Anim. Behav., 1993, **46**, 177–188

Anecdotes, training, trapping and triangulating: do animals attribute mental states?

C. M. HEYES
Department of Psychology, University College London, Gower Street, London WC1E 6BT, U.K.

(Received 17 March 1992; initial acceptance 19 May 1992; final acceptance 8 December 1992; MS. number: 4081)

Abstract. Three methods (anecdote collection, conditional discrimination training and trapping) used to investigate the possibility that primates attribute mental states are evaluated with reference to recent studies employing these methods. No convincing evidence of the phenomenon is found, and it is argued that none of these current methods could provide such evidence. In each case, behaviour consistent with the attribution of a mental state to an interactant could also be the result of associative or inferential learning about observable properties of the interactant's appearance or behaviour. A fourth method of investigation, which triangulates on mental state attribution using conditional discrimination training followed by transfer tests, is recommended as having the potential to provide evidence of mental state attribution in animals. It is argued that progress in this field has been hampered by a lack of full recognition that animals may learn inferentially about observables, and engage in associative learning without explicit training; and by misconstrual of the relationship between 'behaviourism' and mental state attribution.

The question of whether animals 'attribute mental states' to other animals (Cheney & Seyfarth 1990a, b, 1992) is at the core of investigations of 'theory of mind' (Premack & Woodruff, 1978), 'tactical deception' (Woodruff & Premack 1979; Mitchell 1986; Whiten & Byrne 1988), 'Machiavellian intelligence' (Byrne & Whiten 1988), 'politics' (De Waal 1982), the 'social function of intellect' (Humphrey 1976), the 'Panglossian paradigm' (Dennett 1983), and 'mind reading' (Krebs & Dawkins 1984) in animals. These investigations tend to assume that primates have first-order mental, or intentional, states, such as beliefs about the world around them, and seek evidence that they also have certain higher-order mental states, namely beliefs about the beliefs of other animals. Research of this kind is commonly perceived as having been unreasonably stultified by psychological behaviourism, and as enjoying a period of genuine progress in the clean air of late 20th-century cognitivism (Goodall 1971, 1986; Menzel 1974; De Waal 1982; Dennett 1983; see Kummer et al. 1990 for a contrary view).

The renaissance has certainly led to adaptation and refinement of the methods used to assess mental state attribution, and penetrating discussion of its evolutionary implications. However, I argue that there is still no convincing evidence of mental state attribution in non-human animals, and that most current methods of investigation do not have the potential to provide such evidence. Four empirical methods, anecdote collection, conditional discrimination training, trapping and triangulation, will be distinguished and evaluated with reference to recent studies employing those methods. The first three are ineffective because, although they yield behaviour consistent with the attribution of a mental state to an interactant, that behaviour could also be the result of associative or inferential learning about observable properties of the interactant's appearance or behaviour. Confidence in these methods appears to arise from a lack of full recognition that animals may engage in associative learning without explicit training, and in inferential learning, or 'reasoning', about observables. A fourth method of investigation, which triangulates on mental state attribution using conditional discrimination training followed by transfer tests, will be recommended as having the potential to provide evidence of mental state attribution in animals.

COLLECTING ANECDOTES

It has recently been implied that, under certain conditions, a collection of anecdotes may be sufficient evidence of mental state attribution (Whiten &

0003–3472/93/070177 + 12 $08.00/0

Byrne 1988). In this context, 'anecdotes' refers to narratives describing spontaneous social interactions among free-living animals. For example: 'One of the female baboons [*Papio anubis*] at Gilgil grew particularly fond of meat, although the males do most of the hunting. A male, one who does not willingly share, caught an antelope. The female edged up to him and groomed him until he lolled back under her attentions. She then snatched the antelope carcass and ran' (Jolly 1985). This report raises the possibility that primates can 'distract with intimate behaviour' (Whiten & Byrne 1988, page 238), one of many categories of deceptive behaviour that may involve the agent forming beliefs about, or 'representing', the beliefs of a social interactant. Thus, the female baboon may have believed that, while being groomed, the male would be ignorant of her intentions; that he would believe something other than that his meat was in jeopardy.

Whiten & Byrne emphatically rejected the view that single anecdotes of this kind can be anything more than a starting point for more systematic research. However, in stressing the value of 'multiple records' they implied that a collection of anecdotes relating to the same category of behaviour will constitute reliable evidence of mental state attribution provided that (1) the anecdotes come from independent observers, and (2) each provides evidence that the act involved the agent representing the viewpoint and/or beliefs of others. The first requirement is designed to ensure that the type of act in question really took place. The testimony of a single observer has to be viewed with scepticism because he or she may have failed to note, or have forgotten, an important part of the incident.

Unlike the first, Whiten & Byrne's second requirement would be impossible for any anecdote to fulfil. An anecdote could not provide evidence that an act involved the agent representing the beliefs of others because it could not eliminate alternative, plausible interpretations of the act. Consider the example of the carcass-snatching female baboon. As Whiten & Byrne acknowledged, rather than being based on mental state attribution, her act may have been simply fortuitous. It may have been no more than coincidental that the female began grooming the male when he was holding the carcass, and made a grab for the carcass when he was lolling back. One way of evaluating this possibility, without conducting an experiemnt, would involve taking baseline measures and calculating probabilities. For example, one could measure the frequency with which female baboons groom males who are not in possession of a valuable resource, and thereby assess the probability that the female in the narrative groomed the male because he was holding a carcass. If the probability turned out to be low, this procedure would allow the 'fortuitous' explanation to be discounted. However, it is not clear that the evidence would then be 'anecdotal', unless one chose with scant justification to refer to the products of all observational studies as 'anecdotes'.

