

## **Animal cognition and animal minds**

Colin Allen, Texas A&M University

colin.allen@tamu.edu

Psychology, according to a standard dictionary definition, is the science of mind and behavior. For a major part of the twentieth century, (nonhuman) animal psychology was on a behavioristic track that explicitly denied the possibility of a science of animal mind. While many comparative psychologists remain wedded to behavioristic methods, they have more recently adopted a cognitive, information-processing approach that does not adhere to the strictures of stimulus-response explanations of animal behavior. Cognitive ethologists are typically willing to go much further than comparative psychologists by adopting folk-psychological terms to explain the behavior of nonhuman animals.

The theoretical terms of cognitive and folk-psychological approaches are intentional, in Brentano's sense. Contra Brentano, there is considerable optimism among philosophers that it will be possible to provide a naturalistic theory of intentionality, although there is plenty of disagreement about the form such a naturalistic theory will take. Given the various deflationary accounts of intentionality that philosophers are offering these days it is not hard to convince most philosophers and scientists (staunch behaviorists excepted) that intentional terms are respectable ingredients in scientific explanations of the behavior of many different kinds of organisms.

Great apes aside, many scientists are skeptical of claims about animal minds even though they are willing to attribute internal, representational (and hence intentional) states to the same organisms. To illustrate: Gould (1986) has argued that individual honeybees form internal maps of their spatial environments that represent features such as the locations of large bodies of water

relative to the hive. These maps are postulated to explain several phenomena, including that bees are not recruited by dances that indicate a direction and distance to a food source in the middle of a lake. (The dances were cleverly produced by exposing foraging bees to a boat with a food source on board that was gradually moved into the middle of a lake.) Challenges to Gould's hypothesis (e.g. Dyer 1991) do not typically start from the premise that bees are incapable of having cognitive maps; in other words, it is generally considered a reasonable (but perhaps false) hypothesis that bees utilize internal representations that integrate various spatial features of their environments. In contrast, many people would reject out of hand the claim that bees have minds, or that they have beliefs about the locations of bodies of water--thus rejecting mentalistic explanations of bee behavior. The specific challenge to mentalistic attributions that I have in mind might, therefore, be raised by someone who rejects strict behaviorism and accepts that cognitive, intentional explanations are appropriate for explaining some animal behavior. Such a person would accept that more or less complex interactions between internal states with intentional content explain some animal behavior, but would deny that the mental states recognized within folk psychology are appropriately attributed to (most) nonhuman animals.

This attitude presupposes a distinction between cognitive and mental state attributions that is not commonly articulated. Indeed the terms "mental" and "cognitive" are frequently used interchangeably by both philosophers and scientists. Among philosophers, when discussion is focused on intentionality the conflation is unsurprising since both mental states and cognitive states are widely considered to be intentional. When coupled with Brentano's thesis that intentionality is the distinguishing characteristic of the mental the conflation is even easier to understand. The conflation is also evident in Griffin's (1976, 1984, 1992) exhortations to ethologists to become cognitive ethologists by paying more attention to animal minds. And it is

present when Gould and Gould (1994) use the terms "cognitive map" and "mental map" interchangeably (searching their index for the former will locate instances of the latter). If we grant that whatever representation a bee has of its spatial environment may reasonably be called a "map", the further question of whether it is appropriately called a "mental map" or a "cognitive map" may seem to be a mere terminological quibble. But I believe that there is a substantive issue lurking and I hope by the end of this paper it will be clear what that substantive issue is. The fact that it is not obviously incoherent to accept that bees have cognitive maps of their environments while simultaneously denying that bees have minds suggest that there is a distinction to be drawn here.

