Legend. New York: Basic Books.
Weber, M. 1968: *Economy and Society* (eds. G. Roth and C. Wittich). New York: Bedminster Press. (Original work published 1922.)
Wegman, C. 1985: *Psychoanalysis and Cognitive Psychology: A Formulation of Freud's Earliest Theory*. London: Academic Press.
Wilensky, R. 1978: *Understanding Goal Based Stories*. Research report 140. New Haven, CT: Yale University Department of Computer Science.

Mind & Language
Vol. 5 No. 1 Spring 1990
ISSN 0268–1064

Article

The Intentionality of Animal Action

CECILIA HEYES AND ANTHONY DICKINSON

In the preface to his *Psychosemantics* (1987) Fodor tells us that he has a strikingly intelligent cat that goes by the name of Graycat. By 'intelligent' what Fodor means is that Graycat has a mental life built on the rational interactions of intentional states, such as beliefs and desires, and which, Fodor believes, is manifest in Graycat's behaviour. He tells us that 'In the morning, at his usual feeding time, Graycat prowls the area of the kitchen near his feeding bowl. When breakfast appears, he positions himself near his food bowl in a manner that facilitates ingestion' (p. ix). From this and other similar observations Fodor concludes that 'The reason, for example, that Graycat patrols his food bowl in the morning is that he wants food and believes—has come to believe on the basis of earlier feedings, that his food bowl is the place to find it' (p. x).

Although we suspect that Fodor himself would not claim that simple observations, such as that of Graycat's feeding activities, are sufficient to establish the cognitive status of an action, contemporary 'cognitive' ethologists have attributed intentional states to animals on the basis of passive observation of their behaviour under free-living conditions (e.g. Griffin 1984; Whiten and Byrne 1987). We shall argue in this paper that such observation, however careful, can be misleading; that intentionality is not directly 'manifest' in behaviour and that the attribution of intentionality to animals should be based upon specific behavioural criteria that cannot be applied through passive behavioural observation in an uncontrolled environment.

Thus, our purpose in this paper is to specify the behavioural criteria that have to be met if an action is to warrant an intentional account. To do so, we must first provide an analysis of what is involved in claiming that an action is intentional. We have not tried to justify this analysis, to make it comprehensive or even innovative. Instead we have simply developed the account to the point where it provides an adequate introduc-

tion to the second section of the paper in which we discuss the application of the criteria. Finally, we will consider several potential objections to our approach to the study of intentionality in animal action, focussing on the extent to which it is consistent with our understanding of both human intentionality and evolution.

The Intentional Account of Action

In developing this account we have adopted a blatantly 'realist', rather than 'instrumentalist' view of intentionality (Bechtel 1985). For the present purposes this means that we regard an intentional account of action as a variety of causal explanation, and in order to test most causal explanations, including intentional ones, it is necessary to translate them, implicitly or explicitly into counterfactual claims. Thus, in the case of an animal action, an intentional account must be translated into claims about what the animal would have done if its circumstances had been different in certain, specifiable respects from those in which the action actually occurred (e.g. Millikan 1984; Lockery 1989). The central problem with attributing intentionality on the basis of naturalistic observation alone is that such observation seldom provides the opportunity to evaluate these claims. This is a direct and inevitable consequence of the fact that in the field, rather than the laboratory, it is rarely possible to observe an action under more than one set of circumstances, let alone under circumstances that are known to vary in a systematic way.

In order to specify the nature of the counterfactual claims that we take to be involved in an intentional explanation of animal action, we must first outline the general character of an intentional theory of action. Consider, once again, hungry Graycat's approach to his food bowl. An intentional account of Graycat's act has three main components:

(1) *Instrumental Beliefs.* Graycat must believe that approaching the bowl causes access to food. For expository purposes we shall represent the content of mental states in terms of PROLOG, a programming language designed, at least in part, to simulate cognitive processes. The assumption then is that Graycat has a belief with the content *cause (approach,access(food))*..

(2) *Desires.* Graycat must have a desire, and since he is hungry we can assume that he wants access to food, a desire the content of which can be represented as *access(food)*..

(3) *Practical Inference Process.* Finally, an intentional account has to specify how the instrumental belief and the desire interact to produce the action. We shall assume that the presence of a mental state, an intention to perform an action A, is sufficient in the present context for the execution of the action. We shall represent the content of this intention in PROLOG in the form of the command, *perform(A)*.. Thus, two mental states, a belief and a desire, are assumed to interact by a process that conforms to a rule of practical inference to produce a third mental state, an intention. The rule describing the practical inference process can be represented in PROLOG by

perform(A):-cause(A,access(O)),access(O)..

By this process of practical inference, an animal that had the belief *cause(approach,access(food))*. and the desire *access(food)*. would also have the intention *perform(approach)*..