More intractable problems emerge when one considers another possibility that would also have to be ruled out before an anecdote could fulfil the second requirement and thereby provide evidence of mental state attribution. This is the possibility that the behaviour was acquired through associative learning. For example, the female baboon may have snatched the carcass when the male was lolling back, not by chance, and not because she attributed to him a mental state of ignorance, but because in the past similar acts had proved rewarding when executed in relation to supine individuals. That is, the female could have snatched food from conspecifics on many previous occasions, initially without regard to their posture, but if she got away with it when the victim was supine, and not when the victim was upright, she might have acquired an association between the stimulus Supine conspecific + Food, and the response Snatch food, such that exposure to the one would trigger the other.

Whiten & Byrne (1988) implied that field observations could distinguish behaviour that is the product of associative learning and mental state attribution when they recommended that anecdotes should be accompanied by information bearing on the question of whether the behaviour arose 'through a specific mechanism, for example, trial-and-error learning' (Whiten & Byrne 1988, page 244). However, they did not indicate what kind of observations would be relevant. Students of animal behaviour used to assume that trial-and-error learning occurs gradually, while the effects of inferential learning, or reasoning, suddenly become apparent in behaviour (e.g. Kohler 1925). If this were true, and if field notes could indicate reliably whether learning was gradual or abrupt (and if this was still counted as 'anecdotal' evidence), then it may be possible using the anecdotal method to assess whether an animal had acquired a behaviour through associative learning or some inferential process. But it is not true. The many reports of

one-trial food aversion learning in rats, *Rattus* spp., show that associative learning can be abrupt (e.g. Kaye et al. 1988), and evidence that animals acquire beliefs about the relationship between lever pressing and food in the course of instrumental training (Dickinson & Dawson 1988, 1989; Heyes & Dickinson 1990) implies that the consequences of inferential learning may only gradually become apparent in behaviour. It may be possible to distinguish associative from inferential learning using 'reinforcement revaluation' procedures (Heyes & Dickinson 1990), but it is unlikely that these could be implemented under field conditions, and if they could, the data would no longer be in any sense 'anecdotal'.

The questions of how and where associative and inferential learning may be distinguished do not warrant extended consideration here. Even if field data could show that a behaviour has been acquired through an inferential, or reasoning, process, it would have another, more important, hurdle to cross before it could be regarded as evidence of mental state attribution: it would have to show that the subject had been reasoning about the mental states of other animals, and not about their observable characteristics. In relation to studies of mental state attribution, the crucial question concerns what, not how, the animal has learned. What matters is that the female baboon may have been acting on the basis of learning about appearance or behaviour, not about the relationship between these observable properties of other animals and their mental states. In short, she may have snatched merely because the male was 'lolling back', not because she took lolling back to be indicative of a particular kind of mental state, i.e. ignorance.

No single anecdote could provide convincing evidence that an animal's behaviour was influenced, not by a social interactant's appearance or behaviour, but by a mental state attribution based on those stimuli, and strictly interpreted, Whiten & Byrne's second requirement suggests that a collection of anecdotes could do no better. That requirement implies that a collection of anecdotes will provide convincing evidence of mental state attribution only if each anecdote in the collection indicates that the act involved the agent representing the viewpoint and/or beliefs of others. However, their research method belies this conclusion, and suggests that the problem would be overcome if a number of anecdotes, each relating to the behaviour of a different individual, implicate a common mental state attribution. Such data would not, in fact, be persuasive since each animal could have been acting on the basis of observable cues, and this interpretation would be plausible even if the cues available to each animal were different.

Consider three animals observed, on separate occasions, snatching food that was previously available to a conspecific. The first grooms the conspecific and snatches when they are supine, the second presents and grabs when the male is sexually excited, and the third throws a missile and makes his move when the conspecific is giving chase. We humans might feel inclined to attribute the state of 'intending to deceive with intimate behaviour' to all three animals, but the potential to attract the same mental state attribution from us might be the sum of what the three have in common as regards mental state attribution. Even if we could be sure that none of them had simply been lucky, and that all of them had acquired the behaviour through some inferential process, it would be fully plausible that they had learned to snatch from supine, sexually excited, and departing individuals, respectively.

TRAINING: CONDITIONAL DISCRIMINATION

A second method of investigating mental state attribution was used by Woodruff & Premack (1979) in their seminal work on 'theory of mind' in chimpanzees, *Pan troglodytes*. It was construed by them as providing the animals with the opportunity to deceive human trainers in a laboratory setting. More generally, the method consists of subjecting animals to a conditional discrimination training procedure which, it is assumed, will be successful only if the animals are capable of attributing mental states to others. Although designed to overcome the interpretative problems encountered by an anecdotal method, it succeeds in surmounting only one of the two major problems. The discrimination training method can eliminate the possibility that behaviour that is apparently based on mental state attribution is, in fact, merely fortuitous, but it cannot discount the possibility that animals are reacting to observables, rather than mental states assumed by the animal to be correlated with those cues.

