minds however, and so justification for the minds of others must proceed by argument based on inference from analogy.

The problem of other minds regained the spotlight in the literature in 1974 because of a paper written by Thomas Nagel entitled "What is it Like to be a Bat?" Nagel was inspired to write this paper after attending a talk given by Donald Griffin at Rockefeller University on bat echolocation. In order to make his talks more vivid, Griffin used to set loose a bat or two in the lecture hall after he was finished speaking. The paper that Nagel was inspired to write was about the dichotomy between the subjective point of view of the individual and the objective point of view given by science. He used the example of a bat to illustrate the irreducibility of the subjective point of view to the objective point of view contained in scientific theory. Two points from Nagel's paper became topics of discussion in the years following it. One discussion was taken up by scientists concerning the absence of the subjective point of view in scientific theory [for instance Erwin Schrödinger "Mind and Matter" (1958)]. The other point found its way into discussions of animal behavior, that we can never know what it is like to be a bat or other animal, depending on how different our physiologies are, and so we should not bother with this type of investigation. Griffin, credited with starting the discipline of cognitive ethology, was in turn inspired by Nagel's paper to write a book about the subjective experience of bats and other animals, titled "The Question of Animal Awareness: The Evolutionary Continuity of Mental Experience" (1976). Griffin's work was not well received initially, and became the target of many critiques concerned with cognitive ethology as a questionable endeavor.

The problem of other minds is an age-old problem, and I will not pretend to solve it in this chapter. My discussion of it stems from the general fact that it underlies some of the challenges made to cognitive ethology discussed in the other chapters. More

specifically in this chapter, the problem of other minds bears on the search for a subjective point of view in both animals and humans. What shall be done with this problem, which is more of a puzzle or paradox, according to some authors? There are a number of possible reactions to it witnessed in the literature over the years; I'll discuss two of them in this chapter. One reaction is to reject Nagel's original question 'what is it like to be a certain animal?', and maintain that there is no significant answer to the question. Daniel Dennett (1991, 1998) takes this tactic, and offers two 'solutions' to the question that I suspect are more than a little tongue in cheek, although both are currently implemented in the practice of cognitive ethology. Kathleen Akins (1996) also takes this tactic, arguing that the subjective point of view is poorly delineated, based on intuition, and that we should not be too worried about capturing it since we have no clear idea what we are even looking for. The second reaction is to embrace the dichotomy, and acknowledge that perhaps there is an answer to the question. Pursuing the answer to this question is a route taken by many cognitive ethologists and has resulted in the creation of a new area of study for cognitive ethology, the study of subjective or private experience, which I will evaluate in the last section of the chapter. It involves acknowledging that the subjective point of view is not reducible to or even capturable by objective theory, but that it should still be examined because of its link with cognitive capacities. Another implication of studying the subjective experience of animals entails a recognition and subsequent theoretical distancing by cognitive ethologists from what has been up to now anthropocentric perspective, that uses humans as a benchmark or point of comparison, toward a 'bottom-up' approach to studying animals, known as "theromorphism". This approach means taking each species of animal on its own terms and attempting to discover how it represents the environment in addition to what capabilities it might have (Timberlake, 2002:105). In light of the implications this relatively new strategy has, my aim in the chapter is to clarify what this perspective entails and evaluate some of the objections raised concerning whether this strategy is implementable on a practical level. Given the lack of a shared system of communication between humans and animals, is it possible to translate this perspective into a research strategy? If it is found not to be possible, should research in cognitive ethology be given up?

**2. The Problem Posed**

I'll start the discussion with Allen and Bekoff's portrayal of the problem (1997). Their portrayal is interesting because it makes a distinction between the other mind's problem as applied to other humans and as applied to animals. Some authors who make a distinction in the other mind's problem as applied to animals and humans probably do so because they believe, like Davidson, that animals and humans are two distinct categories. I do not include authors such as Allen and Bekoff in the above category. As they mention, the issue, on some views, with the other minds problem is that while we humans might have access to our own mental states, we do not enjoy the same access to the mental states of other human beings. It is safe to say that we do not have direct access to the content of the mental states of others (Allen & Bekoff, 1997:53). Allen and Bekoff note the fact that psychologists often shelve the problem of other minds, even though it is no less of a problem in their field. Behavioral scientists, while admitting that knowledge of other human minds is possible, regard the mental states of other animals as forever closed to us. Allen and Bekoff call this the 'other species of mind' problem, to separate it from the other human minds problem.

Allen and Bekoff offer two forms of the general argument for the claim that we can never have scientific knowledge of other minds, one for other humans and one for animals. Here is the argument in premise form for humans:

1. Mental phenomena are private phenomena.
2. Private phenomena cannot be studied scientifically.

Thus mental phenomena cannot be studied scientifically (1997:53).

This argument depends on the view, according to them, that mental phenomena are private. There are numerous interpretations of the term 'private' and the soundness and thus success of the argument really hinges on this term. Allen and Bekoff choose to interpret 'private' in the sense of 'not directly sensible by others'. Taken in this sense, quarks are also private phenomena. Quarks are nonetheless studied scientifically by inference to the best explanation, which is a selection of the most plausible hypothesis among the competing alternatives for the explanation of observable phenomena. On Allen and Bekoff's view, mental states can be studied in the same manner as quarks. There is only sense of private that would invalidate the using of inference to the best

explanation as a method. That is the sense where the 'privacy of the mental state' means that the mental state has no effects whatsoever beyond the individual possessing the state (i.e., behaviorally or otherwise). Allen and Bekoff argue that even though we are not yet sure in what sense mental states have effects, it is pretty much agreed that mental states have effects on behavior at least. Mental states, because they have visible effects, can thus be studied using inference to the best explanation. Thus, Allen and Bekoff argue, the first premise of the argument is probably untrue if 'private' is taken to mean 'has no effects whatsoever beyond the individual possessing the state', yet the second premise is only true if 'private' is taken to mean just that. Either way, they conclude, the argument is unsound. Since this argument says nothing about animals specifically, it must be taken to include animals as well as humans.

The argument can also be further restricted in order to apply to animals. Here is the version for animals.

1. Mental phenomena are private phenomena.
2. Private phenomena cannot be studied scientifically in non-human animals.

Thus mental phenomena cannot be studied scientifically in non-human animals (1997:540).

This argument makes explicit reference to animals, and while someone, say a behavioral scientist for instance, might accept that we can infer the presence of mental states in humans via language, he may not accept that we do the same for animals. That is, we can't ask an animal what mental state it is in and it cannot answer us. Another common source for the behavioral scientist's view is that in the absence of language use by animals, their behavior is not discriminating enough to allow for the attribution of mental states. The fact that animals themselves also do not use a reasonably humanlike language amongst themselves means that their behavior is not discriminating enough for us to attribute mental states to them.

Allen and Bekoff don't really offer a critique of this argument. They choose not to get into a discussion about whether or not animals do indeed possess some sort of language by citing the latest evidence that some primates have successfully been taught human language. They instead take the tactic of concluding that it is up to cognitive ethologists to break down the notion of mentality, identify its various aspects and show

how each of these aspects is amenable to scientific investigation. This is a good tactic. However, Allen and Bekoff could have taken the same tactic as they did with the first argument, namely to claim that just as progress in cognitive science has contradicted the argument against investigation into human mental states, so will progress in cognitive ethology contradict the argument against investigating animal mentality.

The case could also be made that a separate argument for animals is not even necessary. The wording of the two arguments is exactly the same, save for the additional phrase 'in non-human animals'. The mere act of tacking on the phrase 'in non-human animals' to the premises does not constitute an extra ingredient that makes it different from the human case and that by fiat adds more weight to the argument. Moreover, a separate argument for animals implies that it has already been established that the two species are distinct categories. The behavioral scientist who thinks that scientific investigation is possible for humans but not for animals is likely basing his or her view on a type of distinction between humans and animals that has not yet been demonstrated, such as the category error argument. As we saw in the discussion on anthropomorphism in chapter two, the category error has not yet been demonstrated. Alternatively, the behavioral scientist could offer the lack of language in animals as the reason why investigation into the mental states of animals is not possible. This will not work for two reasons. First, it would have to be shown that animals do not have language. Second, it would have to be shown that animals couldn't be taught to use human language. In fact, if one removes the avenue of language, since some have argued that introspection and verbal reporting in humans are notoriously inaccurate method of accessing the content of mental states, then humans and animals are on equal footing with regard to the other minds problem.

**3. Nagel's Bat**

It would be useful to look at the original question as it was posed by Thomas Nagel (1974). The title of the article and hence the question is: what is it like to be a bat, or what might the phenomenology of a bat be like? The larger theoretical issues that Nagel also discusses in this article are the distinction between the objective point of view of science and the subjective first-person point of view, and how reductionist

explanations cannot reduce the subjective to the objective without losing some aspect of the subjective point of view. The example of a bat is brought in to exemplify the problem of the inaccessibility of the subjective point of view. A further difficulty involved is to construct an objective theory of the subjective point of view of the animal. The issues discussed by Nagel apply to animals as well as to humans. It should be noted that Nagel's essay is not, strictly speaking, about the problem of other minds, nor is it about the other species of mind problem. Rather the problem of other minds underlies the issue discussed by him, that of the inaccessibility of the subjective point of view.

This is how Nagel poses the problem (1974:166-168). He claims that conscious experience is a widespread phenomenon. It occurs at many levels although we cannot be sure of its presence in the simplest organisms. It is very difficult to provide evidence of it. Some extremists deny it in mammals other than man. No doubt it appears in countless other forms unimaginable to us. Nagel then derives the subjective point of view from consciousness. He claims that the fact that an organism has conscious experience at all means that there is something it is like to be that organism. He calls this 'what it is like' phenomenon the subjective character of experience. He then claims that this subjective character of experience is not adequately captured by any present theory. A reductive analysis of the mental doesn't capture it because it is also logically compatible with the absence of it. Explanatory systems of functional or intentional states don't capture it either, for they also are applicable to automata. Explanations in terms of the causal role of experience also fail for the same reason. He then makes an interesting point that both Dennett and Akins, in their critiques, will emphasize. The point is that without some idea of what the subjective character of experience is, we cannot even know what is required of the theory that is supposed to account for it.

He then states that every subjective phenomena is connected to a single point of view, and that any objective theory will abandon that point of view. There is an argument of sorts that leads up to this claim (1974:167). The first statement is that the subjective character of experience appears to be the most difficult phenomenon to explain out of all the things that a physicalist theory must explain. One cannot exclude this phenomenon as one might a by-product of a chemical reaction, that is, by explaining it as an effect on the mind of a human observer. The phenomenon must be given a physical

account. It would seem that this is impossible because every subjective phenomenon is connected with a single point of view, and it seems inevitable that an objective physical theory will abandon that point of view.

Nagel admits that facts about 'what it is like to be an X' are very peculiar, and this peculiarity makes some doubt their reality or claims made about them. His strategy is thus to point out the relation between the subjective and the objective with a view to illustrating the importance of the subjective point of view. To help him make the point he uses the example of a bat. We assume bats have experience, and he has chosen bats rather than some other animal lower down on the evolutionary scale precisely because we would agree that bats have experience, while there might be some question as to whether a slug has experience. He has also chosen bats because their sensory apparatus is so different from ours, he believes that the problem is made exceptionally vivid by this difference.

The argument is as follows (168-9): The essence of the belief that bats have experience is that there is something it is like to be a bat. Bat sonar is not similar to any of our senses, so there is no reason to suppose it is anything like anything we can experience or imagine. In trying to imagine what it might be like, one is restricted to the resources of one's own mind and these are inadequate to the task. For instance, one could imagine oneself with some of the physical transformations that are prominent in the bat's unique experience, such as webbed wings, the apparatus for sonar etc. This will not work, Nagel argues, because there is an unbridgeable difference between a human's experience of behaving like a bat, even complete with some of the transformations that enter into the experience of a bat, and what it is like for a bat to be a bat. The conclusion is that such an understanding may be permanently denied to us by the limits of our nature.

Nagel makes one last noteworthy point that is taken up by ethologists who evince the reaction of embracing the dichotomy and advocate researching the animal's subjective point of view. It is the hypothesis that there might exist humanly inaccessible facts. He puts forward the idea that there could exist facts that could not ever be represented or comprehended by human beings, even if the species lasted forever, simply because our structure does not permit us to operate with concepts of the requisite type

(1974:171). Attempts to reflect on the bat's subjective point of view might fall within this category of facts.

I am of the opinion that Nagel's essay is the most eloquent attempt to illustrate the plight of the subjective point of view with regard to its apparent disappearance in objective theories and to advance the idea that there exists a set of facts that are beyond our human comprehension. However, his arguments, or lack thereof, are not beyond critique. One could argue, for instance, that while it is true that the objective theories that we construct fail to represent a particular subjective point of view, it is unclear that objective theories should endeavor to include the subjective point of view in the first place. He mentions that we could never know what it is like for a bat to be a bat. Is there any significant difference between what it is like for a human to be a bat and what it is like for a bat to be a bat? If there is, would knowing what it is like for a bat to be a bat have any bearing on humans? Dennett and Hofstader interpret Nagel as seeking "...a distillation of that which is common to the experiences of all bats, not the set of experiences of some particular bat." (1981:407). In some sense, what Nagel is after is not a personal point of view of a particular bat, and so we cannot immediately rule out that the objective theory will fail to capture it just because it is a single personal point of view. In Hofstader's view, Nagel is interested not in 'what is it like for me to be X?' but rather 'what is it like to be X?' In other words, Nagel wants to know objectively what it is subjectively like to be an X (1981:409).

The same critique can be applied to the notion of facts beyond human comprehension. How can we even prove that there are facts beyond human comprehension if these facts are, by definition, beyond our comprehension? If knowing what it is like for a bat to be a bat falls into this category of facts, what good does that do us? Indeed the human mind can contemplate the unknowable, but what exactly is the point of the exercise? It doesn't advance our knowledge, or the issue, any further to contemplate such ideas.

Indeed it is true that Nagel has discovered a fact, that the subjective point of view is not found in the objective theories that are constructed from it. I wonder, however, if there is anything significant in the subjective point of view that is therefore missing from the objective theory that is constructed from it. In other words, yes indeed there is a

missing element, but is that missing element of any significant value? In other words, is it possible to know objectively what it is subjectively like to be an X, and is this any different from just knowing objectively what it is like to be an X? This is one of Dennett's main points. He asks "Do we have any reason to believe that there is anything interesting or theoretically important that is inaccessible to us?" (1991:442).

Nagel's conclusions could be taken to imply a constraint on the field of cognitive ethology. That is, in animals as in humans, his conclusions imply that we do not have any access to the subjective experience of others. Investigation into the cognitive capacities of animals is thus possible, but only if the avenue through which the information is gained is not the subjective point of view. Rather than having the effect of preventing ethologists from carrying out their investigations, or even giving pause to the project, Nagel's article has actually been inspirational for people working in the field. Nagel's conclusions don't provide any *prima facie* compelling reason to give up the project of cognitive ethology. After all, questions of phenomenology and subjectivity should present no obstacle to the search for mental states or other cognitive capacities in animals. In fact however, other cognitive capacities probably occur via subjective awareness, or at least in the same vicinity as subjective awareness, and so his idea that there exists a subjective point of view has been applied to animals and has opened up a new research area. This will become obvious in my discussion of the second reaction to the issue.

**4. Solution One: Reject**

Daniel Dennett offers a sustained critique of the problem as Nagel has posed it (1991, 1998). Dennett doesn't in fact believe that there is a problem, but he is willing to entertain Nagel's worries and has two rather comical answers to them, that have in fact already been implemented in research in cognitive ethology. He first notes that the question Nagel posed in his article along with his ensuing response, that the situation is hopeless, has had the curious effect of discouraging subsequent researchers from asking and answering such questions (1991:441). This isn't altogether true, as evidenced by the ongoing debate on qualia, inspired partly by Nagel's essay. In one aspect, however, it is a tension or a puzzle and a swift dissolution of it has not been forthcoming. When a

puzzle is truly a puzzle, a swift dissolution of it is not forthcoming, and it is obvious from Dennett's remarks that he is not comfortable with the idea of an insoluble puzzle. Dennett does note, however, that Nagel didn't really put forward a set of arguments for a conclusion, but rather assumed a conclusion and discussed its implications.

Dennett deems Nagel's strategy to be one of rhetoric, employing one-sided use of evidence (1998:339). One such example, Dennett advances, is the fact that Nagel chose bats as an example, and the fact that he took the trouble to relate a few fascinating facts about this species. Dennett believes that Nagel chose to relate a select few facts about bats for two reasons. One, because these facts support our convictions that bats are conscious, and two, because they support Nagel's conviction that bat consciousness is very much unlike ours. The rhetorical peculiarity of this strategy is the idea that if a few facts can establish the two above contentions, can't a few more facts solve the puzzle? Dennett asks: what kind of fact is it that only works for one side of the empirical equation? This is a good question and I think that Dennett has honed in on one source of the puzzle. A good puzzle, and I think this is a real puzzle, gets much of its intrigue, just like a good joke, from its setup and delivery. The few selective facts about bats that Nagel chooses to use along with the successful establishment of the existence of a subjective point of view serve to set up the puzzle quite well. Dennett is right to then ask, what would a few more facts establish about bat phenomenology? I think he has in mind a list of the commonalities we share with bats, or at least a list of what we do know about bats, since this would also serve to give us knowledge about what it might be like to be a bat. Or, as Nagel admits himself, if he had chosen an animal that shares a sensory structure more similar to ours, the puzzle would not be as vividly reproduced. I wonder if Dennett suspects that if Nagel had supplied different facts about a different species such as a primate, then the puzzle would not be as much of a puzzle. In other words, if Nagel had chosen an animal with a sensory apparatus and physiology more similar to ours, the contemplation of what it might be like to be that animal might not be as far of an empathetic stretch. If Nagel had chosen bricks as an example of something it might be like, there would be no puzzle. Nagel would say that this is because bricks have no experience, but can we be sure that bats have experience?

In asking what a few more facts might establish, Dennett is almost implying that if a few more facts were established, there would be no puzzle at all. This is further established by Dennett's comment to the effect that the reader should beware of being charmed, that emotions and feelings are too easily provoked for them to have any use but to unnecessarily sway the reader into being convinced that there really is a problem.

I don't think that a few more facts would dissolve the puzzle in this particular case, where the goal is to know what it is like to be a certain species of animal. Even if the question were to ask about what it might be like to be a primate, where our genetic makeup is very similar to primates, it would still be quite a stretch to imagine what it might be like to be that animal. This is because it is not merely a matter of knowledge in the form of physiological and neurological facts. Knowing all there is to the species' physiology and neurological makeup will not by fiat give us knowledge of the animal's phenomenological point of view.

To take a mundane example, let us imagine what it might be like to be a dog. One small aspect that goes to make up the animal's point of view is sensory intake that is a function of its spatial location in the environment. A crucial part of the subjective point of view must be supplied by the dog's sensory apparatus, which is fed in large part by its acute auditory and sensory capabilities and the fact that it is a four legged creature. In short, the world doesn't look the same nor feel the same for a dog given its senses and position in space as it does for us humans, and this difference, in part, goes to make up its subjective point of view. In addition to Dennett's facts, other aspects, such as the animal's location in space and the acuity of its senses must also be included to make up the totality of the subjective point of view, in my opinion.

Dennett serves up his own example, a particularly vivid example, to demonstrate the intuitive card in Nagel's strategy (1998:341-3). The example is in the same question form as Nagel's, except that it asks what the smell of a rotting carcass (the vulture's staple food) might be like to a vulture. No amount of third person investigation, Dennett informs us, could ever tell us what the smell of carrion might be like to a vulture. This contention, he adds, is not asserted on the basis of argument, but is an intuitive card that is played. The problem here, he thinks, is the coupling of assertion that there is awareness in the animal, along with no attempt to investigate what this assertion of

consciousness might amount to. This assertion takes the form of the crude conviction "We know what we're talking about even if we can't explain it yet." This conviction underlies the claim that animals have awareness as well as the claim for the existence of the subjective character of experience.

It is here that Dennett offers his first solution to the problem (1998:344). The problem is posed as an issue about how to replace the uncertain and vague assertions about consciousness in animals. His solution is to devise a theory of human consciousness and then to determine which human features, such as memory, problem solving etc., would apply to which animals, if any. Assuming we have a human theory of consciousness on hand, the idea is to then divide this theory up into the various sub-capacities that together make up the phenomenon of consciousness, and finally investigate whether these capacities exist in animals. It's obvious he doesn't believe that this strategy will work, because he goes on to then challenge the notion that consciousness of the human type i.e., an overseer or an inner eye, exists in animals.

Dennett's concluding argument against Nagel's puzzle makes reference to his theory of consciousness developed at length in "Consciousness Explained" (1991). In a nutshell, he argues that only humans have a characteristic that is necessary to consciousness, what he calls 'informational organization' in its most complete form. This characteristic is not innate but rather an artifact of our immersion in human culture. In order to be conscious, or for there to be something it is like to be a creature, he argues, an organism must have that informational organization that includes the power of reflection and re-representation. These characteristics are not automatically present with sentience. Other species may have somewhat similar organizations, but the differences between humans and them are so great that analogies do not make sense. What must be added onto the mere responsivity and mere discrimination present in most species of animals is this further characteristic of informational organization. It is this characteristic of informational organization that gives humans a 'user illusion', "the illusion that there is a place in our brains where the show goes on, toward which all perceptual input streams and whence flow all cognitive intentions to act and speak" (1991:346).

He then challenges the idea that there might be an inner eye of consciousness also present in animals, acting as the overseer of the 'what it is like' feeling. He gives an

example of a snake and asks if we can talk about what the snake itself has access to or just about what its various parts have access to. Nagel's 'what it is like' question, when applied to the snake's parts instead of the snake itself, no longer makes sense to ask. Animals, lacking the characteristic of informational organization, probably also lack the concomitant 'user illusion', and so it makes no sense to ask if there is anything it is like to be that animal.

Dennett's overall position on the matter is the following: he thinks that the idea that there is a dividing line between those animals where 'there is something it is like to be that animal' and mere automata is an artifact of our traditional suppositions. He doesn't believe that consciousness is an all or nothing phenomenon: "there is no principled way of distinguishing when or if the mythic light bulb of consciousness is turned on (and shone on this or that item)." (1991:349). He further claims that if it is this light bulb theory of consciousness that participants in the debate on animal consciousness are carrying around, then the mystery will be maintained.

Dennett thinks that although consciousness is not necessarily an all or nothing phenomenon, the characteristic of informational organization is. Although consciousness has various grades or degrees, informational organization does not, it is present in its complete form in the animal, or not. Animals who are sentient have the capacity for discrimination and responsivity, but not necessarily the characteristic of informational organization. This characteristic, which includes the capacity for re-representation and reflection, is presumably responsible for the "what it is like" part of the phenomenon and is missing in most species of animals.

Dennett seems to be arguing that without the necessary characteristic of informational organization, there is nothing it is like to be a certain creature. Since most species of animals do not have this characteristic, we are forced to the conclusion that there is nothing it is like to be a particular animal.

Dennett takes a slightly different tactic concerning the very same question of 'what it is like to be an X' entertained in a chapter of his book, Consciousness Explained (1991:443-8). Here he puts forth his second solution, that of constructing heterophenomenological narratives for species of animals. In this critique of Nagel Dennett chooses to again question the assumption that there is some leftover

unexplainable feature to consciousness, the 'what it is like' phenomenological 'feel'. He thinks that rather than trying to turn our minds temporarily or permanently into bat minds (a literal interpretation of Nagel's 'what is it like' question), we should instead concentrate on what we do know about bat phenomenology. Heterophenomenological narratives could then be constructed from the knowledge that we do have. These would just be neurophysiological and ecological stories about the animal in question. Dennett takes care to note here that in recommending that we treat bats and other species for interpretation in the same way we do humans, he is not shifting the burden of proof, but merely extending the human burden of proof to other entities. It's obvious by this remark that Dennett is trying to avoid being labeled anthropomorphic yet he ends up in the anthropocentric camp instead despite his efforts. Extending the human point of view into examining the capacities of animals ensures that these capacities will be seen from a human point of view instead of in their own right. Some capacities are bound to be overlooked because they are not sufficiently akin to the human point of view.

These two solutions, to develop a human theory of consciousness and apply it to animals and developing animal heterophenomenologies, are the same solution. They both recommend attacking the problem in terms of what we do know about the animal in question. Both solutions, regardless of how seriously Dennett takes them to be, are already being implemented in cognitive ethology. Above and beyond this, Dennett does not believe that there is anything left over in the manner of 'what it is like' to be the animal in question.

While I do believe that Dennett has aptly characterized Nagel's arguments and shown them to be more intuitive than forceful, I am not convinced by Dennett's conclusion that a series of facts about the animal's neurophysiology will exhaustively capture the full extent of the animal's point of view. In devising a human theory of consciousness and determining which features apply to non-humans, or extending the human burden of proof to animals, there is the potential problem that one will end up in the anthropocentric point of view, thereby masking the actual capabilities the animal does have, and further distancing one from discovering the animal's point of view. Dennett's second solution of constructing heterophenomenologies about the species is a better tactic since it obtains information from a variety of sources and there is less chance that the

human benchmark will interfere. However, there is the possibility that neurophysiological and ecological stories about the animal in question are not the only elements that go into constructing the phenomenology of the animal.

Kathleen Akins (1996) offers a detailed analysis of Nagel's problem, except without a free subscription to her own theory of consciousness, because she doesn't have one and so subscribes to Dennett's. She thus comes to much the same conclusion as Dennett, that the inner eye of consciousness applied to animals is a mistake. Her focus is on the idea that objective theories in science will omit the subjective point of view.

To begin the discussion, Akins is of the same opinion as Dennett, that Nagel's conclusion in his article, the irreducibility of the subjective point of view, has been accepted without much further debate or question. She sees Nagel's article as having outlined the limits of scientific explanation. On her view there is a dichotomy between scientific explanation and phenomenal experience. Phenomenal experience is necessarily an experience from a particular point of view; hence the facts of experience are essentially subjective in nature. Contrast this with the kinds of phenomena that science seeks to explain, which are objective in nature, or viewer independent. Any appeal to scientific facts in explaining the alien point of view will only further distance us from the very property we seek to explain (1996:346).

Akins, like Dennett, believes that intuition is the culprit behind acceptance of Nagel's conclusions. The intuition is that science will necessarily omit the one essential element of phenomenal experience, its intuitive 'feel'. She believes that this negative intuition is grounded in our everyday experiences, manifested in such ways as trying to describe the experience of the pain of a migraine headache. If one extends this difficulty to the phenomenal experience of an alien creature, that shares almost nothing of our sensory and physiological apparatus, the difficulty in imparting the point of view or experience becomes almost insuperable. Her argument is the following: If we can comprehend only those sensations we have experienced, and our own sensations are very unlike the alien creature in question, we will be unable to understand the alien creature's phenomenology. If one then extends these considerations to the efficacy of science, one wonders in what way science could possibly bridge the difficulty and offer us an answer. On the other hand, scientific explanation is not completely irrelevant to the understanding

of an alien creature's experience, at least to the extent to which that experience is based on neurophysiology. So what, she asks, is given and what is not by science? (1996:348).

She employs a thought experiment, where she has come from the future, which is at the point of the end of neuroscience. In other words, all there is to know about animal and human neurophysiology is known. She has two 3 dimensional colour movies in her possession. These two movies are shown to an audience in a theater, the first is of a human hang-glider's experience as he or she flies through the air, with a camera placed on the forehead of the hang-glider like a surgeon's light, capturing the visual scene of the hang-glider's perspective. The second is of the bat's experience as it flies through the air in a dark room, catching mealworms thrown up into the air by an experimenter. The audience watches both movies, and while it is easy for the audience to visually process the human glider's point of view, it is nearly impossible for them to make any sense of the bat film. The human's point of view is easily simulated for two reasons: 1) because the sensory system is the same, and 2) because a human from the audience can artificially simulate the hang glider's visual input. It would be as if you or I were up in the air with a camera placed on our forehead, we would see the exact same scene, in the exact same way, because our visual systems are identical. The bat's point of view, on the other hand, is nearly impossible to simulate, because none of the human senses can simulate echolocation in the bat. Visually, the movie looks to a human like a disorganized mix of colored patches. As Aikens mentions, the bat's auditory experiences have been cued with a visual kaleidoscope of color patches on the screen. So, for instance, color hues are encoded as frequencies of sound waves, brightness is represented as volume or intensity of sound, and configuration of the color patches represents the spatial properties of the sound waves. Not having the same system of echolocation as the bat, and not having a translator in the brain that can immediately translate the visual analogue of echolocation into something that a human can process, the audience cannot make sense of the film. The bat's visual system, which is very impoverished, is far outstripped when compared with the acuity of the visual system of humans. On the other hand, echolocation, which the bat uses to navigate through the air, is completely alien to any of our human capabilities.

What the film demonstrates, on Akin's view, is that even if we were able to successfully simulate in ourselves the 'feel' of the bat's experience with the help of the film, we still would not understand the bat's point of view. The 'feel' of the bat's experience is the qualitative aspect of it and the point of view aspect is the representational aspect. In order to understand the bat's experience in all of its phenomenological splendor, we would need to have access to both the representational and the qualitative parts of the bat's experience. Lacking the auditory representational capacities of the bat, namely echolocation, we do not experience the colored patches on the screen in the same way as the bat does.

Akins is of the opinion that it might not even make sense to ask if we could separate the representational from the qualitative aspects in our conscious experience. Applied to the bat example, this would entail a two-step process. We would first have to strip away the entire representational content of the experience. Then we would have to overlay it with the qualitative content of the bat's representations. Akins submits that we have no idea how to tease these two aspects of experience apart, nor how to put them back together, because our intuitions do not provide a concrete distinction between the qualitative and representational aspects of perception.

With regard to Nagel's main claim, Akins is of the same opinion as Dennett, that it is intuition that is responsible for this nagging feeling that something is left out in the transition from subjective point of view to objective point of view. According to Akins, the problem starts when we intuitively and mistakenly construe understanding the point of view of the bat as analogous to the everyday problem of understanding the phenomenal experiences of each other, such as what a migraine headache feels like. These experiences are generally ineffable, and we mistakenly think it is the qualia of the experiences, their 'feel' that are inaccessible. In other words, according to Akins' view, we treat conscious experience as if it were merely a bunch of qualia.

There are two overlooked points in this intuitive and mistaken construal of the problem, according to Akins. First, one cannot distill qualia from conscious experience. In other words, just because we can sometimes isolate and talk about a particular phenomenal experience, such as a particular shade of red or the taste of a fruit, this does not mean that that qualia exists in vacuo, or that it is possible to isolate phenomenal

experience from the representational content of conscious experience. Second, a point of view is not merely a collection of qualia.

Akins offers a postscript having to do with considerations of the light bulb theory of consciousness on the subjective point of view. Taking up the bat example again, Akins claims that watching the bat film makes us realize that given the special task allocated to the bat's auditory system, as a kind of compensation for its poor visual system, we can say that the bat's experience is different from ours. In imagining how the bat's experience differs from our own, we immediately adopt a hypothesis that incorporates our own visual system into the experience (1996:356). In doing this, we end up in the anthropocentric position. Perhaps we do the same with consciousness and assume that the bat also has an overseer of what it is like, that is successively attending to different mental events. Just because this seems to be the way it happens with us does not mean this is what happens in the case of animals. She concludes that it is possible that this is not what truly occurs in the animal. Perhaps even our own experience is only retrospectively like this. In other words, perhaps it is only in retrospect that we are under the impression that events occur in succession instead of all at once and that an overseer is present, taking note of these events.

Her conclusion is the following. Nagel's original claim is that we can never understand the point of view of an alien creature. That is, we can never know the phenomenal experience of a bat, which is not transmittable by description and which one cannot have without similar personal experience. But if introspection does not yield any distinction between the representational and qualitative parts of experience, we have no idea what we are looking for in the first place, and we certainly cannot therefore say what science has left out of the explanation. Given that we are not sure what the subjective view looks like and we have no reason to believe that it necessarily exists in animals, we cannot begin to construct theories about it. Without a good reason to stop empirical investigation in animals with regard to cognitive capacities, we should continue it. Research into other areas of cognition should nonetheless continue.

This is good practical advice. No one has yet solved this puzzle, and so it is conceivable that research into cognitive ethology could go on for the next thirty years and

we would still not have a solution to the puzzle. After all, the problem of a subjective point of view has no direct bearing on research into cognitive ethology. Or does it?

I am of the opinion that subjective experience has a major role to play in the study of cognitive capacities in animals. I would first like to question Akins' contention that one can only comprehend those sensations that one has experienced, since the force of the argument rests on the truth of this claim. It cannot be the case that one can only comprehend those sensations that have been personally experienced. If it were, it would constitute a pretty severe constraint on the range of foreign experiences a person could contemplate. The term 'vicarious experience', taken to mean that which is experienced imaginatively through another person, would have no meaning if we could only imagine those sensations we have personally experienced. The act of empathy would also be vacuous.

I would also question Aikin's claim that we would not comprehend the bat film, in part because we cannot separate the representational from qualitative aspects of experience. I do not think that we can consciously and immediately perform the dissection and at any rate there is no reason to need to be able to do so. If conditions were manipulated, however, I think it is possible that humans could eventually learn to adapt to processing the colored patches on the screen as sounds. For instance, in Kohler's famous experiment where subjects wore inverted image glasses that made the subjects see everything upside down, the subjects were eventually able through adaptation to ride bicycles in traffic and ski down hills (Dennett, 1991:393).

Second, I think that the subjective point of view problem more than has a direct bearing on research into cognitive ethology, it opens up a whole new research area, based on the mode of empathy as a bridge to the animal's point of view. As mentioned above, there is a great probability that many of the other cognitive capacities will either work in conjunction with subjective awareness or at least in the vicinity of it. Moreover, the mere idea that an animal might have a point of view, regardless of whether it is borne out empirically, at least raises the issue that humans have been thus far conducting research into animals with a human benchmark in place. One of the conclusions that can be drawn from Aiken's work is that she has identified one way in which the anthropocentric view arises. Her claim is that the human adopts a hypothesis of how the bat navigates by

incorporating his or her own visual system into the experience. It is trying to imagine how the bat's experience is different from ours that we incorporate our own visual system into imagining the experience. As we will see in the next section, one of the issues is whether or not this implicit incorporation of our own human experience can be dropped once it is recognized to occur.

**5. Reaction Two: Embrace**

The second strategy, taken by an increasing number of cognitive ethologists, is to embrace the subjective-objective dichotomy and advocate examining the subject's point of view, in this case, the animal's subjective view of the world. The idea that animals might have a point of view originates from the renowned ethologist, Jakob Von Uexkull (Rivas & Burghardt, 2002:10). He coined the German term 'umwelt' to represent the animal's sensory and perceptual world. Before considering the theoretical viewpoint of taking the animal's perspective, I first want to discuss the precursor to this new view, developed by Gordon Burghardt among others, that of anthropomorphism by omission.