There are two features of this explanation that are noteworthy. The first is that it is causal in the sense that it is the interaction of the belief and desire in the practical inference process that determines the content of the intention. Without the appropriate belief and desire, the animal would not have the relevant intention. Secondly, the explanation portrays the action as rational. Thus, the practical inference rule is such that if the belief *cause(A,access(O))*. is true, and the intention *perform(A)*. is executed (i.e action A is performed), then, other things being equal, the desire *access(O)*. must be fulfilled. Of course, there is nothing about PROLOG that would prevent us from specifying a rule of the form

perform(R):-cause(A,access(O)),access(O).

where R is an action unrelated to A and without consequences with respect to O. Such a rule could not, however, be part of intentional account of action because the performance of R would be in no sense rational with respect to the belief and desire that caused its execution. It is this 'rationality assumption' (Dennett 1987, p.185) that gives intentional accounts of animal behaviour explanatory power by making them subject to empirical evaluation. If it were assumed that intentional states participate in non-rational interactions, it would be impossible to predict the outcomes of their interactions and therefore impossible to test those predictions through observation.

Of course, this intentional account is rudimentary in the extreme and represents little more than a starting point for a proper cognitive theory of action. For instance, it takes no account of the quantification of beliefs and desires, nor does it specify how competing intentions are resolved in action. Even so, this primitive theory is sufficient for the present purpose, namely that of specifying behavioural criteria for the attribution of intentionality.

l Criteria

vo counterfactual claims supported by this intentional account, t the target action would not have occurred in the absence of ppriate belief and 2) the appropriate desire. These claims imply rresponding behavioural criteria, the belief and desire criteria, met in order to justify an intentional attribution.

Criterion

to the intentional account, Graycat's approach to his food bowl y a belief *cause(approach,access(food))*.. If this is so, the action persist in a world where approach no longer gives access to owl. Suppose we placed Graycat in the world faced by Alice went through the looking glass; a world in which goals recede walk towards them, but draw nigh when you attempt to retreat The contingencies of such a world would not support a belief make the approach of a hungry animal to its food bowl rational. tly, if Graycat's action is intentional he should, like Alice, adapt vironment by, at least, removing *cause(approach,access(food))*. rpus of beliefs. Thus, we would have grounds for asserting that pproach behaviour in the normal environment is mediated by iate causal belief if he adapts to the looking glass world by no mpting to walk towards his food bowl. This would show that condition for approach is that the environment provide contin- t would support the belief required by an intentional account. e, we cannot know what Graycat would have done in a looking l, but there is evidence to suggest that he might have persisted ing to walk towards his bowl. Hershberger 1986 arranged a ss environment for some chicks: their food bowl receded from vice the rate they walked towards it, and approached them at rate they retreated from it. In spite of the fact that they could e gained access to food by walking away from the bowl, the sisted in chasing the bowl away. After 100 trials the chicks in gaining access to the food bowl only 30% of the time. ction, which appears to be intentional on the basis of passive n, looses this property when there is a simple reversal of the ental contingencies.

be argued that Hershberger's study indicates, not that the chicks belief about the relationship between approach and access to that this belief was highly resistant to change. There is some is argument because outside Hershberger's apparatus the chicks continually contacted contingencies supporting conventional ut the consequences of walking towards objects. This argument, could not account for maladaptive approach when the behaviour nally acquired within the experimental situation itself. It is a

common procedure in the animal psychologist's laboratory to study signal learning by measuring approach during a signal for food. In this procedure hungry rats are presented with some food pellets every few minutes, and each presentation is announced by a signal, say a tone, that comes on a few seconds before the food is delivered into a bowl. Across successive pairings of the signal and food presentations, the rats learn to approach the bowl during the signal so that they are adjacent to it at the time of presentation. An intentional account would argue that the development of this behaviour reflects the acquisition of a belief *cause(approach-tone, access(food))*., namely the belief supported by the tone-food contingency to the effect that approaching the bowl during the tone causes access to food. However, it is easy to show by an implementation of the looking glass world that this account cannot be sustained. If we simply arrange that from the outset of training the presentation of the food is omitted whenever the rats approach the bowl during the tone, the animals can never experience a contingency that would support the appropriate causal belief for approach during the tone. Despite the lack of this experience, the animals do acquire approach under this 'omission schedule'. Holland 1979 found that rats lose a significant proportion of the rewards available to them by acquiring and persisting in approach during the tone, a behavioural pattern completely at variance with the rationality posited by an intentional explanation of approach behaviour.

We know of no evidence indicating how Graycat and his feline fellows would cope with an omission contingency but, if they are anything like rats and chicks, it is unlikely that their approach behaviour would support the intentional explanation that Fodor, and many other cat lovers, would like to give it. However, just because things look bleak for an intentional account of Graycat's approach behaviour, it does not mean that *none* of his actions are intentional. While the rat's approach responses appear to be relatively insensitive to their causal consequences (and therefore appear to be non-intentional), there is no doubt that rats are capable of at least one action that *is* sensitive to its consequences and therefore meets the belief criterion. If rats are simply trained to press a lever for food, they will refrain from pressing when the schedule is subsequently changed so that the food is delivered independently of whether or not they press (e.g. Rescorla and Skucy 1969).

