Mind & Language ISSN 0268-1064
Vol. 9 No. 3 September 1994

Article

Animal Concepts: Content and Discontent

NICK CHATER AND CECILIA HEYES

1. Introduction: The Status of Research on 'Animal Concepts'

1.1 Preliminaries

The psychology of human concepts is concerned with the structure of concepts such as FISH, PERSON, LIKES and so on. The stock of concepts that we are assumed to have corresponds rather directly to the stock of predicates of natural language. The nature of this relationship between human concepts and natural language is central to research on animal concepts, since, unless there is some way of understanding concepts that is independent of their connection with natural language, nonlinguistic animals cannot have concepts.

This problem has not prevented comparative psychologists from describing themselves as studying 'concept formation' in animals. Nor should it have done. Just as intuitions about infinity and the geometry of space have been overthrown by mathematical and scientific advances (Putnam, 1962), so speculations about the impossibility of animal concepts may be overtaken by intellectual progress. Experimental research may have made sense of the claim that animals have concepts; a sense that is simply not evident to pretheoretic intuition. The empirical literature certainly gives

We should like to thank Martin Davies, Stevan Harnad, David Premack and an anonymous reviewer for valuable comments on this paper. Correspondence should be addressed to Nick Chater, who is now at the University of Oxford, Department of Experimental Psychology, South Parks Road, Oxford, OX1 3UD.
Email: nicholas@cogsci.ed.ac.uk.

the impression that animals' concepts have been clearly demonstrated, and ipso facto that concepts need not depend on natural language.

To assess whether this has indeed been shown, we examine the various accounts of concepts to be found in human and animal psychology and in philosophy. In particular, we search for a sense of 'concept' that:

(1) applies to humans, and assigns to them concepts corresponding to terms of natural language;
(2) can be applied to nonlinguistic animals;
(3) allows empirical investigation of animal concepts. In animal concepts research, such empirical investigation has generally taken the form of behavioural experiments, rather than, for example, neurophysiological studies.

If any of these criteria are unfulfilled then it is difficult to see how animal concepts research could be coherent. If (1) and (2) cannot both be fulfilled, then we cannot even sensibly ask whether animals, like humans, have concepts. If (1) and (2) are fulfilled but (3) is not, then it may be possible to formulate clearly the questions of whether and what concepts animals have, but animal concepts research will be unable to answer them. We will argue that there are currently no accounts that fulfil all three of these criteria. That is, the idea of a concept has not been successfully decoupled from natural language, and hence there is currently no coherent account of what animal concepts might be. Notice that we are not adopting the familiar line of stressing that possessing concepts is a major cognitive feat, of which few or no animals may be capable. Rather, we argue that we simply have no account of what cognitive feat possessing a concept amounts to for nonlinguistic agents, and hence cannot assess whether this or that animal possesses concepts or not, still less ascertain the content of its putative concepts.

1.2 Background

In comparative psychology, categorization experiments are variously described as studies of 'discrimination learning' and 'concept formation'. The latter description began to be applied consistently by comparative psychologists about 80 years ago (Bingham, 1914; Hunter, 1913; Johnson, 1914), and achieved some currency in the 1920s and 1930s when experiments by Lashley and others (Fields, 1932; Lashley, 1930; Munn, 1931) suggested that rats that had been trained to discriminate, say, an equilateral triangle from a rectangle, would, with only minimal training, continue to select the triangular stimulus despite variations in its form (e.g. isosceles or scalene), orientation, brightness or background. This led to the suggestion that rats can acquire the 'abstract idea' or 'concept' TRIANGLE (Washburn, 1926). In the 1940s and 1950s the investigation of animal concepts was less vigorous, but since the 1960s it has undergone a revival

and an increasing number of animal studies have been described by their authors as studies of 'categorical concept formation' (e.g. Cerella, 1982, 1986; D'Amato and Van Sant, 1988; Greene, 1983; Herrnstein and Loveland, 1964; Herrnstein, Loveland and Cable, 1976; Lubow, 1974; Malott and Siddall, 1972; Morgan, Fitch, Holman and Lea, 1976; Roberts and Mazmanian, 1988; Schrier, Angarella and Povar, 1984; Siegel and Honig, 1970; Wasserman, Kiedinger and Bhatt, 1988). The current preoccupation is, however, with the possibility that animals have or can acquire 'natural' categories such as TREE and FISH rather than 'artificial' concepts such as TRIANGLE. Whereas artificial concepts experiments use very simple, specially constructed stimuli, such as simple geometric shapes, natural concepts experiments attempt to teach the animal highly complex, real-world categories, generally using photographs of real scenes or objects. For example, the ability of a pigeon to learn to discriminate slides depicting trees from slides not depicting trees has been interpreted as evidence that pigeons can acquire the TREE concept (Herrnstein et al., 1976).

The foregoing is typical of recent experiments on categorical concept formation in that (1) complex visual stimuli are used, (2) subjects are pigeons (monkeys are also sometimes tested), and (3) the procedure is successive discrimination, in which pecking a response key in the presence of stimuli belonging to one category (S+) is rewarded with food, while pecking the same response key in the presence of stimuli belonging to another category (S−) is not rewarded. Subjects are judged to have acquired the discrimination if they respond at a higher rate in the presence of S+ stimuli than in the presence of S− stimuli, but this is seldom regarded as evidence of concept formation unless a difference in rate of responding is maintained when novel stimuli are substituted for those used during training, i.e. unless 'transfer' is observed. Finally, the study of Herrnstein et al. (1976) is typical in using two complementary stimulus categories, with the S− category or negative set defined as consisting of stimuli lacking something that is present in the S+ or positive set.

This may remind nonspecialist readers of behaviourist methods, but in fact most contemporary comparative psychologists working in this area are willing to explain behaviour with reference to internal states and processes. Indeed, many would describe themselves as 'animal cognitive psychologists', or contributors to the field of 'animal cognition', and one of the primary aims of research in this field is to find out precisely what kind of internal structures and processes underlie the categorization performance of animals (e.g. Roitblat, 1982). Research on animal concepts is certainly not unique in this respect; the majority of contemporary research on learning and memory in animals could be described as investigating internal structures and processes. (For convenience, however, we shall use 'comparative psychology' to refer to the study of nonhuman species and reserve the label 'cognitive psychology' for the study of humans.) What is distinctive about animal concepts research is the extent to which it aims to specify not only the structure of animals' internal

representations, but also their content; *what* an animal can discriminate, learn, remember or know. Furthermore, the use of experimental categories such as 'tree', 'person' and 'cat', which are paradigmatic examples of human concepts, suggests that researchers want to find out whether the content of animals' internal structures is, or can be, the same as that of humans.

Thus, comparative psychology has undergone a change not only in theoretical orientation but also in theoretical ambition. This is problematic, we shall argue, because the literature on human concepts has little to contribute to fulfilment of the ambition of animal concepts research; it does not provide good leads in the form of appropriate theoretical tools or empirical methods. Specifically, concepts and natural language are so closely tied together that it is not clear how to make sense of animal concepts.

1.3 *Structure of the Paper*

A natural interpretation of what it would be for animals to have concepts is simply that their categorization behaviour is mediated by mental structures of the same sort that are postulated in theories of human concepts—definitions, sets of exemplars or prototypes. We consider this possibility in Section 2, and argue that it does not provide a foundation for animal concepts research since it ties concepts to natural language. Only the prototype view can be formulated nonlinguistically, and even this is unsatisfactory for independent reasons. Nor does empirical research on human concepts provide methods that can be applied to other species, since the stock of concepts is taken as given (by natural language predicates), rather than being the object of investigation, as in comparative psychology. In consequence, comparative psychologists have been compelled to chart a relatively lone experimental course and have reached an impasse.

Section 3 explores a different potential connection between human concepts and putative animals concepts. Rather than assuming that animals have internal structures or mechanisms in common with humans, perhaps animals may be judged to have concepts because they can learn discriminations that correspond to human categories. This requires some account of what it is to categorize in a particular way, to have a concept with a particular content (e.g. TREE or FISH), which allows humans and animals to be compared. Experiments on animal concepts have presupposed a 'correlational' view of categorization which identifies concepts perceptually rather than linguistically. Considerations derived from informational semantics (Dretske, 1981; Fodor, 1990; Stampe, 1977) reveal theoretical problems with this approach, which are reflected in the practical difficulties that animals concepts researchers have encountered in determining which features of stimuli are the basis for performance in discrimination learning experiments. Viewing these difficulties as symptomatic of the problems identified through informational semantics suggests that they are not

merely technical; that more ingenious experimentation will not make them go away.

In Section 4 we broaden the focus to consider studies of animal communication, so-called 'relational' concepts and more 'naturalistic' studies. Finally, in Section 5, we consider the implications of the problems we have identified in animal concepts research for comparative psychology in general. We suggest that these problems may be avoided insofar as comparative psychologists are able to make only very general, rather bland, assumptions about the content of animals' internal representations.

2. *The Structure of Concepts: Leads from Cognitive Psychology?*

2.1 *Introduction*

One suggestion concerning how concepts can be identified nonlinguistically is that a concept is an internal state with a particular structure (a definition, a prototype, or a set of exemplars), which, in the human case, may or may not have the same meaning as terms of a natural language. According to such a view, in human natural language terms inherit their meaning from the corresponding concepts, but nonlinguistic animals may also be concept-holders in good standing. Equating the structure of animal and human concepts in this way might provide the basis for explaining the content of animal concepts. To explore this suggestion, we now consider the extent to which a range of theories of human concepts can be applied to animals, taking definitional, exemplar and prototype theories in turn, and find none to provide an adequate starting point for animal concepts research.

2.2 *Definitions: Failing on Criteria 2 and 3*

According to the definitional view, to possess a lexical concept is to know a set of necessary and sufficient conditions for category membership. For example, to have the concept BACHELOR is to know that a bachelor must be adult, male and unmarried (although see Lakoff, 1973). In contemporary formulations, this means that lexical items of natural language are represented in terms of complex definitions in a system of internal representation, the 'language of thought' (Fodor, 1975). Opponents of the definitional view (e.g. Fodor, 1981) argue that lexical items of natural language correspond directly to primitives of the language of thought, and hence have no non-trivial definition in that language.

In the literature on human concepts, the definitional view has been tested both by attempting to formulate good definitions for particular classes of lexical items (Miller and Johnson-Laird, 1976; Schank and Abelson, 1975) and has been challenged on the grounds that such definitions are seldom or never adequate (Fodor, 1981; Medin and Smith, 1984; Smith

and Medin, 1981; Wittgenstein, 1953). It has also been tested using sentence comprehension tasks. For example, if the word 'bachelor' is internally represented not simply as BACHELOR, but with the explicit negation—NOT MARRIED—then this explicit negation may be expected to interact predictably with other logical connectives in a comprehension task (Fodor, Garrett, Walker and Parkes, 1980, use such a method to undermine the definitional account).

The definitional view of concepts concerns the relationship between lexical items of natural language and a putative human language of thought. Since animals do not have natural languages, similar or even related questions simply do not arise. It seems that criterion (2) cannot be met.

It might be objected that the definitional view might be applied in comparative psychology by asking whether lexical items of *human* natural language correspond to definitions in an *animal's* language of thought. This assumes that animal concepts correspond closely to the terms of human language (this is denied by, for example, Dummett, 1978; Gillett, 1987, 1988). Even if this anthropomorphic assumption could be justified, this recasting of the definitional view of concepts is problematic. First, the definitional view appears to be a nonstarter from an experimental perspective. It is hard to imagine how the kinds of reasoning and comprehension tasks used to test the definitional view in the human case could possibly be adapted for nonlinguistic agents. Second, the availability of the concepts in the putative definitions will presumably be just as uncertain as the concept to be defined. For example, to suggest that an animal's concept WOMAN is internally defined as FEMALE, ADULT and PERSON presupposes three controversial concepts in an attempt to account for one. Perhaps these difficulties could be circumvented if the primitives in terms of which definitions are couched were perceptual. However, while it is difficult to provide plausible definitions of most natural language terms using *any* primitives, it has proved impossible, despite two centuries of empiricist labour, to provide definitions of a perceptual kind (Austin, 1962; Ayer, 1956; Fodor, 1981). Thus, even if the definitional view could be plausibly adapted to apply to animals (i.e. to meet criterion (2)), it appears unable to support the empirical investigation of animal concepts (i.e. to meet criterion (3)).

Before leaving the definitional view, we should stress that it requires much more than that animals be able to represent definitions of some sort. As Fodor (1981) points out, for humans, phrases of natural language (such as 'green tree') are almost invariably assumed to have definitions in the language of thought (in terms of the internal representations for 'green' and for 'tree'). However, this fact does not bear on the question of whether the definitional view of concepts is correct for humans. The point at issue is whether lexical items of natural language map onto definitions in the language of thought. Thus, even if animals have a language of thought, and can frame definitions within it, the definitional view still does not apply to them.

2.3 *Exemplars: Failing on Criterion (2)*

A wide range of theories of concepts falls under the heading of exemplar accounts (Estes, 1986; Hintzman, 1986; Hintzman and Ludman, 1980; Medin and Schaffer, 1978; Nosofsky, 1984, 1986; Reed, 1972; Vandierendonck, 1990). Rather than attempt a survey, let us simply consider the view in its purest, if least sophisticated, form. A concept is held to consist of a set of representations of particular instances of that concept. For example, to have the concept PODIUM is to have a set or list of stored representations of podia that have been encountered in the past. An object is judged to be, and typically named as, a podium if it is sufficiently similar to a stored instance. Exemplar theory has been supported by experiments showing that membership judgments for novel instances of the category are more rapid, accurate, and consistent when the new instance closely resembles a familiar instance (Medin and Schaffer, 1978).