Woodruff & Premack tested their chimpanzees for both 'production' and 'comprehension' of

deceptive communications but, since the results of both tests are subject to the same range of interpretations, the production test alone will be considered. At the beginning of each trial in this component of Woodruff & Premack's study, a chimpanzee observed the baiting of one of several inaccessible containers. After baiting, a person playing one of two roles entered the room. When playing the 'competitive trainer' they were dressed in white and behaved in a brusque manner in relation to the chimpanzee, but when playing the 'cooperative trainer' they dressed in green and were friendly towards the chimpanzee. The trainer was allowed to search one container for food, and had been instructed to select the container that the chimpanzee appeared to be indicating through its orientation and/or gestures. If a cooperative trainer selected the correct container, the chimpanzee was given the food, and if he selected an incorrect container the chimpanzee ended the trial empty handed. If a competitive trainer selected the correct container, he ate the food himself and the chimpanzee got nothing, but if he selected an incorrect container, the chimpanzee was given the food. Four animals were trained in this way, and in each case after 120 trials the cooperative trainer showed greater choice accuracy than the competitive trainer, and measures of the chimpanzee's orientational responses showed that they were turning towards, and gazing and pointing at, the baited container more often in the presence of the cooperative than the competitive trainer.

The consistency of the chimpanzees' behaviour shows that they were behaving systematically in relation to the cooperative and competitive trainers. It was not by chance that their behaviour heightened the salience of the baited container more often on trials with the cooperative trainer than on trials with the competitive trainer. However, the question remains as to what, as far as the chimpanzees were concerned, was the difference between the two types of trial. Did they, as the role labels suggest, attribute different sets of mental states to the two types of trainer, or did they merely react to differences in their appearance and behaviour? Fleshing out these possibilities, if mental state attribution was involved, the chimpanzees' behaviour might have been a consequence of the following sort of reasoning: 'Both trainers believe the food to be hidden in the container that I indicate. If he finds the food, the one in green will give it to me because he likes me and wants to help me, but if the trainer in white finds the food he will eat it himself because he doesn't like me and wants to get ahead. Therefore, I will indicate the baited container for the chap in green, and an empty container for the one in white; I will deceive him'. This is the reasoning about mental states (RAM) interpretation. If, on the other hand, the chimpanzees were acting on the basis of reasoning about observables (RO), their thoughts might be rendered: 'Both trainers go to the container that I indicate. If he finds the food, the trainer in green will give it to me, but if the one in white finds the food I won't get any. Therefore, I will indicate the baited container in the presence of the trainer in green, and the empty container in the presence of the one in white.' Evidence that animals are capable of reasoning or inferential learning about observables has been provided by studies of both chimpanzees (Premack & Premack 1983) and rats (Dickinson 1988).

The principal problem with this discrimination training method is that it does not allow an RO interpretation to be discounted in favour of a RAM interpretation, because it confounds the two relevant variables. In order to show that behaviour that may reflect mental state attribution does not instead occur by chance, it seeks evidence that different behaviour will occur under conditions that seem likely to support different mental state attributions. However, since we assume that animals' mental state attributions, if any, will be based on their experience of the target's appearance and behaviour, conditions that might support different mental state attributions will also have different observable characteristics, and if these are consistent across trials, discriminative behaviour could be based on them alone.

Rather than attempting to support a RAM interpretation over a RO account of their chimpanzees' behaviour, Woodruff & Premack (1979) argued against a 'conditioning' explanation which also assumes that the animals were sensitive only to the observable differences between the cooperative and competitive trainers. According to the conditioning, or associative learning, account considered, the chimpanzees may have learned a number of stimulus–response associations. For example, they received food each time the following sequence of events occurred; container A baited, green trainer enters the room, chimpanzee orients to container A. Consequently, the chimpanzees may have experienced an 'urge' to orient towards A (the response) whenever they saw container A baited and

the green trainer enter (the compound stimulus). According to this account, the chimpanzees did not engage in any sort of reasoning, or even expect their behaviour to have any particular outcome.

Woodruff & Premack (1979) tried to discount this conditioning interpretation by, for example, pointing out that the chimpanzees learned to indicate the baited container in the presence of the cooperative trainer faster than they learned to indicate an unbaited container in the presence of the competitive trainer. This pattern of results is consistent with a ROM interpretation if one assumes that it is easier or more natural for chimpanzees to formulate honest rather than deceptive intentions, but, as the authors acknowledged, it can also be argued to be consistent with a conditioning interpretation. An associative learning process may, for instance, have been working with an innate disposition to approach rewarding stimuli in the case of the cooperative trainer, and against the same tendency in the case of the competitive trainer. This debate, about the process of learning, is far from trivial, but it is not necessary to examine the various arguments in detail because a resolution could not make Woodruff & Premack's study, or the discrimination training method more generally, provide convincing evidence of mental state attribution. Even if we could be sure that associative learning was not responsible for the chimpanzees' behaviour, we still would not know whether they were reasoning about mental states or observables.