This contrast, between approach and lever pressing behaviour, underlines the fact that intentionality is primarily a property of an agent with respect to a particular action rather than of the agent *per se*. Of course, there may be species of animal that are incapable of any intentional actions, but it is equally clear that the capacity for one kind of intentional action does not guarantee that all, or indeed any, of the animals' other actions are intentional. This is apparent when one considers the human case. Few people doubt that humans are capable of intentional action, but even we have difficulty in adapting to certain omission schedules. This is a claim, however, that we can only substantiate by anecdote. Consider the case of

a friend of ours, Tony, who became amorously involved with a young lady who was particularly susceptible to 'macho' displays. Unfortunately, Tony's disposition was to display affection and courtesy when in love. His girlfriend rapidly became bored by this lack of the machismo spirit and rejected his subsequent advances, a rejection that so hurt and angered Tony that the next time they met by chance he was cool and offhand. This display was, of course, irresistable to the lady in question, who instantly revived the relationship, thereby initiated the whole cycle over again. Tony went through a number of such cycles with her before he realised that his affection was in fact under an omission contingency; the display of affection caused the loss of the desired goal. However, this realisation did not help Tony to cope with the schedule because his affectionate behaviour was not under rational, intentional control. The only solution was to break out of the schedule altogether, thereby ensuring that he was never again exposed to this powerfully attractive stimulus.

In summary, the first criterion that must be met if an action is to warrant an intentional account is a belief criterion. The behaviour must be sensitive to whether or not the environmental contingencies will support a belief with the appropriate causal content. Other things being equal, if an action is acquired under contingencies that would support a contradictory belief, this action does not warrant intentional characterisation.

The Desire Criterion

Although our discussion of the belief criterion has left us with doubts about the intentionality of Graycat's approach behaviour, we have at least one animal action that has been shown to be sensitive to the relevant instrumental contingencies, namely lever pressing for food by hungry rats. However, even this action has a further test to pass before we can conclude that it is intentional. The intentional account assumes that an action is mediated not only by an instrumental belief but also by a desire; in this case, a desire with the content *access(food)*.. Consequently, if the desire is eliminated or, at least, diminished, then the performance of the action should, other things being equal, decline.

At first sight, it appears to be easy to devise a behavioural criterion corresponding to this counterfactual. Surely all we have to do is to reduce the desire for food by simply ensuring that our rat is well fed before we test him? Under these circumstances no one would be surprised to find that the animal has little inclination to press the lever, but a moment's reflection shows that the desire criterion cannot be met so easily. The problem with this test is that it does not ensure that 'other things are equal'. A hungry rat is likely to differ from a satiated rat in a number of respects other than in his desire for food. For example, a satiated rat may be more inclined to sleep, and its slumbers, rather than its weaker desire for food, may result in relatively little lever pressing.

Animal psychologists have thought long and hard about how to demonstrate that a change in the desirability of a goal or an incentive can affect instrumental action, although they have seldom expressed the problem in these terms. One of the procedures they have come up with goes by the cumbersome name of the 'irrelevant incentive test'. First, consider two groups of hungry rats all of which are trained to press a lever and pull a chain concurrently for two foods rewards, dry food pellets and sugar solution. The only difference between the two groups is that lever pressing is rewarded with the pellets in one group but with the sucrose solution in the other. In both groups chain pulling is trained with the other incentive. All the animals are then sated for food and their propensity to press the lever is assessed when they are thirsty, and in the absence of any rewards. According to an intentional account of this action, the group trained with the sugar solution for lever pressing should press more than the one trained with the food pellets. The shift from hunger to thirst should reduce the desirability of the dry food pellets while maintaining that of the sugar solution, an incentive for which rats will work when thirsty. This is precisely the outcome observed under certain circumstances (see below) by Dickinson and Dawson 1988, 1989.

The significant feature of this irrelevant incentive effect is that it demonstrates an apparent effect of changing the desirability of a goal under conditions in which other, non-intentional factors appear to be equated. Both groups are in the same motivational state at the time of test, and both received comparable training under hunger. The latter is demonstrated by the fact that if the animals are tested in the absence of any rewards while hungry, rather than thirsty, they press at similar rates (Dickinson and Dawson 1987a). Thus, the difference observed during the test under thirst must be due to the animals being in that motivational state.

This irrelevant incentive design also ensures that the effect of the motivational manipulation intended to change the relative strengths of the desires for the two incentives is mediated by the instrumental contingency between lever pressing and the incentives. This requirement is implicit in the claim that an intention to act is the product of the interaction of a desire and a belief about this contingency in the practical inference process. Two features of the design meet this requirement. If we had presented the sugar solution and the dry food pellets on test, when the animals were thirsty, any variations in performance could be explained by differences in the immediate impact of the two incentives. By testing in the absence of any rewards, we ensured that any difference in the performance of the two groups was due to the 'information' acquired during training, namely the instrumental belief about the consequences of lever pressing. Secondly, the concurrent training procedure in which both groups received the same exposure to the two incentives, but contingent upon different actions, ensured that the only difference between them was the content of the instrumental beliefs supported their training schedules.