Unlike the definitional view, the exemplar view of concepts appears to have an analogue in the comparative literature, in the form of stimulus generalization accounts of categorization. However, in the comparative field, stimulus generalization is commonly regarded as an *alternative* to concept-mediated accounts of categorization (D'Amato and Van Sant, 1988; Fersen and Lea, 1990; Pearce, 1989; Vaughan and Greene, 1984). Generally speaking, stimulus generalization theories assume that in the course of a typical categorization experiment (outlined in Section 1.2) a representation of all or part of each training stimulus is formed in the animal's long-term memory. Some of these representations become associated with a representation of reward, or of the response that followed stimulus presentation, as a function of the extent to which the stimulus or response predicted reward. When a novel stimulus is presented in a transfer trial, it is compared with all of the representations formed during training and activates each to the extent that it resembles the incoming stimulus. It is assumed that, once it is activated, a representation will excite any reward or response representation with which it is associated to a degree that is determined by the level of its own current activation and by the strength of the pre-existing association. The combined influence of the activated representations is taken to determine the identity and/or vigour of the animal's response to the test stimulus.

Given the similarity between exemplar and stimulus generalization accounts, it may seem puzzling (and somewhat ironic) that while cognitive psychologists regard the former as a theory of concepts, comparative psychologists regard the latter as an alternative to a concept-mediated account of categorization. This is less puzzling when one considers that, according to the exemplar view, each stored representation of a category, say TREE, is labelled as a instance of that category. The stimulus generalization view, in contrast, assumes that stored representations of positive instances have in common only the fact that they have each been independently associated with reward or with a particular response. According to the stimulus

generalization view, instances that are category members from the point of view of the experimenter need not be recognized by the animal as having anything in common.

Experiments suggesting that animals can form 'equivalence classes' may be mistakenly interpreted as evidence that, contrary to the predictions of standard stimulus generalization models, animals' stored representations of category members *do* have something in common beyond the fact that each is independently associated with a common response or trial outcome. Such experiments show, for example, 'transfer of reversal': initially, responding to stimuli $A_1,\ldots, A_n$ is associated with reward and responding to $B_1,\ldots, B_n$ is not. Then the opposite contingency is adopted, so that responding is rewarded for the $B_1,\ldots, B_n$ but not the $A_1,\ldots, A_n$. Exposure to some stimulus items under the opposite contingency facilitates reversal of discrimination performance with respect to the rest of the training set (e.g. Nakagawa, 1986; Vaughan, 1988; Zentall, Steirn, Sherburne and Urcuioli, 1991). The most plausible explanation for transfer of reversal is that associations have been formed between the stored representations of stimuli associated with a common response or trial outcome. In this case, rather than having a common label, as exemplar theories postulate, the stored instances would be connected with one another.

Evidence that even this kind of unification of representations is not the norm in animals comes from the many studies which have failed to find transfer of reversal (e.g. D'Amato, Salmon, Loukas and Tomie, 1985; Preston, Dickinson and Mackintosh, 1986; Sidman, Rauzin, Lazar, Cunningham, Tailby and Carrigan, 1982), and from experiments showing that learning to categorize stimuli using one response does not facilitate their categorization using another response (Bhatt and Wasserman, 1989). Pigeons that had learned to peck four different response keys when shown slides of cats, flowers, cars and chairs, respectively, did not learn any faster than control birds to peck a single response key at a different rate in the presence of each of the four different categories of stimuli.

Even if animals typically showed evidence of having formed associations between stored instances, this would not be sufficient to ascribe them concepts. First, it is not clear how associated sets of instances can be used in inference. Second, the significance of the distinction between symbolic labelling and association is that the same set of exemplars can be labelled by many different labels (so that, for example, a given pair of exemplars can be represented as both being instances of ANIMAL, DOG and FURRY, but as differing regarding FIERCE) whereas association between exemplars is merely present or absent. Therefore while it is possible for different labels to capture many different classifications, which may cross-classify or be arranged in hierarchies, associations can only produce a single partition of exemplars into two or more disjoint sets (related points are made in the context of a critique of connectionist models, by Fodor and Pylyshyn, 1988). In short, stimulus generalization may be able to explain

the ability to distinguish dogs from non-dogs, or furry from non-furry things, but not both at once.

To what extent then does the exemplar view provide a lead for comparative research? Although it is superficially similar to a stimulus generalization account, the exemplar view cannot be directly applied in comparative fields as an account of the kind of mental structure that constitutes a concept. In order to apply this view, it would be necessary to develop a theory suggesting how exemplars might be nonlinguistically labelled as members of the same category, and what principles might govern the assignment of particular exemplars to particular labels. In the human case it is assumed that stored exemplars are assigned internal labels corresponding to the natural language labels literally assigned to the stimuli they represent. Thus, those exemplars which are encountered with the word 'tree', will be marked with a cognitive label 'tree', and become part of the TREE concept. But this view of how exemplar labels are learned cannot, of course, apply to animals.

In view of this, it seems that, to proceed at all, comparative research must provide an alternative account of how exemplars are bound into concepts. It bodes ill that where similar problems have arisen in consideration of the numerical competence of animals it has proved impossible to find consistent evidence of an overt (behavioural) or covert (internal) system of indexing (Davis and Perusse, 1988). Hence, it is not surprising that animal researchers have not attempted to develop a nonlinguistic exemplar view of concepts. Instead they have interpreted exemplar effects (in which the identity of particular training stimuli apparently affects transfer performance) as evidence that stimulus generalization, rather than a concept, underlies categorization performance.

2.4 *Prototypes: Failing on Criteria (1) and (3)*

According to the prototype view, a stimulus will be categorized as a tree if it is sufficiently similiar to the central tendency, or prototype of trees. Since similarity is a matter of degree, membership is typically viewed as a graded, rather than an all-or-nothing property. Evidence for the prototype view has been adduced from a number of experimental methods. People are able to rate reliably the extent to which a member of a category fits the meaning of the word which stands for that category (Rosch, 1973, 1975). Such ratings covary reliably with how quickly category membership can be verified (Rips, Shoben and Smith, 1973), order of acquisition (Anglin, 1970), the speed and probability of generation of an instance when subjects are required to list all the members of a category that they can think of, and so on. Thus, in a task involving the category JEWEL, a diamond will be classified more rapidly, reliably and accurately as a category member than, for example, an opal or bloodstone; in response to a request to name a jewel, 'diamond' is a swift, high-frequency response; and a diamond is rated as a better example of JEWEL than an opal, or a bloodstone.

Unlike the definitional and exemplar views, the prototype view appears to have a natural nonlinguistic formulation. The linguistic version is that instances of concepts corresponding to lexical items will cluster together in an appropriate feature space. However, the existence of such clusters does not depend on there being a lexical item to which each cluster corresponds. Indeed, there are a variety of statistical and neural network methods which find clusters in unlabelled data (e.g. Carpenter and Grossberg, 1988; Hartigan, 1975; Kohonen, 1984). Thus, if category instances genuinely cluster in some perceptual feature space, categories can be learnt without reference to labels provided by natural language.

The majority of the paradigms used to test whether human concepts have a prototype structure are inherently linguistic, and therefore cannot readily be applied to animals. However, prototype theories have also been evaluated (principally in comparison with exemplar theories) by asking people to categorize artificial stimuli (Medin and Schaffer, 1978; Nosofsky, 1984; Posner and Keele, 1968) and there have been isolated attempts to test animals in the same way (Lea and Harrison, 1978; Pearce, 1989).

In Pearce's autoshaping experiment, pigeons learned a discrimination between two groups of histograms which differed in terms of the total height of their three bars. Slides of histograms with a total height of 9 units were followed by reward, while histograms with a total height of 15 units were not. It was assumed that with respect to the dimensions or features in terms of which the pigeons classified the stimuli, the central tendency (in this case, both mean and mode) of each category would be, respectively, three bars of three units each (3-3-3), and three bars of five units each (5-5-5). These prototypical instances were not presented during the training phase of the experiment, but were presented among a number of transfer stimuli. Pearce argued that if differential responding to the two kinds of histogram were mediated by comparison of each stimulus with a prototype, then the pigeons' discrimination performance would be maximal when they were presented with the prototypical stimuli. Specifically, the birds were expected to peck the 3-3-3 slide vigorously, and to respond little to the 5-5-5 slide. Pearce did not observe this pattern of results and concluded that a stimulus generalization account may be appropriate. However, even if Pearce had obtained the pattern of results he predicted on the basis of the prototype view, it is not clear that an account in terms of stimulus generalization could have been ruled out.

The predictions of the prototype and stimulus generalization views may be harder to distinguish than originally thought. It is normally assumed that if there are 'exemplar effects'—if responding to transfer stimuli is a function of their proximity to specific training stimuli—then the stimulus generalization view is confirmed. Conversely, the occurrence of 'prototype effects'—responding to transfer stimuli as a function of their proximity to the central tendency—is taken to support the prototype view. However, there are two considerations that muddy the waters. First, whether a stimulus generalization model generates exemplar or prototype effects

depends on the precise details of the account (Nosofsky, 1989; see Shepard, Hovland and Jenkins, 1961 for a related analysis). Second, simple associative networks may generate exemplar effects without storing exemplars, and prototype effects without computing prototypes (Shanks, 1990).

Thus, as a model of concepts to be used in comparative research, the prototype view has the weakness that it is not readily distinguishable from stimulus generalization accounts, which do not involve concepts at all. This problem, which is a deficiency with respect to criterion (3), may not be insuperable given a more precise formulation of the contrasted views, but there are independent reasons to doubt the utility of the prototype view. Specifically, the prototype view may violate criterion (1): it may fail to apply to the human categories DOG, TREE and PERSON. The prototype view assumes that instances of a category cluster together in some feature space; in terms of some set of features, most members of a category are closer to the central tendency of that category than to the central tendencies of other categories. However, as we hope to show in the remainder of this section, the literature on pattern recognition and computational vision strongly suggests that this assumption is false in the case of human categorization, and arguably also false for animal categorization.

The discussion so far has proceeded on the assumption that the features in terms of which the environment is perceived are known. In practice, of course, not only do we not know what the relevant features are, but we do not know what kind or level of features might be an appropriate basis for natural categorization in pigeons or any other animals. The relevant features might be the kind of things to which a single cortical cell responds (Hubel and Wiesel, 1962, 1968; Kuffler, 1953), or much more complex aspects of the image or environment. Comparative psychologists studying categorization seldom state explicity how 'coarse' they take features to be, but clearly a wide variety of possibilities has been entertained. For example, Morgan et al. (1976) wonder whether the features that pigeons use to discriminate 'A's from '2's concern the presence or absence of an apex, the number of vertical strokes, degree of curvature, and so on. On the other hand, D'Amato and Van Sant (1988) suggest that monkeys discriminate between slides depicting a person and slides not depicting a person using '...conjoint features of red patch/animal' (p.54). Clearly there is a considerable difference in the complexity of putative 'apex' and 'animal' features.

The attraction of relatively low-level features is that neurophysiological research is able to suggest appropriate primitives—for example, blobs, edges, moving bars and the like; and that there is a considerable computational literature that indicates how such features might be used in categorization. However, this literature also suggests that theories of categorization that are directly based on low-level featural representations are fundamentally inadequate. Commenting on featural views of categorization entertained in the 1950s and 1960s (Barlow, 1953; Kruskal, 1964), Marr (1982, pp. 340–1) says:

> The hope was that you looked at the image, detected features on it, and used the features you found to classify and hence recognize what you were looking at. The approach is based on an assumption which essentially says that useful classes of objects define convex or nearly convex regions in some multidimensional feature space where the dimensions correspond to the individual features measured. That is, the 'same' objects—members of a common class—have more similar features than objects that are not the same. It's just not true, unfortunately, because the visual world is so complex... Different lighting conditions produce radically different images, as do different vantage points. Even in the very restricted world of isolated, two-dimensional, hand-printed characters, it is difficult to decide what a feature should be. Think of a 5 turning into a 6—a corner disappears, a gap narrows. Almost no single feature is necessary for any numeral. The visual descriptions necessary to solve this problem have to be more complex and less directly related to what we naturally think of as their representation as a string of motor strokes.

In the comparative literature optimism regarding the prospect of a satisfactory featural analysis of category discrimination (e.g. Lubow, 1974) has given way to a recognition that the variability and complexity of natural stimuli makes a featural approach intractable (D'Amato and Van Sant, 1988; Fersen and Lea, 1990). However, as this passage from Marr indicates, the problem may be more profound than is typically recognized by comparative psychologists. Not only natural stimuli, but also stimuli from relatively simple artificial domains are too variable and too complex to permit featural analysis. Where comparative psychologists anticipate that featural analysis will be difficult, Marr argued that it will be impossible. He suggested that two key tenets of the featural view must be abandoned. First, the assumption that category instances cluster in feature space must be given up. This rules out a prototype view of animal concepts. Second, it must be recognized that a simple featural representation is inadequate for the categorization of natural objects, and that complex, structured descriptions, at a variety of levels, must be computed from the retinal image.