TRAPPING

Several investigators (Dennett 1983; De Waal 1986; Whiten & Byrne 1988; Cheney & Seyfarth 1990a, b, 1992), while acknowledging the virtues of an experimental over a purely anecdotal approach, have erroneously assumed that the principal weakness of the discrimination training method lies in the fact that it provides 'prolonged training histories for the behaviourists to point to in developing their rival, conditioning hypotheses as putative explanations for the observed behaviour' (Dennett 1983, page 348). Consequently, they recommend the administration of experimental tests of mental state attribution to animals that have not been given any explicit training. Following Dennett (1983), I refer to this as the 'trapping' method because it uses a human device, an experiment, to try to identify, or capture, an ability which may have developed naturally. In effect, the trapping method seeks to reduce uncertainty about what an animal has learned (what it believes, represents or knows about other animals) by withholding explicit training.

Cheney & Seyfarth (1990a) studied Japanese and rhesus macaques, *Macaca fuscata* and *M. mulatta*, in one of the very few attempts to put the trapping method into practice. They conducted two experiments but, since both are subject to the same interpretive problems, I will describe only one of them. At the beginning of each trial, female Japanese macaques were allowed to see, either in the presence of their offspring or alone, that food or a predator was hidden in an enclosure (four conditions). The offspring was subsequently released into the enclosure and it was anticipated that, if the females attributed knowledge to a juvenile that had been present during the first stage of the procedure, and ignorance to a juvenile that had been absent, they would make more food or predator calls on juvenile-absent trials. In fact, the mothers called with the same, low frequency when the offspring had been present and when it had been absent at the beginning of the trial.

While acknowledging that, like all null results, the outcome of this experiment is difficult to interpret, Cheney & Seyfarth viewed it as evidence that monkeys, in contrast with chimpanzees, are incapable of attributing mental states. This tentative conclusion may have been valid if the design of the experiment had been such that a positive outcome would have supported the contrary view, i.e. that monkeys can attribute mental states. In fact, a positive outcome (more calling in the juvenile-absent condition) would not have supported this conclusion because it could have been interpreted in at least two other ways. First, the mothers could have varied their calling rate with differences in the behaviour of offspring during juvenile-present and juvenile-absent trials. In support of this interpretation, the results showed that when the hidden object was a predator, juveniles crouched and huddled near their mother more on juvenile-present than on juvenile-absent trials; and when food was hidden, they took longer to find it on juvenile-absent than on juvenile-present trials. Second, the mothers could have varied their calling rate according to whether they remembered their offspring to have been present or absent at the beginning of the trial. Maki (1979), among others, has shown that pigeons, *Columba livia*, can learn to peck one key if they remember being given food several minutes earlier at the beginning of the trial, and a different key if they remember not receiving food at that

juncture. In an analogous way, the female macaques may have learned, by association or through reasoning, that calling to a recently absent offspring affects their behaviour more, or with a more positive outcome, than calling to an offspring that has remained nearby. Thus, Cheney & Seyfarth's experiment would not have provided evidence of mental state attribution even if it had had a positive outcome, and therefore its negative outcome cannot be regarded as even suggestive evidence that monkeys are unable to attribute mental states. This experimental method, and the trapping method more generally, simply lacks the power to provide a positive or negative answer to the question of whether animals attribute mental states.

It is hardly surprising that the trapping method is unsuccessful when one views it as an attempt to reduce uncertainty about what an animal has learned by withholding explicit training. In the case of discrimination training of the kind used by Woodruff & Premack (1979), we cannot assume that the animals have learned about mental states because we know that in their relevant experience putative variations in the social interactant's mental state have been confounded with variations in their observable characteristics. In the case of spontaneous learning, we do not know whether such confounding has occurred in the course of the animals' relevant experience, but we cannot under these circumstances legitimately adhere to the adage 'What the eye doesn't see, the heart doesn't grieve over'. We cannot assume that no such confounding has taken place, simply because we have not taken the opportunity to observe it. Instead, we need evidence that the animals have not learned a straightforward discrimination between observables, and a trap of the kind used by Cheney & Seyfarth could not yield that evidence because the confound that exists in the test (e.g. between recent presence and knowledge, and recent absence and ignorance), could also have been present during learning.

TRIANGULATING

The 'triangulation' method consists of conditional discrimination training followed by transfer testing, and it has recently been used by Povinelli et al. (1990) to investigate mental state attribution in chimpanzees.

At the beginning of each trial in the first, discrimination training, stage of Povinelli et al.'s procedure, a chimpanzee was in a room with two trainers. One trainer, designated the 'Guesser', left the room, and the other, the 'Knower', baited one of four containers. The containers were screened such that the chimpanzee could see who had done the baiting, but not where the food had been placed. After baiting, the Guesser returned to the room, the screen was removed, and each trainer pointed directly at a container. The Knower pointed at the baited container, and the Guesser at one of the other three, chosen at random. The chimpanzee was allowed to search one container, and to keep the food if it was found. If the chimpanzee chose an incorrect container, the Knower displaced the correct cup and showed the contents to the chimpanzee.