In conclusion, the rat's lever pressing behaviour can fulfil the desire criterion; the animal's actions adjust appropriately to manipulations that

should change its desires in a way that depends upon the training schedule supporting the appropriate instrumental beliefs. We have discussed the application of the belief and desire criteria in some detail because we feel that it is important to appreciate that an animal action warrants an intentional explanation only if it fulfils certain quite specific requirements. It is somewhat ironic that the only example of an animal action that has, as far as we know, met both criteria is the behaviourist's prototypical example of a non-intentional, stimulus-response habit—rats' lever pressing in a Skinner box.

Potential Objections

Our analysis suggests that in order to find out whether any given example of animal action is intentional it is essential to measure the effects on that action of changes in the animal's environment which could be expected to alter the content of the animal's mental states. Many behaviours, that appear to be intentional on the basis of simple observation, fail to change in appropriate ways under the influence of new environmental contingencies, and therefore, if our analysis is correct, naturalistic observations of behaviour provide no reliable information about the intentionality of animal action. But is our analysis correct? We will now consider briefly several potential objections to our approach, each of which alleges either that we have not taken sufficient account of what is known about human intentionality, or that we have disregarded the facts of evolutionary biology. Thus, we will attempt to defend ourselves against allegations of both 'species-ism' and anthropomorphism.

Manipulating Mental States

First, it could be argued that inherent in our analysis is an anthropomorphic bias against recognition of intentionality in animals; that it will tend to yield 'false negative' conclusions because it presupposes that, if an animal has any beliefs and desires, then a scientist can reliably identify environmental contingencies and manipulate motivational states that will affect their content. Our short response to this allegation must be 'Guilty, m'lud'. Our approach does require the identification of conditions that will alter the content of mental states, and we recognise that there are major obstacles to such identification. We shall argue, however, that the difficulties associated with the manipulation of desires are greater than those encountered when trying to implement the belief criterion.

In the case of beliefs, the contentious issue is conceptual rather than empirical. In withholding an attribution of intentionality unless an animal adapts to a reversal of the instrumental contingency, we are demanding evidence not just that the animal has beliefs, but that it has veridical or true beliefs. It could be argued that the very fact that the rat approaches

the food bowl during the tone is the strongest evidence we could have for a belief *cause(approach-tone,access(food))*.. According to this view, acquisition of this behaviour under an omission schedule merely shows that the content of the animal's belief does not veridically represent the contingencies in the world. For us, however, this argument undercuts the whole concept of representation. Unless a system appears to be capable of detecting the extent to which the contents of its mental states actually match or represent states of affairs in the world, and of adjusting the content to bring about a degree of correspondence, we should certainly be cautious about according the determinants of its behaviour representational, let alone intentional, status.

The requirement that a system should meet the belief criterion in order to qualify for intentional status is also consistent with the consensus view of certain test cases. For example, in effect it is the application of the belief criterion that leads most people to deny that simple homeostatic mechanisms, such as thermostats, are intentional systems. In the case of a thermostat we could measure performance against the belief criterion by rewiring its connections to the boiler so that the output that previously switched the boiler on, now switches it off. This rewiring would reverse the instrumental contingencies operating on a thermostat, just as an omission schedule reverses the instrumental contingencies operating on the rat's approach response. The fact that a simple thermostat would never adapt to such a reversal is consistent with the widespread intuition that a thermostat does not have a representation of the instrumental contingency between its output and the temperature of its environment.

It should be made clear that we are arguing that it is necessary in *practice*, as part of an empirical programme designed to assess the intentionality of animal action, to treat all beliefs as veridical. Of course, this is not entirely satisfactory because *two*, inseparable possibilities are implicit in the concept of representation: that representations can be true and that they can be false. There is no doubt that both animal and human beliefs about even simple instrumental contingencies are not always veridical. For instance, if we simply introduce a short delay between the performance of an action and its outcome, both animals, in terms of their performance (e.g. Williams 1976), and humans, in terms of their causal judgments (Shanks, Pearson and Dickinson 1989), fail to distinguish this relationship from a truly non-contingent schedule. However, the implications of these contingency judgement studies are rather different from those, described above, showing that animals can acquire a maladaptive behaviour on an omission schedule. In both cases the subjects could be said to have acted on the basis of a false belief, but the origins of the false belief are mysterious only in the case of the omission schedule. When, as in the contingency judgement studies, the presence of a contingency has no effect on behaviour, it suggests that the subject has failed to form certain true beliefs that its environment would, in fact, support. On the other hand, when the contingency *does* affect behaviour, but the behaviour is maladaptive, as in

the omission schedule experiments, it suggests that the subject has formed a belief that is inconsistent with any true belief that the environment would support. How could such a false belief come about? The onus is surely on those who would like to explain such maladaptive behaviour in intentional terms to provide a principled answer to this question.