Just what, however, is being denied when it is denied that bees have minds? There is no simple answer to this question. We have something like the following state of affairs: many philosophers are convinced that a naturalistic theory of mind is in the offing because they have cracked the nut of intentionality. However, in doing so they have given accounts of intentionality that seem to apply to many organisms and devices that do not, *prima facie*, possess minds. One reaction to this state of affairs is to reject the *prima facie* mindlessness of such entities: thus, for example, McCarthy (1980) argues that it is appropriate to attribute beliefs to thermostats. A second reaction is to reject the proffered accounts of intentionality (e.g. Searle 1980, 1992). A third is to attempt to say what else must be added to an organism before we say it really has a mind. This latter task has been tackled by several philosophers. (Bennett 1964/1989 specifically runs the thought experiment on honeybees; see also Fodor 1986; Dretske 1986; Millikan 1993). I shall take a suggestion of Searle's on board (even if he would not approve), but what follows is my attempt to add to the third tradition--which we might call "theoretical comparative

psychology"--of attempting to say what besides intentionality is relevant to mentalistic attributions.

Spacetime limitations necessitate passing rather quickly over some controversial assumptions. First, the question of whether mentalistic (folk-psychological) explanations an adequate theoretical framework for explaining nonhuman behavior takes on its greatest significance against a background of the presumed adequacy of using such terms to explain human behavior (see Lycan, this volume). This is a huge presumption, but nonetheless one I shall make because I wish to open just one can of worms at a time. By making this assumption, I do not mean to be claiming that mentalistic terms *alone* are adequate for explaining human behavior. My claim is only that their use is not fatally flawed. Given this assumption, the issue at hand is whether there special reasons for thinking that mentalistic terms are inappropriate for explaining the behavior of nonhuman animals. Second, I take a realist stance to explanation: an explanation is genuinely explanatory only if the statements comprising its explanans are true. Thus, mentalistic explanations are genuinely explanatory only if the mentalistic attributions involved are literally true of the subjects whose behavior is to be explained, and the question of explanatory adequacy is linked to the question of whether or not animals possess minds. Third, I hope it would go without saying that we are seeking a naturalistic and empirically tractable theory of mental state attributions. In particular, the extent to which questions about animal minds can be brought into the contemporary scientific domain depends on the extent to which it is possible to get beyond the reliance on anecdote that is so characteristic of popular discussion of this topic, and that turned out to be the downfall of early Darwinian comparative psychology (Burghardt 1985; see also Gerald Massey's commentary following this paper).

Concern with the folk psychological notions also serves to bring philosophers back in touch with their scientific colleagues. The reason for this is *consciousness*: the mental states recognized by folk psychology are paradigmatically conscious states. Despite the recent explosion of philosophical and neuropsychological work on consciousness, much remains puzzling about the notion, including even whether there is a single notion here at all. Suspicion about consciousness is, therefore, partially responsible for suspicion about attributing believing, desiring minds to nonhumans. Although Freudian ideas about subconscious beliefs and desires have infiltrated folk psychology, these do not provide reassurance for attributing mental states to nonhumans because subconscious beliefs and desires have explanatory value within a Freudian model of the mind only in the context of competition between the subconscious mind and conscious ego.

Aside from the deliverances of folk psychology, there are various reasons for thinking that the topic of consciousness should not be avoided in context of thinking about animal mentality. Here are two; one is historical, the other conceptual:

(1) Griffin, who coined the term "cognitive ethology", has made consciousness a central issue in his writings (Griffin 1976, 1984, 1983). Critics of Griffin's work have been quick to pick up on this issue and use it to question the scientific status of cognitive ethology (Bekoff and Allen 1996; Allen and Bekoff 1996). For instance, Blumberg and Wasserman (1995, p.133) write: "We submit that it is this very goal of investigating animal consciousness that, although grand and romantic, falls far outside the scope of a scientific psychology that has struggled for the better part of the past century to eschew such tantalizing, but ultimately unsubstantiable, analyses of subjective mental experience." Consequently, whether proponents of mentalistic explanations of animal behavior like it or not, to ignore the issue of consciousness with respect to mentalistic

explanations of animal behavior is not a practical option. Those philosophers (including myself) who have turned their attention explicitly to cognitive ethology have tended to focus on intentionality and have had rather little to say about animal consciousness. It is time to restore the balance.