One may be tempted to assume that Marr's critique, while valid for low-level perceptual features, does not apply if we attribute the use of high-level features (such as 'animal') to the categorizing individual. This high-level approach has certainly been adopted by prototype theorists investigating human concepts (e.g. Glass, Holyoak and Santa, 1979; Smith and Medin, 1981). For example, the features underlying the human concept BIRD are usually thought to be concerned with properties such as being able to fly, having feathers, eating worms, building nests, laying eggs, and the like (Glass et al., 1979). However, the assumption that high-level features are in use has its own problems. First, the perception of candidate

features seems to be just as complex, and just as much in need of explanation, as the concepts they are supposed to underlie. Taking D'Amato and Van Sant's (1988) example, it is not clear that 'animal' detection is any simpler than 'person' detection (though 'red patch' detection presumably is much easier). If a concept such as ANIMAL can be freely assumed, without careful experimentation, then why is such investigation necessary to establish the target concept PERSON? Second, it is unlikely that clusters of even high-level primitives will correspond to natural categories. Consider the case of trees described in terms of volumetric primitives (various classes of cylinder, cone, etc., which correspond to trunks, twigs and the like). What distinguishes trees from, say, a pile of kindling, is not the presence or absence of particular features, but rather the way in which those features are arranged. As in the case of low-level analysis, a schema or, in Marr's terms, a structured description (over which a similarity measure, which is required for a prototype view, cannot readily be defined), rather than a simple feature list, is likely to be required. Although these points were originally made in the context of human perception and categorization, they seem equally applicable to animals.

In this section we have examined theories of concepts from cognitive psychology and found that they do not provide good leads for animal concepts research. The definitional and exemplar views cannot be freed from their links with natural language and therefore fail on criterion (2); they cannot be applied to nonlinguistic animals. The prototype view can be formulated in perceptual rather than linguistic terms but since prototype models are based on highly questionable assumptions regarding the relationship between natural categories and perceptual feature clusters, and are difficult to distinguish empirically from stimulus generalization accounts of categorization performance, the prototype view has serious deficiences with respect to criteria (1) and (3); it appears neither to apply to humans nor to permit experimental investigation or animal concepts.

Since none of the candidate internal structures (definitions, sets of exemplars and prototypes) meets criteria (1)–(3), animals cannot be said to have concepts on the grounds that they possess the same internal structures that underlie human concepts. Hence, it may be argued that it is simply obscurantist to ascribe concepts to animals, on the basis of *any* empirical data, since there is no clear understanding of what is being postulated. However, the possibility remains that what it is to have a particular concept, say TREE, is not a matter of possessing a particular internal structure.

The 'functional' approach to the study of animal concepts seeks to identify animal concepts behaviourally, by studying what animals can learn to discriminate (Keller and Schoenfeld, 1950; Lea, 1984). To pursue this line, the comparative psychologist needs leads concerning what is required for an internal structure (of whatever sort) to have a particular content; leads providing a common basis for the ascription of particular concepts to animals and humans alike. To assess the feasibility of this

strategy, we must turn from cognitive psychology to the philosophy of representation and also to suggestions from within comparative psychology itself.

3. *The Content of Concepts: Leads from Philosophy?*

3.1 *Representation and Language: Failing on Criterion (2)*

As with structure, many accounts of content are fundamentally bound up with natural language, and cannot be directly appropriated by the comparative psychologist to specify the content of animals' internal representations. For example, 'meaning-as-use' approaches to content are typically taken to apply *in principle* only to external, public languages. Equally, the 'reference borrowing' (Devitt and Sterelny, 1987) aspects of the so-called 'New Theories of Reference' (Kripke, 1972; Putnam, 1975), which concern the importance of causal chains between the current use of a word and its original usage, the social character of meaning in linguistic communities, and so on, necessarily apply only to language users. These mismatches between much standard philosophy of representation and a notion of representation appropriate for animal concepts research arise because natural languages are external to the agent, and putative concepts are internal to the agent. But even to the extent that philosophy applies to internal languages, there is still an apparent mismatch with animal concepts research, which focuses on the content of particular internal structures, not that of whole internal languages.

This difference of subject matter may or may not appear to pose serious problems depending on one's philosophical viewpoint. On the one hand, it may be argued that the meaning of particular representational structures may be ascertained independently of the system of representation (if any) in which such structures are embedded. Accordingly, an understanding of the meaning of whole languages is founded on an understanding of the meanings of their parts; and the meaning of these parts can be determined independently. From this perspective, it *is* appropriate to attempt to discover the concepts that animals have one by one, and, more specifically, it may be possible to determine that an animal has a particular concept without knowing much about the conceptual system in which that concept may play a part.

On the other hand, there are those who see meaning as fundamentally a property of languages, and who take the meaning of parts to be defined only within the framework of the whole language (e.g. Davidson, 1984; Quine, 1960; see Fodor and Lepore, 1992 for a critique). According to such a view, the meaning of a particular mental representation cannot be determined *in vacuo* but only as part of a system of representation. Hence, it will not be possible to discover the content of a particular concept without knowing about the animal's entire conceptual system. (One might

suggest somewhat less holistic positions, according to which meaning is a property of fragments of language, rather than entire languages, but this would not substantially change our argument.) To incorporate this into an empirical methodology for animal concepts research seems problematic in the extreme: the relatively modest goal of assessing whether animals have or can acquire particular concepts (PERSON, FISH, TREE) is replaced by the vast and intractable task of empirically uncovering the animal's entire conceptual scheme. Indeed, such a view appears to undermine the notion of a concept as a free-standing entity, independent of a representational system. For example, it is difficult to understand how we can talk of two agents with different conceptual schemes as nonetheless sharing a particular concept (e.g. Dummett, 1973; Fodor and Lepore, 1992). In consequence, philosophers who adopt this perspective tend to be sceptical about concepts, not just on empirical grounds, but as a theoretical construct, whether applied to people or animals (e.g. Quine, 1960). Such philosophers tend to argue that both ascriptions of linguistic content and ascriptions of concepts must necessarily be rough and ready. Hence concepts are not seen as a basis for scientific analysis, in studying either people or animals. Thus, in looking for leads from philosophy the comparative researcher is looking for theories of content that are not holistic, and that do not make ineliminable reference to external natural language.

There are a range of suggestions from the philosophical literature concerning how individual concepts can be fixed without ineliminable reference either to each other or to entire conceptual schemes. For example, Peacocke (e.g. 1989) develops an account based on an attempt to spell out the 'possession conditions' for a concept; Bennett (1976) aims to spell out what it is to 'register' a property, as a foundation for having a concept. Rather than attempting to survey a full range of options, we shall concentrate on a recent and popular line of attack, which meshes particularly well with the experimental methodology of animal concepts researchers. This approach, 'informational semantics' (Barwise and Perry, 1983; Dretske, 1981; Fodor, 1987; Israel and Perry, 1987), matches the approach of comparative psychologists by attempting to ground concepts in terms of perceptual discrimination, rather than linguistic abilities.

3.2 *Perceptual Accounts of Concepts: Failing on Criterion (1)*

3.2.1. The Correlational View

Let us take our starting point not from philosophy, but from comparative psychology itself. The simplest perceptual account of concepts is that possessing a concept consists in being able to discriminate instances from noninstances of that concept. It is this position that appears to underlie the use of the discrimination learning paradigm in animal concepts research. Lea (1984), accepting that a concept is an internal representation while remaining agnostic about the structure of that representation,

expresses this view succinctly. He suggests that an individual with a concept '...has some unique mental structure which is active when and only when an instance of that concept is present in the external, physical environment or when associated concepts are active in the mental environment' (p.270). According to this view, having the X concept is simply a matter of being able perceptually to discriminate X's from non-X's; and such discrimination abilities are just what paradigmatic animal concept experiments aim to test. This 'correlational' account of what it is to have a concept has a counterpart in philosophy as what Fodor (1987, 1990) calls the 'crude causal theory' of meaning. It is also closely related to Dretske's (1981) proposal that conceptual structures carry the information that is their content in 'digital' form.

For all its appeal to the comparative psychologist, this view, in its bare form at least, is inapplicable to human concepts, thus violating criterion (1). Furthermore, refinements made within informational semantics and designed to meet criterion (1) are fraught with problems. In particular, many recent proposals involve recourse to considerations beyond the scope of experimental investigation, and therefore, at best, fulfilment of criterion (1) is purchased at the expense of criterion (3). Let us consider these points in turn.

3.2.2 The Problem of Error

The problem with the correlational view as an account of human concepts is not hard to find. It is unable to account for a ubiquitous aspect of categorization: error. With respect to everyday categories such as fish, tree and person, human judgment is notoriously fallible. Strolling home on a moonless night, it is not uncommon to mistake pillarboxes, trees or dustbins for lurking strangers. Similarly, any person who is visible, but perhaps very still or in a dark corner, may be overlooked as a pattern in the shadows or even mistaken for a tree. In view of these false negative and false positive errors, there could not be a mental structure that is active when and only when a person is in the 'external, physical environment' (Lea, 1984, p. 270). Any mental structure that might be a candidate 'person detector', according to Lea's view, will in fact only be activated by people who are not, say, in camouflage gear or in bizzare fancy dress, who are nearby, in good light, at least reasonably close to foveal vision, and so on.

As Fodor (1987) points out, the programme of informational semantics in philosophy is concerned with attempting to patch up such problems with correlational accounts. A number of proposals have been made (e.g. Chater, 1989; Fodor, 1987, 1990; Papineau, 1987; Stampe, 1977) but none are widely considered to be satisfactory (see Godfrey-Smith, 1989 for a review). Rather than attempting to survey the range of possible responses, we will consider here just one suggestion from philosophy about how the problem of error can be met (Fodor, 1984). This suggestion has been chosen because it is one that may also naturally occur to psychologists.

It may seem from our 'moonless night' illustration above that the problem of error crops up only when stimulus exposure occurs under sub-optimal conditions, and that if some way of distinguishing optimal from sub-optimal conditions could be found, the problem of error could be avoided. This, however, is not the case. Consider as an example the confusion that commonly occurs at night between a star and the lights of a plane. It might be argued that this confusion does not necessarily imply that there can be no mental structure that is a PLANE concept, since planes are only confused in this way when they are viewed from a considerable distance and in the dark. In optimal conditions, it may be argued, the activation of the relevant mental structure would correlate perfectly with the presence of planes. Since this notion of optimality would, on this line of argument, underpin the ascription of the PLANE concept, rather than the PLANE OR STAR ON A DARK NIGHT concept, it is essential that it be possible to specify what are optimal conditions without presupposing the identity of the concept. However, as Fodor (1990) forcefully points out, the notion of optimal viewing conditions is inescapably relative to the concept concerned. Applying this point to our example, while daytime is optimal, and nighttime sub-optimal, for detecting planes, nighttime is optimal for detecting planes or stars (since you can see instances of both at night), and daytime is sub-optimal (since only some instances—planes—are visible). Thus, according to the correlational position, concepts are defined in such a way that there can be no such thing as 'getting it wrong'. Since the content of the concept is *whatever* the activity of the mental structure correlates with, misclassification is impossible. The optimality response to this problem is just one of a number that try to allow for error by attempting to distinguish two kinds of situations: one kind in which performance determines what content the representation has, and hence what concept it corresponds to; and another kind that is 'nonoptimal' (Fodor, 1984; Stampe, 1977), or outside the 'learning period' (Dretske, 1981). In all such cases it is very difficult to see how to define the distinction between the two classes in a noncircular way.

It may be, of course, that a philosophical solution to these difficulties can be found—indeed the project of informational semantics is wedded to the hope that it can. But could a more sophisticated and satisfactory theory of content be tied to some alternative programme of empricial investigation? Certainly, Fodor's (1987, 1990) most recent and ingenious suggestion, relying on what he calls the 'asymmetrical dependence' of counterfactuals underwriting recognition in 'errorless' versus 'error-tolerant' situations, and Dretske's (1988) attempt relying on teleological considerations, appear to put content ascription entirely outside the purview of experimental paradigms (although not necessarily outside the scope of empirical investigation). Another response to the problem of error involves the assertion that certain properties 'carve nature at the joints' and others do not; the former are more 'eligible' for reference (this suggestion has been advocated by Lewis (1984) in a broader context). The problem of

error would be solved if it could be shown that the 'intended' concept is eligible, and that the unwanted alternatives that we have discussed are not eligible. But it seems that we have done no more than relabel the original problem, so that the problem of error is recast as the problem of specifying the difference between eligible and non-eligible referents. In the absence of an account of eligibility, this line does not appear to offer a useful lead to comparative psychologists plagued by the problem of error.

It may seem that these philosophical problems stem from an unnecessarily precise interpretation of the correlational view. Perhaps the correlation may simply be imperfect. An animal may have the concept X if it has some mental structure which is active 'more often when' X is present in the external environment than 'when it is not' (Lea, personal communication). If this looser correlational view were acceptable, there could be little doubt on the basis of current evidence that animals can learn concepts of PERSON, FISH, TREE and so on, simply in virtue of their showing above-chance transfer performance. However, this view cannot be accepted since it entails extreme profligacy in content ascription. Consider the range of properties P that correlate imperfectly with the performance of an animal discriminating slides containing a person from slides not containing a person. These properties might include pinkness, puffiness, height, hairiness and so on. If in each case the property correlates positively with discrimination behaviour, then the animal must be ascribed the corresponding concept P. If the property correlates negatively, then the animal must be ascribed the concept NOT-P. Only if discrimination performance has precisely zero correlation with a property P will neither P nor its negation be ascribed. One might tighten the looser version of the correlational view by introducing a threshold value for the correlation. However, any cutoff point short of 1 (the precise correlational view) would reduce but not eliminate this profligacy.

Rather than attempting to solve the problem of error, perhaps we should simply embrace its premise, and accept that discrimination performance is, in a certain sense, error free. Thus we might accept that the content of a mental structure is, by definition, the precise external correlate of its activation; that the content of a concept can simply be identified with its perceptual base. Suppose, for example, that it has been discovered that in a 'person' discrimination task, monkeys are actually responding to some complex structural property of the image. They might respond positively when presented with any stimulus containing one set of contour relationships, hues and so on, whose instances usually look, to the human eye, like a person. Indeed, we may label, after Fodor (1990), the relevant complex constellation of properties of the stimulus 'that-persony-look'.