The interpretation of failures to meet the desire criterion is far more problematic. What are we to make of an action if its performance does not adjust appropriately to a manipulation designed to change the mediating desire? Does this mean that the action is non-intentional, or that the manipulation has failed to change the animals' desires as anticipated? There seems to be no principled way of deciding this issue. Of course, one chooses ones manipulations in the light of what is known about the animal's physiology, ecology and behaviour, but that this is not necessarily a reliable guide can be illustrated by reconsidering the irrelevant incentive effect. It seems reasonable to suppose that shifting an animal's motivational state from hunger to thirst will lead to a greater desire for sugar solution, a hydrating fluid, than for dry food pellets. However, if that is all one does, then an irrelevant incentive effect will not be observed; the rats trained to lever press for the food pellets on the concurrent schedule will respond just as much under thirst as those rewarded with the sugar solution for lever pressing during training (Dickinson and Dawson 1987a). In order to observe an irrelevant incentive effect, Dickinson and Dawson 1988, 1989 found that the rats had to be given the opportunity to consume the sugar solution and food pellets under thirst at some point prior to the instrumental test. Rats appears to have to learn about the relative desirability of the two incentives under the motivational state of thirst. Only after this incentive learning will a shift in motivational state bring about an appropriate change in behaviour.

Although in retrospect it seems that we should have realised that incentive learning might play a role in controlling desires, there was no *a priori* basis for certainty. Rats could have been, and other species might be, 'pre-wired' to desire certain fluids when thirsty. The dependence of the irrelevant incentive effect upon incentive learning is simply an empirical discovery. The implication of this example is that we must remain agnostic about the intentional status of actions that meet the belief criterion but not the desire criterion because in such cases we may simply have failed to manipulate desires appropriately.

Direct Perception of Intentionality

In the foregoing section we have tried to indicate why we think that the belief and desire criteria are necessary, and how they should be used to avoid false negative bias. Now we turn our attention to a more specific, empirically based, potential objection to our approach. It could be argued that, regardless of the details of our analysis, it must be wrong because it yields a conclusion that can be at odds with the results of recent experiments on human perception. Many of the experiments in question use Johansson's 1973 'patch-light' technique. Subjects are exposed to the trajectories of lights attached to the joints of an actor, and on the basis of this information alone they can report accurately the content of certain of the actor's mental states. For example, Runeson and Frykholm 1983 have shown that when the actor is observed preparing to lift a box, the observer can detect what the actor expects the box to weigh, what the box actually weighs, and what the actor intends the observer to believe about the weight of the box. Our analysis suggests that intentionality is not necessarily manifest in behaviour, while experiments like these apparently indicate that intentionality can be 'directly perceived' even under impoverished stimulus conditions.

If intentionality is defined with reference to rational interactions among mental states, covert states with representational content, and if what is meant by 'direct perception' is that the observer cannot be misled, then experiments of this kind certainly do not show that intentionality can be directly perceived. The potential for illusory attributions of intentionality is apparent when one considers the experimental technique itself. While Johansson and his followers happen to have generated their patch-light displays through the performance of intentional actions, there is no reason why the same displays could not be generated artificially without destroying the phenomenological percept. Under these circumstances, the observed movements would be perceived to be intentional although they would have, at best, 'derived' intentionality; i.e. they would be the products of intentional action on the part of, for example, the person who programmed the visual display.

It could be argued that the perception of intention in an artificially generated display is not illusory, that derived intentionality is intentionality-in-good-standing, or that, while it *is* illusory, such non-veridical attributions do not present any challenge to the claim that intentionality is manifest in natural behaviour under normal viewing conditions. Implicit in the latter argument is an appeal to evolutionary considerations to support the idea that intentionality can be directly perceived. While we agree that it can be helpful to think about the detection of intentionality in evolutionary terms, we suspect that, if anything, an evolutionary perspective casts doubt on the view that *human* observers can directly perceive intentionality in *animal* behaviour. Adopting such as perspective, Runeson and Frykholm 1983 suggest that direct perception of intention is unlikely to be possible when i) it would not bestow any fitness advantage on the perceiver, because the perceptual apparatus necessary to exploit the information in the stimulus array will not have evolved, and ii) there is a reliable cost to the actor of having his intention directly perceived. In the latter case, mechanisms camouflaging the intention are likely to have evolved. Now, with respect to how many animal actions is it likely to be the case that direct perception of their intentionality i) *would* bestow adaptive advantage on a human observer, and ii) would *not* have any cost

to the animal observed? Our guess would be that the answer to this question is 'Precious few'. Consider, for example, species that are preyed upon by humans. Selection pressure might well favour perceptual mechanisms that would allow people to perceive their intentions, but this, in turn, would favour the evolution of camouflage processes on the part of the prey.