Two of the four animals tested in this way quickly acquired a tendency to select the container indicated by the Knower more often than that indicated by the Guesser, but this does not necessarily mean that they attributed knowledge of the food's location to the Knower. The roles of Knower and Guesser were randomly assigned to the trainers at the beginning of each trial and, unlike those in Woodruff & Premack's (1979) study, the trainers did not alter their appearance or general behaviour between roles. However, the successful chimpanzees could simply have learned to choose the container indicated by the trainer who did the baiting, or who was in the room at the time of baiting. The second, transfer, stage of the procedure was designed to test these hypotheses against the possibility that the chimpanzees were attributing knowledge to the Knower.

In each of the 30 trials in the second stage, baiting was done by a third trainer in the presence of both the Knower and the Guesser, but during baiting the Guesser had a paper bag over his head. Povinelli et al. anticipated that if, in the first stage, the chimpanzees had simply learned to select the cup indicated by the trainer who had done the baiting, or who was present during baiting, then their choice accuracy would go down in this second stage. Now that both trainers were present throughout the trial, and neither did the baiting, the chimpanzees would be confused. In fact, the chimpanzees are reported to have shown excellent transfer; no decrement in choice accuracy between stages 1 and 2, training and transfer.

Povinelli et al. interpreted their findings as evidence that chimpanzees are capable of 'modelling the visual perspectives of others'. They interpreted the animals' choice behaviour as having been

guided, not by particular observable cues (active versus absent, or bag versus no bag, during baiting), but by something which the animals took these cues to indicate: whether or not the trainer had seen where the food was placed. In our, human theory of mind, the link between seeing and believing is so strong that, if this interpretation were fully supported, then I would regard the experiment as providing convincing evidence of mental state attribution in chimpanzees.

Unfortunately, two features of the study prevent it from supporting this interpretation. First, transfer was measured in a less than sensitive way, by comparing the relative frequencies of Knower, Guesser and Other choices for the last 50 trials of stage 1 with those for the 30 trials of stage 2. Consequently, there may have been a decrement in performance at the beginning of stage 2, indicative of re-learning, that was not detected. Second, Povinelli et al. did not indicate whether the chimpanzees had ever seen a person with a bag over their head before, or how they reacted to this sight in stage 2. It is very unlikely that the chimpanzees were accustomed to dealing with 'bagged' humans, but if they were, their smooth transfer performance might have been due to prior learning that such people are poor cues. On the other hand, the bagged trainer might have been a novel and rather alarming stimulus from which they averted their gaze. If so, this might also explain the chimpanzees' tendency to continue in stage 2 to select the container indicated by the Knower.

These objections to Povinelli et al.'s study as evidence of mental state attribution are quite different from those raised in relation to the other studies reviewed in this paper. Rather than suggesting that the method is fundamentally flawed, they reflect relatively trivial problems which could be overcome by (1) more careful data analysis, (2) ensuring that the chimpanzees had not had prior experience of people with bags over their heads, and (3) instructing the Knower to put a bag over his head in stage 2, just after baiting and before the two trainers begin to point. They should not obscure the fact that conditional discrimination training followed by transfer tests could provide a powerful method of investigating whether animals attribute mental states.

The power of the triangulation method lies in the fact that it requires animals to distinguish one mental state, X (e.g. knowing where the food is hidden), from another, Y (e.g. not knowing where the food is hidden), in two or more situations which differ in terms of the observable cues that might be correlated or confounded with X and Y. In the training situation, X is confounded with feature A (e.g. did the baiting) and Y with feature B (e.g. absence during baiting) of the social interactants' appearance or behaviour, but in the transfer test, X and Y are correlated with features C (e.g. no bag during baiting) and D (e.g. bag during baiting), respectively. If the animal's behaviour is unchanged despite this shift in observable stimuli, and if the most plausible account of the relationship between A and C on the one hand, and B and D on the other, construes them as indicators or manifestations of X and Y, respectively, then one has evidence of the attribution of mental states X and Y.

Thus, triangulation has the potential to overcome the problem of confounding or correlated cues, not primarily by virtue of the quality of a single test or measurement procedure, but by compounding tests, each of which is highly fallible but in a different way. As the term 'triangulation' implies, two carefully chosen performance measures are necessary and potentially sufficient to triangulate on a mental state attribution. However, given the fallibility of individual measures, it will often be advantageous to use more than two tests. For example, Povinelli et al. could have added a third stage to their experiment, in which both the Knower and Guesser were present while a third party baited a container, but with the Guesser's view obstructed by a screen similar to the one used to hide the containers from the chimpanzee. If the animals had continued under these conditions to select the container indicated by the Knower, then the interpretive problems associated with the stage 2 bag test would, in effect, have been overcome. There would be good reason to believe that the chimpanzees had attributed a mental state to the Knower, despite the fact that performance in each of the transfer tests could, when considered in isolation, be otherwise explained.