Burghardt has identified a further type of anthropomorphism called anthropomorphism by omission. It is defined as a tendency to commit anthropomorphism, to attribute human-like qualities to animals, because of a failure to consider that animals have a different sensory world than ours (2002:10). Burghardt claims that it occurs often in the literature on cognitive ethology, most often where the underlying theoretical perspective of the researcher entails comparing animals to humans, or using human standards as the benchmark. Akin's pointing out of the tendency in humans to incorporate their own visual system into the experience of bats would constitute a good example of one of the possible mechanisms leading to anthropomorphism by omission.

Examples of anthropomorphism by omission abound in the literature, according to Burghardt, and even he and his colleagues have not been immune to it. I will discuss two theoretical examples. The first comes from Colin Beer, who, in his discussion on anthropomorphism, puts forth a point of view reminiscent of Davidson's views in chapter one. It is a good example of the implications that arise when one is speaking from an anthropocentric point of view. Beer makes the claim that "the reach and complexity of

connections attaching to ideas in the human case will usually far exceed what is conceivable for any animal." (Beer, 1997:203). Taking an example of a cat watching a mouse escape down a hole in the floor, he compares a human description of thoughts on the event to the cat's. He states "Only a small part of the network within which mouseness is nested for us extends into the cat's world." (Beer, 1997:203). In other words, from the human's perspective, the cat only has a fraction of the thought network that a human has. This is the same type of argument made by Davidson with regard to the impossibility of determining the exact contents of the beliefs of Malcolm's dog, from a human perspective. The mistake here is aptly summed up by Millikan: "To attempt to express the contents of the cognitions of animals by translating these or correlating them with English sentences would not be accurate." (Millikan, 1997:196). To further conclude from this that the mouse therefore has no network or only a very small network of thoughts is unwarranted. Burghardt notes that it is unfortunate that Beer neglects to consider that the cat has a different worldview from us humans. If he were to consider it, he would find many phenomenological aspects that are absent in our worldview but present in the mouse's, such as those arising from the mouse's sensitive sense of smell and hearing. The case can thus be made that the animal's worldview far exceeds that of the human, at least with respect to the auditory and visual capacities (Rivas & Burghardt, 2002:13).

The same consideration extends to the issue of human and animal language. When the issue arises, it is often stated that animals do not possess a language, at least not in the humanlike sense of possessing a syntax and semantics. When the definition of language is enlarged to allow animal systems of communication to be compared with human language, it is often stated that human language is far superior to animal systems of communication. Yet dolphins have been found to have a system of communication so sophisticated that humans have sought to reproduce it in submarine communication. Sonar is at least comparable to human language in terms of sophistication. It is actually much like the echolocation system in bats, dolphins use sound to determine the distance of objects in the water that they cannot see (Herman & Morrel-Samuals, 1996:290). Here again, in using human standards as the benchmark, those aspects of as yet undiscovered

animal communication that might be quite sophisticated and worthy of further investigation are overlooked.

The above examples demonstrate theoretical occurrences of anthropomorphism by omission, which show the implications that arise when one implicitly takes a human-centered perspective on animals. The perspective of actually taking the animal's point of view is an approach called 'theromorphism', coined by the ethologist William Timberlake (2002). It is defined as using experience-based knowledge to view the world as though one is a particular animal. It involves orienting one's perspective toward not only 'putting oneself into the shoes' of the animal, but also wearing the shoes of the animal (2002:105). The phrase 'wearing shoes' is meant to symbolize the process of embodiment. That is, taking the animal's sensory environment as a function of its spatial location into consideration. Burghardt speculates that this process of imaginative projection is no different, conceptually, from understanding a person who is different from oneself in age, gender, sensory and motor abilities etc (Rivas & Burghardt, 2002:11).

This idea is well illustrated with an example. It was found that researchers were able to gain much more information about what it is like to be a dinosaur by walking around with a weighted suit frame molded in the shape of a particular species. This experiment gave researchers added information about the maneuverability of dinosaurs that they might not otherwise have thought to consider (Rivas & Burghardt, 2002:11).

The eccentricity of this point of view or perspective compels one to ask how it might be possible to implement. In other words, what theoretical research methods have been constructed as a function of this novel perspective? Timberlake thinks that the best way to go about implementing theromorphism into a research plan is to construct models of the various species of animals, much like Dennett's heterophenomenologies. Information would include that already known concerning the mechanisms, function and evolution of cognition of the species in question. He acknowledges that this is not a trivial task, but that humans are well suited to such a task. Contrary to the tendency humans have of viewing other species according to a human benchmark, Timberlake argues that we also have a special ability to use our experience to integrate information about an animal's sensory physiology, behavioral organization and learning to understand

and predict an animal's behavior. This capacity, coupled with talking to people who have such a model of animals in place because it bears on their livelihood, such as fishermen, hunters and trackers, should provide a solid informational base. These models, in Timberlake's opinion, allow an observer to predict behavior by virtually placing him or herself in the position of a specific animal, not as a human, but as the animal.

At issue here is whether or not this new animal perspective is empirically tractable and thus a worthwhile pursuit for cognitive ethologists. The first question to contemplate is whether a subjective point of view even exists in the animal, and whether it is possible to study it. The second question is whether it is possible to remove the human tendency to view other animals through human colored glasses. Is it possible to virtually place oneself in the position of an animal, as the animal? A third question, that bears directly on the work of Nagel, is whether or not we shall still be able to recognize this experience as experience, if it turns out to be as alien as Nagel hypothesizes. With our human comparison benchmark removed or bracketed, how will we recognize these so-called alien experiences if there is no common ground or overlap between the two types of experience? A fourth question, again based on Nagel's thoughts, is whether we will be able to construct any objective theories from this subjective point of view.

Concerning the existence of this subjective point of view in an animal, we saw that Dennett questions whether there is anything left over that would constitute a subjective point of view that is not already accounted for by neurophysiological facts. He questions that there even exists an overseer of what it is like that would be witness to the phenomenological feel of experience. Akins believes that there is such a thing as subjective experience, but that it is so ill-defined, scientists need not concern themselves with whether or not they have captured it in objective theories. The privacy of the mental argument, discussed by Allen and Bekoff, guarantees that we will never directly know what the animal's subjective experience consists in.

Given all these conflicting theoretical points of view on the phenomenon, what are we to think about the existence of the subjective point of view in animals? I am not convinced by Dennett's overseer argument that questions the existence of the phenomenon of 'what it is like'. While the claim that all animals are reflective beings might still require some additional proof, I think that Nagel is right to say that the essence

of the belief that a creature has experience means there is 'something it is like' to be that creature. Aiken's opinion on the matter is accurate: this phenomenon is at present not very well defined. There is no good reason to stop doing research at this point however.

It is my opinion that notwithstanding the privacy of the mental argument, research into the subjective point of view of the animal can go on in spite of the lack of a philosophical resolution to the other minds problem. None of the considerations of the problem of other minds bear directly nor negatively on the proposed examination of the subjective point of view. Even acknowledging that we will never have direct access to the subjective point of view of another human, let alone that of an animal, should not stop us from trying to gain that knowledge through indirect means. The lack of a philosophical theory for phenomena such as consciousness has not held up research into this facet of study in animals, and it should not in the case of phenomenal experience either. Unfortunately, the lack of any philosophical guidelines in the form of a theory makes discussions much more confused, often degenerating into a case of one researcher talking past another. Nonetheless, the lack of a theory shouldn't hinder examination, it should rather inspire it.

**6. Conclusion: A Fifth Aim for Ethology**

Gordon Burghardt, stealing a trick from Dennett's bag, has decided to act 'as-if' the phenomenon of subjective experience does indeed exist and has carved out a research area devoted to studying it. The original four aims of ethology, conceived of by Nico Tinbergen (1957), do not include the study of subjective experience. Briefly, the four aims are to study:

1. Causation: the identification of the internal and external factors underlying behavior.
2. Ontogeny: the identification of patterns and processes in behavioral change.
3. Evolution: the identification of historical patterns and processes in behavioral change.
4. Survival value: the identification of how behavior patterns contribute to reproductive and inclusive fitness in the various species of animals (Burghardt, 1997:257).

The fifth aim, so far inexistent, according to Burghardt, would be called 'private experience' and its concern would be the identification of patterns and processes in life as

it is experienced by the animal. The idea behind the creation of the fifth aim is to return to the effort of Von Uexkull, to attempt to understand the perceptual and inner worlds of other organisms, both human and non-human, and to try to gain some understanding of what it is like to be the animal, to make inferences about private experience and to see what such an understanding can contribute to studies in the traditional four aims (Burghardt, 1997:260).

The next question mentioned above concerns whether or not it is possible to remove or bracket the human perspective in studying the lives of animals. Certainly the human tendency to see animals in terms of humans is a hindrance. In short, it masks us from discovering the true capabilities the animal might have. The problem is that in attempting to gauge the world from an animal's perspective, whether it is by getting down on all fours on the ground or donning a dinosaur suit, one is still a human looking at the world from the perspective of an animal. One is only going to get information on what it is like for a human to be a dog or a dinosaur. The privacy of the mental argument thus guarantees that, because we don't have direct access to other humans' states of mind, we have even less access to animal minds, and thus we will never completely know what it is like to be another animal.

Should the above argument be taken to be the end of the story? I don't think it should. The first step forward on the issue was a recent recognition that investigators were implicitly adopting a human centered perspective in studying animals. This led to the creation of the label, anthropomorphism by omission, and lent credence to the already existing idea of anthropocentrism. From there, Timberlake took an empathetic viewpoint and labeled a new perspective, that of theromorphism, which means to take the animal's point of view. Theoretically, there should be no trace of anthropocentrism in this view, the human point of view should be successfully bracketed. Practically speaking, and considering the nature of the creature carrying out the research, a human, is this possible? I think that it is as possible as many analogous activities that humans successfully perform every day with each other, such as getting advice from a psychotherapist, for instance.

The impossibility of removing the human benchmark seems to also bear on the issue of whether we will recognize the subjective experience of animals or not (the third

question above). Perhaps it is that tendency to see things through human-colored-glasses that causes us to consider the pure subjective point of view of the animal as alien. The experience of the bat is unrecognizable, as Akins pointed out, because we are not bats but humans trying to look through the eyes of the bat, and the bat has poor vision compared to us humans. If we could somehow consciously bracket the anthropocentric tendency to relate to experience as a human, the gap between our perspective and the bat's could begin to be narrowed.

On the other hand, perhaps the subjective experience of animals falls into the set of facts that are beyond human comprehension. If this is the case, then it could possibly be forever unrecognizable. I don't think that this view is anything more than pessimistic. The idea of landing on the moon may have seemed inconceivable many years before it occurred, particularly when we lacked the telescopic apparatus to even view the moon accurately.

In any case, the attempt to implement this idea of the animal's subjective point of view into a research program would seem to be severely constrained if not impossible by Nagel's conclusion that it is not possible to capture the subjective perspective in an objective theory. Even if we were to be able to gauge animal experience by some miracle of modern technology, we wouldn't be able to construct theories about it.

This conclusion may also turn out to be hasty and preemptively pessimistic. First of all, it may turn out that the idea of constructing a theory, any type of theory, from subjective experience is physically impossible. A theory involves organizing bits of disparate information into a set of generalizations. Often, in the construction of the objective theory, the data in its original form is missing, having been turned into a theory through generalizations and collation of data. In this case the raw data containing the subjective perspective would not be found anywhere in the objective theory. There are other possibilities however. One is to construct a theory comprised of the properties of subjective experience. Gordon Burghardt's research perspective of 'critical anthropomorphism' is a good candidate for this type of research plan, since it requires gathering information from a variety of sources that could then be used to compile a set of properties of subjective experience. Some of these sources of Burghardt's proposed method include our own perceptions, feelings, and identification with the animal. These

sources together constitute an anthropomorphic empathy, hence the term 'anthropomorphism'. The practice of anthropomorphism is 'critical' in the sense that these anthropomorphic empathetic intuitions obtained by asking ourselves 'what we would do in the situation if we were the animal in question' are then rigorously tested in the form of hypotheses and predictions of experimental outcomes. The method of critical anthropomorphism is particularly apt for two reasons. One, adopting the method means acknowledging that one is generating possible explanations for an animal's behavior from the anthropocentric stance based on the mode of empathy. Researchers deliberately 'try out' explanations from a human perspective as a heuristic predictive device. These explanations are then empirically tested to determine if they are accurate or not. The method is thus a type of conscious anthropomorphism, because we are using our own experience as a basis, but with an inherent corrective device, that of empirical testing. This method is also apt because it focuses on the creature's internal stimuli and subjective responses to these stimuli. As Burghardt argues, examination of these aspects is necessary to an adequate understanding of behaviors such as problem solving, deception and courtship in the animal (Burghardt, 1994:1).

Burghardt cites a second candidate research model for studying private experience in animals. The common element to his model and the one he suggests is this notion of testing empathetic intuitions obtained by first taking the anthropocentric stance and wondering what one would do in a situation; formulating these intuitions into anthropomorphically based hypotheses and testing them; and also gathering and comparing evidence and data from numerous different objective sources for the purposes of cross referencing and checking for accuracy. This second model comes from Frans De Waal, the famous primatologist and has four components. The first two components are compiled from naturalistic observation of subjects, and entail collecting data in both qualitative and quantitative forms, that later can be cross-referenced for accuracy of observation and to factor out any observer bias. As we saw in Chapter two, quantitative data is numerically coded data, whereas qualitative data is in terms of verbal description, not previously coded. It is possible to compare these two types of data, taken from the same observation period, and check for accuracy. The third component is controlled observation, the strategy taken by Cheney and Seyfarth in their study of vervet monkeys

in Kenya. It basically entails performing an experiment, or manipulating certain variables, in the species' natural habitat. The fourth component is experimentation carried out in a laboratory environment, where greater control over extraneous variables such as environmental conditions is apparently achieved. Included in De Waal's research method are all possible types of objective data, which can then be used to either support or contradict the hypotheses generated by the researchers. For instance, the accuracy of an animals' reaction obtained in a laboratory experiment can be cross-checked with the same data obtained in the animal's natural environment, both with experimental manipulation and without, i.e., pure naturalistic observation.

With the creation by Burghardt of this new fifth aim in ethology, it would thus appear that research into the subjective experience of animals will be carried out regardless of philosophical agreement or a guideline regarding the existence and ontology of the subjective point of view. While the research effort may prove a dismal failure, its existence will at least shed light on and subsequently minimize the tendency to view animal capabilities through human eyes. That accomplishment alone should prove to be quite an advance.

## Chapter Four

## Methodology and Theory of Mind

### 1. Introduction: Experimentation and Interpretation

The final objection to the project of Cognitive Ethology that I will be examining in this first half has to do with methodology. I treat this objection last because in my opinion it is the most damaging objection that could be made to the project of searching for mental states in animals. Methodology is the foundation of the discipline, and if the methodological foundation is found to have cracks in it, the whole discipline rests on thin ice. I also treat this objection last because it has major implications for the second half of the thesis. As mentioned, there is a leftover problem that gets solved at the end of chapter six.

The objection in its most general construal is that the study of mental states in animal minds is empirically intractable or experimentally impossible largely because animals lack language. In an experimental situation, since animals can't speak a human language they can't inform researchers of the state they are in. Humans, in contrast, can speak a human language and thus answer any questions bearing on mental states that are asked by researchers. There are other specific implications on methodology due to this lack of language. One of them is that experimental design must be quite a bit more intricate and sophisticated, to get around the fact that verbal response is impossible. The fact that experiments are more intricate invites the question of interpretation. That is, do the experimental results represent evidence that would justify the attribution of mental states to animals? A second related implication has to do with naturalistic observation and anecdotes, important empirical tools of the cognitive ethologist that get around the lack of language problem. The challenge to naturalistic observation is that results from naturalistic observation and anecdote, because they lack control and rigor, cannot be used to provide evidence for say, a theory of mind in primates.

The philosophical question of whether animals may be attributed mental states does not rest uniquely on empirical evidence. The question also cannot be decided uniquely on theoretical debate alone. As we saw in chapter one with regard to the

language issue, examining the issue purely from a theoretical standpoint leads one to the anthropocentric stance and appears one-sided considering the empirical evidence that is out there with regard to animals. On the other hand, empirical evidence by itself will not decide the issue either, and constitutes the view that I will argue against in this chapter. I believe that both aspects, theoretical considerations and experimental evidence, are necessary for a thorough examination of the question of whether animals may be attributed mental states or not. I thus also hope to show in this chapter what the relation is between empirical evidence and theoretical discussion.

One of the debates going on within the project of cognitive ethology is the investigation of the possibility of a theory of mind in primates. A theory of mind theory is a hypothesis about the type of mechanism underlying an organism's capacity to explain and predict another's behavior. Theory of mind explanations are thus concerned with two-person interactions, and with the mutual attribution of mental states to predict and explain behavior. This is a different aim from that of cognitive ethology, which is concerned with investigation into whether animals possess mental states tout court. On a theory of mind account, mental states are attributed to a creature because they are hypothesized as being the link or intervening variable between the observed behavior and the explanation or prediction given about the behavior. The usual scenario is that creature X will explain or predict creature Y's observed behavior through the attribution of mental states. Researchers are interested in whether or not this capacity exists in primates. To help clarify the issue, Premack (1988:179) has developed a very useful tripartite distinction corresponding to various possible degrees of a theory of mind. The lowest level corresponds to species that make no attributions of mental states of any kind. The second level corresponds to species that make attributions that are limited in a number of respects. The third corresponds to species whose attributions are unlimited. The first level, that of no attributions, probably includes most species of animals, Premack suggests. The second level might apply to some species of primates. The third level corresponds only as of yet to humans.

It could be objected that the second and third levels are not very helpful, that the phrases 'limited in a number of ways' and 'unlimited' do not serve to qualify what types of attributions are made at these levels. One solution is to modify the second and third

levels so that an additional difference between them becomes explicit. One can distinguish between the two upper levels of mental state attribution based on whether it is first order attributions that are made or embedded attributions that are made. First order attributions are attributions of propositional attitudes such as beliefs. Embedded attributions are also attributions of propositional attitudes, but of a multiple order. This second difference is the same as Malcolm's distinction in belief made in chapter one, that between beliefs and beliefs about beliefs. The levels thus become the following: the lowest level remains unchanged, the second level refers to species, possibly primates, whose attributions are of the first order type, and the third level, probably restricted to humans, refers to those species that make embedded attributions.

There is one further modification that could be made that would make the schematic complete. It entails creating an additional level that would correspond to the possibility of representing things other than mental states. This new level would thus includes the possibility of representation of items in the animal from the physical world, but not representations of mental states. The new schematic would be the following:

> First level: No representation of any kind.
> Second level: Representation of the physical world only.
> Third level: Representation of mental states of the first or lower order, i.e., beliefs about trees.
> Fourth Level: Representation of embedded mental states, i.e., "John believes that Mary knows he likes her".

Premack originally titled his schematic as three degrees of a theory of mind. Taken in a strict sense, his schematic should not include as a possible degree of a theory of mind those levels in which no mental states are postulated, since levels in which no mental states are postulated are not, strictly speaking, levels of a theory of mind. It seems to me that the two lowest levels, those of no representation and representation but not of mental states, fall under the minimum condition for a theory of mind. I will thus make reference to this framework in a more general way throughout the chapter, since the rivals to theory of mind therories, such as behaviorist or associationist theories, can also be fit into the framework, namely at the first or second levels. This schematic will thus serve as the framework for the ensuing discussion.

Cecilia Heyes (1998) calls into question certain of the claims made by cognitive ethologists, in particular Donald Premack and Peter Woodruff, who investigate the possibility of a theory of mind in primates (1978). According to Heyes, one of their claims is that there is observational and experimental evidence that apes have mental state concepts, such as 'want', 'know' and 'see' (1978:515). Heyes is of the opinion that we should stop asking the question of whether primates have these concepts, or possess a theory of mind, until we can get better designed experiments that will, in and of themselves, decide the issue (1998:102). In this respect it could be argued that she is of the opinion that whether or not primates have a theory of mind is an empirical question in the sense that the experiments themselves provide a decisive answer to this question. In this chapter I am going to argue that the issue is not strictly empirical. It is not necessary to wait until an experimental design is produced that will decide the issue because such a type of experimentation is not forthcoming. Experiments will not decide the issue because they are not the only consideration. Experiments cannot decide the issue in this particular case due to the special nature of the issue: the postulation of mental states as intervening variables to explain behavior. Since there is the possibility of interpretation of the results, theoretical discussion must also be pertinent. Part of the reason why empirical evidence is indeterminate is not poor experimental design as Heyes thinks, but rather her incomplete presentation of the experimental results. However, even with an improved representation of the relevant research, experimental results are still subject to interpretation. This is in part due to the fact that the intervening variable of a mental state is not something that is observable, it must be inferred.

Heyes makes an interesting preliminary argument against the claim that there is evidence for a theory of mind in primates. She draws a relation of asymmetry between the disciplines of developmental psychology and research on theory of mind in primates, arguing that although much progress has been made in developmental psychology, no substantial progress has been made in the case of primates. She states that thanks to the empirical tractability of a theory of mind in children, researchers have been able to determine the origin, on-line control of, and epistemic status of human folk psychology. The same amount of progress should be evident in studies of primates, since non-verbal young primates are similar to children in age, etc (1998:102). This claim is gratuitous in

my opinion, although it serves on the surface of it to strengthen Heyes' case. There are all sorts of difficulties with studying primates, such as limited availability, cost of housing and raising them that are not present in research with children. The progress made in developmental psychology is relative: some say it is quick, other say it is agonizingly slow. The often-cited problem for lack of quick progress is the lack of language in pre-linguistic children, which is an analogous problem with primate research. Furthermore, as will be obvious from a reading of the four objections in the first half of this thesis and with Heyes' critical comments in this chapter, advances made with regard to evidence for a mental life in animals are for the most part treated with the severest skepticism by detractors. It is no wonder that progress in research is slow. At any rate, nothing is gained for Heyes' arguments by pointing to a lack of relative progress in Theory of Mind research and comparing it with research in human development.

In every case where a theory of mind component has been professed to be found in primates, the experimental results, according to Heyes, are also explicable by three other possibilities (1998:102). These three other possibilities fall into the category of alternative non-mental explanations, and constitute her set of rival explanations to the theory of mind hypothesis. One possibility is that the result could have occurred by chance. By chance, Heyes means the variable that is statistically pre-set in experiments and that could vary from 20 to 50%, depending on the amount of subjects and trials in the experiment. The second possibility is that the result could be a product of non-mentalistic processes, such as associative learning. The third is that the result could be a result of inferences based on non-mental categories. Relative to the schematic outlined above, Heyes' first explanation of chance is not included in the levels. Her second and third explanations, non-mental processes and inferences based on non-mental categories, would fall into the lowest level. That is, these explanations allow for representation, but not of mental states.

## 2. Theory of Mind Defined

There are many substitute terms for 'theory of mind', which is a rather vague shorthand to represent a variety of mental capacities. Other terms used are metarepresentation, mindreading, metacognition, Machiavellian intelligence and mental

state attribution. Heyes employs a particular construal of Theory of Mind in her discussion that is too strong, in my opinion. There are two noteworthy aspects to her construal. The first is that the individual must possess mental state concepts such as 'believe', 'know', 'want', and 'see' and that an individual with these concepts uses them to predict and explain behavior. The second aspect is that the individual must believe that the mental states play a causal role in generating behavior, but does not identify mental states with behavior. Here the implications are that the individual must possess mental state concepts, and hold a theory of causality with regard to mental states or at least possess the concept of cause. This construal of Heyes' is much too strong in my opinion, and moreover not the one held by Premack and Woodruff. On Premack and Woodruff's view, an individual has a theory of mind if the individual imputes mental states to himself and to others. These mental states are much like the ones humans impute, such as purpose or intention, knowledge belief and pretending (1978:515). On any given variety of theory of mind theory, one need not possess the concepts of belief, desire and the like, and one need not believe that these mental states play a causal role in generating behavior. Most theory of mind theories require minimally that the individuals involved attribute mental states to each other in order to explain or predict behavior, and often nothing beyond this.

Now that Heyes has defined the theory of mind camp, she next draws a contrast between the unifying features of theory of mind hypotheses and her non-mentalistic alternatives. The unifying feature of the theory of mind hypotheses is that primates categorize and think about themselves and others in terms of mental states. The unifying feature of the non-mentalistic alternative explanations is that they do not assume that primates represent mental states. Rather, primates respond to or categorize and think about themselves and others in terms of observable properties of appearance and behavior (1998:102). It could be argued here that although mental state representation is not assumed on Heyes' definition, it could still occur. That is, her use of the phrase "does not assume" is ambiguous enough to cause collapse between the two types of explanations. The presence or absence of mental state representation should be the defining feature of theories of mind, otherwise what else is there to prevent collapse between the two explanations? It is then not surprising that current experiments cannot point to either

explanation exclusively, and that Heyes is able to advance a non-mental explanation alongside every theory of mind explanation. Thus the phenomenon of too much overlap between explanations is one reason why both explanations can be advanced for an experimental result.

For the purposes of the ensuing discussion, the theory of mind and non-mentalist explanation must be qualified in order to bring out a usable distinction between them, and to allow for hypotheses to be formulated. There are two problems with Heyes' definitions. First, Heyes' distinguishing features of a theory of mind are not strong enough to prevent collapse between the two explanations. I would thus reject her two distinguishing features, possession of concepts such as belief and desire and belief in mental states as causes of behavior, in favor of a general unifying feature. I am going to assume for the purposes of this discussion that theories of mind have the unifying feature of some sort of mental state attribution. It need not be of the higher order type, i.e., a belief about a belief, but the individual, on a theory of mind account, must attribute some sort of intervening mental state either to himself and/or to another individual in the explanation or prediction of the other's behavior.

Second, Heyes' stipulation that non-mentalist explanations do not assume that primates represent mental states must be modified in order to imply a stronger distinction. That is, non-mentalist explanations must not make reference to mental states *at all*, in order to be properly distinguished from the theory of mind explanations. The non-mentalist alternatives that Heyes cites cannot postulate mental state attribution as an explanation of behavior because otherwise both explanations will be indistinguishable. It will then be impossible to declare that one explanation over the other is able to account for the results. Thus while Heyes' non-mental alternatives can make reference to eliciting stimuli, stimulus-response associations or stimulus-response pairings as an explanation for behavior found in the experiments, they cannot make reference to mental states.

### 3. Six Indicators of a Theory of Mind

Heyes evaluates six different indicators that have been offered as evidence of a theory of mind. It should be noted here that neither a theory of mind nor mental states

are observable entities, and so attempts to demonstrate the existence of a theory of mind is a matter of inference. It is a matter of one individual explaining or predicting the behavior of another through the inference of attributing a mental state to that individual. Moreover, the indicators, such as 'self-recognition' or 'imitation', are not to be identified with the behavior evidenced in the experiment, nor are they mental states. In other words, the indicator of role-taking cannot be identified with the mental state of role-taking, for such a mental state does not exist. An example would clarify the issue. 'Deception' is thought to be an indicator of a theory of mind because it involves one actor causing another actor to either believe something erroneously or act in a mistaken way. The erroneous belief can take any form, but is not to be identified with the mental state of 'deception', for such a state does not exist. Generally speaking, both of these situations involve the attribution of mental states, either to one or both of the actors, in the prediction or explanation of behavior. All of the indicators are thought to be indicators because they involve postulating an intervening variable of mental states as an explanation for the behavior.

The six indicators are imitation, self-recognition, social relationships, deception, role taking or empathy and perspective taking. They are each evaluated as representative of a theory of mind by Heyes based on two criteria, Competence and Validity. Competence is defined by whether there is reliable evidence that the individual has the relevant behavioral capacity that, if present, would indicate a theory of mind. In order to try and ease understanding in the reader of the competence criterion, Heyes states that the competence criterion attempts to establish which environmental cues the primates use to guide their behavior (1998:102). The established presence of this behavior in the experiment might then indicate that a theory of mind is present in the primate. I say 'might' rather than 'would' because the indicator then has to pass the validity test. Validity is understood as: if present, would this behavioral capacity indicate a theory of mind? (1998:102). An indicator would fail this criterion if there were another equally plausible non-mentalistic alternative explanation at work. In other words, if Heyes can show that one of her alternative non-mentalistic explanations also fits the experimental results, then the theory of mind explanation cannot rule out other alternative explanations and fails the validity test. In an attempt to ease understanding for the reader, Heyes

characterizes the validity question as the following: validity asks about the psychological processes that led the primates to use these cues instead of others (1998:102). To sum up, the competence criterion is passed if the behavior is present in the experiment. The validity criterion is passed if there is no rival explanation for the results. It could be said that the validity criterion is passed through the process of elimination of other explanations.

Heyes gives an example that should help to shed light on what job she has in mind for the competence and validity criteria. One of the proposed indicators of a theory of mind is 'self-recognition'. This indicator would pass the competence criterion if there is evidence that the primate uses a mirror to gain information about itself. That is, if the primate shows behavioral evidence such as looking in the mirror and touching some area of its body, the competence criterion is passed. The indicator of self-recognition passes the validity test if Heyes cannot offer some other alternative non-mentalistic explanation for the primate looking into the mirror. The general framework is the same for all six indicators discussed. In each case, the presence of a behavioral indicator is taken to be evidence for a theory of mind, if no other explanations for the behavior exist.

Due to Heyes' less than thorough discussion of the six indicators, I have been able to identify a missing element that, if not discussed, contributes to allowing her to conclude that current research is not decisive. The missing element is hypotheses, and the fact that they are not mentioned means that prediction of experimental outcomes are also lacking. The way an experiment is normally conducted is the following. An experiment is designed. The researcher commits in advance of the performance of the experiment to a possible outcome of the experiment, that is, to a particular set of results. The two possibilities are either that the hypothesis is confirmed, i.e., the behavior is in evidence, or that the hypothesis is not confirmed, i.e., the behavior is not evidenced. In some more complicated experimental designs, there is sometimes one other different potential outcome of the experiment and it often favors the other competing hypothesis. In the simplest of experimental designs, the two possible outcomes are either that the behavior is displayed or it isn't. Most of the experiments discussed by Heyes have been conducted by the theory of mind side, thus the hypothesis mentioned in the experiments is always a theory of mind hypothesis. Theory of mind experiments do not fit into the

simple design mentioned above. Since the experiments have been conducted by the theory of mind side, and are designed to determine if a theory of mind is attributable, they make reference to intervening variables such as mental states to explain the behavioral result obtained in the experiment. In these experiments, the deciding factor is not the presence or absence of the behavior. The crucial distinction between theory of mind hypothesis and non-mental hypotheses is rather whether the behavioral result obtained warrants a higher level explanation or not, i.e., mental state attribution. In Heyes' discussion of the six indicators, she regularly fails to mention the theory of mind hypotheses. This might be why she is able to advance numerous alternative explanations for the results. The fact that she fails to mention predictions but advances numerous explanations could invite the charge that her explanations fit the data because they are all 'after the fact' and so they are *ad hoc* explanations. Perhaps another mitigating factor in Heyes' failure to formulate hypotheses based on non-mental explanations is that she simply cannot because mechanistic explanations have no predictive power. We will come back to this question at the end of the chapter.

There is a second problem that has to do with the validity question, the proposed link between the indicator and the theory of mind hypothesis. The way Heyes puts the issue is to have the experiment answer the question. That is, if she can find no other rival explanation for the results, then the indicator is representative of a theory of mind. I rather think the link should be framed the following way: Is an indicator such as deception, because it implies the attribution of an intervening variable of a mental state, a reasonable indicator of a theory of mind? The only way that deception could be taken to be a reasonable indicator of a theory of mind is because it postulates the attribution of a mental state as an intervening variable. Heyes mistakenly, in my opinion, assumes that experimental results should be able to answer the validity question, that is, be able to show whether the proposed mental state indicator is indicative of a theory of mind only by eliminating all other potential explanations of the results. Proof of this is that the validity criterion is passed if she can't find another explanation to account for the results. Ruling out alternative explanations is thus the only aspect considered in the validity issue. The fact that she can't advance any alternate explanations for the results is only one consideration amongst others, in my opinion. It certainly doesn't tell us whether the

inference to mental states is warranted. For this reason the validity question is also very much a theoretical issue that can be argued for or against completely separately from the experimental results. It cannot be determined merely by a lack of the existence of alternative explanations. Moreover, where is the criterion that evaluates the mental state attribution as an intervening variable? It seems to me that an evaluation of a proposed indicator for a theory of mind should include some kind of determination as to whether an indicator implies the possession or attribution of mental states. In the case of theory of mind, this is the crucial issue, i.e., this is why the indicator is being proposed in the first place, because it requires the attribution of mental states. In other words, the validity criterion, as Heyes has construed it, as a kind of mechanical experiment decider is insufficient. It must at least include an answer to the question: what behavior indicates which mental state? This issue, I maintain, can be debated separately from the experimental results, and this is where theoretical discussion becomes important.

### 3.1. Imitation

The first indicator discussed is that of imitation, defined as the spontaneous reproduction of novel acts yielding disparate sensory inputs when observed and executed (1998:102). The point about imitation, or 'aping', as it is sometimes called, is that the observer reproduces the same action that he or she has just observed or otherwise gleaned through the senses in another individual. An example would help to illustrate this indicator. The action, say, an alarm call, is observed with the eyes and heard with the ears by the watcher and then reproduced by the same watcher with its mouth. It is thought to be an indicator of a theory of mind because it involves ascription of purpose or goals by the imitator to the model (1998:103).

Heyes discusses two experiments on imitation, although the first cannot truly be deemed an experiment. It is the so-called 'Hundredth Monkey Phenomenon' where increasing numbers of Japanese Macaques on Koshima Island have been observed to bring sweet potatoes down to the river to wash off the dirt before eating them. The phenomenon, thought to be started by a single monkey, is claimed to have spread to the entire population through the process of imitation (1998:103). A possible problem with the experiment is that the potatoes were deliberately made available to the monkeys by

the researchers, but without any attempt to experimentally manipulate any conditions. There was also no hypothesis made by the researchers.

Heyes cites another experiment designed to demonstrate the phenomenon of imitation. The aim was to see whether monkeys would imitate the demonstrator's use of a human gardening tool, in this case a rake, to get food out of reach (1998:102). The hypothesis is that the monkeys would reproduce certain of the researchers' behaviors, i.e., obtaining fruit with a rake. This experiment could be charged with being somewhat anthropocentric, equally anthropocentric would be to teach the primates how to use a food processor. Although primates have and use tools in the wild, the rake and the food processor are not in their repertoire of tool use. Interestingly, the results showed that encultured chimpanzees (animals with extensive training history) more than non-encultured chimps did appear to imitate the experimenters' use of a variety of human tools to solve problems such as obtaining food that was out of reach.