Of course by responding in this way we may have mistaken the implications of the direct perception research altogether. The claim inherent there may be, not that certain mental states are manifest in behaviour, but that what were once thought to be mental states are in fact properties of behaviour. Thus, the Gibsonians might well be in the process of redefining intentionality in such a way that it is *necessarily* observable. Tolman, the prophet of cognition in animal behaviour throughout the dark-ages of behaviourism, often seemed to be attempting the same kind of redefinition:

> We, the observers, watch the behavior of the rat, the cat, or the man, and note its character as getting to such and such by means of such and such a selected pattern of commerces-with. It is we, the independent neutral observers, who note these perfectly objective characters as imminent in the behavior and have happened to choose the terms *purpose* and *cognition* as generic terms for such characters. (Tolman 1932, p.13)

Any attempt to 'reduce' intentional states to behaviour is likely to meet strong opposition on many grounds (e.g. Montefiore 1989), and indeed there is good evidence that Tolman himself was not committed to a behaviourist interpretation (see Amundson 1986). For the present purposes, however, we need only note that a claim for the priority of manifest intentionality based upon a behaviourist interpretation of cognition simply misses the point of the present argument, an argument predicated upon a 'realist' view of mental states and their causal properties.

Intentionality in the Lab and in the Field

We have used an evolutionary argument in an attempt to show that our approach to the investigation of animal intentionality is not inconsistent with data on human perception. A cognitive ethologist might be surprised, if not indignant, to find us using an evolutionary argument because, they might claim, the fundamental weakness of our approach lies in its failure to consider the possibility that animals are less likely to provide evidence of intentionality in the laboratory than in the field, i.e. when they are under free-living conditions. Dennett 1983 makes this claim explicitly, and his view has been echoed by Whiten and Bryne 1987. Dennett believes that behaviour, like lever pressing, which is the product of hundreds of training trials is very unlikely to warrant intentional characterisation

because it will probably be explicable in terms of 'rival, conditioning hypotheses' (Dennett 1983, p.348). It is not clear to us why Dennett, and others, hold this view; why they think that the existence of a 'prolonged training history' is inconsistent with the attribution of intentionality. One possibility is that they assume that unlike, for example, S-R habits, beliefs are formed quickly, on the basis of minimal experience.

There may well be some truth in this assumption in certain specific cases. Adams 1982 investigated the effect of changing rats' desire for sucrose after training them to lever press for it. Instead of changing the desire by shifting motivational state, he established an aversion to the sucrose by associating its consumption with the induction of gastric illness. When their propensity to press the lever in the absence of any incentives was subsequently assessed, those that had received only 100 rewards pressed less than a control group who had not been averted to the sucrose, a finding that fulfils the desire criterion. By contrast, after 500 training rewards, lever pressing was unaffected by decreasing the desirability of the sucrose. Adam's 1982 finding should not lead us to assume, however, that overtraining automatically robs an action of its intentional status. In an experiment similar to that of Adam's, Colwill and Rescorla 1985 found that overtrained lever pressing fulfils the desire criterion if trained on a concurrent schedule with chain pulling. What it is about the training schedule that determines the intentionality of an instrumental action remains unclear (but see Dickinson 1989).

Even if it is true that the intentional status of an action is determined by its training history, it is not clear why this would make the field a better place to look for animal intentionality than the lab. Free-living animals may not have been trained by human hand, but their behaviour surely has a 'history'; quite possibly a long one. We can think of only one line of reasoning that would lead one to expect to find more evidence of animal intentionality in the field than in the lab: If intentionality is an attribute that has evolved, then it may only be present in animal species that have experienced certain selection pressures, and only then when the animals are under conditions resembling those that constituted the selection pressure. If this reasoning is correct, it could have turned out to be impossible to study intentionality in the lab., i.e. it may have been impossible to find a laboratory trained animal action that fulfilled stringent criteria for intentionality. The fact is, however, that at least one such action *has* been found.

Causal Judgment and Intentional Action

Finally, our analysis may be faulted on the grounds that it says nothing about the conscious experience of animals. There is no doubt that one reason for endorsing the 'realist' view of the intentional causation of action is that, at least sometimes, we are aware of the instrumental beliefs and desires implicated in this account. This immediately raises the question

Assessing the belief criterion requires that performance of the action is investigated in an environment that differs from the target context only in the content of the instrumental belief that it will support. If action is acquired in a context that will not support the appropriate belief, an intentional account is unwarranted. Secondly, it must be demonstrated that the performance of the action adjusts appropriately to manipulations designed to alter the desire for the outcome. Furthermore, the desire criterion requires that the adjustment is demonstrated to depend upon training contingencies that support the appropriate instrumental belief. Applying these two criteria has revealed that the simple instrumental act of lever pressing performed by rats in the laboratory can support an intentional account.