I refer to this method as 'triangulating' because the use of conditional discrimination training followed by transfer testing is one way of implementing an experimental design to which Campbell (e.g. 1953) assigned that term. This design, and the implementation here described as triangulation, are used widely in both psychology and biology. For example, in the study of animal cognition, conditional discrimination training followed by transfer testing has been used to find out whether

animals identify locations with respect to single cues or configurations of stimuli (e.g. Susuki et al. 1980; Morris 1981); whether they rehearse or use mnemonic strategies to remember events (Maki & Hegvik 1980; Cook et al. 1985); and to investigate the possibility that animals have concepts (Herrnstein et al. 1976; D'Amato & Van Sant 1988). In spite of its widespread use, and the long-standing recognition by both scientists and philosophers that the design has considerable value (e.g. Peirce 1936/1868; Levins 1966; Laudan 1971; Wimsatt 1981), the virtues of triangulation have been most clearly elucidated in Campbell's (e.g. 1953, 1963, 1969) discussions of social science methodology and evolutionary epistemology.

ILLUSORY TRIANGULATION

A further study by Povinelli et al. (1992) underlines the fact that an experiment's potential to triangulate on mental state attribution lies primarily in the logic of its design, rather than its procedural details. The apparatus and procedure used in this study were similar to those of the Knower–Guesser experiments described above, but the design was rather different. After simple, rather than conditional, discrimination training, the animals were given a novel task, rather than a transfer test. That is, their second problem could not be solved using the same cues and responses as the first. Four chimpanzees took part in the experiment, and each was paired with a human trainer. In the discrimination training phase, two of the chimpanzees were given a 'cue detection' problem, and two were given a 'cue provision' problem. On each trial in the cue detection task, one of four containers was baited within view of the trainer, but not of the chimpanzee. The trainer than pointed at the baited container, and if the chimpanzee selected that container by pulling one of four levers, both the chimpanzee and the trainer were given food. On each trial in the cue provision task, one of four containers was baited within view of the chimpanzee, but not of the trainer. If the chimpanzee then behaved such that the trainer could select the correct container using the chimpanzee's gestures or orientation, both parties were given food. When each chimpanzee had attained a high level of accuracy in its task, the problems were switched. In this second phase of the procedure, the chimpanzees that had been given the cue detection problem were confronted with the cue provision problem, and vice versa. In three of the four subjects, this switch did not result in a significant decline in choice accuracy.

The results of this experiment were taken to indicate that chimpanzees are capable of 'cognitive empathy' or 'role taking ... the ability to adopt the viewpoint of another individual' (Povinelli et al. 1992), but they are subject to a much simpler interpretation. The chimpanzees may have quickly achieved a high level of accuracy on the second task, not because the first had allowed them to imagine the situation from another's perspective, but because they had learned most of what they needed to know to solve the second problem during pretraining and outside the experimental situation. During pretraining, all of the chimpanzees had learned to pull the levers to obtain food, and Povinelli et al. (1992) reported that their subjects commonly exhibited and encountered pointing behaviour in their day-to-day laboratory lives. If the results had shown that each problem (cue detection and cue provision) was learned faster when it was presented second than when it was presented first, then there would be reason to believe that some feature of the first task had facilitated performance in the second. However, even in this case, further experiments, varying the requirements of the first task, would be necessary to find out which feature was enhancing second task performance, and it is not clear which manipulations, if any, could provide unambiguous evidence that the opportunity for mental state attribution was responsible. Thus, although this task-switching experiment was superficially similar to the earlier study by Povinelli et al. (1990), it lacked the potential to triangulate on mental state attribution.

The anecdotal method provides another example of illusory triangulation. In their attempt to revitalize anecdotal investigation of mental state attribution, Whiten & Byrne (1988) have emphasized the value of 'multiple records', and the triangulation method involves 'multiple measures'. Viewed in this way, the two appear very similar, and yet triangulation is, and Whiten & Byrne's method of anecdote collection is not, an effective means of finding out whether animals attribute mental states. Why?

Not all sets of records or measurements have the same virtues, and the value of any particular set depends on (1) the relationship between each measure (e.g. test procedure or anecdote) and the hypothetical object of measurement (e.g. a higher-order mental state), and (2) the relationship

between the measures themselves. More specifically, in order to provide a trustworthy 'reading', (1) each measure in the set must have a clear theoretical or empirically established relationship to the object of measurement, and (2) the various measures must be similar enough to detect the same object or phenomenon, but different enough to avoid being subject to the same errors. A physical example of triangulation illustrates these points (Crano 1981). It is possible to identify reliably the origin of a stationary radio signal using two or more mobile, directional antennae only if (1) each is close enough to the source to pick up the signal, and (2) the antennae are distributed in space such that their projections intersect.

In the case of anecdote collections, the measures are informal reports of the behaviour of free-living animals, and the measuring instruments are the people who make the observations and write the reports. The kind of collection favoured by Whiten & Byrne fails to provide evidence of mental state attribution on both of the counts listed above. The link between a measure and the thing-to-be-measured is weak because field observers are looking for and reporting behaviour that appears to be a product of reasoning. As I have argued above, it is not only extremely difficult to distinguish reasoning from other, non-inferential forms of learning in animals, but an animal is no more likely to reason about mental states than about observables. Consequently, a measure of reasoning (even a good one) is a poor indicator of mental state attribution; the antennae are too far from the source.

Turning to the second count, as envisaged by Whiten & Byrne (1988), the measures contributing to a collection of anecdotes are both too close together and too far apart to triangulate on a higher-order mental state; the antennae projections are either parallel, or fail to intersect because they cannot pick up the same signal. Anecdotal measures are too close together whenever the people who make the relevant observations share important assumptions about, for example, the cognitive capacities and social organization of primates, and they are too far apart whenever they are taken from different animals.