If we accept the view that discrimination is error free, then, in this case, the animal cannot be ascribed the concept PERSON but only the conjunctive and disjunctive concept the shorthand for which is THAT-PERSONY-LOOK. This state of affairs may be judged unsatisfactory or disappointing

in view of the early expectation that animals might be able to acquire concepts such as TREE, FISH and PERSON (Herrnstein and Loveland, 1964; Herrnstein et al., 1976; Herrnstein and de Villiers, 1980). Furthermore, if the same account of concepts is applied to people as for animals (in line with criterion (2), then we must accept that people too cannot possess TREE, FISH or PERSON, thus violating criterion (1). After all, humans can also be fooled. It is only because one has been told about Disneyland, and told that Abraham Lincoln is dead, that one does not take his automaton to be a person. It is only because one has been told that Mickey Mouse is a figment of Walt Disney's imagination that one takes the leader of the Disneyland parade to be a person. An individual without this specific sub-cultural information may well be misled, as are monkeys (D'Amato and Van Sant, 1988), by cases similar to these. Using the same argument that we used for the monkey, it follows that a human without this trivial and specific information is responding to, or has the concept, THAT-PERSONY-LOOK rather than PERSON.

In this section we have seen that far from providing useful leads for comparative researchers, a consideration of philosophical theories of meaning or representation indicates that their current strategy for defining the content of animal concepts is deeply problematic. We have reviewed the weaknesses of the correlational approach using Lea's (1984) account as an example, but the strategy is used, usually implicitly, throughout the literature on animal concepts and beyond. Behaviour analysts apply it in defining concepts as 'equivalence classes' (e.g. Bhatt and Wasserman, 1989) or as 'uncommon generalization classes' (Stemmer, 1980). A similar position also appears to be implicit in Gibson's (e.g. 1979) notion that everyday categories corresponding to affordances may be identified with 'higher-order invariants' of the perceptual array, to which the perceiver may or may not be appropriately 'attuned'. Whatever the general merits or otherwise of the behaviour analytic and Gibsonian projects, as accounts of conceptual content they are no more likely to succeed than other correlational theories.

4. Do Alternative Methodologies Fare Better?

4.1 Introduction

We conclude, on the basis of the foregoing survey, that currently there is no adequate theoretical basis for the claim that nonlinguistic animals have concepts in general, or any one concept in particular; that, as things stand, these claims do not make sense. It may be argued, however, that our focus has been too narrow, and that alternative approaches to the empirical investigation of animal concepts can overcome the problems we have raised. In this section we shall consider two such alternatives: The first (4.3) raises the possibility that the traditional methods of comparative psychology can be used to ascribe relational, if not categorical, concepts

to animals. The second (4.4) seeks to ascribe categorical concepts to animals through observing their behaviour under free-living conditions. Both approaches have been pursued, in part, using animals that communicate in an apparently linguistic manner; animals that have been trained (successfully or otherwise) to use language or that seem spontaneously to use the rudiments of a language. In view of this, we will begin the section by discussing animal language.

4.2 *Animal Language*

We started this paper with the reflection that the apparently close connection between human concepts and human language raises problems for accounts of animal concepts. In the subsequent discussion, we have argued that consideration of human cognitive psychology and philosophy does not exorcise, but rather reinforces, this difficulty. The crucial difference between animal and human to which we have appealed is the difference between language users and non-language users, and this suggests that our arguments will not apply to linguistic animals if any. A bland and unfulfilling conclusion looms: that animals can be coherently ascribed concepts if they have language, and that whether or not animals have or can acquire language is somebody else's problem. However, there are two caveats which makes this conclusion considerably less bland.

First, very few animals, or animal species, are currently claimed by ethologists or comparative psychologists to possess the linguistic abilities that would be necessary to qualify them for concept ascription. We assume that an animal would be so qualified—that it would be coherent to *ask* which concepts the animal has—if its linguistic abilities were such as to bring it within the scope of analysis of language-based theories of concepts. These theories assume that concepts (TABLE, PERSON, PERSON WITH RED HAIR) correspond to properties (the property of being a table, a person, a person with red hair), and that in natural language properties are represented by predicate expressions ('is a table', 'is a person', 'is a person with red hair'). This implies that animals would come within the scope of these theories only if they use predicate expressions—that is, if they use expressions which have a particular syntactic/semantic role in the animals language. In particular, this presupposes that the animal's communication system has a syntax and a semantics. Animal's natural communication systems are rarely claimed to have both of these attributes, and even if they were present in the behaviour of all 'language-trained' animals, we would still only have a handful of creatures—several apes, a few marine mammals and a parrot (e.g. Herman, Richards and Wolz, 1984; Pepperberg, 1988; Savage-Rumbaugh, McDonald, Sevcik, Hopkins and Rubert, 1986) to which concepts could, in principle, be ascribed.

Second, identifying *which* concepts a language using animal might have is much more difficult than in the human case. If the language is taught, and it is a human natural language or has been devised by the experimenter, there is a natural temptation to assume that the meaning of formulae of the language used by the animal is the same as the meaning of those formulae for the experimenter. However, the fact that, for example, Pepperberg's (1988) parrot Alex produces and reacts differentially to a variety of sounds corresponding to English words, does *not* imply anything about the meaning of Alex's vocalizations.

Nonetheless, within particular restricted domains at least it might appear that the content of animal utterances can be pinned down. For example, consider the case of 'number' concepts—ONE, TWO, THREE and so on. At first blush it might seem that these are straightforwardly evidenced by the ability to count (Seibt, 1982; Thorpe, 1956), and since several animals, including Alex, have been taught to count, it might seem churlish not to accept that Alex's 'two' vocalization really means two, and that the concept TWO can be justly ascribed to him. However, from within the comparative literature (Gelman and Gallistel, 1978) there has been the demand that number concepts be applied only to agents that can perform arithmetical operations: addition, multiplication, and so on. Without the demonstration of such additional competence, it is argued, Alex's number utterances, for example, simply do not capture the richness of human number concepts. If this line is followed, it is difficult to know what criteria would be satisfactory for translating Alex's number utterances. There appears to be no reason to stop at demanding simple arithmetical operations; what about fractions, percentages, ability to deal with the real numbers, modulo arithmetic, and the like? Thus, even in the prescribed domain of numbers, it is extremely difficult to fix criteria for interpreting putatively linguistic utterances. Attempting to interpret animal utterances outside such clear cut areas will surely be more difficult still.

It is tempting to view 'language-training' as a measuring instrument that reveals concepts the trained animal, and its conspecifics, had all along (Griffin, 1976; Pepperberg, 1987a). However, if our arguments are valid, this view is misleading because it presupposes, wrongly, that we have some sense of what it would be for a nonlinguistic animal to have a concept. Without a theory to provide this sense, the claim that an animal had concepts before it was language trained is no more substantive than the claim that it had the potential to be language trained before it was language trained. A contrasting view, that language training permits an animal to acquire 'abstract' concepts (Premack, 1983), is similarly problematic. If we do not know what it would be for a nonlinguistic animal to have concepts, then it should be acknowledged that while language training may bring an animal within the scope of theories that allow coherent concept ascription, it may or may not do so by virtue of causing some significant change in the kind of internal structures that the animal can form. We simply do not have adequate means, theoretical or empirical, for addressing this question. In our view, it is just these considerations that have prevented resolution of the debate over whether concepts or language are primary in children (Cromer, 1990; Mandler, 1988, 1992; Weiskrantz, 1988).

ical concept formation, comparative psy-
thods to investigate the possibility that
oncepts, such as SAME and DIFFERENT,
e concentrated so far on work aimed at
ause it is in this context that comparative
pts' terminology most consistently and
n category discrimination are commonly
reveal the extent to which animals can
same structure and content as those of
velty—familiarity discrimination exper-
to test theories of recognition memory
search on animals (e.g. MacPhail, 1980;
go and Wright, 1984; Todd and Mackin-
ever, that in ignoring studies of relational
ignored some of the most important and
have concepts. Perhaps those who study
ess concerned about content ascription
n solved.

ibution of the concept SAME is 'transfer
hing task involves the presentation of at
ue or 'sample' (e.g. a red light) and two
d a green light); the animal is rewarded
g or touching, the choice stimulus that
natching is said to have occurred when
ion of a new matching problem (e.g.
measured in terms of first trial accuracy
iterion performance (D'Amato, Salmon
Gordon, 1974; Mackintosh, Wilson and
1988; Zentall and Hogan, 1976, 1978).
s is difficult, however. Although transfer
matching problems, there will be many
bserved. To take an extreme example,
rs of varying numbers of dots, and that
whether the number of dots in a cluster
ot prime, square versus not square, or
. Alternatively, the stimuli might consist
right-way-up versus upside-down let-
s, it seems that to the extent that the
y an animal at all, responses must be
ply do not have the relevant knowledge
f the letters of the alphabet) with which
ifferent property. In consequence, there
raining on the original problem. Thus,
problems (which should bear equally

strongly on whether SAME is present) on which conflicting results will be obtained.

These considerations, it may be argued, merely underline the fact that an animal cannot have the same concept SAME, or the same range of SAME concepts, as a human, but leave intact the claim that animals can be ascribed some sort of SAME concept. Specifically, transfer of matching might be regarded as evidence of the concept PHYSICALLY SAME. However, what if the animal's response is mediated, not by PHYSICALLY SAME, but by BEING PERCEPTUALLY INDISTINGUISHABLE; HAVING A SIMILAR VISUAL STRUCTURE; BEING THE SAME IN RESPECT OF COLOUR AND FORM; or BEING PHYSICALLY ALIKE? If behaviour were dependent on any of these more specific properties, transfer would be observed in our original example—between horizontal versus vertical stripes and red versus green light—and the concept PHYSICALLY SAME would be mistakenly ascribed. The problem of trying to ascertain empirically to which property the animal is responding does not seem noticeably easier than in the case of categorical concepts. In particular, it seems equally difficult to allow for error (to allow that an animal may mistakenly take two stimuli to be, for example, physically alike); and equally difficult to avoid the primacy of appearances. THAT-SAMEY-LOOK appears to be inevitably a better candidate concept than SAME or PHYSICALLY SAME on the basis of transfer experiments, for just the same reason that THAT-PERSONY-LOOKS appears to be inevitably a better candidate concept than PERSON. After all, if two stimuli have the same appearance, then whether or not they actually have the same constitution (whether, for example, one is a real apple and one a wax apple) is irrelevant to the animal's performance. In short, the evidence for relational concepts from transfer of matching experiments appears to suffer all the difficulties which undermine discrimination learning experiments as evidence of categorical concept formation.

Perhaps the performance of certain language-trained animals is immune to these objections. Both the Premacks' chimpanzee, Sarah (Premack and Premack, 1983), and Pepperberg's parrot, Alex (Pepperberg, 1987b, 1991) can solve simultaneous same/different discrimination problems. Among Sarah's lexical items there was a piece of plastic symbolizing (at least as far as the experimenters were concerned) 'same', and another symbolizing 'different'. On being presented with two objects that had not previously been used in training (e.g. two keys, or a glass and a peg), Sarah would reliably select the first of these symbols to place between the objects if they were duplicates, and the second if they were not. Alex's same/different discrimination performance is even more striking. On each trial he is presented with two objects that differ in colour, shape or material (e.g. a blue, wooden triangle and a red, wooden triangle), and the trainer says, in English, either 'What same?' of 'What different?', Alex responds correctly, by uttering one of the sounds 'colour', 'shape' or 'material', both when the objects have been used in training and when they are relatively unfamiliar to him.

The results of such experiments are indeed impressive, but it is not at all clear how they serve to reduce the indeterminacy found in the transfer of matching experiments. Since we cannot take the English translation of Alex's utterances at face value, then these apparently linguistic responses must be treated as uninterpreted responses for the purposes of experimental analysis. Let us, then, reconsider Alex's behaviour with the 'words' of trainer and parrot replaced by arbitrary symbols: Alex is shown two objects that differ in colour, shape or material, and the trainer says either 'A' or 'B'. In response to 'A', Alex responds by uttering 'X' when the objects differ in colour, 'Y' when they differ in shape and 'Z' when they differ in material. In response to 'B', Alex responds by uttering 'X' when the objects are the same colour, 'Y' when they are the same shape and 'Z' when they are made of the same material. Although Alex's differential responding to 'A' and 'B' may be taken to indicate that he has the concepts SAME and DIFFERENT, it is also consistent with the ascription of the concept pairs TWO and ONE, DOUBLE and SINGLE, DUPLICATED and NOT DUPLICATED. Similarly, his pattern of, for example 'Z' responses, could be regarded as evidence that he has the concept MATERIAL, but by the same token it may warrant ascription of concepts such as TEXTURE, RIGIDITY, SMOOTHNESS, SCINTILLATION or, since Alex is often rewarded with the opportunity to sink his beak into one of the stimuli, WARMTH, SMELL, FLAVOUR or SUSCEPTIBILITY TO SWIFT DESTRUCTION. Yes, some of these possibilities could be distinguished empirically with relative ease, but there is no more reason in this case than in the case of categorical concepts to suppose that success in the first few battles would predict ultimate victory. As many investigators of 'relational concepts' would happily acknowledge, research of this kind can and does provide information about discrimination, but not about concepts. That is, it enables humans to define an animal's competence in a way that allows us to formulate and test theories relating to its perceptual and learning processes. However, like research on categorical concepts, it does not indicate how, if at all, the animal defines or conceives of the stimuli to which it is exposed.