Moreover, experimental analysis suggests that a strict application of these criteria is necessary because animals are endowed with non-intentional processes that under simple observation, and even cursory analysis, appear to generate intentional action. We have already noted the operation of such a process in the case of goal approach. It is important to appreciate, however, the subtlety of such non-intentional processes, a point that can be illustrated by reconsidering the application of the desire criterion through the irrelevant incentive effect. To recap, Dickinson and Dawson 1988, 1989 were able to demonstrate that the rat's lever press meets this criterion once they discovered that changing the relative desirability of sucrose solution and food pellets under thirst depends upon giving animals prior experience with these incentives in this motivational state or, in other words, upon the opportunity for incentive learning. The importance of incentive learning was not immediately appreciated, however, because apparent intentional control was exhibited with a simpler training procedure. If one group of hungry rats is simply trained to lever press for the sucrose solution and another for the food pellets, the sucrose-trained animals will press more when tested thirsty (Dickinson and Dawson 1987a), even though they have had no prior opportunity for incentive learning.

At first sight this simple procedure looks like a perfectly good implementation of the desire criterion. This is not so, however; with the simple training procedure the two groups differ not only in the instrumental belief supported by the schedule, namely that lever pressing causes the delivery of sucrose solution in one case and food pellets in the other, but also in their exposure to the two incentives during training. The groups are exposed selectively to either the sucrose solution or the food pellets. As we have already noted, when the exposure is equated by using a concurrent training procedure in which all animals received both incentives but contingent upon different actions, the irrelevant incentive effect disappears in the absence of the opportunity for incentive learning. It turns out that on further behavioural analysis the effect observed with simple training, although apparently rational and goal-directed, is in fact mediated by a non-intentional process operating through the association of the contextual stimuli and the incentives (Dickinson and Dawson 1987b).

These examples clearly demonstrate that there is often a conflict between

the intentionality manifest in behaviour and outcome of the strict application of our criteria. Given this conflict, we argue for the precedence of the criteria because the argument linking these criteria to intentional states and processes has a stronger rationale than the claim that cognition should be manifest in behaviour. Indeed, once we accept that there are complex and subtle non-intentional processes, such as those mediating basic goal-approach and the adjustment to changes in motivational state, that can mimic true intentional control in many situations, we can understand why the propensity to perceive actions as intentional may have developed. Given that either there is nothing in the stimulus input *per se* to distinguish intentional from non-intentional behaviour or that such a discrimination yields little of consequence in most situations, it may well pay the perceiver to treat both classes of behaviour as intentional in predicting the subsequent course of events. Indeed, as Dennett 1987 has argued, the intentional 'stance' can be the best standpoint from which to predict the behaviour of complex, non-intentional systems.

Having waded through our arguments for these conclusions, the reader may well be left wondering why we have concentrated exclusively on simple instrumental acts at the expense of careful, controlled demonstrations of more impressive cognitive feats, such as the ability of Gillan, Premack and Woodruff's 1981 chimpanzee, Sarah, to solve analogical reasoning problems or Pepperberg's 1987 parrot, Alex, to match by attribute. We have no quarrel with a cognitive account of performance in these and other tasks for in each particular instance there are no obvious psychological explanations which do not appeal to intentional or representational states. But any intentional account of these higher 'cognitive' processes must in the end assume that they are expressed in behaviour through an instrumental act; such acts must be the final common pathway in any intentional account of behaviour. Thus, until we are convinced that animals are capable of simple, intentional acts and understand how intentional states and processes can control such actions, any cognitive account of higher functions will remain divorced from the behaviour that is their only form of expression in animals. It is this fact that gives the simple instrumental action, however prosaic it might seem, priority. We suspect that many ethologists and comparative psychologists of a cognitive persuasion believe that the intentional status of such actions is not in dispute. Our analysis demonstrates, however, that it is indeed an issue.

Finally, many readers may be puzzled about why we should be concerned about the intentional status of simple instrumental actions when it is to be supposed that there are many non-intentional explanations of how an animal's behaviour could fulfil the belief and desire criteria. But, as far as we know, there are no such adequate accounts (see Sutton and Barto 1981, for the best attempt). Whether or not it is possible to explain

the intentionality of instrumental action in terms of a psychological *mechanism* must at present remain an open question.

Department of Experimental Psychology,
University of Cambridge,
Downing Street,
Cambridge CB2 3EB,
UK

References

Adams, C. D 1982: Variations in the Sensitivity of Instrumental Responding to Reinforcer Devaluation. *Quarterly Journal of Experimental Psychology*, 34B, 77–98.

Amundson, R. 1986: The Unknown Epistemology of E. C. Tolman. *British Journal of Psychology*, 77, 525–31.

Bechtel, W. 1985: Realism, Instrumentalism, and the Intentional Stance. *Cognitive Science*, 9, 473–97.

Colwill, R. M. and Rescorla, R. A. 1985: Instrumental Responding Remains Sensitive to Reinforcer Devaluation after Extensive Training. *Journal of Experimental Psychology: Animal Behavior Processes*, 11, 520–36.