The importance of taking multiple measurements from each animal can be illustrated by analogy. Imagine that one needed to know whether a fruit bowl, offered to a group of diners one evening, had contained a kiwi fruit. Perhaps one of the diners was poisoned that night, and the enquiry is part of a murder investigation. None of them has a clear recollection, but one diner reports that there was a furry fruit in the bowl, another says that they saw something greenish, and a third remembers something ovoid in shape and approximately 5 cm in diameter. If they were all remembering qualities of the same object, then one could be fairly confident that the bowl contained a kiwi fruit, but if they were recalling features of different objects, it is at least as likely to have contained a peach, an apple and a plum. Similarly, if one animal reliably behaves towards one or several others as if they were ignorant of the location of food when they (1) had a bag over their head, (2) were out of the room, and (3) had their view obstructed by a screen, during food placement, then that animal provides promising evidence of mental state attribution, but three different animals observed under conditions 1, 2 and 3, respectively, do not. In the case of the trio of animals, we have no reason to doubt that each has learned independently to respond to the observable cues in conditions 1, 2 and 3. In contrast, provided that one or two of the situations are novel, the single animal may be extrapolating across them on the basis of a mental state attribution.

Thus, the anecdote collection or 'multiple records' approach gives the illusion of triangulation; it is presented as a means of identifying mental state attribution through convergent operations, but it does not in reality have the potential to do so. However, it is possible that the approach is not intended as a means of finding out whether animals can attribute mental states. Perhaps it assumes that animals, or at least some primates, can attribute mental states, and is designed to find out about one particular use of that capacity, deception. If so, it is jumping the gun. Triangulation provides a potential means of finding out whether animals are capable of mental state attribution, but there is no convincing evidence yet in hand.

The illusion of triangulation has also emerged in the more general literature on primate cognition. It has been argued that the results of studies of mirror-guided body inspection (commonly and misleadingly known as 'self-recognition') and imitation in primates provide convergent evidence that apes can, and monkeys cannot, attribute mental states (e.g. Gallup 1982; Cheney & Seyfarth 1990b, 1992; Povinelli et al. 1990; Whiten & Byrne 1991). The argument is not persuasive for two reasons (see Heyes, in press a, b for reviews of research on self-recognition and imitation similar to the present

analysis of mental state attribution). First, the studies in question do not provide unequivocal evidence that the members of any primate species are capable of either mirror-guided body inspection (Heyes, in press b) or imitation (Galef 1988, 1992; Visalberghi & Fragaszy 1990). Second, even if they showed that mirror-guided body inspection and imitation occur in apes, but not in monkeys, it could not be inferred that apes can, and monkeys cannot, attribute mental states, because mirror-guided body inspection and imitation do not require mental state attribution. Mirror-guided body inspection requires an animal to distinguish sensory inputs generated by the state and operations of its own body from other sensory inputs, and to detect contingencies, within the former category, between direct feedback and novel, displaced visual feedback (Heyes, in press b). It would not require the mirror-user to appreciate that its mirror image resembles another animal's view of its body. Similarly, in order to imitate, an animal must be capable of mentally representing the actions, but not the mental states, of a model (Heyes, in press a). Although it is not necessary, both mirror-guided body inspection and imitation could involve mental state attribution. However, by the same token, mental state attribution could be occurring whenever an animal navigates around, rather than collides with, an active conspecific.

In connection with the hypothesis that apes and other primates differ in their capacity for mental state attribution, it should be noted that Povinelli et al. have used their triangulation procedure (described above) to test rhesus monkeys (Povinelli et al. 1991) as well as chimpanzees (Povinelli et al. 1990). The outcome of this rare and admirable attempt to compare primate species within a common paradigm was, however, ambiguous. Unlike the chimpanzees, the monkeys did not learn during the conditional discrimination phase of the procedure to select the container indicated by the Knower rather than the Guesser (and therefore it was not possible to give them the transfer test). This was interpreted as evidence that chimpanzees can, and monkeys cannot, attribute mental states, but it may instead have been due to the way in which the Knower behaved after baiting a container in the discrimination training phase of the two experiments. In the chimpanzee experiment, the Knower 'stood up, assumed a neutral posture in a predetermined location behind the apparatus, and fixed his gaze on the front of the subject's cage' (Povinelli et al. 1990, page 205), but in the monkey experiment he 'moved to the door, tapped on it, and thus signalled the guesser to return' (Povinelli et al. 1991, page 320) and only then moved, with the Guesser, into the predetermined location. Consequently, the monkeys may have failed to acquire the discrimination because the activities of the Knower in their experiment made it difficult for them to remember which trainer had been the Knower and which had been the Guesser on that trial. The chimpanzees could have simply fixed their gaze on the trainer who did the baiting and kept it there until they were given the opportunity to choose a container, but a monkey using this strategy is likely to have been distracted in the course of the Knower's movements, some of which were co-incident with those of the Guesser.