(2) Deflationary, naturalistic accounts of intentionality do not provide a complete account of mentality. For example, Millikan's theory allows that trees are producers and consumers of intentional icons (Allen and Hauser 1993; Allen 1995). Despite disagreeing with Searle's Brentano-like equation of mentality with intentionality, I think there is something worth pursuing in his suggestion that mentality requires "some *awareness* of the causal relation between the symbol and the referent" (Searle, 1980, p.454; my emphasis). Searle's point cannot quite be correct; thought can occur even when there is no actual causal connection between a symbol token and its referent--for example, when the intended referent does not exist. I shall therefore consider relations between symbol and referent more generally. It is also unfortunate that Searle says little about what gives rise to awareness of such relations. His gesture towards neurophysiology is not particularly helpful, and by rejecting behavioral tests (exemplified by his rejection of the Turing Test for artificial intelligence) it is hard to see his suggestion as empirically tractable for ethologists and others who study behavior directly. Nonetheless, I think there is an important idea that is worth pursuing in Searle's suggestion. Searle is concerned that purely formal systems have no access to their semantics. So, rather than put the point in terms of causal relations, we might rather say that mentality involves consciousness or awareness of the (semantic) *contents* of internal states. Interestingly, Stich (1978), in arguing explicitly for a distinction between mental (belief) states and other cognitive (subdoxastic) states comes to a similar idea. He writes that the abilities of typical adult human believers to report and answer

questions about their beliefs "are themselves associated with an ability to become aware of or to be conscious of the contents of one's beliefs." (Stich 1978, p.504) Stich, rather deliberately, has little to say about what such consciousness consists in. Although neither Searle nor Stich might approve of the use to which I intend to put their ideas, I will take up the challenge of trying to characterize consciousness of content in a way that makes it amenable to behavioral investigation.

This talk about consciousness of content is, admittedly, so far pretty vague, but in what follows I will try to make it less so. Before going on to give the positive account, I wish state my neutrality on the question of whether the phrase "consciousness of content" picks out qualitative properties associated one-to-one with specific belief contents. Stich writes somewhat ambiguously about this issue, viz. (1978, p.504): "in typical cases of belief a subject will have a certain sort of characteristic conscious experience when his attention is suitably directed to the content of the belief." In the context of Stich's discussion, it is unclear whether he thinks this "characteristic conscious experience" is characteristic of occurrent beliefs generally, or specific to the belief's content. In fact, I shall sidestep entirely the issue of "what it's like" to be a particular organism (Nagel 1974), or other puzzles about qualia. Like Dennett (1991), I think there's plenty left to consciousness even if we end up Quining qualia.

The state of play so far is this: I am assuming that we possess a naturalistic theory of intentionality (and its corollary, a naturalistic theory of intentional content); I am assuming that many organisms who possess internal states with intentional content are nonetheless "mindless" in the sense that these states lack what Rosenthal (1993) calls "state consciousness"; the task remains to specify what else is needed to have a mind; and I have identified a proposal, albeit vague, that consciousness of content is a missing ingredient. (Not the *only* missing ingredient, mind you.)

The game plan for the rest of the paper is as follows: I shall describe a framework for understanding naturalistic approaches to mentality. I shall focus on a particular functional aspect of a particularly narrow aspect of consciousness, that I have labelled "consciousness of content" and I shall show how it might be accommodated within this framework. I shall then consider the consequences of this approach for attributing mental states to nonhuman animals.

The framework for naturalizing the mind that I shall adopt involves the notion of *implementation*. The project, according to this conception, is to show how to implement or instantiate mental properties given nonmental parts. Because of multiple realizability, an account of implementation or instantiation does not imply reduction of the properties at one level to those at another, any more than showing how to implement a radio receiver from transistors constitutes a reduction of the property of being a radio receiver to that of being a particular organized collection of transistors. Within this framework, it is reasonable to take a hierarchical approach, a crude version of which is shown in Figure 1.