4.4 *Back to Nature*

So far reference has been made only to studies of animal concepts conducted within the tradition of comparative psychology, a tradition which descends from the work of Pavlov (1927) and Thorndike (1898), and assigns considerable value to precise control and systematic manipulation of experimental variables in the investigation of animal learning and cognition. The control is typically achieved, as the cited studies illustrate, by examining the behaviour of socially isolated members of a limited range of species (primarily rats, pigeons and monkeys), in a standard apparatus (the operant chamber or Skinner box), and in relation to objects or events that are not designed to resemble those that the animals might encounter in their

natural environments. It has often been alleged that in conducting this kind of research comparative psychologists exchange rigour of method for relevance of result (Dennett, 1983, 1986). Ethologists, cognitive ethologists and behavioural biologists have claimed that as a result of detaching animals from their natural environments, and the study of animal concepts from evolution, comparative psychologists have addressed arcane pseudo-problems, and been prevented from recognizing obvious answers to, and legitimate routes to the resolution of, other, genuine, questions (e.g. Campbell and Hodos, 1991; Hodos and Campbell, 1969; Johnston, 1981; Lockard, 1971; Mason and Lott, 1976; Plotkin, 1979). Is the comparative psychological literature on animal concepts an example of this kind of muddle? Can the problems with animal concepts be solved by re-locating the animals, in theory and in practice, in their natural environments?

It does, to be sure, sometimes appear that the methods of comparative psychology are used to demonstrate, through mighty labour, capacities that are apparent even to the relatively casual observer of free-living animals. For example, anyone who has watched a pigeon alternately pecking at some seeds, and then dipping its beak into a puddle (and not pecking the ground in between) should not be surpised to hear that pigeons can discriminate simultaneously between several stimulus categories; one of the major findings of a recent study in the comparative psychological tradition (Bhatt, Wasserman, Reynolds and Knauss, 1988). However, in defending the tradition, a comparative psychologist might well cite another result of the same experiment; namely, that pigeons learn to categorize slides of cats, which are presumably significant objects in the pigeon's natural environment, and of chairs, an 'artifical' category, with equal ease. It may be argued that this is evidence that research using artificial or arbitrary stimuli is likely to yield the same basic conclusions about animal concepts, and to run into the same problems, as would studies involving more naturalistic stimuli. Another comparative psychologist might respond by pointing to similar experiments suggesting that pigeons succeed in categorizing slides of trees, pigeons and persons, where they have failed to categorize slides of wheeled vehicles and bottles (Herrnstein, 1985). However, since an ethologist would almost certainly eschew this kind of debate, questioning the validity of *all* experiments involving slide stimuli, captive animals and/or domesticated species, we will also step outside the comparative psychologist's *Umwelt* (Uexkull, 1934) and consider the question at issue more directly.

On the basis of an elegant and widely cited series of naturalistic studies, Seyfarth and Cheney (e.g. 1980, 1990) have reported that free-living vervet monkeys give at least four acoustically distinct alarm calls, known as the LEOPARD, EAGLE, SNAKE and BABOON calls. The first is sometimes labelled the TERRESTRIAL PREDATOR call, or the MAMMALIAN CARNIVORE call, and contrasted with the second, AVIAN call or RAPTOR call, but, it could be argued, this variation does not reflect any fundamental problem in establishing the reference of the calls or in ascribing concepts

with particular content to vervets. There may be some minor uncertainty resulting, for example, from the fact that vervets have been observed giving the first call in the presence of lions, hyaenas, cheetahs, and jackals, as well as leopards, but, so the argument goes, this uncertainty could be reduced, and other more bizarre hypotheses about the content of the relevant concepts could be eliminated, by appealing to the *function* of the calls. Indeed, Seyfarth and Cheney have already used this strategy in choosing to identify the first call as the LEOPARD call. Of all the species that elicit the calls, the leopard is the only one that is known to prey on vervets in areas like Amboseli National Park, where their vervets live. This fact leads them to the plausible hypothesis that the function of the call is to defend against predation by leopards. However, this kind of functional argument is not persuasive. What is to prevent us from suggesting that Seyfarth and Cheney have established no more than the possibility that the call functions as a defence against AMBOSELI LEOPARDS, AMBOSELI LEOPARDS IN THE 1980s or AMBOSELI LEOPARDS IN THE 1980s THAT WOULD FIND VERVETS TASTY?

It is widely assumed that the answer to this question is 'evolution'; that animals can be ascribed concepts, at least in the case of 'natural' behaviour, through an appeal not just to the function of the behaviour, but to its adaptive, or 'Proper', function (Millikan, 1984, 1986, 1990). If there is good reason to believe that the vervets' alarm call is an adaptation for defence against leopards, then, so the argument goes, the buck has stopped and the vervets can be satisfactorily ascribed the concept LEOPARD. But evolutionary theory itself has problems with adaptation, of just the same kind as psychology has with concepts (Dennett, 1983): the question of whether knowledge should be understood in terms of concepts is mirrored in the question of whether evolution should be understood in terms of adaptations (Gould and Lewontin, 1984; Saunders, 1988). The question of whether concepts can be identified perceptually and/or with reference to components of language, has parallels in the debate about the coherence of defining adaptations in relation to environmental 'problems' and phenotypic 'solutions' (Gray, 1988; Lewontin, 1978, 1983). And, most important in the present context, while we are asking whether the content of specific concepts can be fixed with reference to the function of behaviour, evolutionists are denying that the current function of an attribute is a reliable indicator of its adaptive significance; of what, if anything, it is an adaptation *to* (Gould and Vrba, 1982; Hailman, 1988; Kitcher, 1985).

The critical problems arise out of the necessity to find out about history. A phenotypic attribute is an adaptation with respect to a particular function only if it was the fulfilment of *that* function which resulted in the retention of the attribute through natural selection. Therefore, in most cases, including presumably that of the vervets' 'leopard' alarm call, hypotheses concerning adaptive function cannot be tested at all, let alone tested rigorously, because the necessary information about past conditions and events is lost in the mists of time. It *is* possible to make empirical headway in

defining the *contemporary* functions of an attribute, and there is a temptation to identify adaptive function with one or several of these, but such identifications are quite spurious. An alarm call that currently functions as a defense against leopard predation could, for example, have been favoured by natural selection because it protected users from predation by some now-extinct species, or because, while its own effects were neutral or even deleterious with respect to mean fitness, it came in the same 'genetic package' as an attribute that controlled parasite infestation. The range of possibilities is *not* limitless, and this is important for those philosophers who, following Millikan, wish to secure a realist interpretation of reference by appeal to adaptive or Proper functions. For example, it may be possible to discount as candidate adaptive functions legitimate characterizations of an attribute's contemporary function that refer only to animals' sense experience, or to imaginary objects and contingencies that could not possibly have been part of the ancestors' environment (Millikan, 1990). Moves of this kind may establish that content could be ascertained in principle, but they provide no hint as to how the many remaining alternatives can be distinguished in practice. Like the attempts within informational semantics to resolve the problem of error, Millikan's 'teleofunctional' approach to content ascription fails on criterion (3); it would not support an empirical programme of investigation into animal concepts.

In the light of these considerations we doubt that evolutionary theory currently offers a reliable solution for the problems of ascribing conceptual content to animals. If so, then one might well view comparative psychological and ethological studies as having equal potential (limited or otherwise) to determine the content of animal concepts. Surprisingly, however, the problems encountered by evolutionists may be interpreted as support for the more 'unnatural' methods of the comparative psychologist. If the adaptive function of an attribute depends on its history of natural selection, then, an evolutionary epistemologist might argue, the content of a concept depends on its history of selection through learning (see e.g. Campbell, 1974; Plotkin, 1982, and Skinner, 1981 for related views), and it is in the laboratory, rather than in the field, that we have the best opportunity to record the details of that history.

5. Conclusions: Content and Discontent

We have assumed that studies of discrimination in animals cannot be understood to reveal anything about animal concepts unless there is a sense of 'concept' that (1) applies to humans, (2) applies to nonlinguistic animals, and (3) allows experimental investigation of concepts in animals; and we have hunted such a sense, without success, among theories of both the structure and content of concepts. Thus we have assumed that cognitive terms are useful in comparative psychology only to the extent

that they can be used in the same way in discussion of humans and animals, and we have argued that 'concept' is not such a term.

Before drawing final conclusions let us consider the possibility that our analysis has been fundamentally misguided; that we have somehow missed the point of animal concepts research. According to one authoritative reviewer, our analysis is invalid because it mistakes the purpose of animal concepts research, which is to answer two principal questions: (a) Can animals discriminate stimulus categories that are based on human concepts? and (b) If so, do they use concepts that are in any way like human concepts to do so? The simplest reading of (a) must be incorrect because it would have been satisfactorily answered, and in the affirmative, by the first experiment indicating that animals show transfer with 'natural' stimuli. Under a different and stricter interpretation, (a) asks whether animals are better able to discriminate categories based on human concepts than pseudocategories—arbitrary groupings of the same stimuli. Recent experiments have shown that this is indeed the case (e.g. Wasserman, Kiedinger and Bhatt, 1988), but the observation of such pseudocategory effects could hardly have been the aim of decades of animal concepts research since they were predictable from the existence of transfer effects. After all, to the extent that an animal is able to generalize from instances that it has encountered to instances that it has not, it should equally well be able to generalize, during training, between instances that it has so far learned and instances that it has yet to learn. That is, when the training set is a category rather than a pseudocategory (where there will be less transfer to new items) there will be transfer within the training set itself, leading to faster learning. The only remaining interpretation of (a) is as a question about what categorization results from discrimination training. But this is just the question which we have assumed to be central: What is the content of animal concepts? Consideration of question (b), which asks whether animal and human concepts are the same, supports this interpretation. If they could establish the content of animal concepts, researchers would like to find out whether the internal structures or mechanisms underlying those concepts are the same as those in humans. This is just the question at issue in our discussion of theories of conceptual structure.

It may seem that our conclusion that concepts can be ascribed only to language-users implies radical, and perhaps implausible cognitive discontinuities both in evolutionary and developmental terms. Certainly the thrust of much comparative research, across the behaviour analysis and animal cognition divide, has attempted to find continuities between the psychological processes of nonlinguistic animals and humans. Equally, developmental psychology is concerned to view the cognitive psychology of prelinguistic infants in the same theoretical terms as the cognitive psychology of linguistically competent adults. Hence, from both these perspectives, the postulation of such a discontinuity would go against the grain. However, this conclusion does not follow from our arguments, and explaining why provides a useful vehicle to draw together the main strands of the paper.

First, we are not primarily arguing that there is a cognitive discontinuity between the internal structures of linguistic and nonlinguistic agents. The definitional view does not apply to animals, not because they have impoverished internal structures, but because what counts as a definition is relative to an external natural language. If the exemplar view is correct, then it may be that animals and humans both make categorization judgments based on stored sets of exemplars; but only in the human case can exemplars be given category labels rather than merely associated with each other. Finally, in this summary of Section 2, the fact that there does not appear to be any featural basis according to which natural categories cluster, rules out the prototype view for both humans and animals.

Equally, we are not postulating a drastic semantic discontinuity between the internal representations of nonlinguistic and linguistic agents. Rather we note, in Section 3, that most accounts of meaning are applicable only to external languages, rather than internal representational primitives; and we argue that those accounts of meaning which can be applied to internal representations, without assumptions about their structure—accounts which attempt to individuate concepts perceptually rather than linguistically—are subject to significant prima facie difficulties. It is not clear how such difficulties can be addressed, but the suggestions abroad appear to take the determination of content to depend on considerations which are beyond the scope of experimental investigation. Finally, research on animal communication, relational concepts and cognitive ethology (Section 4) does not appear to resolve, but rather also to be beset by, the problems of applying concept terminology to animals.

We began this paper by noting that human concepts appear to be so directly tied to natural language that it is not clear how to make sense of concepts in nonlinguistic agents. We then argued that, despite tracing possible leads from cognitive psychology and philosophy, and examining a range of empirical methodologies, no clear sense has been provided for the claim that nonlinguistic animals have concepts. If concepts cannot be understood independently of natural language, then it makes no more sense to say that a nonlinguistic animal has concepts than to say that it has verb phrases or lexical entries or that it parses bottom-up rather than top-down. The discontinuity between humans and nonlinguistic animals is not that humans have concepts and that nonlinguistic animals do not. Rather, we simply do not know how to turn the claim that nonlinguistic animals have concepts into an empirically substantive question. If concepts cannot be freed from a linguistic basis, then the conceptual discontinuity, in both evolution and development, is simply a trivial corollary of an underlying linguistic discontinuity. It is *this* discontinuity that must be explained by any theory of human and animal cognition.

If specifying the content of animals' representations is as difficult as we have argued, how is it that considerable advances have been made in understanding cognitive processes in animals? In our view, the pattern of such success indicates, not that we can and have identified the content

of animal representations, but that progress can be achieved in so far as commitments concerning representational content are minimized. The way in which such commitments are routinely minimized may be exemplified with reference to a standard conditioned suppression procedure. Much of contemporary learning theory has been developed using this procedure, in which an animal in an operant chamber is exposed to a contingency between two stimuli such as a tone and an electric shock. As a consequence of this exposure, presentation of the tone results in a reduction or suppression of the animal's ongoing activity, and the animal is said to have formed an association between a representation of the stimulus and a representation of the reinforcer (Dickinson, 1980). Such an account does not characterize the content of the animal's mental state in the presence of either the tone or shock. It does not specify whether the content of the representation is 'tone present', 'high-pitched tone present', 'high-pitched tone inside chamber', or 'it's that bloody noise again!'. *Mutatis mutandis* for the shock. Rather, the conventional terminology assumes, minimally, and with clear empirical justification, that whatever the content of the rat's mental representations, they are different when the tone is and is not sounded, and when the shock is and is not applied. It is symptomatic of this high level of generality that the long-standing debate in comparative psychology regarding 'what is learned' has concerned, not task-specific content, but the question of whether animals acquire a stimulus-response or stimulus-reinforcer association.