Heyes doesn't think that imitation actually occurred in the experiments, so both fail the competence criterion. She further does not believe that imitation is an indicator of a theory of mind, having been able to advance alternative explanations for the results, so the experiments also fail the validity test.

The alternative explanations advanced by Heyes include instrumental learning, matched dependent behavior, coincidence and emulation learning for reward (1998:103-4). Heyes explains the hundredth monkey phenomenon by the claim that the macaques may have observed one particular macaque wash its potato in the water and reproduced the behavior purely by chance by chasing the macaque into the water while holding a potato. This is an example of the supposedly non-mental capacity of acquisition of a behavior through instrumental learning by coincidence. What this explanation lacks is an account of why a macaque would pick up this particular behavior and not the thousand others that occur in similar chance circumstances, in other words, an account of when and how the association was first formed.

Concerning the rake use experiment, Heyes refuses to grant that these animals were indeed imitating the experimenters. Heyes instead claims that the results could have been due to stimulus enhancement, which is where the primate manipulates an object that has been made more salient through contact with the experimenter. This explanation still

leaves out an account as to why this object is more salient than others, that is, what distinguishes those objects that are manipulated from those that are not. As with the potato washing explanation, this explanation is also ad hoc, it only fits the results after the fact, but fails to predict when or explain why the primate appeared to 'imitate' this particular behavior and not the thousand other behaviors of the model that the primate was also witness to. In other words, the non-mentalist explanation cannot predict which of the behaviors of the model that the ape will choose to ape. In short, Heyes' explanation for the results is just that, an explanation after the fact.

Moreover, the potato washing experiment lacks a clear-cut hypothesis by the theory of mind side. It suffers from many problems, and it falls somewhere between a piece of naturalistic observation and an experiment. The phenomenon is not a case of pure naturalistic observation because it involved artificial intervention, i.e., the potatoes were given to the monkeys by a group of researchers. It is not an experiment because the researchers did not try to manipulate any variables and gave no hypothesis of the results. The researchers just made a novel food source available to a group of monkeys to see what would happen. This phenomenon thus cannot provide much weight to argue for the theory of mind side.

Although the second experiment is a better example of a piece of evidence that can be used by the theory of mind side, Heyes fails to mention an interesting phenomenon that was displayed by the uncultured primates. The experiment was, on some accounts, designed to determine whether the primates would use the rake to solve a problem, such as obtaining food that was out of reach. It was found that the cultured primates, those with extensive training history, did imitate the trainers' demonstration of the use of the rake to obtain the food. The group with no training history was able to solve the problem but did not employ the trainers demonstration. They rather 'invented' their own tactics. It seems to me that given the uncultured group's tactics of invention as a contrast tends to give more weight to the hypothesis that the cultured chimps were displaying something akin to imitation.

In my opinion, the competence criterion has been passed in the rake use experiment. Given the contrast in behavior between the cultured and uncultured primates, imitating the antics of the model versus inventing their own solution, there is a

clearly delineated behavioral result that merits a passing of the competency criterion. Concerning whether imitation is a valid indicator of a theory of mind, I agree with Heyes that it is not, because I don't think it requires the attribution of mental states. I do think, however, that imitation, by definition, has a cognitive element to it. The idea of reproducing an action that is a result of disparate sensory input and output is quite sophisticated a feat and while it might not require attributing goals to the model, the primate must still somehow transfer what it has seen in the model into an action that it then reproduces.

### 3.2. Self-Recognition

The second indicator is that of self-recognition. There isn't a clear definition of this indicator. It is also known as 'mirror-guided body inspection' where individuals use a mirror as a source of information about their own bodies. This indicator is thought to be an indicator of a theory of mind because it implies the potential to imagine oneself as one is viewed by others. It is taken by some primate researchers to further imply possession of the concept of self.

In one such type of experiment, primates are anesthetized and painted with a coloured dot on their head that they cannot see without a mirror. The test is to compare how many times they touch the spot, first in the absence of a mirror, with how many times in the presence of a mirror. There is a clear hypothesis made by the theory of mind side in this case: if the primate's rate of favoring the spot is significantly higher in the presence of the mirror than in the absence of the mirror, then the competence criterion should be passed.

Again, Heyes argues that not only is the behavioral indicator not present in the experiments, but that even if it were present, she is able to advance alternate explanations for the results and so it is not a valid indicator of a theory of mind. It thus fails both the competence and validity tests. Heyes' alternate explanations are bordering on implausible, and include the claim that primates who have a longer recovery time from anesthesia will be more active than those with a shorter time. This would explain the discrepancy between activity without the mirror (very low, because the monkeys were still sluggish from the anesthesia) and then with the mirror (higher, because the

anesthesia had worn off completely). This explanation has been invalidated by other researchers, however. They contend that if Heyes had done a thorough review of the literature, she would have found that the test was not performed until sometimes 24 hours after the primates had been anesthetized, plenty of time for the effect to wear off completely. Moreover, in some experiments the primates were not anesthetized at all but were rather marked while they were awake (Gallup, Anderson, Shillito, 2002:328). A second alternative explanation is that the control group of primates (with no marks) are too busy responding socially to their image in the mirror to engage in the grooming behavior that the experimental group had engaged in that led by chance to their touching the marks. Heyes attributes the control group's behavior to the fact that chimps typically exhibit social behavior on initial exposure to a mirror. It turns out that social behavior in this context means that the primates respond as if their image is another primate. It seems to me that Heyes cannot rule out that the labeling behavior of the primates requires mental states, namely the recognition of the other as 'friend' or 'enemy' or at least as different from oneself.

The contrast in behavior between the control and experimental groups in this experiment is, in my opinion, good evidence for passing the competence criterion. The control group's actions indicate that the primates were responding to the mirror as if the image was of another primate. The experimental group acted as if they had gone beyond this level of responding, to a level where they 'recognized' themselves in the mirror and groomed themselves in impossible to see places. Given this contrast, and the fact that mirrors don't exist in the wild, which could explain the latency period from social responding to self-grooming, I think that the competency test is passed.

Heyes' validity question should ask whether the indicator, as evidenced by the presence of a certain behavior, really is indicative of a theory of mind. In this case the question would be whether self recognition or the concept of self is indicative of a theory of mind. This is a question of interpretation that can be discussed regardless of the experimental results. I am in agreement that the primate's actions with the mirror are not a reasonable indicator of the concept of self. I don't think that the concept of self is amenable to experimentation, in part because it is not clear what mental state would indicate the presence of the concept. The concept of self is a vaguely defined and

controversial topic in human theories of mind. Many are not certain of its parameters or even whether it exists, perhaps there is just the illusion that it exists. I do think that the fact that a primate can inspect parts of its body that are not normally visible to it without a mirror is significant of something, although it might be more closely related to self-recognition.

### 3.3. Social Relationships

The third indicator of a theory of mind is social relationships. This would be understood in its most narrow sense as the observation of a structured interaction between one or more conspecifics (1998:105). Primatologists employ a larger sense of the term, however, where an individual acts on an earlier observation of two other interacting individuals, such as starting a fight with the winner of a fight occurring earlier between two other individuals. This might be more aptly described as knowledge of social relationships. Knowledge of social relationships is thought to be an indicator of a theory of mind because a primate acts, for instance, aggressively toward a second primate on the basis of a mental state such as the belief that the other primate is in a higher rank.

One of the experiments to test this indicator involves training a 'privileged' subset of apes to perform an action to obtain a treat that they could then share with the rest of the group who have not been trained and who have no treats. The other two thirds of the group, not knowing how to obtain treats, would have to rely on this special subset of apes to obtain the treats. It was found that those apes that received food from the trained apes spent more time with and groomed these trained apes.

Heyes distinguishes between two types of social relationships, mere knowledge of social relationships, where the relationships are observable properties and awareness of them is obtained though associative learning, and a more abstract sense of the term, where one individual attributes dispositional mental states such as loyalty, dislike or affection to another conspecific. In the former, more behaviorally obvious case only, on her view, it is possible to say that the existing evidence supports the claim that apes know about social relationships. She thus takes the results of the experiments to indicate not that certain individuals seek contact with high-ranked individuals because they are believed to be high ranked, but because the apes made an association between the

preferred activity and the ape who learned it. So although the apes display the behavioral indicator of social relationships and pass the competence test, there are other explanations and so the validity test is failed.

Her alternative non-mental explanations include the claim that the responses of the non-special group are based on an earlier exposure to a contingency, between that of happenstance grooming of the special primate and obtaining a reward. The other explanation has the same basis: the learning of an association between grooming and obtaining food reward, also known as acquired-affiliative-social-responding (1998:106). As with the indicator of imitating, Heyes' explanation is ad hoc, she can explain after the fact how the behavior occurred but her explanation has no predictive power, i.e., it cannot predict when one chance encounter over another will be the one where the crucial association is made. Moreover, if the non-special group never obtained rewards for hanging around the privileged set of primates, Heyes has no situation to point to where the association was first made in the minds of the non-special group.

In my opinion, the initial problem stems from Heyes' distinction between two types of knowledge of social relationship and her subsequent exclusion of the second type from consideration. To reiterate, she distinguishes one construal that requires mental state attribution and is based on both past and present social events, from her sense, where social relations are observable events and the primate shows behavioral evidence of 'affiliate social responding' to a higher ranked individual. After distinguishing the two types, she excludes the first one from the possibilities, presumably because it makes reference to mental state attribution. This has the effect of setting the situation up in advance to be doomed for the theorist of mind. If mental state attribution is not a potential underlying mechanism in the experimental task, then the experimental results won't be justifiably attributable to a theory of mind explanation. The validity test will then automatically be failed. Moreover, at a minimum, any community species is going to have a social or interactive aspect to its behavior, almost by definition. This social behavior, in all its observed nuances and varieties, cannot be explained solely on the basis of stimulus-response behavior. It must make reference to mental states to explain antagonistic behavior, for instance, between a leader primate and one of its subordinates.

I don't think that knowledge of social relationships is particularly amenable to laboratory experimentation without preliminary naturalistic observation. In my opinion, passive observation in a naturalistic setting is a necessary preliminary step in collecting evidence of social relationships, because it requires extensive tracking of the animals to determine whether and what social and status relations are in effect. These relationships, once they are determined, can be manipulated in an experimental setting, but only if a large enough number of the group is brought in to study. One can't study whether mothers recognize their children's alarm calls if one or the other parties is left behind in the wild.

### 3.4. Role-Taking

The fourth indicator is that of role taking. Of all the indicators, role-taking is most definitive of a theory of mind. It is the act of identifying with a model's circumstances in order to predict what the model would do in that situation. This capacity is thought to be an indicator of a theory of mind because it requires that the role taker attribute beliefs and desires to the model (1998:106). In my opinion, this capacity does not necessarily require the attribution of any mental states, at least not on all theory of mind variations, but Heyes does not entertain this consideration.

One set of experiments designed to test this indicator involved showing videotapes of problem scenarios to Sarah, the ape who was taught sign language. The tape was stopped at the end of the problem and two photographs were shown to Sarah to choose between, one that solved the problem and one that didn't. The theory of mind hypothesis was that Sarah should choose the photo that solved the problem, and it was found that Sarah did indeed consistently choose the problem-solving photo.

In another experiment, chimps were divided into two groups and each trained on a different task. One group was trained to choose one item out of a set of four that a trainer was pointing to. The second group observed a trainer bait one of four containers that they then had to accurately select by pointing to the container. The tasks were then switched on the two groups so that in the test session, each group was performing the task of the other group, a task new to them. The theory of mind hypothesis was that the new task should be performed by the chimps without a decline in performance. It was found

that the rate of successful performance of the new task indeed did not decline for three out of four of the primates in each group.

In the first experiment conducted with Sarah the chimp, Heyes cannot dispute the fact that performance did not decline for the most part, and so ends up granting that the competency criterion is passed. The alternative explanations for Sarah's behavior include choosing a photograph based on familiarity, physical matching, and formerly learned associations. All three of these activities have as a common element the fact that Sarah might have been matching or associating an aspect of the problem that appeared in both the problem videotape and the solution photograph, based on familiarity between the two things. Here Heyes ought to say that an aspect of the problem is identical with an aspect in the photo, because familiarity implies that Sarah would have to abstract the two aspects in order to compare them. Abstraction requires more than stimulus response conditioning whereas choosing on the basis of identity, it could be argued, does not. However, nothing in the task of choosing a photograph requires the attribution of mental states, and so the results cannot be said to argue for the theory of mind side.

Heyes makes a strange point at this juncture, claiming that since there exists no single unitary non-mental explanation for their results, Premack and Woodruff's work on role taking is unique in the literature in this respect. This is an odd claim to make, since Heyes has supplied three alternative explanations: familiarity, physical matching and formerly learned associations. Moreover, why is Heyes all of a sudden restricting herself to one alternative explanation, when she has advanced multiple alternative explanations for each of the previous sets of results? Heyes then makes an even stranger move, cites Premack and Woodruff's work as a standard, and claims that no progress has been made since their research in 1978. I presume that Heyes means here, although she fails to mention it, that the work done by Premack and Woodruff passes both the competency and validity tests and is thus a standard in that it is the only research, from her perspective, to have shown evidence of a theory of mind in primates.

### 3.5 Deception

The next indicator is deception, which is taken in a functional sense to mean the performance of a cue by one animal that will lead another to make an incorrect or

maladaptive response. It is thought to be indicative of a theory of mind because the deceiver must cause the deceived to make an error, and this requires some type of imaginative projection on the part of the deceiver, which requires mental state attribution (1998:106).

There are not many experiments on deception in the literature, although there is a great deal of anecdotal evidence from naturalistic observations. One of the only experiments conducted involves the ape watching a trainer hide a reward in one of two containers. Either a cooperative (dressed in green) or a competitive (dressed in white) trainer then comes into the room and searches the container that the ape points to. The cooperative trainer always shares the found food with the ape. The competitive trainer only does so if the ape points to the wrong container (i.e. empty container). The theory of mind hypothesis predicts that the ape will learn to deceive the competitive trainer by always pointing to an empty container.

Heyes cites another, in my opinion, very telling anecdote obtained through naturalistic observation where a female primate approached and began grooming a male who had caught and was guarding a carcass. The male eventually lolled back into a supine position and let go of the carcass, perhaps due to the relaxing effect of the grooming. The female then snatched the carcass and ran off with it.

While Heyes accepts that deceptive behavior is in evidence in the experiment, and so the competence test is passed, she does not accept that deception is indicative of a theory of mind, because there exist alternative explanations for the behavior. The alternative reasons as to why the behavior might have occurred include: by chance, as a result of associative learning, or as a product of inferences about the observable features of the situation rather than about mental states. An explanation based on chance is ad hoc, it does not account for why the female seizes the opportunity during this situation and not during others where another primate is lying in a supine position eating some food. To postulate associative learning as an explanation in this case, one would have to be able to point to the previous situation when the association was first formed. What is the likelihood that just such a situation occurred in the recent past?

In my opinion, deception is one of the most promising and clear-cut indicators of a theory of mind. First, because the behavior involved in deception is quite well

delineated and easily behaviorally manifested. Secondly, because deception requires that one cause a second individual to believe or act on misinformation, it requires the attribution of mental states, such as beliefs. Moreover, the experiment with the cooperative and competitive trainers has a clear theory of mind hypothesis and the results support this prediction.

### 3.6. Perspective-Taking

The last indicator is that of perspective-taking, which is different from role-taking in that the role taking experiments require the animal to predict what a subject might do next to solve a problem, whereas perspective-taking requires the animal to make the connection between 'seeing' and the propositional attitude of 'believing'. In other words, if a primate has visual access to an event or an object, they are likely to behave in consequence of this knowledge if they understand the relation between seeing and knowing (1998:107). Heyes divides the perspective taking experiments into two types: 'seeing and knowing' and 'seeing and attending'.

The seeing and knowing experiment is a two-stage experiment much like the one for deception. The ape watches a trainer bait one of four containers although the ape cannot see which of the four is baited. The trainer then leaves the room. Two other trainers come into the room and each point at a container. One trainer is the knower, who knows where the treat is, and the other trainer is the guesser, who does not know where the food is. The theory of mind hypothesis is that the ape should learn to point at the knower more often than the guesser. There is a second stage to the experiment, where the guesser trainer has a bag over his head. This second stage was added to ensure that the primate's discrimination was based on the trainer's visual access to the baited container, and not an association made by the primate that is based on some visual aspect of the knower's appearance. Results were initially poor, leading Heyes to propose that animals learned a new discrimination between bagged and unbagged trainers.

In the experiment to test seeing and attending, apes were rewarded for making begging gestures in front of a pair of trainers in a variety of poses. In one situation, one trainer wore a blindfold on his eyes while the second wore the blindfold on his mouth. Another trial involved one trainer who was turned completely away from the ape while

the other trainer was also turned away but his head was turned toward the ape. The apes should pick the trainer whose eyes they could see to beg food from.

In these two sets of experiments the results are not very strong, leading the researchers themselves to doubt that they have found evidence of seeing and knowing or seeing and attending. Surprisingly, Heyes defends the results, claiming that the task of the experiment for the apes was ambiguous and does not thereby provide negative proof for the indicator of seeing in apes. There is a possible confounding factor to both these experiments. It is claimed that in the primate world, staring at one conspecific by another is a threat. Given this, it is possible that primates do not use visual gaze to inform themselves of some state of affairs or mind other than impending aggression but rather use some other form of body language.

Heyes then considers hypotheses for both sides in the indicator of perspective taking. This is the first time she has mentioned hypotheses for any of the experiments. I am not convinced that this is because no hypothesis had been stated by researchers for any of the other experiments, as she maintains. I think she has just failed to mention them, thinking they have no bearing on the interpretations of the results. In her view, experimental outcomes alone ought to be able to determine whether theory of mind or non-mentalist explanations are at work, so there is no need to mention hypotheses. She then states that the Povinelli experiments were presented as if certain outcomes would have supported a Theory of Mind interpretation over a non-mentalistic account. Is this statement to be interpreted as meaning that the Theory of Mind hypothesis would predict a different behavioral outcome than the non-mental alternative? I think a distinction must be made between hypotheses and actual experimental outcomes, the two are discussed interchangeably as if there was no difference between them. As I mentioned earlier, all sorts of ad hoc alternative explanations can be given for the results, especially when there is no hypothesis made by the researcher conducting the experiment. Concerning her idea that certain outcomes support one theory over another, is this ever in fact possible without a clear mention of hypotheses? In my opinion, there exist no experimental results that Heyes cannot advance an alternative explanation for.

In any case, Heyes claims that the Povinelli experiments do not represent a true difference in rival explanations. In other words, although the simple discrimination

procedure upon which the experiments were based tell us which cues the primates use, they do not tell us why they use these cues instead of others. It will be recalled that this is the validity criterion: it asks about the psychological processes that led to the primates using these cues instead of others. Would it be practically possible to implement this difference in the two explanations, i.e., why the primates use these cues instead of others? As I mentioned earlier, in the validity criterion there should be some sort of evaluation of the mental state that is inferred on a Theory of Mind explanation, in my opinion. That is, the real distinguishing feature of the Theory of Mind hypothesis is that it makes reference to mental state attribution whereas the non-mental hypothesis does not. Merely demonstrating that the results rule out all alternative explanations says nothing about whether postulating the mental state as an intervening variable of the mental state is justified or not. Is there a way to implement this distinguishing feature of warrant into an experimental design?

**4. Conclusion: Empirically Equivalent Explanations**

I agree with Heyes that the indicators she discussed, with the exception of deception, social relationships and role-taking, are not ideal indicators of a Theory of Mind. However, I think her line of reasoning leading up to this conclusion is mistaken. I have mentioned that her construal of the validity criterion is insufficient as a tool of evaluation. I now want to discuss what I think the source of the problem inherent for the three failed indicators is, that is, the apparent situation of empirically equivalent explanations.

Out of the six indicators surveyed by Heyes, only two were found to pass the competency criterion, i.e., were behaviorally evidenced in the experiments, those of social relationships and deception. As mentioned, her competency criterion runs contrary to the type of experiment involved, where the issue is not the presence or absence of the behavioral result, but whether the behavioral result warrants the attribution of mental states or not. None of the six were found to be valid indicators of a Theory of Mind, that is, none of the indicators could rule out alternative explanations. We thus seem to be in a situation of empirically equivalent explanations for a piece of experimental behavioral data. Given these results, Heyes main claim gains some credibility; perhaps current

experimental design cannot decide between rival explanations. The case could be made that this phenomenon is due to overlap in explanations. On Heyes' definitions, there is indeed some explanatory overlap between the Theory of Mind and the non-mental alternatives. After all, both sides are able to account for the experimental results. However, even if there is some degree of overlap, the two explanations are not identical. The two explanations are deemed rivals in the first place because there is some extra element appealed to by the theory of mind side over the non-mental side, and it has to do with the postulation of mental states as intervening variables.

Heyes doesn't report on any experiments where the two competing explanatory camps really do advance different hypotheses. That is, the behaviorist hypothesis would predict result A and the theory of mind hypothesis would predict result B. The following experiment is a theoretical one conceived of by Daniel Dennett, and concerns a typical behaviorist stimulus-response hypothesis. A rat is trained in a Skinner box to take exactly four steps forward, press a bar with its nose to obtain a food reward. If the bar were to be suddenly advanced so that the rat had to take a fifth step in one of the experimental trials, Skinnerian behaviorism would not be able to predict that the mouse would take the fifth step necessary to get the reward. The laws of behaviorism would dictate that the rat would only take four steps and jab the air with its nose (Dennett, 1978:14). Setting aside the fact that this experiment is not concerned with non-mentalist versus theory of mind explanations per se, the fact remains that the behaviorist camp is at pains to predict the results of this experiment. The mentalist side, treating the rat as an intentional system, would be able to make a prediction about the rat's behavior, and it would be different from the behaviorist prediction, or lack thereof.

It has been claimed that mechanistic explanations cannot explain novel or spontaneous actions or actions that have not been solicited by stimuli. Predictive power of the behaviorist camp will drop off sharply if either the experimental conditions are changed, or the soliciting stimulus is removed altogether. As we saw for many of the experiments described in the chapter, Heyes' non-mental explanation, particularly when it was based on a standard S-R framework, has very little predictive power. Without a history of previous stimulus-response pairings to refer to, behaviorist theory will often resort to explaining a piece of behavior after the fact.

Dennett, in a critique of the notorious behaviorist B.F. Skinner ("Skinner Skinned" 1978), cites two categories that behaviorism has trouble accounting for: those of novelty and generality. Novel or spontaneous behavior is characterized by the fact that it is different from previous behavioral responses to the same situation in the past. Behaviorism has difficulty in accounting for novel behavior because, as the experiment mentioned above demonstrates, the sameness of the stimulus dictates that the animal's response will also be the same. By generality, I think that Dennett is referring to a behavioral response that is a result of generalizing from a previous situation that is not similar enough to be generalized from except by the process of abstraction or learning. Behaviorism has difficulty accounting for this type of response because the stimulus is not the same in both cases, and without previous history of S-R pairings to refer to, behaviorism cannot account for the process of learning.

A second example of an experiment where two rival explanations are translated into different predictions of the results is the following taken from Stephen Budiansky (1998:95). The experiment has been cited as demonstrating a rather complex cognitive capacity in primates, namely the ability to build up complete and correctly ordered lists from pairwise chaining trials and then run through the complete lists to make correct judgements about the relative order on non-adjacent items. In the experiment, primates were trained with rewards to choose E over D, D over C, C over B and B over A. They were then presented with a novel choice, such as D versus B. It was found that the primates consistently chose D, even though both choices had previously been rewarded with equal frequency. In this experiment, Budiansky claims that the behaviorist model would have predicted a totally different result from the one obtained. The behaviorist model would have predicted, given the fact that B and D had been rewarded with equal frequency, that B and D would have an equal chance of being picked.

The two above described experiments are ideal in that each camp had a different hypothesis of the experimental outcome, but they are not Theory of Mind experiments. Is it possible to implement this distinction between rival explanations into a theory of mind experiment? Theoretically we are looking for an experiment where the theory of mind theory would predict a different outcome from the non-mentalist alternative. The difference in result is a function of the fact that theory of mind explanations postulate a

mental state as a kind of intervening variable. Non-mental explanations do not postulate such a variable. Let's consider the experiment on perspective taking that Heyes rejected as a good theory of mind demonstrator even though it was presented as such by Povinelli and Eddy. It will be recalled that Povinelli and Eddy heralded the experimental design as being decisive in the theory of mind versus non-mentalist debate in that it would demonstrate the seeing and attending phenomenon, a phenomenon not accountable for by the non-mentalist side. Heyes neglects to mention the researchers' hypothesis. We may guess that it is the following: if the primates understand the relation between seeing and attending, they will make use of the trainer's gaze or body posture to determine whether or not they should beg food from them. The primates should then only make begging gestures to the trainers who are looking at them or turned toward them. Results were only at chance on some trials, leading the researchers to doubt that primates understand the relationship between seeing and attending. If the chimps had performed better, the researchers would have been able to conclude that primates do understand the relation between seeing and attending.

To sum up, I have found five identifiable problems with Heyes' survey of the theory of mind research, that contribute to her conclusion that current experimentation is indecisive in theory of mind research. The first is that her definition of theory of mind theories is too strong, it includes the unnecessary stipulation that individuals with a theory of mind must possess mental state concepts as well as hold the belief that mental states cause behavior. On the other hand, the second problem is that her distinguishing feature between non-mentalist and theory of mind theories is not strong enough to prevent collapse between the two theories. The third problem is that Heyes fails to mention hypotheses, which I have argued are necessary to an appropriate evaluation of the experimental results. The fourth problem is that Heyes' competency criterion is incompatible with the experimental design employed in theory of mind experiments. As mentioned, the issue is not whether or not the behavior is in evidence in the experiment, for it usually is, but rather whether the behavioral result obtained warrants an explanation that appeals to mental state attribution or not. And fifth, Heyes' validity criterion is also incompatible with the current experimental design employed in theory of mind research.

Since neither side can declare a victory in terms of explanation, we end up in the standoff situation of empirically equivalent hypotheses.

Concerning research into the possibility of a Theory of Mind in primates, I do not believe that the project has no value and should thus be discontinued. On the contrary, what is needed is more theoretical input, hopefully from philosophy, on issues such as what exactly constitutes a theory of mind theory on a behavioral level, on what basis could one design experiments that would produce truly competing hypotheses and explanations as well as other theoretical issues. The most important source of input coming from philosophy would be to develop or discover a trait that would distinguish the theory of mind or other mentalist type explanations from stimulus or mechanistic types. In chapter six, after much looking around, I claim to have discovered a theory that contains just such a trait.

**Interim Summary- Conclusion to Part One.**

I hope I have demonstrated that there is no conclusive reason to halt or abandon the search for mental states in animals. The challenge to the scientific status of cognitive ethology is perhaps not as grave as Bekoff and Allen report, but the situation, in my opinion, certainly required closer examination. Thus the objections that I have treated in detail in the first four chapters, while they might have had prima facie value, turn out not to have the force they appeared to have once examined in further detail at close range. One of the benefits of examining the objections at such close range is discovering the grain of truth that can be retained from each of them. Generally speaking, Allen and Bekoff's opinion on the issue is particularly apt here: the difficulty in determining whether or not animals have mental states should not be taken for the impossibility of doing so.

It is possible to distill a thread running through the four objections, that is to say, two elements common to all objections, having to do with the relation between language and anthropomorphism. The common thread through all the objections is that they all reduce to anthropomorphism construed as an error of categories, with language as the most often cited distinguishing factor between the two categories. I have explored this thread somewhat at the beginning of chapter three, claiming that the lack of human language in animals renders the problem of other minds doubly intractable as compared with humans. With regard to methodology, the lack of language in animals makes experimentation that much more difficult than it is in the case of humans.

What do then we retain from each chapter that represents the moral or 'grain of truth' to each objection? From chapter one the mundane conclusion was that animals certainly don't possess human language and that concluding this doesn't amount to saying much. What we should retain from chapter one is a motivation to look further into animal language because there is good reason to believe that it is a language when construed as a system for the communication of information. We retain from chapter two that, lacking a so-called mid-level language (between mindless and mind-full) to describe

things like animals, computers and the like, perhaps we should exhaust the possibilities of our existing language first, intentionality-imbued though it is. Anthropomorphism, when it is used to develop hypotheses, is thus sanctioned under the circumstances. Examination of the other minds problem in chapter three brought the benefit of uncovering a new research area and method that shows promise with regard to the search for mental states in animals. It also brought into sharp relief the anthropocentric bias that colors the views of many detractors, as well as much of the research in cognitive ethology. Chapter four brought out the most important and damaging objection made to cognitive ethology on a methodological level, that theory of mind or other non-mentalist explanations are indistinguishable from mechanistic ones. Thus there is no need to postulate mentalist explanations because they can't even be demonstrated from a methodological point of view.

I thus conclude that while I may have gotten us around the objections to the search itself for mental states in animals, I have not succeeded in entirely solving the methodological objection. The fact remains that even when experimentation is properly carried out, behaviorist explanations apply equally well to the data as do those that postulate intervening mental states. This lacunae in mentalist explanations, i.e., the inability to rule out other competing hypotheses and explanations, remains with us into the second half of the dissertation.

## Introduction to Part Two

The aim of the second part of this thesis is to examine two areas of current research in cognitive ethology: the attribution of mental states and the attribution of concepts to animals. In the first two chapters of this second half I will examine some of the research actually carried out in cognitive ethology at the present time with regard to theories of intentionality, given a set of four guidelines that I have developed as a result of discussions in the first half. The result should be an intentionality 'guide theory' that navigates the search for mental states in animals on a path that gets neatly around the various limitations imposed by the objections visited in the first half. Two of these guidelines arise from the discussion on language in chapter one, that a good theory should have an account of error and should also be able to identify the content of mental states to a certain extent. The third guideline comes from chapter three's discussion of the empirical tractability of the fifth aim in ethology, a good theory must be empirically demonstrable in animals. The last guideline comes from chapter four's discussion on experimentation and also constitutes the leftover problem from the first half: a good theory of intentionality must be able to distinguish between intentional and non-intentional behavior, that is, demonstrate a clear victory for explanations postulating mental states over mechanistic ones.

What exactly does the word intentionality mean and why have I grouped the four theories examined under the heading of 'theories of intentionality'? The word intentionality has at least two meanings, one corresponds to the ordinary-use of the term and the other corresponds to its technical philosophical sense. Understood in its ordinary use sense, the term connotes purpose, and so we are trying to determine whether the mental states of animals, if they exist, are purposeful. A loose analogy can be drawn between this ordinary-use connotation of intentionality and some terminological distinctions referred to in the first half of the thesis. In the second chapter, reference was made by Pamela Asquith to Purton's distinction between Agent-purposive and Organic-purposive behavior. Both types of purposive behavior would be, generally speaking, considered intentional behavior in the sense that both types are purposeful. As will be remembered, however, the distinction between A and O is that only with A-purposive

behavior, the agent is aware of the goals of the behavior. This notion of agent awareness can be compared with the distinction mentioned in chapter one made by Malcolm concerning two kinds of belief, one referring to the content of the belief itself and the event of having it, and the other type of belief that is higher order, in other words a belief about a belief. In this higher order belief type, the agent is aware of his or her belief. Michel Seymour appears to have captured the distinction alluded to by Purton and Malcolm most succinctly with his distinction in types of belief, namely between material and intentional beliefs. In having a material belief, if I have interpreted Seymour correctly, the agent is not necessarily aware of having the belief, whereas having an intentional belief requires that the agent be aware of having the belief (1999:312-3). This notion of intentional belief can be compared with A purposive behavior and Malcolm's notion of a higher-order belief (although should not be equated with these notions) in that in all these cases, the intentional mental state or behavior is reflexive, there is self-awareness of the state on the part of the agent. This notion of reflexiveness or agent-awareness also corresponds to the technical sense of the term intentionality.

Understood in its most general technical sense, the term intentionality merely connotes a directedness of mental states toward objects. On this reading, we are trying to determine whether mental states in animals are intentional in the sense that these mental states are about something. I will be using the term intentionality in this general sense rather than in the sense alluded to above, i.e., as self-awareness of ones mental states, since at least two of the theories discussed employ the ordinary-sense construal of intentionality as purpose or intention.

I am thus looking for a guide theory to the search for intentional mental states in animals. To this end I start with an examination of a theory conceived of by Anthony Dickinson and Cecelia Heyes, behaviorist in nature, that specifies three criteria that have to be met in order for a bit of behavior to be deemed intentional. From there I look at a theory conceived of by David Beisecker, normative in nature, that carves intentionality on a normative dimension. Following this I next examine Daniel Dennett's Intentional Stance, pragmatic in nature, that claims degrees of intentionality within the intentional realm, and finally I examine Jonathan Bennett's theory that offers a solution to the methodological problem, i.e., that offers a way of distinguishing intentional behavior

from non-intentional behavior. I evaluate these four theories according to four guidelines that I have compiled, also as a result of discussion of some of the objections in the first half. The idea is to delineate the constraints out of which any satisfactory theory must evolve if it is going to help guide the search for mental states in animals. The bonus is that out of this examination, we end up with a solution to the methodological problem left over from the first half.

In chapter seven I am looking for a theory that attributes concepts to animals, with the caveat that it must not depend on language in order to get around the 'lack of language in animals' problem. This time in the search for such a theory I hit a theoretical snag. According to Chater and Heyes, the term 'concept' cannot be understood in a way that is independent from natural language. Many of the theories of concepts on offer in the literature are dependent on language, or at least rely on language for their elucidation. I disagree with Chater and Heyes' opinion on this issue and so take the opposite point of view through an examination of some of the more popular theories of concepts in the literature. I end up proposing a theory mentioned in chapter one, that of the eminent ethologist Colin Allen, that offers three behavioral criteria for the attribution of concepts.

# Chapter Five
# Intentionality I

## 1. Introduction: Two Senses of Intentionality

As I hope to have made clear in the first half of the thesis, there is no good reason not to think that animals might have mental states. Proceeding on the supposition that animals might have mental states, it is natural to ask whether they do have mental states and if so, what is the nature of these mental states. In other words, if they exist, are the mental states of animals intentional, do these mental states have content, and if so, what is the nature of that content?

The goal of the next two chapters is to examine several theories of intentionality that have been specifically developed for applicability to animals, in order to determine what an adequate theory would need to minimally consist of. In the case of humans, some form of intentionality of mental states has already been established, so to speak. There are numerous theories that have thus been developed to account for this already existing intentionality in human beings. Cognitive ethologists have lately wondered if perhaps the behavior and even the mental life, if there is any, of animals also evidences intentionality. Since researchers don't have the luxury of being able to study the verbal locutions of animals to determine if they are intentional or not, their behavior becomes the next most obvious site of examination. According to Colin Beer, the attempt to determine whether the behavior of animals is intentional is completely misguided. He speculates that the use of intentionally imbued language in descriptions of animal behavior constitutes a latent and pervasive kind of anthropomorphism (Beer, 1997:205). I'm of the opinion that this speculation is unduly pessimistic in that it assumes that an error of categories has already been made in the attempt to attribute intentionality to animal behavior.