Place Figure 1 here

The first step within this hierarchical approach (implementation theory 1) is to provide a theory of content that shows how intentional properties can be instantiated by processes described in the nonintentional terms of a natural science (for want of a better term, I refer to these as "mechanical processes"). For the purposes of this paper I am assuming that we possess a satisfactory, but deflationary, account of intentionality along these lines--Fodor's (1990) causal theory, Millikan's (1984) biofunctional theory, and Dretske's (1986) "indicator" semantics are all candidates. Our concern is to specify what else is needed for mentality. Specifically, the suggestion contained in Figure 1 is that mental processes are implemented by intentional processes.



There is much to quibble about the model in Figure 1. One might object to the suggestion that all mental processes are implemented by intentional processes. The objection may be correct, but I wish to see how far we can get by ignoring it. Even so, the model is crude in at least two respects. First, a successful strategy for naturalizing mental phenomena is likely to involve more than two theories of implementation spanning three layers. Secondly, the model is probably too linear--that is, it may be necessary to naturalize different features of mentality separately and then show how mental processes are instantiated by combining several different kinds of process. (A more sophisticated model would therefore appear as a graph containing at least one pathway--through various implementation theories--from every higher level node down to nodes representing mechanical processes.) Nonetheless, the crude model provides a useful starting point for discussion.

Dretske (1981, 1986, 1988) provides the most explicit example of the two level strategy in Figure 1. The lower-level implementation theory is developed in Dretske (1981) from the mathematical theory of information. Turning his attention to belief, Dretske says (p.175): "The first order of business is to understand how higher-order intentional structures can be manufactured out of lower-order intentional states." This is at the level of implementation theory 2, and Dretske's account of mental states is given in terms of their etiology--causation by simpler indicator states--and their executive role in behavior. Particularly important to his theory at this level is the role that learning plays in fixing the content of an internal state and in establishing and maintaining causal connections between internal states and behavior. He also writes (1981, p. 171): "We have provisionally used the concept of belief to distinguish genuine cognitive systems from mere processors of information." Note the coupling of belief and cognition in this quote, thus conflating mentality and cognition. In a later work, Dretske (1986) returns to the question of

distinguishing beliefs from the kind of "primitive representational capacities" (p.32) exemplified by bacterial magnetosomes. Focusing on the idea that beliefs can be false, Dretske explicitly attempts to show how the "clear and unambiguous capacity for misrepresentation" (p.32) can be implemented by simpler intentional processes. According to Dretske, an internal representation R can be said to misrepresent the occurrence of an external condition F only if, as shown in Figure 2, there is more than one pathway from F to R (through different proximal stimuli  $s_1$  and  $s_2$  and internal responses to those stimuli  $I_1$  and  $I_2$ ) and (not shown in Figure 2) the system is capable of associative learning with respect to stimuli that typically (but not necessarily) occur in the presence of F. Given such a set up, it is possible for a particular stimulus to trigger R even when F is not present, thus (omitting some details) R can misrepresent the presence of F.

Place Figure 2 here

Let us grant that this provides the robust notion of misrepresentation that Dretske desires, and that the capacity for misrepresentation is a necessary condition for an organism to have beliefs. Nonetheless (and I expect that Dretske would agree) it is not yet sufficient to lay to rest worries about belief attributions to nonhuman organisms. Associative learning is a very widespread phenomenon; its neural basis has, for example, been studied in the marine snail *Aplysia californica*. If we could say nothing more about beliefs than that they involved the capacity for misrepresentation, then we might be faced with having to bite the bullet and attribute beliefs to *Aplysia*. However, the number of scientists or philosophers willing to do this is probably vanishingly small, and although this is not an issue to be settled by taking a poll, I do think that the intuition that *Aplysia* do not have beliefs is sound.

The distinction suggested earlier between cognition and mentality suggests a slightly more complicated model for naturalizing the mind than that shown in Figure 1.