The same point is exemplified at the other end of a methodological spectrum in some recent research in cognitive ethology. Some investigations of animal intentionality have quite deliberately rejected the minimalist approach and have consequently encountered serious problems. For example, studies of deception in primates (Whiten and Byrne, 1988) involve the ascription of specific beliefs—such as 'that animal does not believe that I am here' or 'that animal believes falsely that I want to groom her'—the contents of which are radically underdetermined (Danto, 1988; Heyes, 1987, 1988, 1993; Humphrey, 1988; Mitchell, 1988; Thomas, 1988). Other studies of animal intentionality, with more modest aims (Dickinson, 1989; Heyes and Dickinson, 1990), both exemplify the problem and suggest that limited progress may nonetheless be possible. These experiments were originally based on the plausible assumption that thirsty rats would have a desire that could be satisfied by drinking sucrose solution. Thus, the object of this desire could be very general—perhaps 'fluid', 'thirst-quenching fluid' or 'watery stuff'. The discovery that thirsty rats did not preferentially make a response that had, in the past, earned them sucrose solution rather than dry food pellets, cast doubt on this assumption about the content of the rats' desires. This finding might have indicated either that the rats did not have a desire at all, that they had a more specific desire, perhaps for water, or that they did desire, say, thirst-quenching fluid, but did not believe that sucrose solution is a thirst-quenching fluid. Dickinson selected the latter explanation, having shown that thirsty rats

that had past experience of drinking sucrose solution when thirsty did favour the sucrose producing response. Thus, the claim that the animals' action was intentional was sustained despite the fact that the initial intentional explanation was rejected, and that the precise content of the rats' beliefs and desires was never identified.

Examples such as these point to the general moral that advances can be made in the study of animal cognition by making pragmatic, flexible, and, as far as possible, minimal assumptions about the content of animals' representational states. However, the efficacy of this rough and ready strategy should not mislead us into supposing that, with just a little more effort, the content of representational states may be specified precisely. This is exactly what animal concepts research has attempted and, we have argued, failed to achieve. Unsatisfactory as it may seem, content ascription is likely to be rough and ready for the foreseeable future, and should be recognized as such.

Department of Psychology
University of Edinburgh
7 George Square
Edinburgh EH8 9JZ
UK

Department of Psychology
University College London
Gower Street
London WC1E 6BT
UK

References

Anglin, J.M. 1970: *The Growth of Word Meaning*. Cambridge, MA.: MIT Press.
Austin, J.L. 1962: *Sense and Sensibilia*. Oxford University Press.
Ayer, A.J. 1956: *The Problem of Knowledge*. Harmondsworth: Penguin.
Barlow, H. 1953: Summation and Inhibition in the Frog's Retina. *Journal of Physiology*, 119, 69–88.
Barwise, J. and Perry, J. 1983: *Situations and Attitudes*. Cambridge, MA.: MIT Press.
Bennett, J. 1976: *Linguistic Behaviour*. Cambridge University Press.
Bhatt, R.S., Wasserman, E.A., Reynolds, W.F. and Knauss, K.S. 1988: Conceptual Behavior in Pigeons: Categorization of Both Familiar and Novel Examples from Four Classes of Natural and Artificial Stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 219–34.
Bhatt, R.S. and Wasserman, E.A. 1989: Secondary Generalization and Categorization in Pigeons. *Journal of the Experimental Analysis of Behavior*, 52, 213–24.

Bingham, H.C. 1914: A Definition of Form. *Journal of Animal Behaviour*, 4, 136–41.
Campbell, D.T. 1974: Evolutionary Epistemology. In P. A. Schlipp (ed.) *The Philosophy of Karl Popper*. LaSalle, Il.: Open Court.
Campbell, C.B.G. and Hodos, W. 1991: The Scala Naturae Revisited: Evolutionary Scales and Anagenesis in Comparative Psychology. *Journal of Comparative Psychology*, 105, 211–21.
Carpenter, G.A. and Grossberg, S. 1988: The ART of Adaptive Pattern Recognition by a Self-Organizing Network. *Computer*, 21, 77–90.
Cerella, J. 1982: Mechanisms of Concept Formation in the Pigeon. In D. J. Ingle, M. A. Goodale and R. J. W. Mansfield (eds), *Analysis of Visual Behaviour*. Cambridge, MA.: MIT Press.
Cerella, J. 1986: Pigeons and Perceptrons. *Pattern Recognition*, 19, 431–8.
Chater, N. 1989: Information and Information Processing. Unpublished doctoral dissertation, University of Edinburgh.
Cromer, R. 1990: *Language and Thought in Normal and Handicapped Children*. Cambridge University Press.
D'Amato, M.R., Salmon, D.P. and Columbo, M. 1985: Extent and Limits of the Matching Concept in Monkeys. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 35–51.
D'Amato, M.R., Salmon, D.P., Loukas, E. and Tomie, A. 1985: Symmetry and Transitivity of Conditional Relations and Monkeys and Pigeons. *Journal of the Experimental Analysis of Behavior*, 44, 35–47.
D'Amato, M.R. and Van Sant, P. 1988: The Person Concept in Monkeys: *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 43–55.
Danto, A.C. 1988: Deception and Explanatory Economy: Commentary on Whiten and Byrne. *Behavioral and Brain Sciences*, 11, 252.
Davidson, D. 1984: *Inquiries into Truth and Interpretation*. Oxford University Press.
Davis, H. and Perusse, R. 1988: Numerical Competence in Animals: Definitional Issues, Current Evidence and a New Research Agenda. *Behavioral and Brain Sciences*, 11, 561–615.
Dennett, D.C. 1983: Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended. *Behavioral and Brain Sciences*, 6, 343–90.
Dennett, D.C. 1986: Philosophy as Mathematics and as Anthropology. *Mind and Language*, 1, 18–19.
Devitt, M. and Sterelny, K. 1987: *Language and Reality*. Oxford: Basil Blackwell.
Dickinson, A. 1980: *Contemporary Animal Learning Theory*. Cambridge University Press.
Dickinson, A. 1989: Expectancy Theory in Animal Conditioning. In S. B. Klein and R. R. Mowrer (eds), *Contemporary Learning Theories: Pavlovian Conditioning and the Status of Traditional Learning Theory*. Hillsdale, N.J.: Erlbaum.
Dretske, F.I. 1981: *Knowledge and the Flow of Information*. Cambridge, MA.: MIT Press.
Dretske, F.I. 1988: *Explaining Behavior*. Cambridge, MA.: MIT Press.
Dummett, M. 1973: *Frege: Philosophy of Language*. London: Duckwork.
Dummett, M. 1978: *Truth and other Enigmas*. London: Duckworth.
Estes, W.K. 1986: Array Models for Category Learning. *Cognitive Psychology*, 18, 500–49.
Fersen, L.V. and Lea, S.E.G. 1900: Category Discrimination by Pigeons using

Five Polymorphous Features. *Journal of the Experimental Analysis of Behavior*, 54, 69–84.
Fields, P.E. 1932: Studies in Concept Formation I: The Development of the Concept of Triangularity by the White Rat. *Comparative Psychology Monographs*, 9, 1–70.
Fodor, J.A. 1975: *The Language of Thought*. New York: Thomas Crowell.
Fodor, J.A. 1981: The Current Status of the Innateness Controversy. In *Representations*. Cambridge, MA.: MIT Press.
Fodor, J.A. 1984: Psychosemantics. Ms. Massachusetts Institute of Technology.
Fodor, J.A. 1987: Psychosemantics. Cambridge, MA.: MIT Press.
Fodor J.A. 1990: *A Theory of Content and other Essays*. Cambridge, MA.: MIT Press.
Fodor, J.A., Garrett, M.F., Walker, E.C.T. and C.H. Parkes 1980: Against Definitions. *Cognition*, 8, 263–367.
Fodor, J.A. and Lepore, E. 1992: *Holism: A Shopper's Guide*. Basil Blackwell.
Fodor, J.A. and Pylyshyn, Z.W. 1988: Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition*, 34, 93–107.
Gelman, R. and Gallistel, C.R. 1978: *The Child's Understanding of Number*. Harvard University Press.
Gibson, J.J. 1979: *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.
Gillett, G. 1987: Concepts, Structures and Meaning. *Inquiry*, 30, 101–12.
Gillett, G. 1988: Learning to Perceive. *Philosophy and Phenomenological Research*, 48, 601–18.
Glass, A.L., Holyoak, K.J. and Santa, J.L. 1979: *Cognition*. Reading, MA.: Addison-Wesley.
Godfrey-Smith, P. 1989: Misinformation. *Canadian Journal of Philosophy*, 19, 533–50.
Gould, S.J. and Lewontin, R.C. 1984: The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. In E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA.: MIT Press.
Gould, S.J. and Vrba, E.S: 1982: Exaptation: A Missing Term in the Science of Form. *Paleobiology*, 8, 4–15.
Gray, R.D. 1988: Metaphors and Methods: Behavioural Ecology, Panbiogeography and the Emerging Synthesis. In M.-W. Ho and P. T. Saunders (eds), *Evolutionary Processes and Metaphors*. New York: Wiley.
Greene, S.L. 1983: Feature Memorization in Pigeon Concept Formation. In M. Commons, R. J. Herrnstein and A. R. Wagner (eds), *Quantitative Analyses of Behaviour: Discrimination Processes, Volume 4*. Cambridge, MA.: Ballinger.
Griffin, D.R. 1976: *The Question of Animal Awareness*. New York: Rockerfeller.
Hailman, J.P. 1988: Operationalism, Optimality and Optimism: Suitabilities Versus Adaptations of Organisms. In M.-W. Ho and P. T. Saunders (eds), *Evolutionary Processes and Metaphors*. New York: Wiley.
Hartigan, J.A. 1975: *Clustering Algorithms*. New York: Wiley.
Herman, L.M. and Gordon, J.A. 1974: Auditory Delayed Matching in the Bottlenose Dolphin. *Journal of the Experimental Analysis of Behavior*, 21, 19–29.
Herman, L.M., Richards, D. and Wolz, J. 1984: Comprehension of Sentences of Bottlenosed Dolphins. *Cognition*, 16, 129–219.

Herrnstein, R.J. 1985: Riddles of Natural Categorization. *Philosophical Transactions of the Royal Society, London*, B308, 129–44.
Herrnstein, R.J. and Loveland, D.H. 1964: Complex Visual Concept in the Pigeon. *Science* 146, 549–51.
Herrnstein, R.J., Loveland, D.H. and Cable, C. 1976: Natural Concepts in Pigeons. *Journal of Experimental Psychology: Animal Behaviour Processes*, 2, 285–311.
Herrnstein, R.J. and de Villers, P.A. 1980: Fish as a Natural Category for People and Pigeons. In G. H. Bower (ed.), *The Psychology of Learning and Motivation, Volume 14*. New York: Academic Press.
Heyes, C.M. 1987: Contrasting Approaches to the Legitimation of Intentional Language within Comparative Psychology. *Behaviorism*, 15, 41–50.
Heyes, C.M. 1988: The Distant Blast of Lloyd Morgan's Canon. Commentary on Whiten and Byrne. *Behavioral and Brain Sciences*, 11, 256–7.
Heyes, C.M. 1993: Anecdotes, Training, Trapping and Triangulating: Can Animals Attribute Mental States. *Animal Behaviour*, 46, 177–88.
Heyes, C.M. and Dickinson, A. 1990: The Intentionality of Animal Action. *Mind and Language*, 5, 87–104.
Hintzman, D.L. 1986: 'Schema Abstraction' in a Multiple Trace Memory Model. *Psychological Review*, 93, 411–28.
Hintzman, D.L. and Ludman, G. 1980: Differential Forgetting of Prototypes and Old Instances: Simulation by an Exemplar-based Classification Model. *Memory and Cognition*, 8, 378–82.
Hodos, W. and Campbell, C.B.G. 1969: Scala Naturae. *Psychological Review*, 76, 337–50.
Hubel, D.H. and Wiesel, T.N. 1962: Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *Journal of Physiology*, 166, 106–54.
Hubel, D.H. and Wiesel, T.N. 1968: Receptive Fields and Functional Architecture of Monkey Striate Cortex. *Journal of Physiology*, 195, 215–43.
Humphrey, N. 1988: Lies, Damned Lies and Anecdotal Evidence. Commentary on Whiten and Byrne. *Behavioral and Brain Sciences*, 11, 257–8.
Hunter, W.S. 1913: The Question of Form Perception. *Journal of Animal Behaviour*, 3, 329–33.
Israel, D. and Perry, J. 1987: What is Information? Technical Report, Centre for Study of Language and Information, Stanford, California.
Johnson, H.M. 1914: Hunter, on the Question of Form Perception in Animals. *Journal of Animal Behaviour*, 4, 134.
Johnston, T.D. 1981: Contrasting Approaches to a Theory of Learning. *Behavioral and Brain Sciences*, 4, 125–73.
Keller, F.S. and Schoenfeld, W.N. 1950: *Principles of Psychology*. New York: Appleton-Century-Crofts.
Kitcher, P. 1985: *Vaulting Ambition*. Cambridge, MA.: MIT Press.
Kohonen, T. 1984: *Self Organization and Associative Memory*. Berlin: Springer-Verlag.
Kripke, S. 1972: Naming and Necessity. In D. Davidson and G. Harman (eds), *Semantics of Natural Language*. Dordrecht: D. Reidel.
Kruskal, J.B. 1964: Multidimensional Scaling. *Psychometrika*, 29, 1–42.
Kuffler, S.W. 1953: Discharge Patterns and Functional Organization of Mammalian Retina. *Journal of Neurophysiology*, 16, 37–58.
Lakoff, G.P. 1973: Hedges: A Study in Meaning Criteria and the Meaning of