The philosophical use of the term "intentionality" must first be distinguished from the ordinary use of the term. I suspect that two of the four theories visited in the next two chapters employ a sense of intentionality that conflates the ordinary-use term with its

philosophical counterpart, and this leads to unnecessary confusion. The verb 'intending' and the noun 'intention' in ordinary language are most often meant as synonyms for purpose. This ordinary-use connotation refers to just one type of intentional state, understood in its philosophical technical sense, among others such as believing, desiring wishing etc. The ordinary language terms of intending or intention have no priority amongst the various attitudes, they are no more basic or important (Searle, 1983:3). However, one can understand the technical sense of intentionality by taking the ordinary use of the term as a jumping off point.

In order for the reader to properly follow the discussion in the next two chapters it is necessary to flesh out more of the various features that have been associated with intentionality in its technical philosophical construal. The description of features that follows borrows largely from Ruth Millikan, John Searle, Colin Beer and David Beisecker's understanding of the term. Intentionality is that property of many mental states and events by which they are directed at or about or of objects and states of affairs in the world. This feature of 'aboutness' or 'directedness' is intentionality (Searle, 1983:1). In Millikans's opinion, the word intentionality understood in its technical sense is used by philosophers to refer to items that are 'about' other things (Millikan, 1997:194). As a first pass, we may say that intentionality thus encompasses the propositional attitudes (Beer, 1997:21). However, not all mental states are intentional (Searle, IBID). For instance, some mental states such as sensations or anxiety or dread do not have an object, are not about anything and are thus not intentional. The commonality amongst intentional states is that they are attitudes toward or about something, they have content (Beer, IBID).

One consideration that will help the reader to understand the term is to ask what the relation is between the intentional state and the object or state of affairs it is in some sense directed at. The answer, according to Beisecker, is that intentional states are objects of the mind that are directed at (are about) things and happenings in an external world (Beisecker, 1999:282). Every intentional state consists of a representational content in a certain psychological mode (Searle, IBID). For instance in the sentence "John believes that Jack will leave the room", the representational content, often

introduced by a 'that-clause', is " Jack will leave the room" and the psychological mode is "believes".

The reader might get the impression based on the above discussion that mental states are intentional states, that intentional states encompass the propositional attitudes, that a paradigm propositional attitude report is of the form 'X believed that Y' and thus that intentional states might take the form of linguistic sentences inside the head. The impression is further reinforced by the fact that John Searle's theory of intentionality applies primarily to speech acts. For this reason Ruth Millikan's understanding of intentionality is included, since it is more general and applies particularly well to animals with an as yet undiscovered language. Her version is not so much different from Searle's, but rather more apt for animals since her target is not restricted to speech acts.

According to Millikan, external items that exhibit intentionality are called representations. All cognitions, including beliefs, hopes and desires are inner representations. To attribute intentional purposes to an animal is to attribute to it some kind of inner representational system, some way of mapping the world and its goals, which serve as its means of achieving those goals (Millikan, 1997:194). Notice here that on Millikan's view, intentional states are not necessarily construed uniquely as propositional mental states, although the possibility is there. Intentionality applied to the animal comes in some form of a representational system, that will be used by the animal to achieve its goals.

In this chapter, I will examine two theories of intentionality. I will first look at Cecilia Heyes and Anthony Dickinson's theory, in the form of behavioral criteria for the attribution of intentional states to animals (1990). They are motivated by the following argument: Contemporary cognitive ethologists have attributed intentional states to animals on the basis of passive observation of their behavior under free living conditions. Since intentionality is not directly manifest in behavior, such observation, however careful, can be misleading. Their aim is thus to specify the behavioral criteria that must be met if an action is to warrant an intentional account. In their opinion these criteria cannot be applied through passive behavioral observation in an uncontrolled environment. An interesting point to note here is that implicit in their argument is a bias against naturalistic observation. Heyes and Dickinson mention that one of the downfalls

of naturalistic observation is that intentionality is not directly manifest in behavior, and so without intervention in the form of experimentation, intentionality will not be capturable. The question is how will experimental manipulation better capture intentionality if it is not directly manifest in behavior in the first place? Furthermore, how can one develop behavioral criteria for Intentionality at all if it is not directly manifest in behavior?

I will next look at Allen and Bekoff's critique of Heyes and Dickenson's theory (1995). Allen and Bekoff find fault with their emphasis on laboratory manipulation to the detriment of naturalistic observation. The second constructive theory that I will look at is an improvement on Heyes and Dickinson's. David Beisecker's normative theory offers a different set of criteria for intentionality, arguing that a failure to distinguish different kinds of intentionality is the main problem with theories concerned with the attribution of intentional mental states to animals (1999). He initially considers and rejects other attempts, such as Millikan's biological account and Dennett's intentional stance, before eventually advancing his own theory that is based on normativity, and the importance of learning.

My aim in this chapter and the next is to evaluate four theories of intentionality that are on offer according to a set of guidelines that I have compiled from a discussion of the issues presented in the first four chapters. Each of the four theories emphasizes a different aspect of intentionality, and can be grouped according to how intentionality should be measured. For instance, Heyes and Dickenson offer a theory that has criteria that are behaviorist, whereas David Beisecker offers criteria that are normative in nature. In the next chapter, Daniel Dennett offers a theory that is pragmatic, and Jonathan Bennett offers one based on behavioral patterns and explanatory power. As to the source for the guidelines that I have developed, the first two result from chapter one's examination of the issue of language and Davidson's comments. The first issue is to account for error, and one of the guidelines is thus that a theory of intentionality should contain an account of error. The second issue from that chapter is that a theory should be somewhat able to identify the content of propositions, and this becomes the second guideline. From chapter three the challenge is put forward to find a theory or method that is practically implementable or empirically tractable, given the fact that animals are 'other minds' and that we have only indirect access to the contents of those minds. The

third guideline is thus that a theory should be applicable to animals, and implementable on a practical level. Chapter four's discussion on methodology and Heyes' comments on the issue raises the issue of finding a way to decide between competing hypotheses for a given bit of behavior. This translates into determining whether a given bit of behavior is intentional or not. The fourth guideline is thus that a theory of intentionality should be able to discriminate between behavior that is intentional from behavior that is not. This guideline ensures that candidate theories of intentionality do not beg the question. The two theories discussed in this chapter as well as two more in the next will be evaluated according to these four guidelines.

**2. Behaviorist Criteria**

Heyes and Dickinson inform the reader at the beginning of the article that they will adopt a realist view of intentionality rather than an instrumentalist view (1990:88). An instrumentalist view maintains that beliefs and desires do not have an existence of their own. A realist view entails that the intentional account of action is a type of causal explanation, and also that these states have a separate existence of their own. Adopting a realist view, on Heyes and Dickinson's account, means that it is necessary to translate intentional explanations into counterfactual claims. Counterfactual claims entail what the animal would have done if circumstances had been different from those that actually occurred. This makes clear the reason for their attitude toward naturalistic observation. It is their view that the main problem with naturalistic observation regarding the attribution of intentionality is that since no manipulation is involved one cannot view more than one set of environmental circumstances at a time and one cannot vary any of the conditions systematically in order to evaluate the counterfactual claims (1990:87). Naturalistic observation is thus unhelpful because one cannot manipulate environmental conditions to find out if the animal's actions are intentional. This point becomes important in Allen and Bekoff's critique later on in the chapter. Another preliminary point to be made about Heyes and Dickinson's theory concerns one of the main characteristics of intentional actions. They state that intentionality is a property of an agent with respect to a particular action rather than of the agent per se (1990:91). Thus some but not necessarily all of an agent's actions may be intentional.

There are three components to the theory behind Heyes and Dickinson's intentional account of action: instrumental beliefs, desires, and a practical inference process. In order to warrant an intentional account, a behavior or an action must be represented by an instrumental belief that has a content similar to: action A causes B to occur. The animal must also have a desire, the content of which includes the goal of the action. The practical inference process will then specify how the instrumental belief and the desire interact to produce a third mental state, an intention (1990:89). The content of this intention is often represented as an action verb such as 'perform' or 'approach'. Heyes and Dickinson list two noteworthy features of the theory. The first is that the explanation is causal in the sense that it is the interaction of the belief and the desire that determines the content of the intention. If one of these elements, the instrumental belief or the desire, is missing from the account, the relevant intention will not be produced and the action will not result. This entails that every intentional action, on this account, must have at least one instrumental belief and at least one desire to produce it (1990:89).

The second feature is an assumption of rationality. As we will see in the next chapter, this assumption comes from Dennett's account of the Intentional stance (Dennett, 1987). The assumption is that, by and large, an animal's behavioral patterns, like a human's, are rational. Applied to Heyes and Dickinson's account, the rationality assumption requires that the action must be a rational outcome of the belief and desire interaction. Their motivation for this feature is the same as it is for Dennett: with the rationality assumption, predictability and empirical tractability of the animal's behavior is possible. Without the rationality assumption, it would be impossible to predict the outcomes of the animal's intentions and impossible to empirically evaluate these same intentions. Heyes and Dickinson admit that this theory is little more than rudimentary. They note for instance, that it is a single factor analysis of intention in that it isolates only a single belief and desire in each case of action. It thus fails to account for how competing desires are resolved in action. They also note that it lacks an account of the individuation of the contents of belief and desire (1990:89).

In order to get from the rudimentary theory described above to a proper behavioral account of intentionality, there is a middle step, and it is the following. There are two counterfactual claims supported by the theory, namely that an action would not

have occurred in the absence of the appropriate belief, nor would it have occurred in the absence of the appropriate desire. These two counterfactuals suggest two corresponding behavioral criteria that must be met in order for an action to warrant an intentional account, and they are the belief criterion and the desire criterion (1990:90).

Belief is much easier to translate into a behavioral criterion than desire, according to Heyes and Dickinson, especially when one is attempting to design an appropriate experiment that will isolate the desire variable while holding all other factors constant. Heyes and Dickinson model their belief criterion on the idea behind the Looking Glass World from the book Alice in Wonderland (Carrol, 1916). In a Looking Glass World, things tend to retreat when you run towards them and run after you when you attempt to retreat from them. Designing an experiment with this idea, where a food bowl retreats when the animal approaches it, Heyes and Dickinson predict that the animal should modify or at least remove from its repertoire the belief that approaching the food bowl will give access to the food. The removal of the belief is contingent upon whether or not the behavior of the animal is sensitive to the environmental contingencies that support a belief with the appropriate causal content. If an action is acquired under contingencies that would support a contradictory belief, then the action does not warrant an intentional account. In this case, if the animal persisted in approaching the retreating food bowl, then the environmental contingencies would support a contradictory belief, namely that running after a retreating food bowl will give access to food. The action of the animal would not warrant an intentional characterization in this case (1990:92). On their account, generally speaking, if a behavior appears to be relatively insensitive to its causal consequences, then it is non-intentional.

In order for an action to warrant an intentional account, it also has to pass the desire test. In the case of desire, Heyes and Dickinson have determined that if desire is significantly reduced or diminished in the animal, the performance of the action to satisfy the desire should decline (1990:93). This is great in theory, but designing an experiment where desire is isolated and manipulated, and all other factors are held constant, is extremely difficult. This is because one cannot be sure that desire has been properly identified and isolated and that all other factors are held constant. Heyes and Dickinson have designed an experiment based on what is known as the 'irrelevant incentive test'

(1990:93). The variable that is meant to represent desire is thirst. It should be noted here that employing a variable such as thirst, which is often thought to be one of the only instincts in animals and thus not under conscious control, could be problematic as a representation of a desire. The reasoning behind this experiment is that change in the desirability of the goal or incentive should affect instrumental action. The experiment is as follows: two groups of hungry rats are trained to press a lever and pull a chain concurrently for two rewards, either food pellets or a sucrose solution. One group is rewarded with food pellets for pressing the lever and the other is rewarded with a sucrose solution for pressing the lever. Both groups are then trained with the same rewards this time for pulling a chain. The test portion entails satiating the rats so that they are not hungry, waiting until they are thirsty, and seeing which group will press the lever to obtain the sugar solution in the absence of any rewards. According to their theory, the group that has initially been trained to press the lever to obtain the sucrose solution should press more than the group initially trained to press the lever to obtain the food pellets. This is because the motivational state of the group has shifted from hunger to thirst, thus the group that has been rewarded with a thirst-quencher for their lever pressing efforts in the past will be the group to press the lever more during the test period. The desire for food should become reduced in the test phase, since the rats have just eaten, whereas the desirability for sugar solution should increase. The group that had been trained to get the reward by performing the appropriate action in the training phase (even though they were hungry at that time as opposed to thirsty) should then perform this same action in the test phase in the absence of rewards, presumably because they are now thirsty. Heyes and Dickinson remark that, to their knowledge, the only animal that passes their behavioral criteria for intentionality is the behaviorist's prototypical example of a non-intentional, stimulus-response habituated creature, a rat engaged in lever-pressing in a Skinner box (1990:94).

In my opinion, this experiment does not demonstrate that changes in the desirability of a goal can affect instrumental action, although it might appear to support this claim. In the training phase, rats were trained to execute an action in order to obtain a reward or incentive irrelevant to their state of hunger, which was sucrose solution. The fact that they performed the appropriate action in the test phase given their state of thirst

does not allow Heyes and Dickinson to conclude that it was indeed a change in the desirability of the goal as a potential thirst quencher, and not the pre-training, that cause the rats to press the lever. The fact that the sucrose solution was an irrelevant incentive in the training phase and not so in the test phase does not rule out other possibilities as to why the rats performed the right action. They are certainly not warranted to then conclude that the only case of intentionality in animals comes from stimulus-response training rats. The most obvious problem with this experiment is that the two groups being tested were thirsty to the same degree in the test phase. If behavior is indeed contingent on thirst then a more obvious way to test changes in thirst would be to have the two groups thirsty to different degrees. Moreover, this experiment seems to lack a control or comparison group, in which no variables are manipulated. If a control group was included, the idea would be to not manipulate this group thirst, so that changes in the test group's thirst levels could be compared with this test group.

Heyes and Dickinson's conclusion of sorts is the following: Their method suggests that in order to find out whether an animal's actions are intentional it is necessary to measure the effects of changes in the environment which bear on the animal's mental states. Many behaviors that appear initially to be intentional fail to change under the influence of new environmental contingencies. Therefore, naturalistic observations of behaviors provide no reliable information about the intentionality of animal action (1990:94).

**3. Objections and Evaluation**

Heyes and Dickinson treat three potential objections to their theory, but as we will see, most miss the mark of what is truly problematic about it. The first possible objection that they have identified is that their theory has an inherent anthropomorphic bias against identifying intentionality in animals (1990:94). That is, The method will tend to yield false negative conclusions because it presupposes that scientists can reliably identify environmental contingencies and motivational states that will affect the content of the animal's intentional states. To this objection they agree that, indeed their account requires the identification of conditions that bear on the content of mental states, but that this identification is easier done in the case of belief than desire. In the difficult case of

desire, failure of environmental contingencies to change the desire state is subject to two interpretations. Either the behavior is truly not intentional or the experimental manipulations have failed to change the desire state. Heyes and Dickenson admit that there is no principled way of deciding the issue. They are of the opinion that in the situation where the belief has been manipulated but the desire fails to be, they should err on the side of caution and remain agnostic about the intentional status of the animal.

It could be further objected that in the case of belief the fact that the account requires all beliefs to be true or veridical is too stringent a criterion. In other words, it should be enough that an animal approaches the food bowl at the sight or sound of a stimulus in order for the animal's behavior to be deemed intentional. The act of approaching a food bowl is presumably indicative of the belief that approaching the food bowl will give the animal access to food. The second part of the experiment, where stimulus contingencies are reversed or changed, is deemed necessary by Heyes and Dickinson to prove that changes in environmental contingencies affect an animal's mental states. In their view, changes in environmental contingency should produce changes in beliefs, which are necessary to demonstrate intentional behavior. In Heyes and Dickinson's opinion, it is not enough to show merely that the animal has a belief, for this would be too difficult to demonstrate empirically, since it would be hard to isolate the belief and show that it is there. It is much easier to demonstrate the existence of a belief empirically by showing changes in that belief.

As a rejoinder to this objection, it has been shown even in humans that certain behaviors will persist in the face of absent or negative reinforcement. That a human's behavior will persist in the face of negative reinforcement does not mean that the human's beliefs are not intentional. Evidence of false negatives in the animal's behavior will tend to miss attributing intentionality where it should be attributed. Heyes and Dickinson's account will thus fail to attribute intentionality where it should be attributed, namely in the case of behavior that is intentional and that persists in the face of absent or negative reinforcement. Their answer to this is to maintain that it is necessary to insist that all beliefs be veridical in order to be able to test intentionality empirically. Merely testing if the animal approaches the food bowl, on an omission schedule at least, would only show that the content of the animal's belief does not veridically represent the

contingencies of the world. For Heyes and Dickinson, this is not enough to warrant an intentional attribution. The system must additionally be capable of detecting the extent to which the contents of its mental states match the states of affairs in the world, and also be able to adjust the content of these mental states to bring about eventual correspondence with the state of affairs in the world (1990:95).

The second objection anticipated by Heyes and Dickinson is that their theory goes against current research into human perception. It has been found that intentionality is apparently directly perceivable at least in some behavioral situations involving humans. At the beginning of their article they claim the contrary, that intentionality is not directly perceivable in behavior. The reference for direct perception is to a famous set of experiments conducted by the psychologist Gunnar Johannson in 1975 (Goldstein, 1975:307-8). Previous to the experiment a series of point-light walkers were created by outfitting several people with a string of Christmas lights attached to their limbs and filming them as they move about in the dark. The films of these walkers were then shown to a group of subjects. It was found that subjects could guess whether an object about to be picked up by one of the walkers was heavy or light, based on the walker's actions. These guesses are taken to indicate that subjects could directly perceive intentions, therefore intentionality, in the walkers. This is at odds with Heyes and Dickinson's claim that intentionality is not directly manifest in behavior. To this objection they answer that if direct perception is understood in a certain way, then the experiments on direct perception do not show that intentionality can be directly perceived. They construe direct perception as meaning that the observer cannot be misled, in other words, that direct perception is error-free perception, not subject to error on the part of the observer. With this construal of direct perception in mind, they can maintain that it is possible for the subject to attribute illusory intentionality to the walker, if the subject is mistaken about the walker's actions. Under optimal conditions, when the subject successfully guesses the walker's intentions, the intentionality is still of a derived form, since it comes from the designer of the lights.

The term 'direct perception' has nothing to do with the perception of intentionality and moreover should not be construed in the sense of errorless, that the observer cannot be misled. The term originally was coined by the perception researcher

J.J. Gibson and was taken to mean that no cognitive processing is required on the part of the observer when observing a scene (Best, 1986:90-7). The phrase 'direct perception of intentionality' if it were to have a meaning, would mean that an observer could directly perceive intentionality in behavior. If the term were to exist it would negate the entire present discussion. There would be no debate on whether a creature's behavior was intentional, since intentionality could presumable be read straight from the behavior. Moreover, Johannson's experiments were not designed to demonstrate that intentionality either can or cannot be directly perceived in behavior. The experiments were designed to demonstrate only that subjects can detect apparent uniform human movement when the lights on the walker are in motion, whereas the lights are perceived as disparate and motionless when the walker is motionless. In my opinion Heyes and Dickinson are using the ordinary use connotation of intentionality construed as purpose or intention, to interpret these experiments. They seem to have taken the subject's ability to guess whether the object a walker picks up is heavy or light to mean the same as the ability to guess the intention of the walker. They also equate being able to guess an actor's intentions with being able to perceive intentionality directly.

The third objection relates to the debate over the value of naturalistic observation versus experimental manipulation. The objection to Heyes and Dickinson's theory, which is based uniquely on laboratory experimentation, is that one is less likely to find intentionality in the lab by their methods than in the field through naturalistic observation. Contrary to their main claim that naturalistic observation is unlikely to provide evidence of intentionality, some authors argue that it is only through naturalistic observation that one can find such evidence. As we will see in the next chapter, authors such as Daniel Dennett argue that the hundreds of training trials that animals undergo in a typical experiment are hardly worthy of an intentional characterization, but are rather more representative of heavily pre-trained stereotypic behavior explicable in terms of rival conditioning hypotheses (1983:250). Heyes and Dickinson's rejoinder to this is to wonder why such authors would believe that the existence of a prolonged training history is incompatible with the attribution of intentionality. They speculate that perhaps such authors assume that, unlike S-R habits, beliefs are formed quickly, on the basis of

minimal experience. While they accept that this might be the case in some training situations, they wonder why over-training should rob an action of its intentional status.

The issue is not whether stimulus-response over-training is incompatible with intentionality, for it is compatible. That is, stimulus-response training can produce intentional behavior in the animal. The issue here is whether repeated attempts to create or modify a unit of behavior somehow mask or remove the real spontaneous behavioral reaction in the animal, which might be found to be intentional. The worry is also not that over-training an already intentional behavior would rob it of its intentional status, because this seems an impossible feat, but rather that training would create a new intentional behavior that is artificial to the animal's repertoire. An even more counterproductive feat would be to train a behavior in the animal that is artificial to its repertoire and not even intentional. One can train an animal to act intentionally, just as one can train an animal and thereby modify the animal's original intentional reaction into one devoid of intentionality. The creation of so-called false positives, where the intentional behavior is not within the animal's repertoire of behaviors but is rather created through repeated stimulus response training, is just as bad as false negatives, where the behavior really is intentional but failed to be labeled as such. The aim is to see if any of the naturally occuring behavior in the animal can be considered intentional and it is only through noninvasive naturalistic observation that such an aim can be carried out.

Heyes and Dickinson's rejoinder to this objection is to wonder why, even if intentional status is a product of training history, one would find more evidence of intentional action in the field than in the lab. The short answer to this is that since no experimental manipulation is involved in naturalistic observation, the chances of producing artificially induced intentionality through training history are minimized. This is precisely why ethologists insist on naturalistic observation: to determine what an animal's natural reactions are in view of certain natural environmental constraints that are found in their natural habitat. It is one thing to unobtrusively note how the environment and evolution constrain the animal's range of behavior, be it intentional or not, it is quite another to artificially constrain the animal's behavioral reactions through experimental manipulation and create intentional behavior in the animal. Ethologists are interested in

discovering whether the naturally occurring behavior of animals is intentional. They are not interested in creating or training behavior to be intentional.

Allen and Bekoff, in their critique of Heyes and Dickinson's theory (1995, 1997), correctly sum up the main argument of their article into four premises. The point of Heyes and Dickinson's argument is that an animal's approach to a food source does not warrant an intentional account because the animal fails to modify its behavior given opposite feeding contingencies. A schematic of argument is as follows:

1) An action A warrants an intentional account only if it is caused by an (instrumental) belief of the form "Action A causes access to some desired object O."
2) If an action A would be acquired or persist under contingencies that do not support the instrumental belief that A causes access to O, then A is not caused by that belief.
3) The action of approaching food (A) is acquired (by rats) and persists (in chicks) under contingencies that do not support the belief that approaching food (O) causes access to the food.

Hence, the action of approaching food performed by chicks, rats and by other species does not warrant an intentional account (Allen & Bekoff, 1997:167).

Allen and Bekoff assess the degree to which each premiss supports the conclusion. The problem with the first premiss is that it is possible that other intentional states besides the single instrumental belief and the single desire may be causally implicated in an action. For instance, an animal may be moved to act by a number of beliefs in addition to the single instrumental one, that are not necessarily instrumental, as well as numerous other mental states whose content involves attitudes other than belief or desire. Heyes and Dickinson's account is thus overly restrictive and overly rudimentary in the sense that it covers only simple instrumental acts. It should be recalled that Heyes and Dickinson have admitted that their account is rudimentary in the sense that it only considers the very basics of intentionality. They justify this simplicity by appeal to the fact that all inferential processes including higher cognitive abilities will eventually reduce to simple instrumental beliefs and desires. It could be argued in reply that this type of reduction, although characteristic in behaviorism, might not apply to intentional states. Perhaps

some higher-order intentional beliefs and desires and other propositional attitudes cannot be reduced to simpler, albeit also intentional, states. In the reduction to basic beliefs and desires, the issue is not whether intentionality is lost along the way, it is rather that the content of those higher order states may be implicated in the execution of the end result behavior, and that their causal implications would disappear in the reduction to simpler states.

It is possible to dispute the second premiss if the case can be made that the persistence of an irrational behavior is still compatible with causation of the behavior by a veridical instrumental belief (Allen &Bekoff, 1995:319). Allen and Bekoff advance the idea that what may seem irrational from one perspective may seem rational from another. They claim that belief persistence, despite a change in evidence, at least from an evolutionary perspective does not necessarily provide evidence of irrationality. They cite Gilbert Harman's article (1986) on belief persistence despite conflicting or negating evidence as further proof that irrational behavior can still be caused by instrumental beliefs. Reasons abound as to why beliefs persist despite evidence to the contrary, from the view that beliefs are like old habits, hard to break, to the idea that a link has not been made between the new evidence as discrediting the old beliefs (Harman, 1986:326-330). It has been experimentally tested and verified in humans that beliefs do in fact persist in the face of discrediting evidence. This belief persistence should not be equated with irrationality, nor should belief persistence be considered non-intentional. It is the view of Allen and Bekoff that animals may also evidence this tendency for belief persistence, which would invalidate premiss 2.

To invalidate premiss 3, Allen and Bekoff offer the idea that most prey runs away from the animal that is trying to catch it. In fact, with the exception of certain domesticated breeds, the food sources for most carnivorous wild animals is another animal who will try to prevent itself from being captured and eaten. This point would explain why Heyes and Dickinson got the results they did in the experiments: the chasing behavior of the animal persists when confronted with a retreating food bowl. And so, interestingly enough, the result obtained in the Heyes and Dickinson experiments, although contrary to their hypothesis and constitutive of a failure of the animal to exhibit intentionality actually exactly matches the behavior of the same animal observed in the

wild. Given this result, the behavior of chasing a retreating food bowl should be considered intentional. This further underlines the need for naturalistic observation at least as a preliminary to experimental manipulation.

The idea that in natural conditions most prey runs away from its predator reinforces the point made earlier about the value of naturalistic observation. Here is a good example of the contrast between naturalistic observation and experimental manipulation with regard to intentional behavior. It would be found through observation in a naturalistic setting that most prey runs away from its predator. Through experimental manipulation, Heyes and Dickinson have attempted to modify the original intentional behavior into some other behavioral response, especially when one contrasts it with what occurs in the animal's natural habitat. The result is intentional behavior according to them, but it appears to be artificially induced 'intentional' behavior in my opinion. If Heyes and Dickinson somehow fail to train the new behavior in the animal, the proper conclusion to draw is not that the animal fails to behave intentionally, but rather that the behavior modification attempts have failed. Through out all this, the animal's natural intentional reaction to run after prey, which is what cognitive ethologists are interested in, is masked.

How well does Heyes and Dickinson's theory fare against the set of guidelines mentioned at the beginning of the chapter? It would be helpful to examine the guidelines first in a little more detail in order to see why I have chosen them. The first guideline results from my discussion of Davidson's views in chapter one. While I don't think he succeeds in making the case that conscious awareness of error and thus knowledge of true and false belief is a condition for thought or rationality, I do think that any theory of intentionality that will be applied to animals must account for the possibility of making mistakes, otherwise it is incomplete in a fundamental way. In my opinion, the stipulation that all beliefs be veridical doesn't allow for the possibility of false belief, or error, in the animal. Heyes and Dickinson's theory cannot account for error, and there are a number of reasons offered by them as justification. The most important reason likely stems from their stipulation that all beliefs in the animal must be veridical. They defend the stipulation by claiming it necessary for empirical testability. But empirical testability would not necessarily be compromised if the theory were to allow for error on the part of

the animal. One can still develop a theory that contains an account of error that is empirically testable on animals. The requirement that all beliefs be veridical would have to be gotten rid of however, since it doesn't allow for error, in and of itself. They justify the fact that all belief must be veridical by the claim that there isn't another way to ensure that the animal's beliefs are a true reflection of states of affairs in the world. I think it is possible to construct an experiment that will better track whether the animal's beliefs match the state of the world without having to assume that they are all true. An example of such an experiment would be to set up a counterfactual situation, and if the animal fails to change its behavior, it can be interpreted as having committed an error. At any rate, their theory, lacking an account of error, fails to be adequate in a fundamental way.

The second guideline has to do with identification of propositional attitude content, particularly that of beliefs. It is designed to meet the challenge that non-verbal animals are unable to be sensitive to fine-grained distinctions in content in the way that language-speaking humans are. In one of its variations, the challenge takes the form of the underdetermination argument. It will be recalled that Davidson discussed the example of Malcolm's dog chasing a cat into a tree. He claims that the dog cannot have any beliefs about the cat in the tree since we wouldn't know what particular belief to attribute to the dog, and there are too many to choose from. A good theory of intentionality should be able to narrow down the content of beliefs and desires to a reasonable degree. Heyes and Dickinson's theory, while it might attempt to answer whether an animal's behavior is intentional or not, does not go the further step of narrowing down the content of belief and desire. This lack of an ability to narrow content stems from, in my opinion, the rudimentary variables they have chosen to manipulate such as thirst and hunger, and with the simple experimental design, a Skinner box. Moreover, the requirement that all beliefs be veridical means that identification of the content of false beliefs will not be considered. The testing of rudimentary states such as thirst and hunger ensures a restriction on the complexity and the variability of the content of the supporting beliefs. It may turn out to be the case that thirst and hunger do not translate into desire states. Thirst and hunger are not themselves mental states. 'Thirst' would have to be translated into 'desire to drink' and 'hunger' with 'desire to eat'. Moreover, the simple experimental design puts unnecessary constraints on the

variability of the actions of the animal, to the point where all responses are stereotypic and mechanical.

As we saw in chapter four, a recurring problem with data interpretation in experiments is the failure of the experiment to be able to decide between competing hypotheses. There are difficulties in establishing that an animal's actions are intentional, because the data doesn't offer a clear-cut victory of intentional explanations over non-intentional ones. As mentioned, this is in my opinion the greatest obstacle to establishing any sort of intentionality in animals. Heyes and Dickinson's theory takes us no further in this regard. If anything their experiments have established that it is possible to train artificially-induced intentional behavior in animals. They admit that their difficulty in isolating the desire criterion makes it impossible to decide whether the behavior is non-intentional (a product of stimulus-response) or whether the behavior is intentional and environmental contingencies have failed to modify the rat's desire.

The fourth guideline is practical implementability. If a theory of intentionality is going to advance the issue of animal mentality it must be applicable to animals and be empirically tractable, considering the fact that animals are 'other minds' and we don't have the access point of human language. Heyes and Dickinson, in developing behavioral criteria for the attribution of intentionality, appear to have practical considerations in mind. Given that researchers do not have the common communicational path of language with animals, they have chosen the next obvious indicator of intentionality which is behavior. Notwithstanding the fact that they could have chosen better experiments, their theory is in principle very applicable to animals and compensates for the lack of language, thus their theory passes the implementability guideline.

**4. Normative Criteria**

An important motivation for Heyes and Dickinson to come up with behavioral criteria for the attribution of intentionality, aside from the idea that it is amenable to a lab setting, is the fact that the creature under study lacks language. Any creature that lacks language, the argument goes, also lacks sensitivity to the fine-grained intensional

contexts or states of mind that human language is a perfect vehicle for, and thus cannot be attributed intentionality.

David Beisecker (1999) thinks that this point is significant for the argument that animals possess some sort of prelinguistic primitive form of intentionality, a form possibly matching an earlier evolved form in ourselves. There is more than one form of intentionality, on Beisecker's account. He thinks that what has prevented the idea that different forms exist from being considered is the popular entrenched view that there is only one type of intentionality known as 'original' and all other forms are 'derived' from the original form. This popular view is often associated with evolutionary theories such as that of Millikan and Dennett. The view is that all intentionality is of a derived form and comes from mother nature, the ultimate designer. Beisecker has been led to consider the possibility of different forms of intentionality by the implausibility of the claim that heat-seeking missiles and sunflowers have the same type of intentionality that animals have, and that this in turn is the same type of intentionality that thinking humans have. The challenge is then to find a way to distinguish the special sort of directedness possessed by bona fide thinkers from the more primitive kinds exhibited by these simpler systems (Beisecker, 1999:283).

On Beisecker's view, there exists at least two types of intentionality, one for linguistic humans and another type for non-linguistic creatures. I suspect that his motivation for creating a type of non-linguistic intentionality to apply to animals is the idea that if animals could talk, we would not hesitate to attribute intentionality to them. Just because animals don't have language, there is no reason to think that they don't also have intentionality of another sort. Otherwise the case could then be made that intentionality is dependent on language. On the other hand, there is no good reason to think that animals do have intentionality of another form. I think that Beisecker makes an assumption in order to get his theory off the ground, one which not only renders his theory somewhat question begging, but also causes it to fail one of the four guidelines. The assumption is that there exists this second form of non-linguistic intentionality and that it is associated with mental capacity. Instead of asking whether animals could be considered intentional beings, which is what the point of all the theories discussed is, he assumes intentionality in humans as well as in animals, creates a separate type of non-

linguistic intentionality and shows how it might operate in animals according to a normative criterion. As we will see, the main component of his theory, that of an expectation, is already intentional.

Beisecker thinks it makes sense to talk about different varieties of intentionality by focusing on the normativity of intentional phenomena. The hallmark of intentional states would be their susceptibility to evaluation. According to his account, a system is credited with intentional states only if it can be judged as correct or mistaken with respect to some standard or purpose. The question he begins the discussion with is: In what sense could a non-linguistic animal be said to be mistaken about the way things are? There are a variety of options. The first option is Millikan's biological account (1984, 1993). Briefly, she offers an account of how creatures of natural selection exhibit a genuine biological sort of intentionality, based on what the proper function of the organism's organs and internal mechanisms is. Whether or not a creature exhibits intentionality on this account is a matter of whether its organs and internal mechanisms are carrying out their proper function, determined according to the evolutionary history of the item (Beisecker, 1999:285-6).

Beisecker lists a number of problems with this account. One problem corresponds to a failure of my third guideline, having to do with the identification of content. The problem with theories based on biological function is that determinations of proper functions are too ad hoc or too indeterminate to underwrite ascriptions of belief or other states with propositional content. Beisecker thinks that underwriting ascriptions of belief is a necessary component of intentional theories. For instance, the proper function of most creatures from the point of view of evolution is said to be propagation of the species. It would be immensely difficult to somehow work this function into the content of a belief in a particular situation, especially if the situation appears to have no relation to propagation of the species, for instance, that of two animals playing together. The ultimate weakness with the account on his view is that it appeals to the wrong sort of normativity to be a compelling account of mental representation. The sort of intentionality ascribed to these creatures, since it has to do with natural selection, could be deemed as a form derived from the original designer, in this case, Mother Nature. Beisecker's problem with this is that the norms or standards by which the behavior is

evaluated are not set by the animals themselves, but rather by the designer of the animal, in this case, mother nature or evolution. The intentionality of these animals is thus of a derived form since they are assessed as correct or mistaken relative to the standards set by natural selection. Beisecker thinks it makes more sense to look for a type of intentionality that is original, perhaps intrinsic to the animal, and thus not evaluated only by biological or proper function.