Place Figure 3 here

In terms of this model, I would argue that Dretske's account of misrepresentation belongs to implementation theory 2a for two reasons. First, as I made explicit at the beginning, I am interested in those uses of mentalistic language that apply only when state consciousness is involved, but associative learning in the kind of system shown in Figure 2 can plausibly occur unconsciously. Second, misrepresentation in Dretske's robust sense appears at the "merely" cognitive or subdoxastic level. For example, the familiar Müller-Lyer illusion (Figure 4) involves misrepresentation of the relative lengths of two lines. Susceptibility to this illusion seems to depend on experience with rectangular three-dimensional objects. Visual patterns of the form <--> are associated with protruding, hence closer and smaller edges between two surfaces, and patterns of the form >--< are associated with receding, hence more distant, larger edges.

Place Figure 4 here

Illusions such as this are extremely interesting for understanding a difference between mere cognition and full blown belief. Our visual systems are such that, given typical exposure to buildings and other rectangular objects, we cannot help but see one line as longer than the other when presented with the Müller-Lyer figures. There is no choice in this matter. But at the level of conscious understanding we are capable of representing that the lines are really the same length even though they do not appear that way. Put more generally, we are capable of recognizing that our visual systems have misrepresented the world.

I want to use this idea of a capacity for detecting one's own misrepresentations to flesh out the notion of consciousness of content. There are two separate projects here: one is to argue that the functional capacity to detect misrepresentation is necessary (although perhaps still not sufficient) for consciousness of content; the other is to provide an account of how such a capacity

might be implemented. I shall concentrate on the first in the hope that a reasonable functional characterization is an important first step towards an implementational theory.

Organisms act in ways that are appropriate to the contents of their representations. Attributions of misrepresentation are explanatorily useful, because they can help to explain why an organism does something that is not appropriate given the actual state of affairs. Consider a worker ant confronted with a dead nestmate already beginning to decompose. Responding to oleic acid, a chemical byproduct of decomposition, she goes into an internal state R that causes the invocation of a remove-from-nest routine. Experiments show that ants will engage in the remove-from-nest routine even when the oleic acid is emanating from something other than a dead ant, such as a piece of paper, or even a moving live ant. I do not wish to argue against the view that, as described so far, R is a meaningless intervening variable because this behavior is amenable to a strict stimulus-response explanation according to which the stimulus of oleic acid invariably produces the remove-from-nest response. But suppose that there are organisms similar to the ants for which there are other pathways to R that do not involve the same proximal stimulus and that these organisms are capable of learning to associate other stimuli with dead nestmates. That is, these hypothetical organisms conform to the structure laid out by Dretske (Figure 2) and their tokens of R are capable of misrepresenting the external condition of the presence of a dead nestmate (F). For such an organism, it would be appropriate to explain her removal of a live nestmate or a piece of paper on the grounds that she had misrepresented these objects as dead nestmates.

If one identified the capacity for belief with the capacity for misrepresentation, one might loosely say that our imagined organism removes the corpse, paper, and live nestmate because she

believes them all to be dead nestmates. Alternatively one might say that she removes these objects because they appear to her to be dead. Someone might respond to this loose talk by asking "Which is it--does she believe those things to be dead, or do they just appear dead to her?" This question presupposes a distinction between appearance and belief that makes sense only when applied to an organism that is capable of making a distinction between the way things appear and the way they are. As described so far, there is nothing to suggest that ants or our imagined ant-like organisms are capable of making such a distinction for they are entirely locked into responding as if R has true content. In contrast, consider our own responses to the Müller-Lyer illusion. We are not locked into responding to the deliverances of our perceptual systems as if they had true content, but we are capable of comparing the contents of those perceptual states to other representations of the same states of affairs. Thus we can believe that the lines are the same length even though they appear to have different lengths.