Fuzzy Concepts. In C. Corum, T.C. Smith-Starkand and A. Weiser (eds), *Proceedings in the Ninth Annual Meeting of the Chicago Linguistic Society*. Chicago.
Lashley, K.S. 1930: The Mechanism of Vision III. The Comparative Visual Acuity of Pigmented and Albino Rats. *Journal of Genetic Psychology*, 37, 481–4.
Lea, S.E.G. 1984: In What Sense Do Pigeons Learn Concepts? In H. Roitblat, T. G. Bever and H. S. Terrace (eds), *Animal Cognition*. Hillsdale, N.J.: Erlbaum.
Lea, S.E.G. and Harrison, S.N. 1978: Discrimination of Polymorphous Stimulus Sets by Pigeons. *Quarterly Journal of Experimental Psychology*, 30, 521–37.
Lewis, D. 1984: Putnam's Paradox. *Australasian Journal of Philosophy*, 62, 221–36.
Lewontin, R.C. 1978: Adaptation. In E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. Cambridge, MA.: MIT Press.
Lewontin, R.C. 1983: Gene, Organism and Environment. In D. S. Bendall (ed.), *Evolution from Molecules to Man*. Cambridge University Press.
Lockard, R.B. 1971: Reflections on the Fall of Comparative Psychology. *American Psychologist*, 26, 168–79.
Lubow, R.E. 1974: Higher-Order Concept Formation in the Pigeon. *Journal of the Experimental Analysis of Behavior*, 21, 475–83.
Mackintosh, N.J., Wilson, B. and Boakes, R.A. 1985: Differences in Mechanisms of Intelligence among Vertebrates. *Philosophical Transactions of the Royal Society, London*, 308B, 53–65.
MacPhail, E.M. 1980: Short-Term Visual Recognition Memory in Pigeons. *Quarterly Journal of Experimental Psychology*, 32, 521–38.
MacPhail, E.M. and Reilly, S. 1989: Rapid Acquisition of a Novelty versus Familiarity Concept by Pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 242–52.
Malott, R.W. and Siddall, J.W. 1972: Acquisition of the People Concept in Pigeons. *Psychological Reports*, 31, 3–13.
Mandler, J. 1988: How to Build a Baby: On the Development of an Accessible Representational Scheme. *Cognitive Development*, 3, 113–36.
Mandler, J. 1992: How to Build a Baby: II. Conceptual Primitives. *Psychological Review*, 99, 587–604.
Marr, D. 1982: *Vision*. New York: Freeman.
Mason, W.A. and Lott, D.F. 1976: Ethology and Comparative Psychology. *Annual Review of Psychology*, 27, 129–54.
Medin, D.L. and Schaffer, M.M. 1978: Context Theory of Classification Learning. *Psychological Review*, 85, 207–38.
Medin, D.L. and Smith, E.E. 1984: Concepts and Concept Formation. *Annual Review of Psychology*, 35, 113–38.
Miller, G.A. and Johnson-Laird, P. N. 1976: *Language and Perception*. Cambridge, MA.: Harvard University Press.
Millikan, R.G. 1984: Language, Thought and other Biological Categories. Cambridge, MA.: MIT Press.
Millikan, R.G. 1986: Thought without Laws; Cognitive Science without Content. *Philosophical Review*, 95, 47–80.
Millikan, R.G. 1990: Truth Rules, Hoverflies, and the Kripke–Wittgenstein Paradox. *Philosophical Review*, 99, 323–53.
Mitchell, R.W. 1988: Ontogeny, Biography and Evidence for Tactical Deception.

Commentary on Whiten and Byrne. *Behavioral and Brain Sciences*, 11, 259–60.
Morgan, M.J., Fitch, M.D., Holman, J.G. and Lea, S.E.G. 1976: Pigeons Learn the Concept of an 'A'. *Perception*, 5, 57–66.
Munn, N.L. 1931: An Apparatus for Testing Visual Discrimination in Animals. *Journal of Genetic Psychology*, 39, 342–58.
Nakagawa, E. 1986: Overtraining, Extinction and Shift Learning in a Concurrent Discrimination in Rats. *Quarterly Journal of Experimental Psychology*, 38B, 313–26.
Nosofsky, R. 1984: Choice, Similarity and the Context Theory of Classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 10, 104–14.
Nosofsky, R. 1986: Attention, Similarity and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General*, 115, 39–57.
Nosofsky, R. 1989: Exemplars, Prototypes and Similarity Rules. Research Report 17, Cognitive Science, Indiana University.
Papineau, D. 1987: *Reality and Representation*. Oxford: Basil Blackwell.
Pavlov, I. 1927: *Conditioned Reflexes*. Oxford University Press.
Peacocke, C. 1989: Possession Conditions: A Focal Point for Theories of Concepts. *Mind and Language*, 4, 51–61.
Pearce, J.M. 1989: The Acquisition of an Artificial Category by Pigeons. *Quarterly Journal of Experimental Psychology*, 41B, 381–406.
Pepperberg, I.M. 1987a: Interspecies Communication: A Tool for Assessing Conceptual Abilities in the African Grey Parrot. In G. Greenberg and E. Tobach (eds), *Language, Cognition and Consciousness: Integrative Levels*. Hillsdale, N.J.: Erlbaum.
Pepperberg, I.M. 1987b: Acquisition of the Same/Different Concept by an African Grey Parrot: Learning with Respect to Categories of Color, Shape, and Material. *Animal Learning and Behavior*, 15, 423–32.
Pepperberg, I.M. 1988: The Importance of Social Interaction and Observation in the Acquisition of Communicative Competence: Possible Parallels between Avian and Human Learning. In T. R. Zentall and B. G. Galef (eds), *Social Learning: Psychological and Biological Perspectives*. Hillsdale, N.J.: Erlbaum.
Pepperberg, I.M. 1991: A Communicative Approach to Animal Cognition: A Study of Conceptual Abilities of an African Grey Parrot. In C. Ristau (ed.), *Cognitive Ethology: The Minds of Other Animals*. Hillsdale, N.J.: Erlbaum.
Plotkin, H.C. 1979: Brain-Behaviour Studies and Evolutionary Biology. In D. Oakley and H. C. Plotkin (eds), *Brain, Behaviour and Evolution*. London: Methuen.
Plotkin. H.C. (ed.) 1982: *Learning, Development and Culture: Essays in Evolutionary Epistemology*. Chichester: Wiley.
Posner, M.I. and Keele, S.W. 1968: On the Genesis of Abstract Ideas. *Journal of Experimental Psychology*, 77, 353–63.
Premack, D. 1983: The Codes of Man and Beasts. *Behavioural and Brain Sciences*, 6, 125–67.
Premack, D. and Premack, A.J. 1983: *The Mind of an Ape*. New York: Norton.
Preston, G.C., Dickinson, A. and Mackintosh, N. 1986: Contextual Conditional Discriminations. *Quarterly Journal of Experimental Psychology*, 38B, 217–37.
Putnam, H. 1962. It Ain't Necessarily So. *Journal of Philosophy*, 59, 658–71.
Putnam, H. 1975: The Meaning of 'Meaning'. In K. Gunderson (ed.), *Minnesota*

Studies in the Philosophy of Science, Volume 7, Language, Mind and Knowledge. University of Minnesota Press.
Quine, W.V.O. 1960: *Word and Object*. Cambridge, MA.: MIT Press.
Reed, S.K. 1972: Pattern Recognition and Categorization. *Cognitive Psychology*, 3, 382–407.
Rips, L. J., Shoben, E.J. and Smith, E.E. 1973: Semantic Distance and the Effect of Semantic Relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1–20.
Roberts, W.A. and Mazmanian D.S. 1988: Concept Learning at Different Levels of Abstraction by Pigeons, Monkeys, and People. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 247–60.
Rosch, E. 1973: On the Internal Structure of Perceptual and Semantic Categories. In T. Moore (ed.), *Cognitive Development and the Acquisition of Language*. New York: Academic Press.
Rosch, E. 1975: Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, 104, 192–233.
Roitblat, H.L. 1982: The Meaning of Representation in Animal Memory. *Behavioral and Brain Sciences*, 5, 353–406.
Santiago, H. and Wright, A.A. 1984: Pigeon Memory: Same–Different Concept Learning, Serial Probe Recognition Acquisition, and Probe Delay Dffects in the Serial-Position Function. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 498–512.
Saunders, P.T. 1988: Sociobiology: A House Built on Sand. In M.-W. Ho and P.T. Saunders (eds), *Evolutionary Processes and Metaphors*. New York: Wiley.
Savage-Rumbaugh, S., McDonald, K., Sevcik, R.A., Hopkins, W.D. and Rubert, E. 1986: Spontaneous Symbol Acquisition and Communicative Use by Pygmy Chimpanzees. *Journal of Experimental Psychology: General*, 115, 211–35.
Schank, R.C. and Abelson, R.P. 1975: Scripts, Plans and Knowledge. In P. N. Johnson-Laird and P. C. Wason (eds), *Thinking: Readings of Cognitive Science*. Cambridge University Press.
Schrier, A.M., Angarella, R. and Povar, M.L. 1984: Studies of Concept Formation by Stumptailed Monkeys: Concepts, Humans, Monkeys and Letter A. *Journal of Experimental Psychology: Animal Behavior Processes*, 13, 136–43.
Seibt, U. 1982: Zahlbegriff und Zahlverhalten bei Tiereu. Neue Versuche und Deutungen. *Zeitschrift für Tierpsychologie*, 60, 325–41.
Seyfarth, R.M. and Cheney, D.L. 1980: The Ontogeny of Vervet Alarm Calling Behavior: A Preliminary Report. *Zeitschrift für Tierpsychology*, 54, 37–56.
Shanks, D.R. 1990: Concept Learning in an Associative Network. Paper presented to the Easter Meeting of the Experimental Analysis of Behaviour Group, University of York, U.K.
Shepard, R. N., Hovland, C.I. and Jenkins, H.M. 1961: Learning and Memorization of Classifications. *Psychological Monographs*, 517, 75, 1–42.
Sidman, M., Rauzin, R., Lazard, R., Cunningham, S., Tailby, W. and Carrigan, P. 1982: A Search for Symmetry in the Conditional Discriminations of Rhesus Monkeys, Baboons, and Children. *Journal of the Experimental Analysis of Behavior*, 37, 23–44.
Siegel, R.K. and Honig, W.K. 1970: Pigeon Concept Formation: Successive and Simultaneous Acquisition. *Journal of the Experimental Analysis of Behavior*, 13, 385–90.

Skinner, B.F. 1981: Selection by Consequences. *Science*, 213, 501–504.
Smith, E.E. and Medin, D.L. 1981: *Categories and Concepts*. Cambridge, MA.: Harvard University Press.
Stampe, D. 1977: Towards a Causal Theory of Linguistic Representation. In P. French, T. Euhling and H. Wettstein (eds), *Midwest Studies in the Philosophy of Science, Volume 2*. Minneapolis: University of Minnesota Press.
Stemmer, N. 1980: Natural Concepts and Generalization Classes. *The Behavior Analyst*, 3, 41–48.
Thomas, R.K. 1988: Misdescription and Misuse of Anecodotes and Mental State Concepts (Commentary on Whiten and Byrne). *Behavioral and Brain Sciences*, 211, 265–6.
Thomas, R.K. and Nobel, L.M. 1988: Visual and Olfactory Oddity Learning in Rats: What Evidence is Necessary to Show Conceptual Behavior? *Animal Learning and Behavior*, 16, 157–163.
Thorndike, E.L. 1898: Animal Intelligence: An Experimental Study of the Associative Process in Animals. *Psychology Review Monographs*. No. 8.
Thorpe, W.H. 1956: *Learning and Instinct in Animals*. Harvard University Press.
Todd, I.A. and Mackintosh, N.J. 1990: Evidence for Perceptual Learning in Pigeons' Recognition Memory for Pictures. *Quarterly Journal of Experimental Psychology*, 42B, 385–400.
Uexkull, J. von 1934: *Streifyuge durch die Umwelten von Tieren und Menschen*. Berlin: Springer.
Vandierendonck, A. 1990: Rule Structure, Frequency, Typicality Gradients, and the Representation of Diagnostic Categories. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie and G. Erdos (eds), *Lines of Thinking: Reflections on the Psychology of Thought, Volume 1*. Chichester: Wiley.
Vaughan, W. 1988: Formation of Equivalence Sets in Pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 36–42.
Vaughan, W. and Greene, S.L. 1984: Pigeon Visual Memory Capacity. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 256–71.
Washburn, M.F. 1926: *The Animal Mind*, 2nd edn. New York: Macmillan.
Wasserman, E.A., Kiedinger, R.E. and Bhatt, R.S. 1988: Conceptual Behavior in Pigeons: Categories, Subcategories, and Pseudocategories. *Journal of Experimental Psychology: Animal Behavior Processes*, 14, 235–46.
Weiskrantz, L. (ed.) 1988: *Thought without Language*. Oxford University Press.
Whiten, A. and Byrne, R.W. 1988: Tactical Deception in Primates. *Behavioral and Brain Sciences*, 11, 224–44.
Wittgenstein, L. 1953: *Philosophy Investigations*, 2nd edn, trans. G. E. M. Anscombe. Oxford: Basil Blackwell.
Zentall, T.R. and Hogan, D.E. 1976: Pigeons can Learn Identity or Difference, or Both. *Science*, 191, 408–9.
Zentall, T.R. and Hogan, D.E. 1978: Same/Different Concept Learning in the Pigeon: The Effect of Negative Instances and Prior Adaptation of the Transfer Stimuli. *Journal of Experimental Analysis of Behavior*, 30, 177–86.
Zentall, T.R., Steirn, J.N., Sherburne, L.M. and Urcuioli, P.J. 1991: Common Coding in Pigeons Assessed Through Partial versus Total Reversals of Many-to-One Conditional and Simple Discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, 17, 194–201.