It should be noted here that Beisecker's stipulation that theories underwrite ascriptions of belief with content is a little strong. There is no need to stipulate that belief contents should refer directly to proper function, on most biological theories of intentionality. It would suffice to be able to trace the function back to evolution, even in an indirect way. Note also that it is not necessary, although it is for Beisecker's account, that the norms or standards by which the behavior is evaluated intentionally to be set by the creatures themselves.

Looking next at Dennett's theory of the Intentional Stance (1987), Beisecker notes that Dennett is reluctant to distinguish between varieties of intentionality, for instance, between the type that humans have and the type that animals have. Dennett employs two different arguments to convince us that there is only one type of intentionality and that there is no difference between original and derived intentionality. Intuitively, and as will be seen in the next chapter from an examination of Dennett's theory, it is obvious that Dennett will not see the point of distinguishing between varieties of intentionality, since his Stance recommends that one treat all systems 'as if' they are intentional. If the question of whether systems actually are intentional or not is un-important, then there is no reason to further distinguish between its different varieties.

The first argument that Dennett advances undermines both of Beisecker's claims above, that the content of the belief must contain the source of intentionality and that the standards or evaluation be set by the creature itself. The argument, known as the "lack of intrinsic content determination" argument, states that belief contents are not completely determinable by reference to the mind alone. The usual argument offered to make this point is the famous 'twin-earth' thought experiment, where two otherwise indistinguishable beings are imagined, one from earth and one from another planet. These two beings can be shown to nonetheless entertain thoughts with different content,

due to differences in their respective environments. This argument is designed to demonstrate that propositional attitude contents are not completely determined by activity inside a subject's head (Beisecker, 1999:291). More generally it refers to a recurring problem with trying to attribute specific intentional states to creatures. The worry is the lack of a clearly identifiable and definable content in the creature's so-called intentional state, in other words the identification of content problem. Briefly, Dennett offers the reasoning that if other creatures don't use the same distinction-making method of human language to conceive of their circumstances, trying to apply this distinction-making method of language to them is not going to give up a one to one correlation, hence the identification of content problem. Beisecker is not in agreement that a slight indeterminacy of content should slide into a rejection of the idea that there exist different varieties of intentionality. In other words, just because one cannot determine with complete accuracy the contents of a dog's thoughts does not mean that there is only one type of intentionality, of a biological or artifactual sort. Beisecker is of the opinion that Dennett is able to claim the above because he mistakenly equates "intrinsic" (in the head) intentionality with "original" (Searle's term for non-derived) intentionality. So, claiming that thought content is determined exclusively inside the head amounts to claiming that intrinsic (construed as original) intentionality is the only true form of intentionality. In Beisecker's opinion, to claim that intentionality is determined in the head does not include the claim that it is the only type that exists.

The second argument of Dennett's is based on the idea that we humans also have only a derived intentionality from Mother Nature, just as any other creature, because we are all products of natural selection. As Dennett humorously maintains, the variety of intentionality that we possess is frog intentionality from frogs all the way up to humans, "(human) belief and desire are like froggy belief and desire all the way up", (Dennett, 1987:112). Since we humans also share the same form of intentionality as frogs, being creatures of Mother Nature, there is no reason to make a distinction in types of intentionality between any of Mother Nature's creatures. There would be no work for this other type of intentionality to do (Dennett, 1996:54). Beisecker is not convinced by this claim of Dennett's either, which boils down to the claim that rival accounts of intentionality preclude one another in favor of that of Mother Nature. As Beisecker

notes, accepting that we are all creatures of natural selection and thus have derived intentionality from Mother Nature doesn't preclude us from having other types of intentionality as well. These other types could also ultimately derive from Mother Nature. The content of all our attitudes cannot all be directly related to evolution and thus evaluated according to proper function. Although Beisecker does not deny that humans and other creatures can be evaluated according to proper function from an evolutionary perspective and thus partake in a form of derived intentionality from Mother Nature, he is unwilling to end the story here. He thinks that there exist different varieties of intentionality, and that these varieties exist on a normative dimension.

Beisecker's own theory begins with the process of identifying rational patterns. He is looking for several types of possible rational patterns corresponding to different ways in which one might adopt an intentional stance. One of these types is educable capacity. His theory, in a nutshell, is basically that the flexibility in behavior of an educable creature gives rise to a special sort of accountability or evaluation. This should explain how a non-biological form of intentionality could still be a product of natural selection, as Millikan's and Dennett's accounts suggest, but goes one step further. Millikan's and Dennett's accounts focuses on how a creature's educable capacity enables it to fulfill its natural purpose in the face of environmental contingencies and constraints. Beisecker's account goes one step further in that it can account for the intentionality an animal might have above and beyond that related to proper function and natural selection. In other words, he is seeking to demonstrate how animals with the capacity to learn from their mistakes could be intelligible apart from the animal's biological purposes (Beisecker, 1994:294).

The phenomenon of educability in animals can be explained by the abstract adjustment of certain of the animal's cognitive structures called expectations. "Expectations" is meant here in the everyday sense of predictions based on the outcomes of previous similar situations. Animals revise their responsive dispositions over time by being sensitive to the consequences of their responses in certain situations. In other words, an animal will eventually revise its expectations if environmental conditions repeatedly fail to respond accordingly. This educable capacity, or the ability to learn

from one's mistakes, is also the basis for Colin Allen's theory of concept attribution, as will be seen in chapter seven.

All expectation-based theories have the same abstract structure, although they might differ in the details (1999:297). There are three basic components. The first component is a condition of activation and de-activation that specifies when the expectation should be turned on and off. The second component is a consequence condition that will pick out the expected state of affairs associated with the activation of the expectation. The third component is the response component, that specifies the responses expected to bring about the consequence condition. The activation of the expectation condition can be evaluated as correct or mistaken, depending on whether it brings about the desired consequence or not. Expectation based creatures can be defined as those whose responses are governed in part by the consequence conditions of their currently activated expectations (1999:297). Beisecker gives no account of how the activation/de-activation, consequence or response components are manifested physically in a creature. Although he gives examples of two phenomena that are explainable by expectations, the problem is basically that it is difficult to identify the three components in any given example and this becomes a real problem for the theory.

There are two phenomena found in the literature of learning theory in psychology, 'blocking' and 'latent learning', that resist explanation in terms of classical stimulis response learning theory. These two phenomena are easily explained on an expectation account, according to Beisecker. The first phenomenon of blocking refers to the apparent failure of a new stimulus to become associated with an already existing stimulus-response pairing in the animal. For instance, if the S-R pairing between a bell and the delivery of food has already been established in the animal, it is very difficult if not impossible to then get the animal to respond if a new stimulus, say a tone, is then paired with the food delivery. It is as if the first associated stimulus blocks or prevents the second stimulus from being associated into a new pairing. On a classical conditioning account, the new stimulus should eventually be paired with the original stimulus, but this is not what happens, and so the theory is at a loss to explain the phenomenon. This phenomenon of blocking can be explained by expectations. The explanation is that the animal fails to generate the new expectation, that it should respond at the sound of the tone, because it is

using the original one, the bell, to predict an outcome of events with reasonable success. The success at prediction with the original stimulus prevents new pairings from being made. Although Beisecker does not spell it out, we can infer what the process is: the success of the original response component prevents its de-activation and subsequent activation of a new response that is associated with the tone, because the original consequence condition is successful for getting food.

Expectation-based accounts can also better explain the phenomenon known as latent learning. Classical conditioning theories are also unable to explain the phenomenon of latent learning, again because the pattern of explanation falls outside the power of classical conditioning. Latent learning is learning which occurs usually in the absence of reward or reinforcement, which is why classical conditioning is at a loss to explain the phenomenon. The learning is not manifested in the performance of the animal until some period afterwards, hence the term 'latent'. Examples of latent learning usually occur in pre-training phases of experiments, such as in the case where a rat that is allowed to explore a maze prior to the experiment is found to be able to navigate it more quickly in the experimental phase than a rat who has not had such previous exposure. Classical conditioning cannot explain latent learning because it cannot explain any type of learning in the absence of rewards or reinforcement. All learning is a result of reward or conditioning on their account, and if there is no reinforcement or reward to point to as responsible for soliciting the behavior in the animals, then the animal cannot be said to have learned anything. Expectation-based accounts can explain latent learning in part because they are not constrained by a pure stimulus response structure. In the experimental situation of a rat navigating a maze, the rat's previous exposure to the maze allows it to form expectations that then allow the rat to exploit this expectation-based knowledge to the pursuit of new goals, such as food rewards.

It is in attempting to generate a concrete example where the three components of the theory are identified that brings out the main problem with Beisecker's theory. Let us take the example of a primate standing at the bottom of a tree that has bananas out of reach in the higher branches. The primate forms the expectation "Shaking this particular branch of the tree will cause the bananas to fall to the ground". Presumably the condition of activation/de-activation will then be activated. The response component is presumably

that the primate begin shaking the tree branch. The consequence condition is presumably that bananas will fall from the tree. The problem here is that it is difficult to determine in advance whether or not the activation condition has been activated, and also it is difficult to determine if one has correctly identified the other two components.

Although Beisecker's account is difficult to implement on a practical level, it does not, like Millikan's or Dennett's, rest on a determination of the purposes for which a creature has been designed or selected. In fact, as Beisecker notes, the goals of the organism may even collide with Mother Nature's purposes. The bonus is that we can identify the expectation configurations that are likely to hinder a creature's attainment of its goals, and call these configurations expectation errors. Errors can be of two forms. An error of commission occurs when a creature's expectation is activated in a situation in which the expectation's response would fail to bring about the satisfaction of its consequence condition. In ordinary terms this would correspond to any situation where a creature executes a movement or an action to accomplish a goal that does not in fact accomplish the goal. A mundane example would be a baseball player who swings too late on a pitch. An error of omission occurs whenever the response of an expectation that is not activated would in fact bring about the satisfaction of its consequence condition. In this situation a creature would fail to execute an action that would bring about the desired result. To take the baseball example again, the player would not swing at a pitch that turned out to be well enough placed to be potential home run.

There are two appealing features to the account, as Beisecker points out, both of which answer challenges raised by Davidson in the first chapter. The first has to do with Davidson's claim that only a creature with language can evidence the fined grained sensitivity to intensional contexts associated with the attribution of genuine intentional states. On Beisecker's account, a creature need not have language and yet still evidence by its behavior relatively fine-grained distinctions in intentional content. The distinctions in the content are realized by differences in the creature's expectations. Distinct expectations may share the same circumstances of appropriate application, since the situations in which one expectation would be satisfied may happen to line up with situations in which another would be satisfied. The same is not true for the content of the expectations themselves however. The particular means by which the circumstances are

picked out would differ for each expectation, since the circumstances are each comprised of different expectation components. Distinctions in expectations can be demonstrated by changing the conditions of the animal's environment. So the fine distinctions in language that humans are easily able to verbalize are analogous to distinctions in expectations that an animal activates, depending on the circumstances (1999:299). The only small problem with this is that environmental conditions are what dictate the fine-grainedness of expectations. The only way to test the theory is with counterfactuals, just as it was with Chater and Heyes' theory.

The theory also accounts for the rational responsiveness to error that Davidson claims is required for any creature to be deemed rational. As we saw in chapter one, Davidson claims that rational responsiveness to error complete with the element of surprise is necessary for any attribution of rationality. On Beisecker's account, a creature with educable capacity displayed in the form of expectations has the capacity to revise expectations in the event of error, and even to take steps to avoid such errors in the future, which would indicate a capacity to learn from one's mistakes. Implicit in the recognition of error is the element of surprise, the recognition that things did not turn out as one predicted. Both these elements are present in expectation-based behavior.

**5. Evaluation**

The motivating factor for Beisecker's theory is to come up with different varieties of intentionality. The intuition is that sunflowers, computers, animals and humans cannot all have the same single type of intentionality. Generally speaking, however, I think that rather than justify the need for different varieties of intentionality, Beisecker has merely assumed different varieties of intentionality evaluated on a normative dimension and shown how this story might be played out in the case of animals. Beisecker's theory is certainly an improvement over Heyes and Dickinson's, although there are a few parallels that give it the same problems as theirs does in the realm of practical implementability.

With regard to my four guidelines, Beisecker's theory appears to pass two of the four. It will be recalled that two of the guidelines come from my discussion of Davidson's views in chapter one. A theory must have an account of error and should be able to narrow down the content of mental states to a reasonable extent. As Beisecker

points out, an expectation-based theory can account for errors made by the animal. If the animal's expectation does not bring about the required response, it can revise that expectation in order to then bring about the desired response. The theory goes even further in that it gives an account of how creatures can learn from their errors.

There remain two problems with regard to error on an expectation based account however. One is the difficulty in identifying errors, and it ultimately stems from a difficulty in identifying expectations. As mentioned, Beisecker cites two types of errors, those of commission and those of omission. While errors of commission are easy to identify; the animal executes an action or movement that does not result in its expectation being fulfilled, errors of omission are harder to identify and at the least are revisionist in the sense that we have to know what the animal's expectation was in the first place in order to note if it was fulfilled or not. The animal's lack of execution of a movement or action will obviously not give any clues. This problem stems ultimately from a failure on the part of the observer to be able to identify expectations in the animal. As with Heyes and Dickinson's account, on Beisecker's account one can identify expectations only by varying environmental conditions and thus only demonstrate changes in expectations. My guideline only asks that the theory be able to give an account of error, however, and so Beisecker's theory does pass the error guideline, although it is not a huge improvement over Heyes and Dickinson's theory in this respect.

As for the ability to narrow down the content of mental states to a reasonable degree, Beisecker points out that expectation based theories can demonstrate a certain amount of fine-grained distinctions in content. Differences in goals of the animal correspond directly to differences in the expectations of the animal. Distinctions in content of the expectations are realized by distinctions in expectations. However, as mentioned above, there is no way to identify particular expectations except by demonstrating changes in them. It is only by varying conditions in either the environment or the goals of the animal that will provide distinctions in content of the animal's mental states. Beisecker's theory thus passes the error guideline, but only barely.

Looking next at how the theory fares with regard to the third and fourth guidelines, the ability to discern intentional from non-intentional behavior and practical

implementability, it is here that problems are found. Regarding the ability to discern between intentional and non-intentional behavior, a theory is already halfway there if it can narrow down the content of beliefs and desires, and Beisecker's theory can do this. That is, if the theory can say something specific about the actual content of the animal's beliefs and desires, then it should also be able to show that explanations in terms of intentional attributions are different from the non-mental alternative explanations. Unfortunately Beisecker's theory is concerned with demonstrating that varieties of intentionality exist. It thus begs the question of whether a state is worthy of being deemed intentional in the first place. This is due mostly to the fact that his notion of expectation is itself intentional. It thus fails the third criterion.

Beisecker's theory unfortunately also fails the guideline of practical implementability. He gives no possibilities for experimentation in his elaboration of the theory. He relegates justification for the lack of this feature to a footnote. In it he states that he is not trying to show that any particular creatures are expectation mongerers because that is the work of ethologists, not philosophers. On his view, he has constructed the philosophical theory, and ethologists should be able to take this theory and apply it to animals. Judging from my difficulty in implementing his theory in a concrete example, it does not appear to be easily applicable. The problem, also stemming from his lack of furnishing concrete examples, is the difficulty in identifying and distinguishing between the three components. Moreover, according to his account of error and content identification and the fact that errors of omission and expectations can only be demonstrated by showing change in them, his theory would have to be tested with counterfactual conditions. As we saw with Heyes and Dickinson's theory, counterfactual conditions make constraints on the variety of responses an animal could display, as well as masking the possibility of error commission. In the next chapter, I will be looking at an example of a theory that has actually been put to the implementability test in the natural environment of the species, that of The Intentional Stance conceived of by Daniel Dennett.

## Chapter Six

## Intentionality II

### 1. Introduction

In this chapter I will be examining two more theories of intentionality that apply to animals and submitting them to my four guideline evaluation. The first is Daniel Dennett's famous Intentional Stance. His theory is particularly interesting because it has been adopted by cognitive ethologists and even tested 'in the field'. It thus represents a major advance in the field of cognitive ethology with regard to investigation into the mental states of animals. However it does suffer from problems, most stemming from the idea that the theory has been touted as being pragmatic. It is thus weak on matters such as predictive power and the issue of whether the mental states of animals can really be deemed intentional. The last theory I will be examining in the chapter, that of Jonathan Bennett, is an improvement on these two matters and represents the most far-reaching advance in applicability of theories of intentionality to animals to date, in my opinion. Bennett's theory has the further bonus of solving the methodological problem encountered in chapter four, that of discerning between intentional and non-intentional states in the animal.

### 2. Intentionality à la Daniel Dennett

Daniel Dennett has had a life-long interest in the notion of intentionality, dating from his first book on the topic published in 1969 called "Content and Consciousness". In that book Dennett was engaged in developing a scientific theory of the mind. He employed an abridged notion of intentionality to get him out of a well-known philosophical dilemma regarding scientific theories of mind, the presumed incompatibility of mental discourse with scientific theories. The dilemma is briefly to try to find the point of interaction between physical properties and properties of the mind, which are non-physical. If one says the two spheres do interact, one is at a loss to explain how mental events, being non-physical, can cause changes in the physical world (1969:3). In attempting to develop a scientific theory of the mind, Dennett treated the

most important obstacle to this endeavor, the intentionalist thesis. The thesis, originally conceived of by Franz Brentano, makes a distinction between mental and physical phenomena, that mental phenomena exhibit intentionality and physical phenomena do not, thus the mental mode of discourse is ultimately incompatible with the physical mode, and no translations, reductions or unifications are logically possible (1969: x). This distinction is a major obstacle to Dennett's attempted construction of a scientific theory of mind.

Brentano borrowed the term intentionality from medieval philosophy. The original conception of it is captured by his statement "Every mental phenomena is characterized by what the scholastics of the middle ages called the 'intentional inexistence' of an object, and what we would call the reference to a content, a direction upon an object." (Dennett, 1969:20). The phrase 'direction upon an object' means that one cannot want without wanting something, one cannot hope without hoping something, and yet the object in all these cases need not exist in the sense of physical objects existing. The phrase 'reference to a content' refers to the fact that in addition to the direction upon an object, intentionality can also manifest itself as a relation to a proposition. Brentano's intentionality thesis thus divides into two parts: some mental phenomena are directed upon an object, and other mental phenomena are related to a content or proposition or meaning (Dennett, 1969:20). Since then, another caveat has been added to the thesis: no statement or statements about non-intentional phenomena can have the same truth conditions as any statement about intentional phenomena.

Dennett modifies Brentano's original conception of intentionality in two ways. First, the realm of applicability that the thesis applies to will be enlarged from the more obvious mental terms to include the entire realm of psychological mental terms. This is because there are a host of terms that are not obviously mental, such as the term 'hunt' and 'search', but that still appear to warrant intentional characterization. Second, intentionality will no longer be a construct that divides phenomena from phenomena, but rather sentences from sentences. Dennett raises the level of discussion from phenomena to talk about phenomena, that is, to a discussion of how we describe or allude to certain phenomena in our ordinary language, in other words, sentences. This coincides nicely

with Dennett's view that there are probably not always actual phenomena for intentional sentences to be about (Dennett, 1969:22).

**3. The Intentional Stance**

Dennett borrows this newly reformed conception of Intentionality and incorporates it into his theory of Intentional Systems, out of which arises the famous Intentional Stance (1971,1978,1987). Intentional systems are those systems whose behavior can be explained and predicted by relying on ascriptions of beliefs and desires to the system. The defining characteristic of intentional systems is that a particular thing is an intentional system only in relation to the strategies of someone who is trying to explain and predict its behavior. That is, an object is not labeled as an intentional system except in the service of predictions and explanations made about it. Any ontological questions about the true nature of the object, for instance, are not answered by taking the intentional stance.

There are three possible stances to take toward a system: the physical stance, the design stance and the intentional stance. The physical stance is based on knowledge of the physical constitution of the object. One can make predictions in the physical stance by determining the physical constitution of the object and the physical nature of the impingements upon it, and use one's knowledge of the laws of physics and nature to predict outcomes of input (Dennett, 1987:16). For instance, taking the physical stance toward a tree informs us that pruning the lower branches of a tree will stimulate denser branches and thicker foliage. Information about the malfunction of objects can be gotten from taking this stance. For instance, one can determine if mechanical breakdowns have occurred from this level.

The design stance is adopted and works best when one is making predictions about mechanical systems. One relies on the notion of function as the basis for one's predictions. Function is understood as a teleological notion, an answer to the question of what purpose has the object, under optimal conditions, been designed to carry out (1978:4). As we saw in the last chapter, evolutionary theories such as Millikan's ascribe intentionality on the basis of the design stance. Design predictions are made based on the assumption that the parts and/or system are functioning properly. Predictions are made

based on knowledge of the object's functional design, irrespective of the physical constitution of the object. One can only make predictions of these systems based on what the designed behavior of the system is. One cannot predict what the behavior of a system will be if one is using it for something other than what it has been designed for. For instance, one cannot predict how a computer monitor will serve as a fish tank, since it has not been designed for this purpose.

The third possible stance to take toward an object is the intentional stance. An object viewed from this stance is an intentional system. Predictions and explanations from the intentional stance are gotten based on the ascription of two things to the system: the possession of certain information and the idea that the system is directed by certain goals. It is then a short step to call this possession of information the system's beliefs and the goals its desires. Predictions made in this stance are based on the idea that the agent will act to further its goals in light of its beliefs (Dennett, 1987:17). Adopting the intentional stance also carries with it an assumption of rationality. Rationality for Dennett's purposes means optimal design relative to a goal (Dennett, 1971:89).

Dennett claims that our commonsense predictions and explanations of behavior in humans and in animals are intentional, and we regularly assume rationality in both humans and animals. In fact, he claims, most experimental psychologists would have trouble designing experiments were it not for the implicit adoption of the intentional stance, with its assumption of rationality, toward lab rats and other such creatures, in order to predict how they would react. He is quick to add, however, that even though we might view both a computer and an animal as an intentional system, we do not adopt the same attitude toward the computer as we do toward the animal that is conscious and rational. The rationality of the computer is said to be pinched and artificial relative to the animal's rationality (1978:8).

The objection could be made here that belief in logical truths is a necessary component of rationality, and yet we do not honestly believe that animals share our belief in logical truths. Dennett does believe that animals share our beliefs in logical truths, or at least they can be said to follow the truths of logic. There is quite a difference between merely following the truths of logic and possessing beliefs in logical truths. If following the truths of logic is a component of rationality, then the animal can be said to follow the

truths of logic by fiat. The animal may not have the further capacity of formulating these truths into proper beliefs. If truths of logic are a matter of discovery, which implies that they already exist, then it is not a far stretch to say that animals follow them. The animal does not need to follow all of the logical truths, just as it need not display perfect rationality. The justification is the same as it is with regard to rationality: an actual intentional system is imperfect, i.e., it does not always follow logical truths, just as it need not always display perfect rationality (1978:10-11).

If the animal does not necessarily have logical truths in its repertoire of beliefs, what beliefs can the animal be said to have in its repertoire, on Dennett's account? With regard to the attribution of beliefs in a more general manner in the intentional stance, there is a rule of thumb, according to Dennett. It is the following: attribute as beliefs all the truths relevant to the system's interests or desires that the system's experience to date has made available. This rule of thumb will of course tend to unrealistically inflate the repertoire of actual beliefs in a creature. Even humans do not have access to all possible true beliefs about a phenomenon at any given moment, given lapses in memory, imperfectness of cognitive processing and the like. More importantly however, this rule will tend to overlook the possibility of error, or false belief occuring in the animal. Dennett claims that false belief always arises from true belief. As he puts it "the falsehood has to start somewhere, and the false beliefs that are reaped grow in a culture medium of true beliefs." (Dennett, 1987:18). He later states that an implication of the intentional strategy is that true believers mainly believe truths (IBID:19). As we shall see later, Dennett has merely sidestepped the issue of error.

There is one more aspect to add to the above description of the Intentional Stance, and that is the different levels or scales of 'embedding' that are possible within the intentional stance (1987:244-7). The first possibility is zero-grade intentionality, that is, no intentional states are ascribed to the system. This level is not included within Dennett's scale, since it is not intentional in nature. It corresponds to the level of description referred to in the terminology of the behaviorists. The first actual grade in the intentional stance is first-order intentionality, where the system has beliefs and desires, but no beliefs about beliefs. Norman Malcolm's distinction between mere beliefs and beliefs about beliefs is an example of first and second-order intentionality respectively.

There is no embedding of belief in the first-order intentional belief, and there is one level of embedding in the second-order. A second-order intentional system, in addition to this level of embedding, can entertain beliefs about its own beliefs as well as beliefs about the beliefs of others. Third-order intentional systems have two levels of embedding and are of the nature of 'Jack hoped that Jill didn't know that he liked her'. Fourth and fifth order intentional systems are theoretically possible, but difficult to follow even for humans.

Dennett entertains a possible objection here. If the system being described as intentional is an inanimate object such as a computer, the question arises as to whether the computer *really* possesses beliefs and desires (1978:7). Dennett's answer is that this question is beside the point, it doesn't matter whether the computer really has beliefs and desires or not, we ascribe beliefs and desires only in order to predict and explain the behavior of the computer. One ascribes beliefs and desires to a system only in so far as this allows one to predict the behavior of the system or explain the behavior of the system. The intentional stance does not treat of the question whether the system really possesses beliefs and desires, and so Dennett's view is often referred to as 'as-if intentionality'. Thus the decision to adopt the intentional stance is a pragmatic one.

Another possible objection could be raised here, that the tactic of imposing the human categories of belief, desire, rationality and the like on other species in order to then predict their behavior is anthropomorphizing (1978:9). Dennett's answer to this is to agree that it is indeed anthropomorphizing, but that it is conceptually innocent anthropomorphizing. All that is being transported into the other species' world are the three categories of rationality, perception and action. While it might be anthropomorphic to attribute hopes, fears, attitudes or outlooks to an animal, the same cannot be said for action or perception since the animal already has these capacities. Moreover, these capacities are not uniquely human and these capacities are not propositional in nature. Dennett adds that we are not assuming that the other species shares any of what might be later discovered to be peculiarly human attributes, such as particular beliefs and desires etc (Dennett, 1971:93). Moreover, as noted in chapter two, it has not yet been established that there are any attributes that are peculiar to humans, except perhaps the capacity for human language.

Dennett anticipates yet another potential objection when he gets to the question of what systems can be properly described as intentional and which cannot (1987:23). In other words, are there any exclusions to the realm of the intentional? One would be reluctant to attribute intentionality to a lectern, for instance. What disqualifies an object from being intentional? The deciding factor, according to Dennett, is if we get no more predictive or explanatory power from taking the intentional stance than we antecedently had for the object when we took the design or physical stances. In such cases where no gain in explanatory power is achieved, it would be better to instead step down a level or two and take either the physical or the design stance toward the object. Deciding whether or not to move up to the level of taking the intentional stance has to do with behavioral complexity. If the behavior of the system under scrutiny is sufficiently complex as to not be adequately captured or exhaustively explained by either the design or the physical stances, one moves up to the level of the intentional. Dennett is making reference here to the same type of phenomena that occurred in chapter three having to do with explanatory range of a theory. The theory of mind explanation for an experimental outcome was often different from the non-mental explanation in that it had a further reaching explanatory range. This is presumably due to the fact that the animal's behavior in the experimental results is sufficiently complex as to require explanation by attribution of mental states rather than by mechanistic explanations.

Now that the above description of the intentional stance is complete, let's see how it fares when applied to the project of cognitive ethology. Here is how Dennett sees the current situation in cognitive ethology. He believes that the new generation of cognitive ethologists is looking for a new theoretical vocabulary in which to carry out and describe their research and findings, the current behaviorist one being too confining (1987:237). Dennett is not, however, opting that researchers create a new mid-level language as Davidson envisions. The situation is analogous in philosophy concerning the status and role of folk psychological terms (intentional or mental state terms). Although the common folk use folk psychology to great success in interpreting behavior, it is the opinion of some that folk psychology will be replaced with more exact methods once all is discovered about humans from a neurological point of view. The question for cognitive ethology is: could the everyday language terms of belief, desire, understanding

etc. also serve as the suitably rigorous abstract language in which to describe the cognitive competencies of animals? We have seen the dangers implicit in taking two possible paths as a result of a positive answer to this question. In answering 'yes' and applying these everyday human language terms to describe the competencies of animals, ethologists first risk being labeled anthropocentric in their application of inappropriate terms to the animal because these terms are for humans only and humans are a separate category from animals. Secondly, ethologists are left with the objection that these terms unnecessarily inflate the competencies of animals because the terms themselves carry meanings and connotations that just aren't within the capabilities of animals. If they choose not to use everyday language terminology, ethologists are left to invent a new terminology to better describe the competencies of animals that would fall midway between the mind-full and mind-less classification levels. As we saw in chapter two, why should ethologists move straight to creating a new vocabulary when the current ones have not yet been eliminated as possibilities?

Dennett believes that the answer to the question of whether we should risk using ordinary language terminology to apply to animals is a qualified 'yes', provided we are careful and understand the assumptions and implications of the strategy we must adopt (namely the Intentional Stance) when we use these words. The most obvious implication is that nothing is stated from an ontological point of view about whether the animal really is an intentional system. One only attributes beliefs and desires in order to be able to predict or explain the system's behavior.

To illustrate the application of the intentional stance to a system, Dennett takes the example of a vervet monkey giving an alarm call (1987:242-9). Vervet monkeys are a species of monkey found primarily in Kenya and the topic of research of two eminent cognitive ethologists, Dorothy Cheney and Robert Seyfarth. Theoretically, the vervet's behavior in a particular situation could be described either by ascribing zero-order intentionality or by first or even second order intentionality. The language of zero-order intentionality, used by behaviorists and often occurring in journal articles in animal behavior journals, ascribes no beliefs, desires or any type of mental states whatsoever. In this case a zero-order explanation would state that the monkey is subject to three flavors of anxiety: snake-anxiety, leopard-anxiety and eagle-anxiety. When one of these three

predators enters the field of vision of the monkey, the monkey associates the predator with danger and has the reaction of uttering a loud vocalization of one of the three flavors. This vocalization has one of three possible triggering effects: sending fellow monkeys into the trees if it's a leopard, out of the trees if it's a hawk, or fleeing on foot if it's a snake. The above scenario could also be described in terms of stimulus-response chains. The stimulus of either a snake, leopard or eagle in the monkey's field of vision elicits the response of an alarm call in the monkey, and the alarm call of the monkey acts as a stimulus to his conspecifics, whose response is to run up into the trees.

A first-order intentional explanation would be that the monkey gives the alarm call to warn other monkeys that a predator is near, or that the monkey wants his friends to act accordingly in order to escape the predator. Morgan's canon, according to Dennett, dictates that we should choose the least sophisticated of the explanations. The original expression of the canon by Loyd Morgan is the following: In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands lower on the psychical scale (Budiansky, 1998:xxx). The problem with the canon is that the terms higher and lower have not been defined, nor have criteria been given for judging one explanation as lower or higher than the other, except in relation to each other. The most glaring problem is thus to determine which of two competing interpretations is the higher and which is the lower, so that a choice can then be made. Dennett thinks that following the canon is the wrong approach, and has in fact prevented the case that higher order explanations are at work from being made for many years (Dennett, 1987:246). This has perhaps contributed to the present situation where cognitive ethologists are looking around for another vocabulary, the present behaviorist one is too confining. Additionally, it should be noticed that perhaps the behaviorist and intentional interpretations correspond to two different levels of explanation that are possible for the same piece of behavior. Human behavior can also often be described in terms of zero-order intentionality, but we do not engage in that practice because the explanations would be severely complex and long-winded. It is nonetheless possible to explain a good portion of human behavior in either intentional or behaviorist terms. Why should we bother with the complexity and lengthiness of behaviorist explanations when a convenient shorthand intentional term is

available? Dennett seems to think that the fact that animal behavior can always be explained in terms of a lower-order theory is no longer interesting to ethologists. He speculates that they would rather find out what gains in predictive and explanatory power would accrue if they were to venture into intentional characterizations of behavior (Dennett, 1987:247)

In addition to the fact that current explanations are too confining, Dennett suggests that the popular laboratory experimental method rules out everything but the stereotypic behavior of a species that most often shows no evidence of intelligence at all. It shows no intelligence precisely because behavioral responses are heavily pre-trained, stereotypic, and not spontaneous in the least. As an example of this situation, Dennett notes the difficulty Premack and Woodruff have had in establishing that primates have a theory of mind. Alternative reasons given by antagonists (such as Heyes) for the experimental results often refer to extensive pre-training in the primate (1987:250).

Dennett also notes that anecdotes have a less than stellar reputation in the scientific community. Anecdotes are not an accepted form of evidence, primarily because they consist in a single unobtrusive observation trial in the natural habitat of the species. Because no experimental intervention is involved, they lack the appropriate controls and therefore rigor. However, Whyten and Byrne, in an effort to tighten up the status of the anecdote, have developed two criteria that serve this purpose. The first is to compile a series of anecdotes, taking care to ensure that each anecdote provides evidence for the same proposed phenomenon, and the second is that the anecdotes must come from independent observers (Heyes, 1998:110). Providing a number of anecdotes that all refer to the same piece of behavior or phenomenon should get around the objection that a single instance does not amount to evidence. The fact that each anecdote comes from a different observer should give some rigor to anecdotes. For instance, one cannot appeal to the fact that since all the evidence comes from a single observer that the observer must be biased or primed to find that sort of evidence.

Taking note of the above considerations on the strengths and weaknesses of anecdotes and experiments, Dennett has come up with a research method that is a hybrid of the two methods of investigation (1987:250-3). Known as the Sherlock Holmes method, it is an intentionality litmus test in that it is an engine for generating or designing

anecdotal circumstances and predicting the outcomes of these situations. The method is anecdotal in nature since it only contains one trial and is usually conducted in the animal's natural habitat. However, it is an improvement on anecdotes in that controlled manipulation of variables is carried out. Additionally, a prediction is made as to how the results will turn out, which is a classic characteristic of experimentation. In addition to the characteristic of hypothesis testing, Dennett's method also is conducted under controlled or artificial circumstances, and some aspect of the target subject's behavior is manipulated, just as with laboratory experimentation. The main advantage this hybrid method has over pure experimentation is the fact that the animal's performance cannot be attributed to intense training, since only a single trial is conducted. Most importantly, the animal's performance is not restricted and artificial, since the test trial is conducted in the animal's natural habitat as opposed to an artificial laboratory environment.