Dennett, D. C. 1983: Intentional Systems in Cognitive Ethology: The 'Panglossian Paradigm' Defended. *Behavioral and Brain Sciences*, 6, 343–90.

Dennett, D. C. 1987: *The Intentional Stance*. Cambridge, MA.: MIT 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, NJ, Lawrence Erlbaum Associates: 279–308.

Dickinson, A. and Charnock, D. J. 1985: Contingency Effects With Maintained Instrumental Reinforcement. *Quarterly Journal of Experimental Psychology*, 37B, 397–416.

Dickinson, A. and Dawson, G. R. 1987a: The Role of the Instrumental Contingency in the Motivational Control of Performance. *Quarterly Journal of Experimental Psychology*, 39B, 77–93.

Dickinson, A. and Dawson, G. R. 1987b: Pavlovian Processes in the Motivational Control of Instrumental Performance. *Quarterly Journal of Experimental Psychology*, 39B, 201–13.

Dickinson, A. and Dawson, G. R. 1988: Motivational Control of Instrumental Performance: The Role of Prior Experience of the Reinforcer. *Quarterly Journal of Experimental Psychology*, 40B, 113–34.

Dickinson, A. and Dawson, G. R. 1989: Incentive Learning and the Motivational Control of Instrumental Performance. *Quarterly Journal of Experimental Psychology*, 41B, 99–112.

Dickinson, A. and Shanks 1985: Animal Conditioning and Human Causality Judgment. In L.-G. Nilsson and T. Archer (eds.), *Perspectives on Learning*

and Memory. Hillsdale, NJ: Lawrence Erlbaum Associates, 167–96.
Fodor, J. A. 1987: *Psychosemantics*. Cambridge, MA.: MIT Press.
Gillan, D. J., Premack, D. and Woodruff, G. 1981: Reasoning in the Chimpanzee: I. Analogical Reasoning. *Journal of Experimental Psychology: Animal Behavior Processes*, 7, 1–17.
Griffin, D. R. 1984: *Animal Thinking*. Cambridge, MA.: Harvard University Press.
Hammond, L. J. 1980: The Effect of Contingency upon Appetitive Free-operant Behavior. *Journal of the Experimental Analysis of Behavior*, 34, 297–304.
Hershberger, W. A. 1986: An Approach Through the Looking-glass. *Animal Learning and Behavior*, 14, 443–51.
Holland, P. C. 1979: Differential Effects of Omission Contingencies on Various Components of Pavlovian Appetitive Conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 5, 178–93.
Johansson, G. 1973: Visual Perception of Biological Motion and a Model for its Analysis. *Perception and Psychophysics*, 14, 201–11.
Lockery, S. 1989: Representation, Functionalism and Simple Living Systems. In A. Montefiori and D. Noble (eds.), *Goals, No-goals and Own Goals*. London: Unwin Hyman, 117–58.
Millikan, R. 1984: *Language, Thought, and Other Biological Categories*. Cambridge, MA.: MIT Press.
Montefiore, A. 1989: Intentions and Causes. In A. Montefiori and D. Noble (eds.), *Goals, No-goals and Own Goals*. London: Unwin Hyman, 58–80.
Pepperberg, I. 1987: Acquisition of the Same/Different Concept by an African Grey Parrot (*Psittacus erithacus*): Learning With Respect to Categories of Color, Shape, and Material. *Animal Learning and Behavior*, 15, 423–32.
Runeson, S. and Frykholm, G. 1983: Kinematic Specification of Dynamics as an Informational Basis for Person-and-action Perception: Expectation, Gender Recognition, and Deceptive Intention. *Journal of Experimental Psychology: General*, 112, 585–615.
Shanks, D. R. 1987: Acquisition Functions in Contingency Judgment. *Learning and Motivation*, 18, 147–66.
Shanks, D. R. 1989: Selectional Processes in Causality Judgment. *Memory and Cognition*, 17, 27–34.
Shanks, D. R., Pearson, S. M. and Dickinson, A. 1989: Temporal Continguity and the Judgement of Causality by Human Subjects. *Quarterly Journal of Experimental Psychology*, 41B, 139–59.
Sutton, R. S. and Barto, A. G. 1981: An Adaptive Network that Constructs and Uses an Internal Model of its World. *Cognition and Brain Theory*, 4, 217–46.
Rescorla, R. A. and Skucy, J. C 1969: Effect of Response-independent Reinforcers During Extinction. *Journal of Comparative and Physiological Psychology*, 67, 381–9.
Tolman, E. C. 1932: *Purposive Behaviour in Animals and Man*. New York: Applet-on-Century-Crofts.
Whiten, A. and Byrne, R. W. 1987: Tactical Deception in Primates. *Behavioural and Brain Sciences*, 11, 231–73.
Williams, B. A. 1976: The Effects of Unsignalled Delayed Reinforcement. *Journal of the Experimental Analysis of Behavior*, 26, 441–9.