There are at least two reasons for thinking that such an ability to detect representational error is an ingredient of consciousness of content. First, misrepresentation occurs when the normal relationship between symbol and referent is absent. Thus, the ability to detect misrepresentation constitutes an ability to detect the failure of this normal relationship. This ties in rather closely with Searle's idea of awareness of the relation between symbol and referent. Second, if we take seriously the claim that this capacity involves drawing distinctions between appearance and reality, then we are required to take seriously that there is a way that things appear to the organism. (Note that this does not entail that the organism is capable of making *the* appearance reality distinction. For that, it is necessary to have the concepts of appearance and reality. The capacity under discussion here requires only the ability to detect specific cases of misrepresentation, not the ability to think about misrepresentation generally.) Although we might

be able to speak of the way things appear to an organism that cannot detect its own misrepresentations, the theoretical motivations for doing so are much weaker than when the organism itself makes the distinction. Appearance is a key component of consciousness, thus the present characterization helps to explicate a role for consciousness in mental life.

As I mentioned earlier, I am not going to delve into implementational details, which are likely to be complicated. However, it is possible to identify some necessary elements (in no way do I maintain that these elements are sufficient). In addition to the Dretskean capacity for misrepresentation, for an organism to be conscious of a specific content that F has occurred, it must have alternative ways of representing the occurrence of F and the non-occurrence of F. Consider the Müller-Lyer illusion once more. Detecting the illusion depends on at least having the visual representation that one line is longer than the other while simultaneously being capable of representing the relative length of the lines nonvisually (in humans, this is plausibly accomplished linguistically). If one was capable only of the visual representation, one would be bound to react as if the lines were of different length. One can detect the visual misrepresentation only because one is capable of representing simultaneously and nonvisually that the lines are of the same length.

The capacity for representing instances of misrepresentation is minimally a second-order capacity because a state that represents such a misrepresentation involves intentional notions in its content. It is important, however, to realize that this capacity does not entail either self-awareness or a conscious theory of mind. For just as an organism may have states with first-order intentional contents yet fail to be conscious of those contents, so too may an organism have states with second-order contents and fail to be conscious of those contents. Having a belief in virtue of

being conscious of the content of some intentional state is not tantamount to having a belief about a belief.

We are now at the last stage of my plan, which is to draw out some of the implications of these ideas for the attribution of mental states to animals. Of prime importance, I believe, is that the idea of sensitivity to representational error is an empirically tractable notion that can be assessed in many ways. Reactions such as surprise, embarrassment, and rapid learning (often involving just one or a few experiences with the conditions that caused the error) are all characteristics that might be shown by organisms who have epistemic access to their own errors. It is an empirical question to what extent various organisms can detect and respond to their own errors, and there is plenty of room for matters of degree. There are various promising areas to look. One such area is social play, which involves responding to behaviors that would elicit different responses in nonplay contexts (there is also an interesting developmental connection in human children between the concepts of pretense and appearance; see Flavell et al. 1987). Another area is social communication, particularly where signals are used deceptively or withheld in conditions where they would be appropriate (see Allen and Hauser 1993 for extensive discussion of an interesting case of signal-withholding by rhesus macaques).

The present approach helps to account for why some features that are often suggested as evidence of nonhuman mentality (or the lack thereof) are, indeed, relevant. Three such features are behavioral flexibility, the integration of information from multisensory sources, and language abilities. *Behavioral flexibility* is relevant to mentalistic attributions because it is connected to an organism's monitoring of its own performance. An organism that cannot detect when its states misrepresent its environment will be limited to adjusting its behavior only when the proximal causes of those states are removed. Merely cognitive systems are typically inflexible in this way;

for instance, the visual system is subject to illusions that no amount of ancillary of information can override as long as the stimulus is present. Whole organisms too can be similarly inflexible. There is much room for variation here, with different organisms, including humans, being relatively inflexible with respect to some stimuli, and more flexible with respect to others. *Multimodal integration*, the ability to access a common representation through different sensory pathways, is relevant because, as in Figure 2 above, having separate sensory pathways is a necessary condition for misrepresentation, which in turn is necessary for the capacity to detect misrepresentation. Furthermore, the ability to compare an integrated representation of the world against the representation provided by a single sensory modality would support the detection of perceptual error in an obvious way. *Language* is relevant because it provides a representational medium that would support the capacity for different internal states with the same contents. However, in the absence of linguistic abilities, it is still possible for an organism to have more than one way to represent a given state of affairs as just indicated with respect to multimodal representation.