The Sherlock Holmes method entails setting up a one-shot experiment that will predict a particular and peculiar outcome, due to the variables that are manipulated in the experiment, based on what the intentional stance predicts the system under observation will do. The method gets its name from the fictional stories in which the crime-solving detective Sherlock Holmes sets up artificial circumstances in order to find out who committed the crime he is attempting to solve. A popular example involves attempting to catch a thief who has hidden the items he or she has stolen in various other rooms of the house. In this situation Mr. Holmes would gather all the suspected guests in the house into the same room and have his assistant yell 'FIRE!'. He would then watch very carefully to see which of the suspects attempted to move the items, thereby leading him to the perpetrator of the crime.

There is a theoretical downside to this particular method however, stemming from the fact that animals, just like humans, are not perfect intentional beings. Dennett predicts that animals will fail some rather baseline tests and yet pass other rather sophisticated ones (1987:255). The fact that the experiments can only be performed once is a further problem. If the animal fails the test one cannot attribute it to chance or misunderstanding and then just perform the experiment again. The urge to demote the species on the basis of it having failed a rather easy test should be suppressed, however.

**4. Intentionality in the Field**

The interesting thing about Dennett's theory is that it did not remain entirely 'in the armchair', within the realm of the theoretical, he has had a chance to test it 'in the field' with two researchers that study Vervet Monkeys in Kenya. Dorothy Cheney and Robert Seyfarth, familiar with Dennett's work, approached him, thinking that his theory provided a good theoretical framework in which to describe their investigations. They thought that perhaps some of his suggestions could even be implemented in the field.

His proposal to Cheney and Seyfarth was to carry out the Sherlock Holmes Method. The first task is to make a tentative catalogue of the vervets' needs. These needs include those having to do with survival, such as getting food and shelter, as well as informational needs, such as knowing where young family members and possible predators are located. The next step is to adopt the intentional stance complete with the assumption of rationality, acting as if the vervets are rational beings with beliefs and desires, based on this compiled catalogue of needs. The idea is to translate this list of needs into a set of beliefs and desires. This would entail translating the vervet's need for food, for instance, into the desire to obtain berries from a tree, or translating the informational need for the location of predators into the belief that there is an eagle circling overhead. The next step is to frame some hypotheses, based on how the vervets ought to behave given the beliefs and desires ascribed to them. A possible hypothesis based on beliefs about predators would be that the vervet should warn his family members and cohort that the eagle is nearby. The last task is to test the hypothesis and see if the outcome matches the predictions that were made based on ascribing intentionality to the vervets. That is, do the vervets behave as they ought to have behaved under the circumstances?

Applying this method to research in cognitive ethology has produced some rather ingenious experiments. The most popular one developed by the researchers involves a direct adaptation of the Sherlock Holmes 'fire' situation. It involves setting up a stimulus from an artificial source, in this case playing back pre-recorded alarm calls from speakers hidden in the bush nearby. The hope is that the monkeys will respond to the fake warning call in the same way they would if it came from a live conspecific instead of a speaker, otherwise the experiment is invalidated.

The downside to the method, according to Dennett, not immediately apparent in the armchair building of the theory, is that the information that needs to be imparted to the subjects in the 'set-up' portion of a typical experiment is often imparted verbally (1998:303). The mindset that Sherlock Holmes must cajole his suspect into in order to bring out the guilty response is often accomplished verbally. When non-verbal animals are the subjects of study it is impossible to set up the 'trap' properly when your subjects do not have language. This point is well taken, but I do not think that it is quite the downside that Dennett makes it out to be. Researchers of non-verbal animals are well aware of this fact and regularly take it into consideration in designing their experiments. Certainly many of the experiments performed on animals do not entail trying to discover intentional states; many of the experiments are designed rather to discover the presence of other cognitive capacities. On the other hand, if intentionality is somehow manifest or detectable in behavior, then the fact that more ingenuity is required in designing these experiments should not be taken for the impossibility of designing experiments that will capture intentionality.

A second problem, much more difficult to mitigate, is shielding the vervets from certain information in the setting up of circumstances in a fake situation (1998:303). For instance, when Cheney and Seyfarth set up fake alarm calls that come from a speaker instead of a fellow vervet, what could potentially ruin the experiment is if the vervets catch on to the fact that the calls are not coming from a fellow vervet, but from a speaker. The following is a typical scenario. Cheney and Seyfarth set up the experiment by hiding the speaker in a bush and cueing the tape to the exact call they wish to play by a target vervet. They then have to wait until the target vervet moves out of sight before playing the call. What often happens is that the target vervet re-emerges into view before they have a chance to play the call for the other vervets, and so they have to stop the experiment and wait for another opportunity. If the other vervets figure out that the call is coming from a speaker and not from the fellow target vervet, the current experiment as well as future experiments will be invalidated.

Cheney and Seyfarth have a series of 'out the door' experiments that they would perform if the political situation in Kenya ever got so dangerous as to force them to leave (1998:306). They are called 'out the door' experiments because they would prevent

further experiments of a like nature to be conducted, for the reason given above: some aspect of their conduction would give away the source of the trickery, and future experiments would be invalidated. Note that it does not matter that the vervets' 'catching on' behavior could also be considered intentional, thus giving weight to the hypothesis that the behavior is intentional. The fact that experimental protocol has not been upheld takes precedence over any interesting results that may have been obtained. If the vervets catch onto the trick all future behavioral results obtained are potentially tainted by this knowledge in that the behavior of the vervets may become altered. These experiments are interesting in that they illustrate the possibility that some aspect of control is possible even in naturalistic observation. First, there is the idea just mentioned that the vervets must never know that the alarm calls come from speakers and not fellow vervets. The fact that the vervets have been exposed to the source of the trickery will more than likely alter their future behavior. The fact that the vervets 'believe' that the vocalization comes from a fellow vervet is thus a controlled variable. Second, there is the idea that the vervets must never 'notice' Cheney, Seyfarth and Dennett: the three observers must never be construed as a source of potential payoff for paying attention to them. For this reason, many weeks are spent in the presence of the monkeys without any attempt at experimentation so that the monkeys can habituate to the observers. It is hoped that once the observers begin experimentation, they will continue to be ignored by the monkeys. If one of the observers somehow informs the monkeys of some unknown as yet danger, then future experimentation is potentially confounded.

**5. Evaluation**

Dennett's intentional stance is not without its critics. Most of the criticism has to do with the idea that it is a pragmatic stance: one takes the intentional stance toward an organism if it shows relative complexity in its behavior, and it remains an open question as to whether or not the animal really is an intentional being. One could argue that Dennett's intentional stance does not advance the issue, for the question we really want the answer to is left unanswered.

Let us see how his theory does with regard to my four guidelines. First of all, does it have an account of error? As we saw, Dennett claims that most creatures are true

believers, they believe mainly truths, and false belief must ultimately arise from a large set of true beliefs. This is more an account of the source of false belief than how it might occur in the animal given his theory. Perhaps an account of error could be distilled from the experiments themselves, in other words if an animal fails one of the Sherlock Holmes method experiments, perhaps the source could be due to an error on the part of the animal. This will not work, however, because these tests are designed to predict whether an animal's behavior is intentional or not, in other words, whether the animal will behave according to predictions made based on the attribution of a list of beliefs and desires and a modicum of rationality to the animal. The content of the beliefs and desires cannot be whittled down sufficiently to be determined as true or false except through further testing following a particular response in the animal. If beliefs and desires cannot be sufficiently identified, the content of erroneous beliefs will also not be identifiable. Notwithstanding Dennett's claim that true believers believe mainly truths, which is probably accurate, his theory does not have an account of error, and so it fails the guideline.

The second guideline is that a theory of intentionality must be able to narrow down the content of beliefs and desires to a reasonable degree. I say reasonable here because I think we should heed Millikan's suggestion that the content of animal states will probably not translate exactly into sentences of English. Dennett's theory is helpful in this way, at least in theory. His Sherlock Holmes method applied to vervet communication systems means that Cheney and Seyfarth can narrow down the possibilities in content to a reasonable degree for any one of the alarm calls. If an initial experiment still remains subject to multiple interpretations, content can be narrowed down even further by performing a family of related experiments, depending on the vervet's response to the initial one.

The fact that vervet monkeys in particular are the subjects of Dennett's applied theory is an even bigger bonus, for according to Dorothy Cheney, vervets are a species with laser beam intelligence. This means that they have brilliant narrowly specialized cognitive talents, with almost no carry-over of skill to other topics (Dennett, 1998:298). One might think that since they have narrowly specified cognitive talents that the content of their beliefs should be relatively easy to narrow down, either because the beliefs are less sophisticated or less numerous. Either the hypothesis about laser beam intelligence

is wrong, or there is something about Dennett's method that prevents the narrowing down of the content of the beliefs of the vervet. Either way, the content of the beliefs of the vervets cannot be identified without multiple additional experiments being performed. The list of potential beliefs and desires for a single communicative call remains too large for Dennett's theory to pass the narrowing of content guideline. Moreover, Dennett helps to sink his own ship in the case of this guideline. According to him, all intentionality is 'as-if' intentionality. One attributes a set of 'as-if' beliefs and desires to the vervet and one determines whether the vervet behaves in an intentional manner in the experiment. Even if one were able to narrow the content of these beliefs and desires enough to actually identify them, they would still be deemed 'as-if' beliefs and desires. If all beliefs and desires are of an 'as-if' nature, they will give us no usable information as to their content.

Concerning the third guideline of practical implementability, Dennett's theory passes with flying colors. Dennett's is one of the only examples, to my knowledge, of a philosopher's so-called 'armchair' theory that has been applied rather successfully in the field to a particular species of animal. Dennet's attitude seems to be one of ambivalence with regard to this success, he mentions two big problems with his theory when applied to the field, both having to do with a lack of language in the species under scrutiny. As we saw, one problem is that the set-up portion of the Sherlock Holmes method is usually done verbally. The other problem is the need to shield the vervets from certain information that, if known, would invalidate the experiment. Notwithstanding Dennett's own pessimistic evaluation of his theory, it still passes the guideline in my opinion.

The last guideline is whether or not the theory can discriminate between intentional and non-intentional behavior. This guideline is important in order for the theory of intentionality to avoid begging the question, in other words, the theory must not assume that which it is trying to demonstrate in the animal, namely that the animal has intentional mental states. Dennett's theory automatically fails this guideline due to the fact that it is a pragmatic theory. He insists that he is not interested in answering the question of whether the animal's behavior is really intentional or not. If we look past his insistence that there is no 'really' question beyond as-if intentionality, we see that his Stance is at least able to distinguish between zero-level or the absence of intentionality

and first order intentionality. The categories of zero-order intentionality versus first-order intentionality are both mentioned in the Stance, but is the theory actually able to distinguish between intention-less or non-mental behavior and intentional behavior?

Dennett's own opinion on this question does not inspire much confidence. As noted above, once an experiment has been performed on an animal, the results are often still subject to multiple interpretations, as was the case in chapter four. In order to eventually get a yes or no answer to the question of whether the animal's behavior is intentional, one must perform large families of related experiments on the animal to 'narrow down the field' as he puts it (1998:298). Thus, while his theory can distinguish between the various levels of intentionality, it cannot tell us more than that the behavior is ultimately as-if intentional. It thus fails the fourth guideline.

**6. Bennett's Guiding Rule**

Jonathan Bennett is motivated by precisely this lack in Dennett's Intentional Stance. In Bennett's opinion, Dennett's theory lacks an account of what it is for a given hypothesis to explain a range of behavioral data. Following from this, it lacks a principled way of deciding between intentional and non-intentional behavior. Bennett makes the humorous claim that were we to rely on Dennett's theory, when an ethologist comes up with a mentalistic hypothesis that he fails to find any rivals to, he merely sits trembling, hoping that no more ingenious and mean-minded colleague would be able to come up with a rival (Bennett, 1990:45). The primary motivation for Bennett's theory is thus to develop a set of generalizations that will explain a range of behaviors. His theory will show how to bring a class of behavioral episodes that make reference to mental states under a single explanation. His theory bears a striking resemblance to Heyes and Dickinson's in some respects except with improvements in problem areas.

According to Bennett, what is needed for a theory of intentionality that will be applicable to non-verbal beings is a set of fairly reliable generalizations relating beliefs and desires to behavior. These generalizations will provide an explanation of behavior in terms of belief and desire, although the explanation will not necessarily be causal in nature. It will be remembered that Heyes and Dickinson immediately couched their account in terms of causality, demonstrable in terms of counterfactuals and

contingencies. Their justification for this move is that experimental manipulation is made easier. One major problem with this move is the impossibility of accounting for error on this type of causally based account. Bennett thinks that causal explanations must always ultimately be couched in neurological terms, and that mentalistic explanations can be explanatory without being causal (1990:39). In fact, and this will become the defining difference between mentalistic and non-mentalistic explanations, mentalistic explanations involve patterns that would be missed altogether by neurophysiological explanations.

It could be asked why Bennett doesn't immediately go the 'causal' route, as many of the other theories examined thus far have done. After all, causal explanation is the most popular variety of explanation. In addition to the reason that causal accounts of intentionality tend to exclude the possibility of error on the part of the animal, there is another more general reason to avoid causal accounts of explanation, on Bennett's view. According to one construal of causal explanation, one can only attribute thoughts to others if their behavior could not have been caused purely by the physical states of their bodies. If one subscribes to this type of explanation, one would probably end up concluding that the physical causes of animal behavior suffice to explain it all, leaving no gaps to be filled from outside the physical realm (Bennett, 1990:38). One would then end up giving up research in cognitive ethology because there would be no point to it. Ethologists don't want to give up the search for mental states in animals until it can be soundly proven that there is no likelihood that animals have them.

I suspect that Bennett's reference to this defining feature between mentalistic and neurophysiological explanations, i.e. the notion of patterns, originates from a sentence in Dennett's article titled True Believers: "It is the patterns in human behavior that are describable from the intentional stance, and only from that stance." (Dennett, 1987:25). Dennett perhaps fails to exploit the notion of patterns since all intentionality, according to him, is ultimately all of an 'as-if' nature. He is thus not interested in differentiating between true intentional behavior and non-intentional behavior. This is unfortunate, for the notion of patterns is, in my opinion, the defining feature of intentional explanations that helps get a theory of intentionality off the ground, and circumvents the objection of begging the question. As we will see, Bennett's theory is the most comprehensive of the four being examined because he is able to offer a defining feature, based on patterns, that

will serve to set intentional explanations apart from other non-intentional or non-mentalist alternatives.

The core of Bennett's theory of Intentionality is a belief-desire-behavior triangle. This triangle should be seen as a mathematical equation, containing the three variables or components of belief, desire and behavior. Some of these three variables are either given or need to be solved for. There is a caveat to the solving of the unknown variables in the triangle. The caveat is that if both the belief and desire variables are unknown, solving these two variables must be tackled at the same time. This is because behavior is only indicative of an animal's beliefs if its wants are assumed, and only indicative of its wants if its beliefs are assumed. There is no possibility of determining one of the elements first and then going on to study the other (Bennett, 1990:41). Thus beliefs and desires must be studied and solved for simultaneously. This might appear an impossible situation, how does one solve the two variables at once if there is seemingly no handle on knowing either variable? One must make a single temporary assumption, and that is that the animal's desires do not change much over time.

The triangle at this stage very much resembles Heyes and Dickinson's belief and desire criteria. It thus suffers from the same problem as their criteria in that it is too libertine. That is, one cannot identify belief or desire content, or even narrow it down to a reasonable extent. Any content can be made to fit in the belief-desire pairs. There are so many possible contents that will fit the pattern of behavior that almost any one will work. Two implications follow from the triangle as it stands: one is that it has no predictive power and the second is that it cannot account for error. It has no predictive power because the procedure cannot connect what the animal thinks or wants at one time with what it thinks or wants at another (Bennett, 1990:41). The procedure also cannot account for the possibility of error in the animal because in order to do so, it must be able to identify the content of the belief as an erroneous one. Amidst the myriad of true and false beliefs, how would one go about choosing one and then justifying this choice over all the other possibilities? As it stands, the triangle also resembles the theory of Dennett, in that his theory also cannot account for error. The triangle thus needs to be grounded in some manner to allow for the identification of a particular belief-desire content. Grounding the triangle will give it predictive power and account for error in that it should

allow some way of identifying the content of belief. What form should this act of grounding take?

Bennett thus adds a fourth variable, that of sensory-input, to the triangle, making it into a square. The square now has as its four components: sensory input, belief, desire and behavioral output. The new relation forged between the animal's sensory apparatus and beliefs serves to ground the square. With regard to the identification of content, the addition of a sensory input component doesn't seem at first glance to bring us any closer to the identification of particular belief or desire contents. However, with the temporary assumption that the animal's desires don't change very much over time, we can hypothesize what those relatively static desires might be, and then determine what the animal's beliefs truly are relative to the desires and as a function of the various environments that the animal finds itself in.

According to Bennett's theory then, two variables need to be identified, and each has its own constraint. The beliefs of the animal must relate systematically to the environment, desires remain relatively static over time. All that is needed to get the theory off the ground is to assume that desires don't change much and identify a single desire. From that identification as well as the animal's subsequent behavior one can hypothesize what beliefs the animal might have given the assumed desire relative to a particular environment. We can then notice if the desires of the animal do change over time, because they will be reflected in the changes in behavioral response of the animal. Having determined what the animal's beliefs are, we can then go back, drop the static desire assumption, and identify the animal's real desires.

The theoretical sequence put into the context of a concrete example is thus the following:

1. Assume that desires don't change much over time and identify a single desire in the animal, to get food, for instance, obtain bananas in a tree.
2. Fix a belief according to that desire, for instance that shaking a particular branch on the tree will make the bananas fall.
3. If the primate shakes that branch, replace the assumed desire with an actual one. In this case the assumed desire turned out to be accurate.

4. If the primate instead approaches a female primate and starts to groom her, the assumed initial desire will prove to be inaccurate, and replaced with the new desire that the primate's desire is to engage in grooming behavior.

Now that it is possible to identify particular belief and desire content, we see that the square has predictive power. Having identified the environments in which an animal has a particular belief-desire content, the square can now predict how the animal will behave under a particular environmental contingency. That is, future behavior can be predicted on the basis of the known variables of belief and desire as well as environment and past behavior, and the idea that desires can be solved for as a function of beliefs given changes tracked in the environment. There will remain some level of indeterminacy of content, but this is to be expected.

Bennett feels that although Dennett's theory is quite good overall, the one problem it suffers from is the inability to rule out rival explanations. That is, it fails one of my most important guidelines, the ability to distinguish between intentional and non-intentional behavior. In addition to begging the question, explanations generated from the theory are thus forever vulnerable to being equally well explained by alternatives such as those of the behaviorist. This problem is similar to that occurring throughout my discussion in chapter four on methodology, although not exactly. There are two possible scenarios involving rival hypotheses that I believe are used interchangeably. The first is a situation of rival hypotheses where the two rivals represents different camps, such as mentalist versus non-mentalist. All of the experiments seen in chapter four concerned rival explanations belonging to different camps. In cases like this it is a matter of eliminating the non-mentalist hypothesis, and in doing so an answer to the question of whether the behavior is intentional or not is also obtained. The other possible situation is of rival hypotheses where both are mentalist hypotheses, and the difference is one of degree, say of a 1$^{st}$ order hypothesis versus a 3$^{rd}$ order hypothesis. In this case, both hypotheses are within the intentional realm. Bennett calls this second situation a case of 'empirically equivalent' hypotheses. Dennett's theory does well regarding the situation of empirically equivalent hypotheses, since his Stance can discern differences of degree within levels of intentionality. His theory does suffer from the first problem mentioned above, however. His theory is unable to rule out non-intentional or stimulus-response

(S-R) explanations in favor of intentional ones.

When faced with either of the situations above, Dennett's prescription is to appeal to some sort of economy rule, either the principle of parsimony or Lloyd Morgan's canon. Rules of this type advise one not to attribute cognitive states to an animal whose behavior can be explained without them. Within levels of intentional explanation, the rule also prescribes that cognitive attributions do not go higher than what is needed to explain the behavior (Bennett, 1991:98). As will be seen, with Bennett's theory it is unnecessary to appeal to this type of rule in trying to eliminate rival hypotheses, since the different patterns of behavior given the two types of explanations suffice to determine a victor.

At this juncture I believe it is necessary to make a distinction between the notion of hypotheses (or experimental predictions as they are sometimes called), and the notion of explanations, in order to clear up any confusion based on conflation of the two terms. On a purely physical level, there is a distinction to be made between hypotheses and explanations, namely that hypotheses are made *before* an experiment is conducted. Hypotheses constitute a prediction of the experimental results, that is, they venture a guess as to how the animal will perform in the experiment. Hypotheses cannot be offered or made after an experiment has been conducted. Explanations, on the other hand, are usually offered *after* the experiment has been conducted and pertain to the actual experimental results obtained. The issue is usually that two rival explanations are often on offer, one representing the non-mentalist or non-intentional or behaviorist camp, and the other representing the mentalist or the intentionalist camp. Given an experimental result, it is always possible to explain it after the fact by either rival explanation. In my opinion, it is in large part the sanctioning of the practice of advancing ad hoc explanations that allows for either explanation to account for the results of an experiment. As I claimed in chapter four, the fact that Heyes could account for the experimental results in every case by a non-mental explanation is precisely because her explanations are explanations rather than hypotheses, and also that they are after the fact. Dennett also claims that any behavior, human or animal, is subject to explanation in terms of behaviorist terminology, and that this fact is now uninteresting to us because it does nothing to advance the issue (1987:247).

Given that rival explanations can always explain the experimental results equally well, we should not look to explanations as giving us any tie breaker victories, unless they are translated into hypotheses. The notion that rival hypotheses will rarely make the same predictions of experimental results should interest us however, because it is at the level of prediction that a tie-breaker to the stand-off will be found. Rival hypotheses, such as non-mentalist versus mentalist, will not make the same predictions of the results because they contain different patterns of explanation. So a non-mentalist hypothesis will not make the same prediction of a set of experimental results that a mentalist hypothesis will make, because it seeks to explain behavior at a different level than the mentalist hypothesis. I gave an example of this type of situation that was originally conceived of by Dennett at the end of chapter three. If we put a rat in a Skinner box and train it to take exactly four steps forward to press a bar to get a food reward, when we retract the bar so that it now takes five steps to get the food, the Skinnerian is forced to hypothesize that the rat will still take the same four steps and end up jabbing the air with its nose (Dennett, 1978:14). Putting the issue in terms of patterns of behavior, the pattern referred to by the non-mentalist is a different pattern than that of the mentalist. This point is further reinforced by the claim that mechanistic explanations, i.e., those of the non-mentalist, behaviorist or stimulus response variety have no predictive power in the absense of eliciting stimuli, that is, one cannot make behavior predictions on these varieties unless there is a stimulus to point to.

According to Bennett's guiding rule, there is a way to decide between two rival explanations. This is because there is a difference between the two types of explanations, having to do with the explanatory ground that each type of explanation covers. We are here interested first in the difference between mentalistic or intentional explanations on the one hand and behavioristic or mechanistic or stimulus-response explanations on the other. The crux of the guiding rule is that mentalistic explanations have a different pattern of explanation from which an extra ingredient emerges. Building on this notion of a pattern, Bennett claims that mentalistic and behavioristic explanations contain elaborations of different patterns of behavior. Stimulus response behavior can be explained by the principle 'Given a certain kind of stimulus-input, the animal produces a certain kind of motor output.' Mentalistic explanations can be covered by the following

principle: 'The class of behaviors to be generalized over involves inputs whose only unifying description is that in each of them the environment is such that there is something the animal can do that will, for instance, bring it food'. The class of behaviors involves outputs that are united only in that in each of them the animal moves in such a way that results in it getting food. The common factor to all mentalistic explanations is that the behavior they explain can be unified by one generalization. The fact that several pieces of different behavior can be grouped under one unifying idea and generalization means that intentionalistic explanations have the extra ingredient of a larger based explanatory power.

Frank Dreckmann has captured the difference between the mechanisms underlying the two kinds of explanations in a comprehensible manner (1999:96-8). Mentalistic behaviors, according to him, abstract from particular tokens of behavior. They leave out various details that an S-R explanation would include, but possess more explanatory range because they group together various differently executed behaviors under a common unified idea. S-R explanations might include mechanistic movements that can be described in a kind of token manner, and that are fully exhausted by a stimulus response type of explanation. Moreover, S-R explanations cannot group a series of slightly different behaviors together under one common idea.

This distinction can then be used to develop the tie-breaker guiding rule for ruling out one of two rival explanations. Let's say we are given a set of disparate behaviors performed by a group of vervet monkeys following an alarm call uttered by a member of the group. One monkey climbs into a tree, another runs along the ground, a third freezes in its tracks and remains motionless like a statue and a fourth falls to the ground and appears dead. These behaviors cannot be explained in an exhaustive manner by the S-R explanation, since there is no pattern common to all of the behaviors of the type "given a certain stimulus input S, the animal produces a motor output R". The stimulus is the same for all scenarios: an alarm call. The response is not the same; each monkey produces a different behavioral output to the alarm call. Since the disparate behaviors can be unified under one common notion or idea, that of 'getting to safety' for instance, this series of disparate behaviors can be successfully explained by a single mentalistic explanation, for instance, the belief that the monkeys wants its friends to safely escape

the predator. The rule of thumb as to whether or not content can be attributed is whether or not one can justify the need for the notion of that particular content in characterizing the class of environments in which the behavior occurs. If there were a single stimulus type that could capture and thereby explain all of the animals' behavior, content based on beliefs and desires won't be justified. In this case, however, there is no single stimulus that would cover the series of disparate behaviors in the monkeys.

If a theory can discriminate between S-R explanations and mentalistic ones, then that theory has the ability to rule out S-R explanations. This translates into being able to determine whether beliefs and desires warrant being attributed to an animal or not. In addition to being able to state whether beliefs and desires can be attributed to an animal or not, Bennett's account can also specify what content to attribute to these beliefs and desires. In terms of the question of what content to specifically attribute, it should be possible to read it straight from the unifying idea. Belief content is thus gotten through what is perceived as common to all the environments in which the behavior occurs.

Bennett's theory has the ability to rule out S-R explanations with the help of his guiding rule. What about the second situation mentioned above, that of empirically equivalent hypotheses, where both hypotheses are from the mentalist camp? Does his theory allow us to be able to choose one level of attitude attribution over another within the realm of intentional attributions? He claims that the theory can help to decide between hypotheses that are empirically equivalent. Let us first pin down the issue at stake in this case. One is trying to decide between two hypotheses, both of which fall into the intentional realm except that one might make reference to a $1^{st}$ order embedding while the second might make reference to either a $2^{nd}$ order or even a $3^{rd}$ order embedding of belief or desire. Taking up the predator example mentioned earlier, a fellow monkey utters an alarm call, and all other members of the group are observed to flee in different manners. Interpreting the situation with reference to a $1^{st}$ order embedding would be: a vervet gives an alarm call because it believes there is a predator nearby and desires that its friends should act accordingly to get to safety. An example of a $2^{nd}$ order embedding is the following: the vervet gives the alarm call because it believes there is a predator nearby and desires that its friends should believe the same thing. Under the $1^{st}$ order interpretation the alarm call itself solicits the different fleeing behaviors of the other

vervets. Under the 2nd order interpretation, the alarm call solicits the belief in the vervets (that there is a predator nearby) that then solicits their own fleeing behavior.

Morgan's canon in this case would suggest that, within mental state attributions, we should not go higher on the scale than is needed to explain the behavior. This translates into picking the less extravagant of the two hypotheses unless a justification can be found for the need for an extra level of embedding. In this case, following Morgan's canon we would have to choose the interpretation with a 1st order level of attribution. In terms of Bennett's theory, to justify the extra level of belief or desire found in the 2nd order interpretation, the vervets must have a variety of uses for the belief that there is a predator nearby that came from the alarm caller vervet. If we go with the first interpretation, we find that in trying to include all of the various reactions of each of the fellow vervets we end up attributing a single thought to the alarm caller of "implausible complexity", according to Bennett (1991:105). The thought would have to include the fact that the alarm call solicits monkey X to run along the ground, monkey Y to climb a tree, and so on. On Bennett's theory, we would rather group all of the vervets' various reactions to the call, and find that we could simplify the interpretation into a unitary thought 'behave appropriately to the belief that there is a predator nearby'. The clause 'that there is a predator nearby' within the attribution constitutes a 2nd order level of embedding. Having justified the need for the 2nd order embedding, we then have a method to distinguish between empirically equivalent hypotheses on Bennett's theory, and a reason to choose the 2nd order interpretation, because it is the simpler of the two.

Bennett's theory, as it stands, still lacks an account of error. None of the examples discussed so far make mention of the possibility of error in the animal. It is the input and output principles (elaborated on above) that must be amended to account for error, because they are what dictate how the animal will react. The original input principle was the following: "The class of behaviors to be generalized over involves inputs whose only unified description is that in each of them the environment is such that there is something the animal can do that will allow it to achieve X". This is an error free principle, for it assumes that the environment will always be such that the animal can act to achieve its goals. The new input principle, amended to include the possibility of error, would thus read: "Each of the relevant environments, given the animal's perceptual

apparatus and space, is significantly similar to the ones in which there is something the animal can do to allow it to achieve X." A comparison set of environments is created to allow for comparison between what the animal does, and what really needs to occur in order for the goal to be achieved. Thus if the animal misperforms in some way or fails to execute a movement, the error will be recognized as such by the difference in the way the animal has behaved and the way it ought to have behaved to achieve the goal. This difference will be reflected in the comparison of the two environments, i.e., between the comparison and the actual. The output principle is also amended to include this comparison set of environments. The original principle was "The class of behaviors to be generalized over involves outputs that are united only in that in each of them the animal moves in some way that results in it achieving X". The new principle would be "On each occasion, the animal moves in a way that would allow it to achieve X if the environment were a member of the comparison set". Stated in this way, the principle allows for error- the possibility that the environment is not a member of the comparison set.

It should be apparent to the reader already that Bennett's theory of intentionality is a strong theory. With regard to the four guidelines I have developed, he developed his theory with two of the guidelines already in mind. That is to say, he became interested in developing a theory that would not only state when it is reasonable to attribute beliefs and desires to an animal, but also what the content of those beliefs and desires might be. His theory thus passes two of the guidelines, the theory is able to rule out rival hypotheses and thus discern when an animal is acting intentionally from when it is not, and the theory is also able to narrow down the content of those beliefs and desires to a reasonable degree. He is explicitly concerned to include an account of error in his theory, and thus creates a set of comparison environments that should point out when the animal has erred. The last guideline is empirical tractability. Bennett's theory automatically passes this guideline due to his interest in ameliorating the lacunae in Dennett's theory. As mentioned, Dennett's theory left out a way to rule out rival hypotheses, a phenomenon originally manifest in the discussion of experiments in chapter four. Bennett's theory offers a way to rule out rival hypotheses using the experiments themselves in their supposed ambiguous state.

## 7. Conclusion: Empirically Equivalent Explanations Solved

The leftover problem from chapter four, it will be remembered, was the inability of the mentalist or theory of mind theory to declare a victory over the non-mentalist explanations. With Bennett's theory it is possible for mentalist explanations to eliminate the non-mentalist camp. Generally speaking, because each explanation manifests a different pattern in the behavior, the mechanistic explanation is unable to account for a variety of different behaviors in the way that the intentional theory can. As it turns out, the situation in chapter four was not the more difficult case of empirically equivalent hypotheses, where rival hypotheses are made that both pertain within the realm of the intentional, but rather a simple case of determining whether the behavior in question is intentional or not, which was one of my four guidelines. As mentioned, the fact that Heyes regularly neglected to mention hypotheses for any of the experiments is one source of the problem. The situation was also exacerbated by the experimental design characteristic to theory of mind experiments and the fact that her competence and validity criteria were ill-suited to this design. In theory of mind experiments and unlike ordinary experiments in psychology for instance, the presence or absence of the behavior in the experiment is not the deciding factor. If Premack and Woodruff had been able to point to the mere presence of a particular behavior as demonstrative of a theory of mind in primates, the situation would be quite simple and Heyes would not be able to tack on her ad hoc explanations as equally accountable for the results. It is precisely because of the fact that the behavioral result is a given, and the issue is whether or not an intervening variable of a mental state is justified, that allows Heyes to try and account for the results according to a non-mentalist explanation.

Cognitive ethology cannot seem to get out from under the weight of the objections of behaviorism. I think I have been successful in advancing the case for the search for mental states in animals in at least one important respect in these two chapters, and that is by finding a theory of intentionality that does not beg the question. In being able to show whether a bit of behavior is intentional or not, Bennett's theory has the further bonus of quieting the most damaging behaviorist objection to a certain extent. This notion of patterns, first discovered by Dennett but that he failed to exploit I suspect because of his insistence on the unimportance of the 'really' question, was picked up on

by Bennett and constitutes, in my opinion, another much needed 'shut-down' argument against behaviorism and the idea that animals cannot have mental states for their behavior is explainable without recourse to mental states.

## Chapter Seven
## Concept Attribution

### 1. Introduction : Three Empirical Questions

One of Davidson's claims in the first chapter makes a link between the three notions of concepts, propositional attitudes and language. It is the following:

1. In order to have a belief, one must have the concept of a belief.
2. In order to have the concept of a belief, one must have language.

This claim has two issues implicit to it. One that I will not be concerned with in this chapter is 'Can one have propositional attitudes, such as beliefs and desires, without having the concept of these propositional attitudes?' The other issue of which I will be examining one aspect in this chapter is 'Must one have language in order to be said to have concepts?'

Theories of concepts can be loosely distinguished in at least three different ways, according to what question investigations are designed to answer. Some theories will attempt to provide answers to more than one question, and all three questions are interrelated. One type of theory is primarily interested in the question 'what is a concept?' Many of the psychological theories of concepts are interested in answering this question. The first theories examined in this chapter, those of Definitional, Prototype, and Exemplar, could be classified as attempting to answer this first question. A second type of theory is interested in answering the question 'what is it to possess a concept?' Many such theories will postulate a set of necessary and/or sufficient conditions that must be met in order for a creature to be considered to possess a concept. The Definitional theory mentioned above would also fall into the category of theories interested in this second question for it postulates a set of necessary and sufficient conditions for possession of a concept. Christopher Peacocke also is interested in this question in his book "A Study of Concepts" (1984). A third possible question is 'When is it reasonable to consider that a creature has or is operating with a concept?' This third question motivates theories that are concerned with the issue of concept attribution. Ruth Millikan views concepts as abilities, and so her theory based on evolutionary adaptability is a good example of a theory interested in this third question (Millikan, 2000). For the

purposes of this chapter I will be particularly interested in theories concerned with this third question in its applicability to animals, for these theories take performance in the animal, i.e., behavioral evidence, sometimes also in the form of conditions, as primarily indicative of when concepts might be attributable to the animal. Using behavioral criteria as opposed to verbal response as indicators means that these types of theories will be less dependent on language and more easily applicable to animals.