Mind & Language ISSN 0268-1064
Vol. 9 No. 3 September 1994

Article

Systematicity in Connectionist Language Learning

ROBERT F. HADLEY

1. Introduction

It is by now widely acknowledged by cognitively-oriented connectionists that human thought and language production display both *compositionality and systematicity*. In large part, recognition of this fact has been fostered by Fodor and Pylyshyn (1988), who argue that prevailing methods of connectionist representation cannot engender the *combinatorial* syntax and semantics necessitated by compositionality. Further, since combinatorial syntax and semantics *cause those systematic relationships* of thought and language that concern Fodor and Pylyshyn (F&P), the very features of connectionist representations that preclude compositionality also preclude systematicity (or so it is argued). Much of F&P's discussion addresses the question whether connectionists can achieve structure-sensitive processing without, in effect, creating connectionist implementations of classical symbol processing systems. F&P argue for a negative answer.

Since the appearance of F&P's paper, a number of connectionists have produced seeming *counterexamples* to the F-P thesis. In the present work I examine six of these apparent counterexamples, which are due to Elman (1990), St. John and McClelland (1990), Chalmers (1990), McClelland and

The author gratefully acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (Grant No. A0899) during the period of this research. I am especially grateful to Steven Pinker for his invaluable remarks on a previous version of this paper. Thanks also to Martin Davies, Allan Bennett-Brown, Tony Plate, and an anonymous reviewer of this journal for their helpful comments. Of course, my indebtedness does not imply agreement on the part of these commentators with all or most of my position. Address for correspondence: School of Computing Science, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada. Email: hadley@cs.sfu.ca

I shout, "No, not yet!") To find the interpretation that best fits her environment, her other behavior, and any other available relevant information, we need more than imitation. One possibility is that, rather than contenting ourselves with the initial first person match to Mary's behavior, we (whether aware of it or not) try out, or try on, various *alternative* motives, using hypothetico-practical reasoning (Gordon 1986) to determine the best fit. How do we try them out? Rather than just watching her arm and noting her direction of gaze, we mentally put ourselves into her environment, further tweaking our own cognitive, motivational, emotional, and decision-making systems until they generate in imagination not just her present behavior but her earlier and later behavior. In addition, we (beginning at age 4) also block out information that would be unavailable in Mary's epistemic situation, a feat that would place us at B&M's level 4. A distinct possibility, of course, is that people have and use a theory that allows them to determine, by hypothetico-deductive reasoning, prototype matching, or some other method, which motive best fits the information. (Because I am not clear how a theory might make use of imitation or first person information, I am inclined to think B&M would lean toward the simulation account.)

My general point here is that B&M give insufficient attention to the need for a methodology, whether it be simulation-based or theory-based or both, for choosing among alternative possible "imitations" of object-directed activity, that is, alternative ways of matching first person information with information about the other's current behavior. In older children and adults, even what appears to be simple mimicry or imitation, such as picking up another's emotional response, may involve the testing of competing hypotheses and a methodology – presumably a first person methodology, like simulation, that is sensitive both to observed facial expression and contextual information of the sorts mentioned earlier (Gordon 1995a; 1995b).

(2) Apart from either simulation or theory, representations linking first and third person information would be of little use. For one thing, a capacity to imagine in the first person what is behind another's behavior would not help one predict or anticipate the other's future behavior. Suppose I pick up from your facial expression your emotional response to the object you are looking at. My first person representation of your emotional response will help me predict your behavior only if I possess a mechanism for moving from it (together, if possible, with available contextual information) to a representation of the likely ensuing behavior. Or suppose I pick up from the Maxi story a first person representation of Maxi's desire to fetch his chocolate, his recollection of its past location, and his ignorance of its current location. Without a mechanism to connect these representations with a representation of behavior, these will lead to no prediction. Simulation provides such a mechanism: First person representations of another's emotions, desires, and beliefs would influence my first person representation of the other's decision making in the same general way that my own emotions, desires, and beliefs influence my own decision making. Possibly a commonsense theory could also provide such a mechanism, although it would have to be an odd sort of theory that identified mental states in first person terms. My general point is that merely to represent the inner aspect of another's present or past behavior will not of itself foretell the other's future behavior.

Imagination and imitation: Input, acid test, or alchemy?

C. M. Heyes

Department of Psychology, University College London, Gower Street, London WC1E 6BT, England; **c.heyes@ucl.ac.uk**

Abstract: Immediate imitation is likely to be a major, direct input to Barresi & Moore's level 2 competence, but deferred imitation is unlikely to play a key role in the transition to level 3, because (1) the attribution of first person knowledge is neither a necessary cause nor an obvious consequence of deferred imitation, and (2) deferred imitation does not correlate phylogenetically with capacities that more plausibly either yield or reflect a concept of intentional agency.

Stage models of ontogeny, however ingenious, often leave one wondering: "But *how?* What is it about stage X competence, combined with the events supposed to occur at or just after that stage, that transforms the system into one with the characteristics of stage X + 1?" One, among many, of the strengths of Barresi & Moore's (B&M's) target article is that it addresses this kind of question head on; it postulates, more explicitly than any other model I have seen, plausible mechanisms of stage transition. But still there seems to be some alchemy afoot. How is the addition of some imagination between levels 2 and 3 supposed to convert a creature that can represent but not recognize a shared intentional relation into one that has "a concept of an intentional agent," that is, that can distinguish first and third person information, and attribute first person knowledge to another?

Because B&M identify deferred imitation as critical with respect to the level 2–level 3 transition, this question can be given a more specific formulation: What is it about the experience of imitating a remembered, rather than a concurrently observed, action that might provoke the recognition that others have first person knowledge? B&M assume that, at level 2, concurrently observed actions are imitated without such recognition, and I see no reason to suppose that imitation of remembered actions would change this situation. Why couldn't past actions be imitated in the same "automatic" fashion?

Those who have difficulty in imagining deferred imitation without level 3 knowledge of intentional relations may be able to make the gestalt switch after considering some evidence of imitation in rats (e.g., Heyes et al. 1992). We allowed each of number of naive "observer" rats to face a conspecific "demonstrator" as the latter pushed a joystick to the left or right of the actor's body for food reward. Subsequently, when the demonstrators had been removed and the observers were given access to the joystick, we found that, although they were rewarded for all responses, the observers tended to push the joystick in the same direction relative to their bodies as had their demonstrators. This resulted in the joystick moving in the opposite direction within the observers' visual field, and, in some experimental conditions, in relation to compass points, to that in which it had moved during demonstrator observation.

Whether or not one can accept this as evidence of true, deferred imitation (Byrne & Tomasello, in press; Heyes, in press a), these data encourage us, as users of an adult human theory of intentional relations, to imagine rats imitating a previously observed action. Compared with infants and nonhuman primates, the tendency to project mature human competence onto rats is minimal. Therefore, this imaginative exercise can show that there is no necessary or obvious relationship between imitation and the attribution of first person knowledge. The idea that the former yields or inevitably reflects the latter needs more explanation.

Perhaps I have misunderstood B&M. Although they suggest that immediate imitation generates level 2 competence, they may believe that deferred imitation betrays rather than provokes level 3 understanding. Deferred imitation may be an acid test of level 3 competence, rather than a critical input to its acquisition; it could be such a test if its emergence coincided, phylogenetically and ontogenetically, with that of capacities that plausibly require or yield level 3 understanding. However, on the phylogenetic side, there are no reliable correlations of this kind and, at the ontogenetic level, some of the correlates are more plausible acid tests than imitation itself.

B&M argue persuasively that monkeys' behavior is likely to reflect no more than level 1 competence. Had they applied the same incisive analysis to the data on social cognition in chimpanzees, they would probably have reached the same conclusion about great apes. Many comparative psychologists now doubt that

apes can "ape" (e.g., Galef 1988; Tomasello, in press) and, even if we could be sure that they are capable of deferred imitation, since all of the chimpanzee data cited by B&M in section 3.2 can be readily explained without attributing even level 2 competence (e.g., Heyes 1994a; submitted), there are no valid indicators of level 3 competence with which imitation could be correlated in nonhuman primates.

Turning to ontogeny, B&M point out that children begin to engage in deferred imitation at about the same time they begin to use personal pronouns and to show mirror self-recognition, self-conscious emotions, and empathy. Self-recognition, measured by mark tests, is no more likely to reflect the attribution of first person information than is imitation (or, for that matter, collision-free locomotion; Gallup et al., in press; Heyes 1994b; in press b); responding to emphatic distress by comforting another (e.g., Hoffman 1977) is plausibly a result of instrumental learning. That leaves pronoun use and self-conscious emotion and, even if the emergence of these behaviors correlates with the onset of deferred imitation (contra, e.g., Meltzoff 1988), surely, by virtue of their greater construct validity, pronoun use and self-conscious emotion are more plausible acid tests than imitation itself.

B&M have provided a valuable insight by showing that the products of immediate imitation are likely to be a major, direct input to level 2 competence. However, it is not clear how deferred imitation could play a comparable role in level 3 understanding of intentional relations, or why, if deferred imitation is not causal, it should be given criterion status.

Understanding minds and selves

R. Peter Hobson

Developmental Psychopathology Research Unit, Tavistock Clinic, London, NW3 5BA, England; **rejurph@ucl.ac.uk**

Abstract: Barresi & Moore provide a welcome focus on children's abilities to integrate first and third person information about intentional relations but they pay insufficient attention to the origins of children's understanding of the nature of subjective orientations vis-à-vis a shared world and the potential significance of such understanding as a source (rather than an outcome) of domain-general information-processing capacities.

Barresi & Moore's (B&M's) account of the early development of human social understanding has many strengths. I am especially in agreement with the view that "sharing" experiences of the world antedates and is a precondition for awareness that self and other are distinct but have minds in common; that infants' interpersonal understanding at the end of the first year of life does *not* yet amount to their having "concepts" of themselves and others with mental states; and that an account of the child's growing understanding of mental states entails and requires a parallel account of the processes of self–other coordination and differentiation (Hobson 1993).

Here I shall focus on the relationship between interpersonal relatedness and social understanding, or, in modern parlance, on the development of "representations" as they concern people and their mental states. I am doubtful whether B&M succeed in accounting for the emergence of new levels in children's self- and other-awareness and understanding at around 18 months and then at 4 years of age by their appeal to an increase in general information-processing mechanisms. To evolve a more satisfactory alternative account, we may need to introduce some refinements in the way that B&M characterise the basic forms of human interpersonal relations.

To begin with, a point of agreement: Certain forms of interpersonal coordination are necessary foundations for, rather than sequelae to, children's concepts about mental states. For example, infants need to experience intersubjective coordination with others in order to have a basis for understanding what it means to share experiences; *understanding* what it means to share or not share experience (i.e., to be able to conceptualise that other people may have subjective experiences that are like one's own but that may be different at any moment) is constitutive of concepts to do with self, other, and the nature of mental states. If all this is correct, then the critical issue becomes that of explaining which forms of interpersonal coordination (1) can take place on a noninferential and non-symbolic-representational basis, and (2) are of a kind that can ultimately yield concepts (or symbolic representations) that concern minds and selves.

B&M group together "various means" by which a child may bring together first and third person information about intentional relations, and they cite emotional empathy, joint attention, and (most important for B&M) imitation by one agent of another's actions on objects. So far so good; but B&M do not analyse the different ways in which such processes may contribute to the development of social understanding, nor how they may be differentially impaired in early childhood autism. For example, my colleague Tony Lee and I have some studies in progress which appear to demonstrate that, whereas autistic children are relatively adept at imitating the goal-directed actions of a model (see also Charman & Baron-Cohen 1994), they are impaired in assuming the behavioural style and/or bodily-affective expressiveness of someone else (see also Loveland et al. 1994). So, too, autistic children appear to register when their own actions are being imitated (e.g., Dawson & Adams 1984). This suggests that, whereas autistic children perceive correspondences between their own and others' *actions*, there may be something unique about autistic children's difficulties in assuming the attitudes or subjective orientations of other people – and this may be critical for their impaired development of social understanding.

All this relates to the issue of whether domain-general cognitive advances in imagination can explain the growth of nonautistic children's social understanding at 18 months and at 4 years. In this regard, B&M elaborate upon Olson's formulation about the child simultaneously "holding in mind" representations of nonpresent or noncurrent objects or events. An alternative view is that at 18 months, a new form of representation – symbolic representation – emerges on the basis of advances in children's understanding of minds, and that this is what accounts for their increase in processing capacity. I suggest that children come to grasp that objects and events exist in their own right, and under a description (or representation) that is independent of people's current actions and attitudes toward those objects and events, by identifying with others' attitudes toward a shared world and toward the child's own attitudes (yielding self-reflective awareness). What is "held in mind," namely, the conceptual distinction between different individuals' attitudes to objects and events and the objects and events themselves, corresponds with the distinction between symbolic representations and the referents with which the representations are concerned. True, this symbolic-representational ability is domain-general insofar as it can be widely applied, but it may be domain-specific in its origins. So, too, at 4 years of age, it is not just that children represent more, but also that they understand that *what* a person represents is (truly or falsely) taken to be "reality." The acquisition of the concept of reality will not be analysable solely with reference to children's increasing powers of imagination. In each case, the quality of the new "representational abilities" seems to reveal that more than an increase in processing capacity is needed for a satisfactory developmental account of their emergence.

In summary, therefore, I do not think that B&M have gone far enough in analysing the sources and implications of young children's ability to coordinate and subsequently to understand perspectives *as* subjectively experienced perspectives. In my view, they overemphasise the significance of children's imitation of actions vis-à-vis children's identification with attitudes, and they do not really provide a convincing argument that stepwise increases in a child's information-processing capacities at 18 months and 4 years are what explain the increases in social understanding, rather than vice versa.