The notion that mental states involve consciousness of content has a further consequence for our understanding of the utility of mentalistic explanations of animal behavior. Good explanations have counterfactual power, which is to say that they not only account for what did happen but they can account for what would have happened under different conditions. One charge commonly raised against mentalistic explanations applied to nonhuman animals is that they seem to lack this kind of power. If one takes a weak interpretation of mental state attributions, then the charge has some merit. The claim that ants remove dead nestmates because they believe them to be dead turns out to have very little explanatory power since ants are wholly insensitive to all sorts of manipulations that do not involve dead ants at all. On the other hand, in



a good mentalistic explanation, we learn a lot by being told what the contents of the relevant beliefs are. Being told that John went to the station because he believed his friend would be there allows us to understand all kinds of reactions that John would have if his friend did not show up--reactions that he would not have had if he had gone to the station for another reason. I think that it is no accident that Chisholm, in his classic discussion of Brentano and intentionality (Chisholm 1957), suggested that the most promising place to look for intentionality in nonhuman animals is expectation. There is, of course, a sense of expectation that is compatible with behavioral conditioning. An organism pressing a lever for a food reward may be said to expect the reward. However, this alone does not entail consciousness of the content of the expectation. An organism that is aware of the contents of its own expectations will react differently when those expectations are violated than an organism that is not aware, in which we may see nothing more than a slow extinguishing of the behavior.

I would like also to tie the present idea into two earlier suggestions that Marc Hauser and I have explored. First, in discussing concept attribution to nonhuman animals (Allen and Hauser 1991) we suggested that the capacity for distinguishing the appearance of death from death was criterial for possession of the concept of death. Given that concepts are constituents of the contents of beliefs and other mental states, it is not, therefore, surprising to find that this capacity should be criterial for mental state attributions. Second, we have attempted to provide a schema for the analysis of a notion of "strong information content" (Allen and Hauser 1993) that builds in the requirement for an organism to be able to represent misrepresentation. For reasons mentioned by Sterelny (1995) I am not entirely happy with our exact formulation of the analysis, but the basic idea, I am convinced, remains a good one: creatures with minds are sensitive not just to changes in the world, but are sensitive to how well they manage to represent those changes.

## **Acknowledgements**

Anderson Brown, Eric Sidel, Larry Shapiro, Kim Sterelny, and Ann Wolfe, all provided helpful written comments on an earlier version of this paper. I also thank Marc Bekoff and the members of an audience at the University of California, Irvine, for discussion. This research was conducted with the support of NSF fellowship SBR-9320214. I am grateful to the organizers and sponsors of the Third Pittsburgh-Konstanz Colloquium in the Philosophy of Science for the opportunity to participate in an excellent conference, and I thank my commentator Gerald Massey and members of the Konstanz audience for a stimulating discussion.