There are not many theories of concept attribution in the literature that have been specifically applied to animals. While there are numerous theories in both the psychological and philosophical literature that been offered for humans, a theory by Colin Allen (1999) is the only other existing theory apart from Millikan's, to my knowledge, that has been designed specifically for animals. One reason has been offered: most human theories of concepts cannot apply to animals because they hinge too much on possession of a natural language. Nick Chater and Cecilia Heyes (1994) base their views on this reason. As we saw in chapter five, Heyes along with Anthony Dickinson have developed a theory of intentionality, listing two behavioral criteria that have to be met in order for an animal's action to be considered intentional. In this chapter, I will be discussing Heyes and Nick Chater's search for a theory of concepts that would apply to animals. The theory must contain a construal of 'concept' that meets three criteria in order to be considered as potentially applying to animals. The caveat, according to Chater and Heyes, is that unless there is some way of understanding concepts that is independent of their connection with natural language, non-linguistic animals cannot have concepts. I will also discuss Allen and Hauser's (1998) reply to their article, as well as their subsequent development of a minimal constraint on concept ascription that is based on evolutionary theory. The point of the minimal constraint is to isolate and identify abstract concepts by the fact that they contain characteristics that are not perceptually available. This minimal constraint, albeit difficult to implement experimentally, nonetheless represents a positive step forward in the eventual development of a theory, since it goes beyond the usual contentious constraint of stimulus generalization. I will end the chapter with Colin Allen's (1999) theory of concept attribution. In my opinion it is the best example of a theory of concepts that can applied to animals because it provides a behavioral demonstration of concept-mediated behavior.

The detection, recognition and modification of errors made by the animal are the crux of the criteria.

**2. The Search for a Theory of Concepts**

The main claim in Chater and Heyes' aptly titled article "Animal Concepts: Content and Discontent" is that unless there is some way of understanding concepts that is independent of their connection with natural language, non-linguistic animals cannot have concepts (1994:209). I am of the opinion that it seems a bit defeatist to discuss whether human theories of concepts that are linguistic in manifestation would apply to non-linguistic animals in the first place. If it has been established that language is so intimately tied to these theories, and that animals do not have a reasonably humanlike language, then what is the use in examining whether these theories apply to animals? Nonetheless, one way to get around Chater and Heyes' claim is to demonstrate that there exist theories of concepts that are not dependent on language.

Given that Chater and Heyes' aim is to search for a particular sense of 'concept' that meets three criteria, it could be said that they are interested in the 'What is a concept?' question. Rather than developing a set of criteria that would have to be met for possession of a concept they instead ask 'Is the nature of concepts such that they are intimately tied to language?'

Chater and Heyes examine the various theories of concepts on offer in the literature. They are specifically searching for a sense of concept that satisfies three desiderata.

1) Applies to humans, and assigns to them concepts corresponding to terms of natural language.
2) Can be applied to non-linguistic creatures.
3) Allows for empirical investigation of animal concepts (1994:210).

These can be summarized for ease of discussion as the human applicability criterion, the animal applicability criterion, and the empirical tractability criterion. As we saw from the discussion in the two previous chapters on intentionality, crucial to any theory applicable to animals is that it is able to generate testable hypotheses on animals, and so criterion three is absolutely necessary to a search for a theory that would apply to

animals. Criterion two is also necessary, it is the main motivation behind a search for a theory that would apply to animals. Criterion one, however, is only present for comparative purposes. It is there to allow comparisons to be made between animals and humans, otherwise, it is argued, we would be at pains to call whatever items we find in animals 'concepts'. However, it could be argued that criteria one and two when taken together, raise a problem similar to my point made above. How can one search for a theory that applies to humans and assigns them concepts that correspond to terms in natural language that will also apply to non-linguistic animals? There will only be very few theories to choose from that will be able to satisfy this linguistic condition of being able to correspond to items in natural language and also apply to non-linguistic animals.

Chater and Heyes discuss the implications that would arise if some but not all of their criteria are met for any given theory of concepts. In the first such scenario, where the first or second criteria have not both been met, then we cannot ask whether animals, like humans, have concepts (1994:210). They seem to be referring to my point made above but offering a counter justification to it: that without a theory that applies both to humans as well as animals, there is no basis for comparing the two species. I think that this criterion is a little stringent for the following reason. Chater and Heyes' overall claim is that unless a theory can be found that understands concepts independent from language, there is no way in which theories of concepts can be applicable to animals. If such a theory was indeed found that applies to animals and did not depend on language, it risks failing the first criterion, applicability to humans. This is in fact the counterintuitive conclusion arrived at by Chater and Heyes with regard to perceptual theories, as will be seen.

In the second scenario entertained, if the empirical tractability criterion is not met, then although we might have found a theory that is in principle applicable to animals, we cannot test whether it does in fact apply to animals (1994:210). This speculation is again a little defeatist, relying on a lack of ingenuity on the part of designers of experiments. It assumes that even if we were to find an apparently applicable theory, it would not be testable because ethologists would be unable to come up with an experimental design that circumvents the lack of language problem in animals.

Chater and Heyes don't specifically discuss the implication if only criteria two and three are met. What could be wrong with having a theory that applies to animals, is empirically tractable, but doesn't apply to humans? In some sense then, failure of the first criterion results in an automatic failure of all three criteria, or at least removal of the theory from consideration, since it doesn't apply to humans. This seems to run counter to the endeavor of finding a theory of concepts that applies to animals. I think that the first criterion should be removed from the list for two reasons. The first is that keeping the first criterion ensures that the sense of concept being sought must correspond to natural language items and this unnecessarily restricts the range of theories being looked at. Second, human applicability is not necessary for a theory of concepts that will be applied to animals. Human applicability forms a useful base for comparison and could comprise one thread of research into the area of comparative cognitive ethology. But as we have seen, when it's the only aim, it might prematurely close the door on animal abilities that are not humanlike. The counter argument to the idea of removing the human applicability criterion from the list is usually that we would not be able to call the theory a theory of concepts. I think the only real conclusion to be drawn in such a case is that the theory cannot be deemed a theory of concepts that applies to humans. The aim with regard to investigating concepts in animals, i.e., comparative versus non-comparative, must be gotten straight from the beginning. The non-comparative aim of investigating animal concepts, that these theories should stand on their own, appears to be missing from Chater and Heyes' examination. This point becomes obvious in their two interpretations considered next.

There are two interpretations of the relation between human and animal concepts that underlie Chater and Heyes' search for a theory. The first interpretation is that the animal's categorization behavior is mediated by mental structures of the same sort that are postulated in human theories of concepts, such as definitions, sets of exemplars or prototypes. The second interpretation is that animals may be judged to have concepts because they can learn discriminations that correspond to human categories. The first interpretation is the stronger of the two, for it assumes strong similarity of mental processes between the two species. This strong interpretation is unlikely to be met given the lack of as yet identification of a lexical system of communication in animals. It also

prematurely narrows the range of possible theories to be examined, in that they must relate to natural language. The second weaker interpretation shows more promise at being met by animals. However, if animals were found to make discriminations that don't correspond to human categories, they could not be fit into either interpretation above. Although the second weaker interpretation seems to move away from the aim of comparison a little, it is not entirely divorced from it. Thus neither interpretation on Chater and Heyes' view is completely devoid of comparison aims and this is a problem, because they don't allow for examination of a theory that fails criterion one, i.e.,that doesn't also apply to humans.

Chater and Heyes evaluate three types of theories under the strong interpretation, those of Definitional, Exemplar and Prototype. The first criterion, applicability to humans, is automatically passed for all three theories since the theories under scrutiny are theories of human concepts. The only possibilities for failure are thus criteria two and three, animal applicability and empirical tractability.

According to the definitional view of concepts, to possess a lexical concept is to know a set of necessary and sufficient conditions for category membership. For instance the concept 'chair' would have as one of its necessary conditions 'four leggedness'. Lexical items of natural language are represented in terms of complex definitions in a system of internal representation, a so-called language of thought (1994:213). Since this theory concerns the relationship between lexical concepts and a proposed language of thought, and animals do not have natural languages, it cannot be applicable to animals. It thus fails criterion two, applicability to animals. Even if, on Chater and Heyes' view, one were to assume a language of thought in animals, there are experimental difficulties that would then make the definitional view violate criterion three. It would be impossible to translate experiments done on humans that demonstrate the definitional view, which include reasoning and comprehension tasks, to animals (1994:214). Both these tasks involve questions with lexical items and verbal answers on the part of the tested subject.

It could be objected that Chater and Heyes' description of the definitional view is too restricted, that not all concepts on the definitional view must necessarily correspond to lexical items. Rising to the objection that there is not a definition for every concept, it is claimed on some versions of the theory that some concepts correspond to complex

lexical items or representations that break down ultimately to sensory primitives which are themselves undefined (Laurence and Margolis, 1999:9). It could be argued that animals possess these sensory primitives. At any rate, the view has been unable to meet other objections made by opponents, in particular concerning a failure to agree on a particular definition for many concepts, and so the definitional theory of concepts is probably not the mechanism underlying human concept possession either.

On the Exemplar Theory, a concept consists in a set of representations of particular instances of that concept. For instance, to have a particular concept such as that of 'chair' is to have a set or list of stored representations of chairs that have been encountered in the past (1994:215). The Exemplar view is different from the definitional view in that the membership set of a concept is a list of representations rather than a lexical item from a presumed language of thought. In other words, one does not need to have a language of thought for this theory to apply. One must, however, demonstrate how an item belongs to a particular concept, and this is made much more difficult by the fact that the item is not in the form of a word, but in the form of a set of representations. In the human case, stored exemplars of concepts are assigned internal labels corresponding to the natural language labels literally assigned to the stimuli they represent. This mechanism cannot reasonably be translated to non-linguistic animals. There is a trade-off of sorts. The theory shows promise in its applicability to animals since no language is assumed. However, it loses on empirical tractability since it is difficult to empirically demonstrate that particular representations are concepts without appeal to the mechanism of natural language. Without such a theory of what makes one representation rather than another a concept, the theory cannot be applied to animals. The theory thus fails criterion two as well as criterion three.

One would think that the Exemplar theory is indeed empirically tractable especially since it is easily demonstrable experimentally by stimulus generalization. In a typical experiment, animals are rewarded for choosing one item from a set of two that falls under the target concept. For instance, a pigeon would be rewarded for choosing a barstool over a loveseat if the target concept was 'chair'. However, on Chater and Heyes' account, stimulus generalization is not a proper experimental corollary of concept-mediated behavior. Stimulus generalization often is presented as a forced choice

situation where the subject must choose one of only two items, and the objection cannot be ruled out that the subject is merely responding to paired associations as opposed to choosing an item because it is an instance of a particular concept. Moreover, potential ambiguity results from the fact that concept mediated accounts, however they might be manifest behaviorally, cannot be sufficiently distinguished from the act of stimulus generalization in the testing situation. That is, it is impossible to determine with certainty that animals are not in fact responding to paired associations (a case of stimulus generalization) instead of picking out an instance of a particular concept (true concept-mediated behavior). In order for the animal's behavior to be truly deemed concept mediated, instances that correspond to concepts would have to be recognized by the animal as having features in common. On a forced choice paradigm such as in stimulus generalization experiments, this aspect is not necessarily demonstrated. Additionally, cross-referencing between concepts that have overlapping features is impossible to demonstrate experimentally. For instance, stimulus generalization may be able to explain the ability to distinguish dogs from non-dogs, and furry from non-furry, but not both features at once (1994:216). Briefly, the view cannot be applied to testing procedures without an adequate account of how exemplars are linked to concepts.

If there is no account of how exemplars are bound to concepts, how then is the view applicable to humans? The theory thus fails criterion one as well. The theory thus fails all three criteria, on Chater and Heyes' view. I happen to think that the view, with its experimental corollary of stimulus generalization, is most promising as a potential theory of concepts in the animal. The fact that on this view concepts are not necessarily lexical items but rather representations, and that demonstrations of concept possession can be done non-verbally in an experimentally tractable manner points to the fact that this type of theory should be further explored for its potential applicability in animals.

On the third theory that Chater and Heyes examine, the prototype view, concepts are complex representations whose structure encodes a statistical analysis of the properties their members tend to have (Laurence & Margolis, 1999:27). An item is categorized as falling under a concept if it is sufficiently similar to the central tendency or prototype of the concept. This view differs from the definitional view in that the properties in question are not necessary for possession of the concept, and it is not

necessary to satisfy all of the relevant features in order for the item to fall under the concept. This is again a non-linguistic theory and so could in principle be applicable to animals. It suffers from the same problem as the theory on Chater and Heyes' view, however: the process of categorization, however it might be manifested behaviorally, is not sufficiently distinguishable from the behavior of stimulus generalization in the testing situation, so it fails criterion three.

The prototype view, like the Exemplar theory, may even fail criterion one, on Chater and Heyes' view. The idea is that category or concept discrimination is probably not based on feature analysis because the visual world is too complex for the brain to be able to compute features from the visual array. The variability and complexity of natural stimuli and even of artificial ones such as geometrical shapes makes the approach improbable in the human case. It should be clear at this point that Chater and Heyes' criteria are too stringent and are preventing theories that are not so intimately tied to language that would ideally be applicable to animals from being considered, often because they fail criterion one, applicability to humans.

On the basis of the above albeit very brief discussion of the three types of human theories of concepts that fall under the first interpretation, Chater and Heyes conclude that the categorization behavior of animals cannot be mediated by the same type of mental structure that is thought to underlie human theories of concepts. They then claim that the argument could be advanced that it is obscurantist to claim that animals possess concepts on the basis of any empirical data, since it is not clear what is being postulated (1994:221). The possibility is left open however, that what it is to possess a concept is not a matter of possessing a particular internal structure. They next examine theories of concepts based on the weaker interpretation of the mechanism underlying concept possession. That is, that animals may be judged to have concepts because they can learn discriminations that correspond to human categories. One would think that examining theories where language is not a pre-requisite would be the logical starting point for Chater and Heyes, since animals don't have language. Perceptual theories will have a much better chance at passing criteria two and three, being more applicable and empirically tractable in animals, than theories that require a language.

Chater and Heyes survey the literature on perceptual theories of concepts next, starting with the correlational view. The correlational view on concept possession states that possessing a concept consists in the subject's ability to discriminate instances from non-instances of that concept. Perceptual theories differ from the more 'lexical' theories described above in that a concept is described as simply an internal representation with no commitment as to what the structure of the representation is. This type of theory, as will be seen, is very similar to Allen's theory of concept attribution. The phenomenon of stimulus generalization mentioned earlier would also be a good experimental demonstration of this category of theory. The correlational view is inapplicable to humans, on Chater and Heyes's account, and so fails criterion one. The problem with the human version of the theory is that it is unable to account for error in categorization. That is, one object may be mistaken for another, or an object may be failed to be taken as such. A possible reason for human error has been cited as failure to categorize correctly because of sub-optimal perceptual conditions. Proposals put forth to account for error include the division of perceptual conditions into those that are optimal and those that are sub-optimal. Categorization errors only occur in sub-optimal perceptual conditions, such as when light is insufficient, or the subject is too far away from the object. If a principled way of distinguishing the two types of conditions could be found, the problem would apparently be avoided. According to Chater and Heyes, it is difficult to see how to define the distinction between the two conditions in a non-circular way. Furthermore, even if researchers were to work on the theory to make it applicable to humans, empirical tractability in animals would then be compromised, and so it would then fail criterion three.

There are two problems with Chater and Heyes' reasoning above. First, it is unlikely that humans do not categorize or recognize concepts on a perceptual basis at least occasionally. Second, it is unjustified to fail the theory on applicability to humans just because it lacks a completely worked-out account of error. Accounting for error hasn't been raised as an issue thus far with regard to other theories of concepts. Third, it is difficult to see how, if researchers could improve the theory to make it applicable to humans, it would therefore become less applicable to animals. If we accept Chater and Heyes' conclusion here, what does that say about their classification scheme? I think it

makes it abundantly clear that criterion one should be removed from the scheme. If the theory is applicable to humans on their reasoning, empirical tractability in animals is automatically compromised. This reasoning renders the whole exercise of finding a theory of concepts that applies to animals as well as humans counterproductive.

At this juncture Chater and Heyes entertain a potential objection to their overall claim. The claim is that unless a way is found to separate language from concepts, non-linguistic animals cannot be said to have concepts. It could be objected that since such a close connection is posited between concepts and language, an implicit division is thereby imposed between linguistic and non-linguistic animals. It could then be argued that the above claim does not apply to linguistic animals, if such a category of animals exists. Certain animals could possess concepts if they also possessed a language, and it is up to someone else to find out whether animals have a language or not (1994:228). Chater and Heyes do not believe that research into this category of so-called linguistic animals should be pursued. They state two caveats that should dissuade anyone from thinking that research into the linguistic capacities of animals is a viable research path to take. The first is that very few species of animals, if any, possess the linguistic abilities to be properly described as concept possessors. Such animals would have to use predicate expressions, which have a particular syntactic/semantic role in the animal's language. This presupposes that the animal's language has a syntax and a semantics, and most languages of animals do not have both these attributes, according to them. There are thus only a few species of animals to which concepts could, even in principle, be applied.

The second caveat is that, on a practical level, it would be much more difficult to identify which concepts an animal has than in the human case. We have seen this problem as it was outlined by Davidson in chapter one, and also with regard to theories of intentionality and the difficulty of identifying or at least of narrowing down the content of mental states. The concept of number has often been cited as a relatively easy concept to identify in the animal, and many have claimed that certain species of animals can count. The rejoinder to this is to claim that a number concept can only be applied to creatures that can perform additional feats such as the arithmetic operations of multiplication and division. Here again, the strategy is to 'up the ante', i.e., additional corollary capacities

are smuggled into the class of things that would imply concept manipulation and possession. Why would a creature possessing the number concepts, one, two, three, and four have to be able to perform the range of human arithmetic operations on these numbers in order to be said to be using these concepts?

The problem with both these caveats is the same. With language as well as numbers, a much too enriched definition of language and numbers is being used. With such an enriched notion of language, a holistic view of language acquisition must also be held; it must be assumed in the case of a child that the child either knows language, or not. It is the same with numbers, the child either knows how to count and perform arithmetic operations on numbers, or not. If one does not wish to espouse this extreme 'all or nothing' view of concept and language acquisition, in particular because it contradicts the way that both these capacities are acquired in the child, the alternative is to think that there are levels or degrees of 'sophistication' to language and concepts. In this case then, the child need not have all that human possession of a number concept entails and yet still be operating with the concept of 'four' in some minimal sense. In terms of language, as the case of children shows one could be operating with a much impoverished syntax and semantics and still be attributed concepts.

Chater and Heyes next anticipate an objection that takes the form of one of the main issues of discussion in cognitive ethology, that of experimental versus naturalistic observation. In view of the fact that naturalistic observation is the main observatory tool of the ethologists, they next look at attempts to test concept possession in a naturalistic setting as opposed to the laboratory situation, and on animals that communicate in an apparently linguistic manner.

Chater and Heyes admit that so far, they have been discussing experiments in a laboratory setting where control is achieved by examining the behavior of socially isolated members of a limited range of species, in a standard apparatus, and in relation to objects that are not designed to resemble those that the animals might encounter in their natural environments (1994:232). This is a most succinct expression of the contrast between naturalistic observation and experimental manipulation. They note in their article that Dennett claims that in the situation of laboratory experimentation rigor of method is exchanged for relevance of results. Dennett's flippant remark is an

understatement of the general idea that while greater control over variables is achieved in the environment of a laboratory, the lack of the animal's natural habitat brings up questions as to how representative the animal's reactions are. This dense claim made by Chater and Heyes must be unpacked to elucidate the issues involved. First there is the setting. The issue is how much the animal's environment constrains its behavior. If it turns out that the animal's natural environment plays a large role in shaping the behavior of the animal, then the argument could be made that a laboratory setting is artificial and so constrains the animal's behavior as to restrict it to unitary stereotypic movements. The practice of using only socially isolated members of a species could also have a negative impact on the generalizability of results, especially if notions like 'social intelligence' continue to gain viability at least in those cases where the species lives in a community environment. Isolated members of a species don't have the prior and maybe crucial experience of interacting with the other members. The issue of using only a limited range of species to study will certainly impact issues of generalizability to other species, and maintains a bias toward only studying certain species of animals that is not characteristic of the field of cognitive ethology. The employment of an artificially conceived of standard apparatus to train all species with might mask or contradict the true natural abilities of each species, abilities which might differ from species to species. Last is the issue of using objects not found in the animal's natural environment. It will be more difficult to assess the animal's degree of tool-use for instance, if one is only employing novel tools not found in the environment of the species under study. Given the above, I am of the opinion that Chater and Heyes should have started their search for a theory of animal concepts with data from naturalistic observation in the wild as opposed to from the laboratory.

Chater and Heyes begin their survey of the wild by asking: can the problem with animal concepts be remedied by relocating the animal back to its natural environment? They have chosen to discuss Cheney and Seyfarth's research on vervet monkeys, one of the most thorough naturalistic studies to date, in their opinion. Cheney and Seyfarth have identified four separate alarm calls given by vervets in danger situations. One of these calls they have identified as the leopard call, and its function is to defend against predation by leopards. Chater and Heyes claim that in order to get around the

indeterminacy of content objection, i.e., how can we be sure that the content of the call contains the concept 'leopard', the researchers appeal to evolutionary explanation. Evolutionary explanations explain concept possession by appeal to the proper or adaptive function of the behavior. If the leopard call is an adaptation for defense against leopards, then the vervets could reasonably be described as possessing the concept 'leopard'. Chater and Heyes find a problem with this, however. An attribute is an adaptation with respect to a particular function only if it was the fulfillment of that function which resulted in the retention of the attribute through natural selection. There is a question, according to them, as to whether the current function of an attribute is a reliable indicator of its adaptive significance. The point is that reliance on history to tell us whether an attribute has been retained through natural selection does not often pan out because the information about past conditions and events is lost in the mists of time (1994:234). All this to say that evolutionary theories fail on criterion three, they would not support an empirical investigation into animal concepts.

I think it is ironic that Chater and Heyes would fail the theory on this particular criterion, since Cheney and Seyfarth's work represents one of the first examples of successful experimental manipulation within the confines of a naturalistic environment. Cheney and Seyfarth were able to mitigate the indeterminacy of content problem precisely with experimental manipulation. Certainly the theory is empirically tractable; in fact, this is how the researchers were able to narrow down the possibilities for the content of the alarm calls. The way that they went about narrowing the content through experimental manipulation brings out the necessity in using a combination of the two types of investigation, rather than choosing laboratory experimentation over naturalistic observation. The manipulation employed was to play previously recorded alarm calls in certain circumstances yet still within the confines of the animal's natural environment. Moreover, Chater and Heyes seem to ignore the most important aspect of evolutionary theory, that survival is indicative of an adaptive trait. If the alarm call was not adaptive, all of the members including the caller would perish. Finally, I think it is also ironic that Chater and Heyes would choose Cheney and Seyfarth's work with vervet monkeys in particular to discuss. Let us try to imagine for a moment how we would go about setting up this experiment in a laboratory situation. It hinges on discovering the nature of alarm

calls in monkeys. Imagine the difficulties involved just in capturing a leopard to use for the experiment, and then providing all the apparatus necessary for the monkeys' escape routes, such as trees or bushes. What validity can the lab experiment possibly possess if it doesn't allow the subjects to escape in the manner they would normally when in the presence of a predator, i.e., by climbing trees?

The problems with Cheney and Seyfarth's research cited above are used to argue for a completely contrary claim by Chater and Heyes: good reason for a return to the more unnatural methods of laboratory investigations. Their argument is as follows. If the adaptive function of an attribute depends on the history of natural selection, then the content of a concept depends on its history of selection through learning, and it is in the lab, rather than the field, that we have the best opportunity to record this history (1994:235). In light of the considerations discussed above concerning experimental manipulation versus naturalistic observation, this argument, with its emphasis on learning, is not very convincing. The first thing to note is that on most accounts, a history of selection through learning cannot be considered as evolutionary in the strict sense of the term. On evolutionary accounts, it is evolution or natural selection and not learning that shapes the animal's behavior. Moreover, while the case could be made that experimental manipulation brings out the sought out capacity in a much more rapid manner than waiting and observing through the process of naturalistic observation for it to occur, the importance of registering the animal's reaction in its natural environment is equally important to an evolutionary explanation, in order to ensure that it is the animal's natural reaction that is being recorded, and not some artificial stereotypic movement. The appeal to rapidity is not justified in this case, especially when we consider what is lost in the process.

Chater and Heyes entertain a possible objection to their general claim before making their final conclusion. It will be remembered that their general claim is that unless a sense of 'concept' can be found that is independent from natural language, there is no plausible sense in which non-linguistic animals can be said to possess concepts. Their conclusion is that since they have not been able to find such a sense, that the term 'concept' cannot be used in the same way in discussions pertaining to humans and animals. The general objection entertained is that Chater and Heyes are mistaking the

true aim of investigation into animal concepts, which is twofold. First, investigations aim to determine whether animals can discriminate stimulus categories based on human concepts, and if so, the second aim is to determine whether the concepts used are in any way like human concepts. Note that the comparison aim is secondary, contrary to Chater and Heyes' two interpretations mentioned above where it appears as the first interpretation. This objection also makes the more general point that Chater and Heyes are mistaking the failure to assimilate animal concepts with human concepts for a total failure to find concepts in animals. The situation is not quite this dreary. All they can conclude is that certain human theories of concepts, particularly those that rely on language possession, probably do not apply to animals. Claiming that animals do not possess human concepts does not mean that animals possess no concepts at all. Just as we saw in chapter one, claiming that animals do not possess a human language does not thereby mean that they possess no language at all. The main point of investigations into animal mentality is not to vindicate the existence of human attributes in animals, although this may be a corollary aim.

There is a point of comparison to be made between the twofold aim of cognitive ethology with regard to concept investigation and the strong and weak interpretations of the link between human and animal concepts in Chater and Heyes' discussion. The first aim of cognitive ethology corresponds to the weak interpretation. That is, ethologists are primarily interested in whether animals can discriminate stimulus categories. The fact that the categories are based on human concepts is because there are no other types of concepts known to humans that could serve as a comparison basis at this early stage of investigation. The mention of comparison to human concepts in the second aim constitutes a branch of comparative cognitive ethology, where studies are made of animals with the underlying aim of elucidating human capacities.

This issue of confusion in aims in cognitive ethology has been discussed in an article by Colin Allen (1999). In this article, Allen suggests that Chater and Heyes' strategy is a rather anthropocentric investigation of concepts in animals, because it focuses too much on assimilating animals to humans. He then goes on to explain that his 1991 article, where he and Marc Hauser elucidate a minimal constraint on concept ascription, could have been misconstrued by authors such as Chater and Heyes. The

reason is that a comparative aim is explicitly stated by Allen and Hauser: that they are attempting to render plausible the claim that animals can be shown to be operating with internal representations that function rather like human concepts. They also employ two thought experiments on human reactions to death with the idea that these experiments would allow for comparison between human and animal reactions to death. In Allen's later work, particularly in his working out of three criteria for the attribution of a concept, he is trying to move away from what he views as the anthropocentric angle that a comparative approach offers, i.e., justifying the attribution of concepts to animals by using human behavior in similar circumstances as the benchmark. The ultimate aim in his later investigations is that research into animal concepts should be able to stand on its own (1999:34).

The issue here, in my opinion, is what specific purpose the comparative approach is being used to accomplish. There are two possibilities. The comparative approach is either being used to advance and elucidate theories of human capacities, or it is being used to justify the existence of animal capacities. The first purpose is beneficial, when explicitly stated as such, for advancing our knowledge of human beings. This second purpose is where trouble arises, for it is where the challenges discussed in the first four chapters are allowed to enter in.

**3. A Minimal Constraint**

Colin Allen, as a precursor to his later work on a theory of animal concepts, and in conjunction with Marc Hauser, has developed a minimal constraint on concept ascription. The basis of Allen and Hauser's research is a theory-theory approach within an evolutionary framework. Generally speaking on a theory-theory construal, a concept's identity is determined by its role within a theory (1991:49). In answer to the question of what theoretical role mental ascriptions might play, Allen and Hauser believe that mentalistic terminology allows a mode of description that enables explanation within an evolutionary framework. This type of framework is being adopted by more and more cognitive ethologists. If the project of cognitive ethology is viewed as the study of behavior within an evolutionary framework, then an animal's behavior is examined in light of its function and evolution. Mentalistic terms provide a level of description that is

appropriate to the functional level of description that is the concern of evolutionary hypotheses. A mental state relates organisms to their environments through its content. A mental state will be adaptive insofar as its content makes links between the environment and the organism's behavior. Mentalistic terms thus provide a natural vocabulary for ethologists to frame their hypotheses. If one were to apply non-mentalistic terms, or purely behavioristic terms to this framework, one would not get the same result.

Allen and Hauser make two preliminary distinctions that are central to the elucidation of a minimal constraint on concept ascription. The first is to posit a difference between a concept and an internal representation (1991:50). One can attribute an internal representation without attributing a concept. They make use of an example in the literature to illustrate the difference. Herrnstein's (1976) work on alleged concept-possession in pigeons has been widely cited both for and against the idea of animals possessing concepts. In a typical experiment, pigeons were shown numerous pictures corresponding to the categories of trees, water and persons, along with other pictures that were considered 'near misses', i.e., not in the category. The pigeons were able to pick out pictures corresponding to their respective categories. Herrnstein concluded that pigeons possess concepts corresponding to certain natural categories. Allen and Hauser do not believe that concept possession is warranted in the case of the pigeons. While they allow that the pigeons could indeed recognize features of certain categories that were present in the pictures and then use these properties to recognize a general class or category, they do not believe that the pigeons were operating with a concept. They believe that this experiment illustrates the difference between category discrimination and concept-mediated behavior, a difference that hinges on the act of recognition.

The second point they make is thus to distinguish between two forms of recognition, between the ability of recognizing an X and the ability of recognizing something as an X (1991:51). The ability to recognize an X can be thought of as an extension of a discrimination ability, and corresponds to the behavior of the pigeons in the experiment. That is, the individual may have the ability to classify things into two categories or classes, that of X and non-X, but this ability can also arise as a result of accidental co-extension. The second ability, recognizing something as an X, requires an

internal representation that abstracts away from the perceptual features that enable one to identify X. This distinction basically functions to discriminate between stimulus generalization and actual concept possession. The act of stimulus generalization is recognizing an X and the act of recognizing something as an X is an instance of concept-mediated behavior.

This distinction in senses of the term 'recognition' is interesting and one can see how it might advance the issue. We can now see how the act of recognition might be pivotal in distinguishing mere discrimination from concept possession. What the distinction lacks however, in my opinion, is a description of the behavioral manifestation of concept possession. Category discrimination is clearly behaviorally manifested. To recognize an X is to pick it out from an array of objects. The notion of choice, or picking the object out as opposed to another or other objects, is crucial to the act of discrimination. What would be the behavioral analogue to concept-possession? Recognizing something as an X is not easily behaviorally manifested other than verbally, i.e., by naming the concept it corresponds to. This is probably why most experiments that test for concept possession in animals are discrimination tests; we lack a test that would evidence concept-mediated behavior in non-verbal animals. As we shall see below, Allen and Hauser have come up with two thought experiments that they think should demonstrate concept-mediated behavior.

Allen and Hauser make a third distinction, having to do with types of concepts, although they do not explicitly state it as such. They make a distinction between concepts that are perceptually direct but that might still involve some abstraction of features, and what they call abstract or higher order concepts, where presumably few or no perceptual features are present. "Square" is considered a perceptually available concept, since many of the features necessary to its identification are perceptually available such as four sidedness etc. The concept of "death", the topic of investigation in the two thought experiments, is considered a higher order concept, since few or none of its identificational properties are visually available. There is no actual criterion that must be passed in order for a concept to be considered abstract per se, rather the rule of thumb is that if the concept has more features that are not perceptually given than features that are given, then it is an abstract higher-order concept. On this distinction, many concepts

find themselves as falling midway on the scale, because many concepts have a combination of both types of features. A potential problem with this type of classification scheme is that any concept that has only perceptually available features and that is successfully discriminated will always be susceptible to explanations based on category discrimination rather than concept mediation.

Regardless of whether this distinction carves concepts into the right types or not, if Allen and Hauser are going to make this distinction, they need to allow for the possibility that concepts with perceptually available features are going to be evidenced by the act of discrimination, and that the discrimination ability in those cases is nonetheless a case of concept mediated behavior. One can see that in distinguishing abstract concepts they are trying to rule out the possibility of accidental co-extension, which is coincidental in nature. However, there must be a way to demonstrate the possession of abstract concepts experimentally without completely disqualifying discrimination behavior.

Based on the above distinctions, Allen and Hauser have come up with a constraint on cognitive representations for them to count as concepts. It is the following:

> An abstract concept could reasonably be attributed to an organism if there is evidence supporting the presence of a mental representation that is independent of solely perceptual information (1991:54-5).

One can see what Allen and Hauser are getting at here: they are trying to rule out acts of discrimination based on perceptually available cues. However, given the distinction between types of concepts outlined above, one has to conclude that this constraint applies to higher-order concepts more so than the lower-order perceptually available type. Moreover, the situation remains the same for perceptually available concepts. There is no way to demonstrate them since stimulus generalization has been ruled out as a possibility.

Allen and Hauser have developed two thought experiments that put this constraint into effect, and thus should help formulate the beginnings of an empirical program into concept attribution. They first suggest two features of behavior that would evidence the constraint, that is, they'll outline two possible ways in which concept mediated behavior could be manifested. One is that an organism whose internal representations are concept-like should be able to generalize information obtained from a variety of perceptual inputs

and use that information in a range of behavioral situations. Second, organisms that can be said to possess a concept should be able to alter what they take to be evidence for an instance of that concept (1991:55).

The two experiments have been designed to be conducted in the species' natural habitat, and although they could be modified for the lab setting, it would be better to run the experiments in the environment where natural selection has shaped the animal's behavior. The first experiment, performed on Vervet monkeys, would test mothers' responses to distress calls made from a loudspeaker that have been previously recorded of their own offspring. The offspring belonging to the mothers being tested have recently died. This experiment should test whether mothers are able to generalize the information of seeing the death of their offspring to the situation of hearing a distress call from the offspring. There are three possible reactions to the call, on Allen and Hauser's account.

1. They might respond as they did when the infant was alive (i.e., look towards the speaker).
2. They might respond in a more agitated fashion (i.e., initiation of searching behavior).
3. They might not respond at all, continuing the activity they were engaged in prior to the playback (1991:56).

If the mothers reacted in the first way, then it seems reasonable to conclude that these Vervets do not have the concept of death, because they react as if the infant is still alive. (There is the possibility that the initial reaction is one of shock. A human might react the same way upon hearing the voice of his or her recently deceased conspecific. The delay in reaction would be crucial in deciding the issue and should thus be recorded in this experiment.) The second reaction has two interpretations: either the mother believes the infant is still alive, or the call has elicited some kind of surprise reaction. Whether or not the mother has witnessed the actual death of her infant is relevant to her reaction in this case. Reaction 3 is more decisive as evidence for the possession of a concept, according to Allen and Hauser, in that the ability to 'turn off' a response seems to indicate that the animal has recognized the finality of the disappearance of the infant. There are other interpretations to this reaction, or lack of a reaction, however. Possibilities include that the mother did not hear the call, that she heard the call but did not associate it as

belonging to her infant, that she believes the call cannot have come from her infant because her infant is dead, etc.

What constitutes the 'right' response in this experiment? Theoretically, if the mothers were operating with the concept of death, they would realize the finality of death and either fail to respond when the fake distress call was played, or react with surprise. Since two responses are equally acceptable, this experiment suffers somewhat from ambiguity. Moreover, the case can be made that some species of animals operate with the concept of death, for instance, some animals feign death in the presence of predators. There is no allowance in this type of experiment for the fact that feigning death is an adaptive response to predation, and that perhaps animals are operating with the concept of death in cases where they pretend to be dead. Instead reactions to artifical situations regarding death are investigated here. Should we take the reaction of the animal when her dead offspring 'comes back from the dead' as central to the concept of death, rather than the behavioral adaptive success of it?

The second thought experiment investigates the second testable feature of behavior mentioned above, the creature's ability to alter what they take as evidence for an instance of a concept, in this case death. This experiment involves administering a drug that makes an animal seem dead to all appearances, placing the 'dead' animal in a cage with its conspecifics, and recording their behavior once the drug wears off and the animal begins to revive itself. The second trial involves a repeat of the first, and the third trial involves choosing another individual animal as a target and putting it in turn through the same two trial procedure. The point is to see if a change in response in the conspecifics' behavior occurs, and then whether this change in response is generalized to subsequent 'dead' individuals. There are two possible reactions on the part of the animals, one is that they fail to modify their behavior, and the other is that they modify their behavior the second time the same animal is drugged and the first time the new animal is drugged. The first reaction does not seem to warrant the explanation that the animals are displaying concept mediated behavior, since they would be reacting as if the animals were really dead, thus failing to alter what they take to be evidence for it. However, the second reaction does suggest that they are operating with the concept of death and that they are able to alter what they take to be evidence for it (1991:58).

In this second experiment, the theoretically ideal response is if the other animals alter what they take to be symptoms constitutive of death and do not remove the conspecific. The right response in this experiment, not taking evidence of death to be final because of certain evidence to the contrary, is opposite to the right response in the first experiment, taking evidence of death to be final, regardless of evidence to the contrary. From an adaptive point of view, I don't think that either of these experiments are compatible with what really goes on in nature. Consider what havoc would be wreaked on the animal kingdom if animals truly were observed to second-guess their judgements made about death. Certainly Burghardt's snakes would no longer be able to use feigning death as an escape route from predators, since the predator would be prone to checking if the snake was really dead or not.

These two experiments are only thought experiments. Is there any way to turn them into real performable experiments? Ethical considerations most likely would prevent these experiments from being carried out. Even setting ethical considerations aside in the animal case, ethical considerations in the human case would prevent the human analogue from being carried out, and we would thus have no comparative ability. In the case of testing for an abstract concept such as death, it would be helpful to have the human results already established in order to have a point of comparison. After all, the human benchmark is already in place in imagining the three possible reactions of the animals.

In my opinion, this minimal constraint on concepts, while it might be very interesting on a theoretical level, raises problems in its practical application. In principle, Allen and Bekoff can be seen as advancing the issue on concept attribution by coming up with a constraint that taps into possession conditions for abstract concepts, concepts that don't have perceptually distinguishable properties. In practice there are a few problems with this idea. One is that concepts with uniquely perceptible properties get disqualified. Many concepts are actually a mix of perceptually available properties and abstract properties. Moreover, identification of the abstract properties is rendered more difficult on a practical level by the fact that they are perceptually unavailable. As an example, consider what might be the abstract properties of death. Allen and Hauser offer no clues as to what properties they have in mind. The second problem is the lack of a behavioral

analogue for concept possession. How is it possible to demonstrate that one possesses or is operating with a concept independent of perceptual information? The method of forced choice is ruled out since it conflates with discrimination and discrimination is not sufficient to be characteristic of concept mediated behavior. I think that the source of the problem is that Allen and Bekoff have carved concept types at the wrong joints. There is something to the idea that some concepts have abstract or perceptually unavailable features, but I don't think that this is the most central feature of concepts in general and thus the wrong aspect to focus on with regard to concept attribution in animals.

**4. Behavioral Criteria**

In Colin Allen's later work, he actually develops and advances three criteria for the attribution of a concept to an animal. He offers a reason why none of the current theories apply to animals, that constitutes a difference outlined in the beginning of the chapter, that between what it is to have a concept and when it is reasonable to attribute a concept. Most of the current theories offer a philosophical analysis of what it is to possess a concept. His three criteria will rather stipulate when it is reasonable to attribute a concept to an animal. These three criteria were mentioned in chapter one as being a theory that goes some way to showing how it could be demonstrated that a non-linguistic animal could be attributed concepts.

Allen first makes a distinction between two notions of 'concept', that is, between the social and the individual notion (1999:35). Allen notes that there is a tension within the notion of a social concept, marked by the phrase 'the concept of X' that remains unanalyzed in the psychological literature. Use of the definite article 'the' implies that there is such a thing as a single construal of X that all individuals share. From the perspective of the individual, marked by the phrase 'Fred's concept of X', it seems unlikely that there is a single such construal of X, but rather that there exist several different but overlapping strands of the concept X. This is an interesting observation to make, and the influence of Wittgenstein's notion of 'family resemblance' is clear. The idea is that there is no essence to a particular concept, that is, concepts do not have a given set of properties that can be identified, rather the set of often overlapping properties constitutes a family resemblance type of classification.

Social and individual notions of 'concept' play different roles in a theory of concepts. Social concepts play a role in explaining cooperation and communication among individuals. Individual concepts are implicated in the structure of individual behavior and differences between the behavior of individuals. Allen claims that philosophical discussions are plagued by failure to heed the distinction between the social and the individual construal of concept. For instance, in the views of Davidson examined in Chapter one there is a failure to heed the difference. It will be recalled that one of Davidson's reasons for not wanting to attribute the belief to Malcolm's dog that the cat went up the tree is that the dog lacks the constituent concepts of 'cat' and 'tree'. Allen takes it that by including the article 'the', that Davidson must be referring to the human and social construal of the concept. Allen claims that there is no reason to think that the dog must have that particular human social construal of the concepts cat and tree, since it might not exist even at the level of humans, and further that the dog must have no concepts whatsoever if it lacks that particular canonical construal of the concept.

The underlying basis for Allen's three criteria for concept attribution is this notion of individual concept and its relationship to perception. From his work with Hauser on developing a minimal constraint in concept attribution, it is possible to demonstrate, with the evolution of his work, a possible connection from perception to concept formation in animals. The issue that Allen is confronting in the evolution of his work is the following. A person with a concept is able to discriminate between an array of items. Discrimination thus constitutes one of the various abilities of a concept holder. However, a person who can discriminate is not necessarily operating with a concept. The issue is thus to establish when it is reasonable to attribute a concept to an individual above and beyond the mere ability to discriminate, since discrimination is not a decisive indicator. Recalling his establishment of a minimal constraint on representations, it can be seen from that work that he was stating a condition necessary in the organism to distinguish the act of mere stimulus generalization from concept using. That is, somehow the animal must be able to abstract features from the perceptual situation that will not be present in every perceptual situation. This constitutes an additional step up from stimulus generalization, where the animal merely uses the features perceptually available to it to discriminate between two items. Allen thus believes that some animals can construct

category schemes that transcend particular perceptual stimuli. Concepts are the nodes of such category schemes.

Thus an organism may be reasonably attributed a concept X whenever

1. The organism systematically discriminates some X's from non-X's.
2. The organism is capable of detecting some of its own discrimination errors between X's and non-X's.
3. The organism is capable of learning to better discriminate X's from non-X's as a consequence of its acquisition of capacity 2 above (1999:37).

Central to this set of conditions are the detection and recognition of error, and the subsequent modification of behavior as a result of this recognition. Allen notes that the above three capacities are empirically tractable in languageless animals. This is a particularly interesting bonus, for it provides a solution to the issue of how concept mediated behavior could be evidenced aside from through verbal response. On this set of criteria, concept-use is directly behaviorally evidenced.

According to Allen, capacity 1 has already been extensively investigated. As we saw earlier, experiments on pigeons have shown that they are able to choose items out of an array based on their belonging to a particular category. Over and above this category selection capability, Allen notes that capacity 1 is also systematic, that is, not based on rote memorization of stimulus response training trials by the animal. Evidence for this is shown by the fact that the trial stimuli in the experiment were different from the training stimuli. For instance, if the pigeons had to choose faces from an array that included other body parts, the test trials contained faces not seen in the training trials. Capacity 1 is also seen as more complex than forced choice discrimination (choice between two items only), in that forced choice discrimination requires only the ability of stimulus generalization. However, capacity 1 demands less than what is required by the minimal constraint, that is to choose items as belonging to the same concept category despite the items not, from a visual point of view, appearing to belong to the same category. This last ability of abstraction is really what is characteristic to concept possession. For this reason, Allen has included the other two conditions. The other two conditions together should allow investigators to settle questions about the content of the representations, although they have not been empirically tested as of yet.

Condition 2 concerns the detection of error. I mentioned in chapter 5 and 6 that an account of error is important for a theory of intentionality. Error is also important in a theory of concept attribution. Allen means to take the notion of error to a more complex level by making the detection of one's own errors a condition of concept attribution. He gives an example of an experimental phenomenon that could be interpreted as personal error detection and recognition. In the experiment, a group of pigs were tested for making same/different choices on pictures of faces and other body parts. In a particular trial for instance, the pigs would be presented with either of two situations, one would be two pictures that are both of the same body part or both of a face, the other possible situation would involve two pictures of different body parts or a body part and a face. The subjects would have to choose either the 'same' or the 'different' answer choice, depending on if the photos represented sameness or difference in the two photos. The pigs performed at about 90 % accuracy. Most interestingly however, during the commission of errors, it was noted that the pigs physically backed away from their incorrect choices 22 of 23 times, demonstrating to Allen that the pigs were aware that they had made an error.

Condition 3, according to Allen, is the hardest to articulate and defend. He thinks the difficulty is not so much a matter of empirically demonstrating condition 3, but rather that satisfaction of condition 3 provides a link between the first two conditions. The reasoning is the following. If detection of ones own errors, condition 2, bears on the capacity to make discriminations, condition 1, then this demonstrates that the animal may be comparing the stimuli to an abstract representation or a concept. These abstract representations are worthy of the label 'concept' because they are independent of the perceptual representations.

Concerning the pessimistic conclusion by Chater and Heyes, that there seems to be an unbreakable link between concepts and language, Allen thinks that many have been seduced into this conclusion by the fact that languages provide a structure that has a vast number of degrees of freedom with respect to immediate perception. Linguistic representation is the most fine-grained system of conceptual representation that we know. But it would be premature to conclude that it is the only one (1999:39). Not only do I agree wholeheartedly with this remark, but also I think it sums up perfectly the problem

with searching for a theory of concepts that applies to animals as well as the search for intentional mental states in animals. The fact that humans have language has many benefits: thoughts and concepts are easily demonstrated in the human through the avenue of language. However, this ease of demonstrability also masks other potential avenues for these items to be demonstrated, and it is in the search for these items in animals that this difficulty is encountered and must be circumvented.

## Conclusion

As mentioned in the beginning of the chapter, there are at least three questions that researchers might be interested in with regard to concepts and their applicability to animals:

1. What is a concept?
2. What is it to possess a concept?
3. When is it reasonable to attribute a concept?

A theory attempting to answer question one is concerned with what the nature of a concept itself is. The other two questions are interested in concept possession or attribution. Not only is there much overlap between the three questions, particularly between the second and third questions, but also many researchers end up interested in more than one question, hence the three questions are interrelated. Colin Allen, it will be remembered, offers this distinction (between nature and possession) and the fact that many theories offer a philosophical analysis of what it is to possess a concept rather than offering criteria for when it might be reasonable to attribute a concept as a possible reason why many of the current theories cannot be applied to animals.

Chater and Heyes, as discussed in the chapter, are looking for a sense of concept that meets three criteria and claim that unless a sense of concept can be found that is independent of natural language, animals cannot be said to possess concepts. Usually, a creature is said to possess or be operating with a concept if they pass some criteria that demonstrates mastery or possession, criteria that could be behavioral or verbal or both in nature. The criteria that Chater and Heyes list apply to the term 'concept' itself rather than the creature's meeting certain conditions. The case can be made that Chater and Heyes are really investigating two questions, the first having to do with the nature of the

term 'concept' and the second with applicability of certain theories to animals. In this case, their first criterion, or at least the second part of it, is unnecessary. Their first criterion, it will be remembered, is that the sense of concept must apply to humans and assign them concepts corresponding to terms of natural language. The stipulation that the sense of concepts must assign terms corresponding to terms in natural language is unnecessary, because any theory that does not meet this stipulation will fail to be considered, and this is ironic considering that any theory whose members correspond to natural language items is unlikely to apply to animals in the first place. With the elimination of this second part of the criterion, many theories, especially those not based on lexical items or a language of thought, might otherwise be promising places to search from the point of view of research.

In my opinion the constraint conceived of by Allen and Hauser, although theoretically interesting, is not empirically tractable. The aim in a search for a theory of concepts, as with a theory of intentionality, is that above all it should be empirically tractable. Colin Allen's three criteria for attribution of a concept meet this condition as well as not depending on language, and thus constitute a good starting point for research into animal concepts.

## Conclusion

In the interim summary at the end of the first half of the thesis, I noted four so-called 'grains of truth' that could be retained from each of the chapters. These grains of truth can be expanded upon to generate possible future research paths that could be taken by cognitive ethologists.

From chapter one and the discussion of the 'no thought without language' argument I looked at a second strategy that could be taken to circumvent Davidson's arguments that would require construing the term 'language' in a larger sense to mean a system of communication. Taken in this wider sense, the animal's system of communication could be considered a language. One possible research path would entail discovering the nature of the system of communication of various species of animals, as well as it's various functions and uses. Within the realm of function, the 'intention' of certain vocalizations could be studied in much the same way as Cheney and Seyfarth have done with vervet monkeys, except with a wider range of species.

From the discussion on anthropomorphism in chapter two I concluded that the practice of anthropomorphism, when used for heuristic purposes, was well suited to test possible hypotheses concerning mental states in animals. As noted at the end of chapter three, it forms the basis for the development of a new fifth aim in ethology, concerned with studying the private experience of animals. Since the private experience of another creature can only be known at best through indirect means, as the discussion on Nagel made clear, one promising way to tap into this experience is to generate hypotheses about how the animal might react in a given situation by asking ourselves how we might react given the same situation. Getting over the obstinate opinion that the private experience of another creature is an area forever closed to scientific investigation, I think this area should be studied, even if only to find out that perhaps Dennett is right, that animals do not have anything that might be deemed subjective experience because they lack an inner eye or an overseer of this experience. Certainly advances soon to be made in the area of cognitive neuroscience will help to determine if animals have any kind of experience, if indeed they do, and also what the content of that experience might be.

As mentioned in chapter four, experimentation is central to the discipline of cognitive ethology. However, the objection that experimental investigation into animals is nearly impossible and made even more difficult by their lack of human language is, in my opinion, a naïve perspective that is the result of an inability to adapt to the special nature of the object being investigated. Certainly wavelengths are not investigated in the same empirical manner as chemical reactions. Extending the analogy, it would be naive to assume that animals should be studied by the same experimental means as humans. Add to this the fact that more often than not in human experiments, because humans are so capricious and have the common communicational ability of language, the true aim of the experience has to be cloaked or masked by some other only slightly related aim in order not to contaminate the results. Looking at the situation in this way, it turns out that experimentation on animals is made easier by the fact that subject-deception is, or has so far been found to be, unnecessary with animals. However, as was made abundantly clear in chapter four's discussion and especially with regard to higher-order mental state investigation such as a theory of mind, there is still a long way to go with regard to refining this type of investigation. This is in large part due to the nature of the topic under investigation, namely mental states, and their unobservability. The fact that mental states must be inferred makes experimental results open to interpretation. This situation is what allows other types of explanation such as those behaviorist or mechanistic in nature equally able to account for results, a seemingly hopeless situation from the point of view of those attempting to demonstrate that animals may have some type of theory of mind that mediates their behavior.

It is exactly this frustrating situation or 'stand-off' between types of explanation that initially led me to investigate, in the second half of the thesis, theories of intentionality and concept-attribution. The case could be made, although it was not my initial intention, that both topics in the second half are also motivated by two objections. The objection that the chapters on intentionality respond to is that even if animals have mental states, we have no way of finding this out because there is no theory out that can tell us how to look for them or when we have found them. The theories of intentionality that I examined are concerned with investigating whether animals possess mental states that are intentional, and because at least one theory can discern intentional from non-

intentional behavior, an answer to the objection is obtained. As for theories of concepts, the objection could be made (and is investigated by Chater and Heyes) that since language is so intimately bound up with concepts, then animals cannot possess concepts because they lack human language. Believing that concepts needed to be divorced from language before any other worthwhile investigation can be embarked upon, I decided to treat this objection in the chapter.

From the standpoint that I arrived at as a result of my investigations in this thesis as well as the many issues that I did not examine, many potential research paths present themselves. With regard to theories of intentionality, the case could be made that although Bennett's Guiding Rule breaks the stand-off situation with regard to explanations, how empirically tractable his theory is constitutes an entirely different issue. Since Bennett mentions himself that Dennett's theory provided the starting point for his own reflections and subsequent development of his Guiding Rule, it might be a viable endeavor to combine the two theories, given that Dennett's theory is very empirically tractable, and Bennett's has the bonus of mitigating the lacunae found in Dennett's theory.

I did not treat the topic of consciousness in any great depth except for a mention in chapter three of Aiken's and Dennett's views on the matter. Aside from the relation between self-awareness and consciousness, a vast topic of which an entire separate thesis could be constructed, there are many other facets of this phenomenon that could be treated. More input from philosophers and psychologists on the task of dividing consciousness up into levels and or types would be welcome as a starting point (Ristau, 1992). Part of the reason that I chose not to examine consciousness directly is the sheer vastness and variability of the literature on the topic. Another issue related to consciousness is the question of higher order mental states and whether animals possess them, in other words theory of mind research. In my brief examination of this issue I have concentrated mainly on defending this branch of research against objections and criticisms. There is still much to be discovered in this area, such as other possible indicators of a theory of mind, particularly within the realm of the recent creation of the notion of 'social intelligence'. Also of particular interest is the discovery and subsequent claim by Marc Hauser that certain species of animals, in particular dogs, have been

observed to engage in play behavior. This claim is interesting because play is not considered to have much, if any, evolutionary function. The discovery of the possibility of play behavior in certain species of animals runs very slightly contrary to the presently dominant view that all behavior in animals has evolutionary significance.

As for theories of concept attribution, it will be remembered that I cited the major flaw in Chater and Heyes' search for a theory as being the second part of criterion one, which stipulates that the sense of concept that they are searching for must correspond to humans and assign them concepts corresponding to terms of natural language. If this second part was dropped, a good comparative search of the literature could then be conducted with the remaining two and a half criteria: that the theory be applicable to humans, to animals, and be empirically tractable in animals. Moreover, if the relation to natural language were dropped, many of the theories of concepts surveyed by Chater and Heyes would have been good candidates for animals, especially the perceptual theories. If it was found that none of the theories applying to humans were also applicable to animals, then the comparative aim could be dropped altogether, the first criterion removed, and the search performed again. I don't think that this would even be necessary however, because there must be some overlap between humans and animals, at least within the realm of perceptual concepts.

Concerning research into concept possession in animals, there is a lot that is already being done, such as Hernstein's work with pigeons, and Allen's work. On a philosophical theoretical level, a new and interesting line of research is presently being examined. Since theories of concepts applying to humans are not so easily applicable to animals, in part because of the lack of language in animals, some authors have recently wondered if perhaps the notion of non-conceptual content could be applied to animals. Representational content that is non-conceptual means that the possessor of the content need not possess the concepts for the properties, objects and relations that are included in the representational content. This is an attractive proposal in its applicability to animals for two reasons. The first is that the propositional content of mental states is difficult to pinpoint in the animal, especially given the lack of language in animals. Perhaps the situation is even as dreary as Millikan prophesizes, that the content of animal states cannot be translated into sentences of a human language. The second reason that non-

conceptual content is an attractive proposal in that it circumvents the objection that one cannot attribute mental states to animals without the caveat that they must possess the requisite concepts. If the content of the animals' states is non-conceptual, there is no need to demonstrate that the animal possesses the concepts named in the content.

Central to this proposal are three issues: the first is to establish the existence of non-conceptual content. The second issue is to determine what the relation is between non-conceptual content and conceptual content, if there is one. A third issue concerns what type of explanatory or causal role, if any, the non-conceptual content plays in the content of experience. To argue for the existence of non-conceptual content, the most popular argument usually appealed to is the fine-grained nature of experience argument which is briefly that perceptual experience outstrips the conceptual resources of the perceiver so that it becomes almost necessary to posit the notion of non-conceptual content to pick up the slack. Another argument that has been used is the idea that experience is independent of belief. The phenomenon of perceptual illusions is often used to provide a rationale for this idea, i.e., the perceiver sees a waterfall image on a sheet of paper and the water appears to be moving even though he or she knows that the waterfall is not actually in motion.

As for the relation between non-conceptual and conceptual content, the issue is whether there a relation between the two types of content or is non-conceptual content autonomous, as some authors have argued. Christopher Peacocke has advanced the Autonomy Thesis, stating that it is possible for a creature to be in states with non-conceptual content even though the creature possesses no concepts at all. Some authors, although they accept the notion of non-conceptual content, deny that it can be autonomous (Bermudez and Macpherson, 1998).

Concerning the explanatory role that non-conceptual content plays, it is obvious that this will be difficult to establish, given the so-far indirect arguments employed to establish its existence, i.e., picking up the slack created by outstripped conceptual resources. That is, if there is no way to demonstrate the existence of non-conceptual content except by derivation, demonstrating the explanatory or causal role it plays will be that much more difficult.

The above potential research path is a good example of the main principle of this thesis, which is that cognitive ethology can benefit from philosophical input. As is the general opinion of many researchers in cognitive ethology, philosophical input provides the necessary theories, and cognitive ethology ideally should pursue an empirical investigation of these theories. One thing is certain: topic areas such as investigation into mental states and concepts in animals can only proceed in so far as they are theoretically well-developed from a philosophical point of view.

Akins, Kathleen. (1996). "A Bat Without Qualities?" In Bekoff and Jamieson, 1996, Readings in Animal Cognition, 345-358.
Allen, C. (1999). "Animal Concepts Revisited: The Use of Self-Monitoring as an Empirical Approach". Erkenntnis, 51, 33-40.
Allen, C. and Bekoff, M. (1995). "Cognitive Ethology and the Intentionality of Animal Behavior." Mind and Language, vol 10, (313-328).
Allen, C. and Bekoff, M. (1997). Species of Mind: The Philosophy and Biology of Cognitive Ethology. MIT Press.
Allen, C. and Hauser, M (1991/1996). "Concept Attribution in Non-Human Animals." In Bekoff and Jamieson, 1996, Readings in Animal Cognition, 47-62.
Asquith, Pamela. (1984). "The Inevitability and Utility of Anthropomorphism in Descriptions of Primate Behavior." In Harré and Reynolds (eds.), 1984, The Meaning of Primate Signals, 138-174.
Asquith, Pamela. (1997). "Why Anthropomorphism is Not Metaphor: Crossing Concepts and Cultures in Animal Behavior Studies." In Mitchell, Thompson and Miles (eds.), 1997, Anthropomorphism, Anecdotes and Animals, 22-34.
Beer, Colin. (1991). "From Folk Psychology to Cognitive Ethology." In Ristau (ed.), 1991, Cognitive Ethology, 19-33.
Beer, Colin. (1997). "Expressions of Mind in Animal Behavior." In Mitchell, Thompson and Miles (eds.) Anthropomorphism Anecdotes and Animals, 198-207.
Beisecker, David. (1999). "The Importance of Being Erroneous: Prospects for Animal Intentionality." Philosophical Topics, vol. 27 (281-308).
Bekoff, M. and Jamieson, D. (1996). Readings in Animal Cognition. MIT Press.
Bekoff, M., Allen, C., and Burghardt, Gordon M. (eds.) (2002). The Cognitive Animal: Empirical and Theoretical Perspectives on Animal Cognition. MIT Press.
Bekoff, M and Allen, Colin. (1997). "Cognitive Ethology: Slayers, Skeptics and Proponents." In Mitchell, Thompson and Miles (eds.), Anthropomorphism Anecdotes and Animals, 313-334.
Bennett, Jonathan. (1976). Linguistic Behavior. Cambridge University Press.
Bennett, Jonathan. (1987). "Thoughtful Brutes." Proceedings and Addresses of the American Philosophical Association, 62, 197-210.
Bennett, Jonathan. (1990). "How Is Cognitive Ethology Possible?" In Ristau, (ed.), 1991, Cognitive Ethology, 35-49.
Bennett, Jonathan. (1991). "How to Read Minds in Behavior: A Suggestion from a Philosopher." In Whiten, (ed.), 1991, Natural Theories of Mind, 97-108.
Best, John B. (1986). Cognitive Psychology. West Publishing Company.
Budiansky, Stephan. (1998). If a Lion Could Talk: Animal Intelligence and the Evolution of Consciousness. Free Press.
Burghardt, Gordon M. (1991). "Cognitive Ethology and Critical Anthropomorphism: A Snake With Two Heads and Hog-Nose Snakes That Play Dead." In Ristau (ed.), (1991), Cognitive Ethology, 53-90.
Burghardt, G. (1994). "Evolution and the Analysis of Private Experience." Psycoloquy, December 1994.
Burghardt, Gordon M. (1997). "A Fifth Aim for Ethology." In Mitchell, Thompson and Miles (eds.), 1997, Anthropomorphism, Anecdotes and Animals, 254-276.
Byrne, R.W. and Whiten, A. (eds.) (1988). Machiavellian Intelligence: Social Expertise

and the Evolution of Intellect. Oxford University Press.
Carrol, Lewis. (1916). Alice in Wonderland. Rand McNally & Company.
Carruthers, Peter. (1989). "Brute Experience." The Journal of Philosophy, vol 86, 258-269.
Chater, N. and Heyes, C. (1994). "Animal Concepts: Content and Discontent." Mind and Language, vol. 9, 209-246.
Davidson, Donald. (1975/1984). "Thought and Talk." In Inquiries into Truth and Interpretation, Oxford:Oxford University Press.
Davidson, Donald. (1982). "Rational Animals." Dialectica, 36, 317-327.
Davidson, Donald. (1984). Inquiries Into Truth and Interpretation. Oxford: Oxford University Press.
Davidson, Donald. (1997). "Seeing Through Language." In Preston, (ed.), 1997, Thought and Language, 15-27.
Davidson, Donald. (1999). "The Emergence of Thought." Erkenntnis, 51, 7-17
Davis, Hank. (1997). "Animal Cognition Vs. Animal Thinking: The Anthropomorphic Error." In Mitchell, Thompson and Miles, (eds.), (1997) Anthropomorphism Anecdotes and Animals, 335-347.
Dennett, Daniel (1969). Content and Consciousness. Routledge and Kegan Paul.
Dennett, Daniel. (1971). "Intentional Systems." The Journal of Philosophy vol 8, (87-106).
Dennett, Daniel (1978). Brainstorms: Philosophical Essays on Mind and Psychology. MIT Press.
Dennett, Daniel. (1987). The Intentional Stance. MIT Press.
Dennett, Daniel. (1991). Consciousness Explained. Little Brown and Company
Dennett, Daniel. (1996). Kinds of Minds. Basic Books.
Dennett, Daniel. (1998). Brainchildren: Essays on Designing Minds.
Dreckmann, Frank. (1999). "Animal Beliefs and Their Contents." Erkenntnis, 51, (93-111).
Evnine, S. (1995). "On the Way to Language." In Hahn, (ed.), 1995, The Philosophy of Donald Davidson.
Fisette, D. (ed.) (1999). Consciousness and Intentionality: Models and Modalities of Attribution. Kluwer Academic Publishers.
Fisher, John. (1996). "The Myth of Anthropomorphism." In Bekoff and Jamieson (eds.), 1996, Readings in Animal Cognition, 3-16.
Gallup, Gordon G., Anderson, James R., and Shillito, Daniel J. (2002). "The Mirror Test." In Bekoff, Allen and Burghardt, (eds.), The Cognitive Animal, 325-333.
Glock, Hans-Johann. (2000). "Animals, Thoughts and Concepts." Synthèse, 51, 35-64.
Goldstein, E. Bruce. (1996). Sensation and Perception. Brooks-Cole Publishing Company.
Griffin, D.R. (1976). The Question of Animal Awareness: Evolutionary Continuity of Mental Experience. Rockefeller University Press.
Griffin, D.R. (1984). Animal Thinking. Harvard University Press.
Guthrie, Stewart. (1997). "Anthropomorphism: A Definition and a Theory." In Mitchell, Thompson and Miles (eds.), 1997, 50-58.
Harman, Gilbert. (1994). "Positive versus Negative Undermining in Belief Revision." In Kornblith (ed.), Naturalizing Epistemology, 317-336.

Harré, R. and Reynolds, V. (eds.) (1984). The Meaning of Primate Signals. Cambridge University Press.

Herman, Louis M., and Morrel-Samuels, Palmer. (1996). "Knowledge Acquisition and Asymmetry between Language Comprehension and Production: Dolphins and Apes as General Models for Animals." In Bekoff and Jamieson, (eds.), Readings in Animal Cognition, 289-306.

Heyes, Cecilia M. (1998). "Theory of Mind in Nonhuman Primates." Behavioral and Brain Sciences, 21, 101-148.

Heyes, Cecilia M. and Dickinson, Anthony. (1990). "The Intentionality of Animal Action." Mind and Language, vol. 5 (87-104).

Hofstader, D. and Dennett, D. (1981). The Mind's I. Bantam Books.

Honderich, T. (ed.) (1995) The Oxford Companion to Philosophy. Oxford University Press.

Hurley, Patrick J. (1982). A Concise Introduction to Logic. Wadsworth Publishing Company. Fourth Edition

Kennedy, John. (1992). The New Anthropomorphism. Cambridge University Press.

Kiriazis, J and Slobodchikoff, C.N. (1997). "Anthropomorphism and the Study of Language." In Mitchell Thompson and Miles, (eds.), Anthropomorphism, Anecdotes and Animals, 365-369.

Kornblith, Hilary. (1994). (ed.) Naturalizing Epistemology. MIT Press.

Laurence, S and Margolis, E. (1999). "Concepts and Cognitive Science." In Margolis and Laurence (eds.), Concepts: Core Readings, 3-81.

Malcolm, Norman. (1972) "Thoughtless Brutes". Proceeding and Addresses of the American Philosophical Association, 46 (1972-73), 5-20.

Margolis, E. (1999) "How to Acquire a Concept." In Margolis and Laurence, Concepts, 549-567.

Margolis, E and Laurence, S. (1999). Concepts: Core Readings. MIT Press.

Millikan, Ruth. (1984). Language, Thought and Other Biological Categories. Bradford Books/MIT Press.

Millikan, Ruth. (1993). White Queen Psychology and Other Essays for Alice. Bradford Books/MIT Press.

Millikan, Ruth. (1997). "Varieties of Purposive Behavior." In Mitchell, Thompson and Miles, (eds.), Anthropomorphism, Anecdotes and Animals, 189-197.

Mitchell, R.W., Thompson, N.S., and Miles, H.L. (1997) "Introduction: Taking Anthropomorphism and Anecdotes Seriously". In Mitchell, Thompson and Miles, (eds.), Anthropomorphism, Anecdotes and Animals, 3-11.

Nagel, Thomas. (1974/1979). What is it Like to be a Bat? In Nagel, Mortal Questions,165-180. (1979). Cambridge University Press.

Peacock, Christopher. (1992). A Study of Concepts. A Bradford Book.

Premack, D. (1988). "Does the Chimpanzee have a Theory of Mind? Revisited." In R.W. Byrne and A. Whiten (eds.), Machiavellian Intelligence: Social Expertise and the Evolution of Intellect, 160-179.

Premack, David, and Woodruff, Guy. (1978). "Does the Chimpanzee Have a Theory of Mind?" Behavioral and Brain Sciences, 4, 515-526.

Ristau, C. (1991). Cognitive Ethology: The Minds of Other Animals. Lawrence Earlbaum Associates.

Ristau, C. (1992). "Cognitive Ethology: Past, Present and Speculations on the Future." Philosophy of Science Association, vol. 2, 125-136.

Rivas, Jesus, and Burghardt, Gordon M. (2002). "Crotalomorphism: A Metaphor for Understanding Anthropomorphism By Omission." In Bekoff, Allen and Burghardt, (eds.), The Cognitive Animal, 9-17.

Savage-Rumbaugh, S, Shanker, S.G. and Taylor T. (1998). Apes, Language and the Human Mind. Oxford University Press.

Searle, John. (1983). Intentionality: An Essay in the Philosophy of Mind. Cambridge University Press.

Searle, John. (1994). "Animal Minds." Midwest Studies in Philosophy, vol XIX, 206-219.

Seymour, M. (1999). "Two Concepts of Belief." In Fisette, Consciousness and Intentionality, 311-344.

Shusterman, R.J. and Gisiner, R.C. (1997). "Pinnipeds, Porpoises and Parsimony: Animal Language Research Viewed From a Bottom-Up Perspective." In Mitchell, Thompson and Miles, (eds.), Anthropomorphism, Anecdotes and Animals, 370-382.

Silverman, Paul S. (1997). "A Pragmatic Approach to the Inference of Animal Mind." In Mitchell, Thompson and Miles (eds.), Anthropomorphism, Anecdotes and Animals,170-185.

Spada, Cenami. (1997). "Amorphism, Mechanomorphism and Anthropomorphism." In Mitchell, Thompson and Miles, (eds.), Anthropomorphism, Anecdotes and Animals, 37-49.

Stephan, Achim. (1999). "Are Animals Capable of Concepts?" Erkenntnis, 51, 79-92.

Taylor, T. (1998) "Rhetorical Inclinations". In Savage-Rumbaugh, Shanker and Taylor, Apes, Language and the Human Mind, 139-180.

Whiten, Andrew. (1991). (ed.) Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading. Oxford: Basil Blackwell.

Timberlake, William. (2002). "Constructing Animal Cognition." In Bekoff, Allen and Burghardt, (eds.), The Cognitive Animal, 105-113.