## References

- Allen, C. (1995), "Intentionality: Natural and Artificial", in J.-A. Meyer and H. L. Roitblat, (eds.), *Comparative Approaches to Cognitive Science*. Cambridge, MA: MIT Press, pp. 93-107.
- Allen, C. and Bekoff, M. (1996), "Cognitive Ethology and the Intentionality of Animal Behavior", submitted.
- Allen, C. and Hauser, M. D. (1991), "Concept Attribution in Nonhuman Animals: Theoretical and Methodological Problems in Ascribing Complex Mental Processes", *Philosophy of Science* 58: 221-240.
- Allen, C. and Hauser, M. D. (1993), "Communication and Cognition: Is Information the Connection?", *PSA 1992*, vol. 2. East Lansing: Philosophy of Science Association, pp. 81-91.
- Bekoff, M. and Allen, C. (1996), "Cognitive Ethology: Slayers, Skeptics, and Proponents", in R. W. Mitchell, N. Thompson and L. Miles, (eds.) *Anthropomorphism, Anecdotes, and Animals: The Emperor's New Clothes?* Albany, NY: SUNY Press, forthcoming.
- Bennett, J. (1964/1989), *Rationality*. Indianapolis, IN: Hackett.
- Blumberg, M.S. and Wasserman, E.A. (1995), "Animal Mind and the Argument From Design", *American Psychologist* 50(5): 133-144.
- Burghardt, G. (1985), "Animal Awareness", *American Psychologist* 40: 905-919.
- Chisholm, R.M. (1957), *Perceiving: A Philosophical Study*. Ithaca, NY: Cornell University Press.
- Dennett, D. C. (1983), "Intentional Systems in Cognitive Ethology: the 'Panglossian Paradigm' Defended", *Behavioral and Brain Sciences* 6: 343-390.
- Dennett, D. C. (1991), *Consciousness Explained*. Boston, MA: Little, Brown, and Co.
- Dretske, F.I. (1981), *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.

- Dretske, F.I. (1986), "Misrepresentation", in R.J. Bogdan, (ed.), *Belief*. Oxford: Clarendon Press.
- Dretske, F.I. (1988), *Explaining Behavior*. Cambridge, MA: MIT Press
- Dyer, F.C. (1991), "Bees Acquire Route-Based Memories but Not Cognitive Maps in a Familiar Landscape", *Animal Behaviour* 41: 239-246.
- Flavell, J., Flavell, E, and Green, F. (1987), "Young Children's Knowledge about the Apparent-Real and Pretend-Real Distinctions", *Developmental Psychology* 23(6): 816-822.
- Fodor, J.A. (1986), "Why Paramecia Don't Have Mental Representations", in P. French, T. Uehling, Jr., and H. Wettstein, (eds.), *Studies in the Philosophy of Mind*, Midwest Studies in Philosophy, vol. 10. Minneapolis: University of Minnesota Press, pp. 3-23.
- Fodor, J.A. (1990), *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.
- Gould, J. L. (1986), "The Locale Map of Honey Bees: Do Insects Have Cognitive Maps?", *Science* 232: 860-863.
- Gould, J.L. and Gould, C.G. (1994), *Animal Mind*. New York, NY: Scientific American Library.
- Griffin, D. R. (1976), *The Question of Animal Awareness: Evolutionary Continuity of Mental Experience*. New York, The Rockefeller University Press.
- Griffin, D. R. (1984), *Animal Thinking*. Cambridge, MA: Harvard University Press.
- Griffin, D. R. (1992), *Animal Minds*. Chicago: University of Chicago Press.
- Lycan, W.G. (1995), "Folk Psychology and Its Discontents", this volume.
- McCarthy, J. (1980), "Beliefs, machines, and theories", *Behavioral and Brain Sciences* 3: 435.
- McLaughlin, B.P. (ed.) (1991), *Dretske and His Critics*. Oxford: Blackwell.
- Millikan, R. G. (1984), *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.

Millikan, R. G. (1993), *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Nagel, T. (1994), "What is it Like to be a Bat?", *Philosophical Review* 83: 435-50.

Rosenthal, D.M. (1993), "State Consciousness and Transitive Consciousness", *Consciousness and cognition* 2: 355-363.

Searle, J.R. (1980), "Minds, brains, and programs", *Behavioral and Brain Sciences* 3: 417-424.

Searle, J.R. (1992), *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Sterelny, K. (1995), "Basic Minds", *Philosophical Perspectives*: forthcoming.

Stich, S.P. (1978), "Beliefs and Subdoxastic States", *Philosophy of Science* 45: 499-518.