BEHAVIOURISM AND MENTAL STATE ATTRIBUTION

The relationship between behaviourism and mental state attribution is a complex one and has been much misunderstood. First, it has been suggested that if animals do not attribute mental states, then they are 'behaviourists' (Premack & Woodruff 1979). Although witty and interesting, this analogy could, when converted into an 'intuition pump' (Dennett 1980), seriously mislead: 'The ape could only be a mentalist. Unless we are badly mistaken, he is not intelligent enough to be a behaviorist' (Premack & Woodruff 1978, page 526). We may well accept this view if we assume, implicitly, that an animal who does not attribute mental states must have mastered the contents of the *Journal of the Experimental Analysis of Behaviour*. Furthermore, if we are not ourselves behaviourists, we may be motivated by the analogy to judge animals capable of attributing mental states because that would seem to deprive scientists who are behaviourists of the moral high ground of parsimony. However, once the analogy's spell is broken, it becomes apparent that we have no way of comparing how much 'intelligence' it takes to interact with other animals on the basis of their observable characteristics and their putative mental states. In view of this, and given that we know that animals can detect and learn about observables, the null hypothesis must continue to be that animals do not attribute mental states.

Another set of misunderstandings centres on the role of 'behaviourist' hypotheses. Alternative

explanations of behaviour that some regard as indicative of mental state attribution are described as 'behaviorist hypotheses', or 'rival, conditioning hypotheses' if they suggest that the animal was reacting to observable features of another animal's appearance or behaviour (Dennett 1983). This is misleading for two reasons. First, one certainly does not have to be a behaviourist to generate or entertain such hypotheses. A thorough-going cognitivist, even one who believes that animals have mental states such as beliefs and desires, should not automatically assume that they also have higher-order mental states. It is possible, for example, that autistic humans (Frith 1989) and young children (Leslie 1987) have one without the other. Second, the equation of behaviourism and conditioning in this context implies that the question of what an animal has learned (about mental states or observables) is inextricably bound to the question of how the learning took place (through an inferential or an associative process). In fact, there is considerable slippage between the two (an animal can learn about observables through either an associative or an inferential process), and failure to recognize this may have inhibited research progress. Two ineffective methods, anecdote collection and trapping, retain credibility in large part because they are thought to overcome the problem of conditioning, when conditioning is not the most intractable problem. Conversely, an effective method, triangulating, has been little used quite possibly because it resembles the methods used to investigate conditioning.

Conditional discrimination training and transfer tests are used to study associative learning, but they are in no sense the tools of behaviourism. Indeed, it was this method of triangulation that yielded some of the earliest and strongest evidence that behaviour must be explained with reference to not-directly-observable states and processes (Dennis 1929; Macfarlane 1930; Campbell 1953). It should, therefore, be no surprise to find that triangulation can now be used to address questions quite antithetical to the behaviourist tradition, including the question, which remains open, 'Can animals attribute mental states?'

ACKNOWLEDGMENTS

This work was supported by a grant from The Leverhulme Trust. I am grateful to Tony Dickinson and Henry Plotkin for incisive comments on the original manuscript, and to Don Campbell for his generous tutelage.

REFERENCES

Byrne, R. W. & Whiten, A. 1988. *Machiavellian Intelligence*. Oxford: Oxford University Press.

Campbell, D. T. 1953. Operational delineation of 'what is learned' via the transposition experiment. *Psychol. Rev.*, **61**, 167–174.

Campbell, D. T. 1963. Social attitudes and other acquired behavioral dispositions. In: *Psychology: A Study of Science Vol. 6* (Ed. by S. Koch), pp. 94–172. New York: McGraw-Hill.

Campbell, D. T. 1969. A phenomenology of the other one: corrigible, hypothetical and critical. In: *Human Action: Conceptual and Empirical Issues* (Ed. by T. Mischel), pp. 41–69. New York: Academic Press.

Cheney, D. & Seyfarth, R. 1990a. Attending to behaviour versus attending to knowledge: examining monkeys' attribution of mental states. *Anim. Behav.*, **40**, 742–753.

Cheney, D. & Seyfarth, R. 1990b. *How Monkeys See the World: Inside the Mind of Another Species*. Chicago: University of Chicago Press.

Cheney, D. & Seyfarth, R. 1992. Precis of "How monkeys see the world". *Behav. Brain Sci.*, **15**, 135–182.

Cook, R. G., Brown, M. F. & Riley, D. A. 1985. Flexible memory processing by rats: use of prospective and retrospective information in the radial maze. *J. exp. Psychol.: Anim. Behav. Proc.*, **11**, 453–469.

Crano, W. 1981. Triangulation. In: *Scientific Inquiry and the Social Sciences: A Volume in Honor of Donald T. Campbell* (Ed. by M. B. Brewer & B. E. Collins), pp. 317–344. San Francisco: Jossey-Bass.

D'Amato, M. R. & Van Sant, P. 1988. The person concept in monkeys. *J. exp. Psychol. Anim. Behav. Proc.*, **14**, 43–55.

Dennett, D. C. 1980. The milk of human intentionality (commentary on Searle). *Behav. Brain Sci.*, **3**, 428–430.

Dennett, D. C. 1983. Intentional systems in cognitive ethology: the "Panglossian paradigm" defended. *Behav. Brain Sci.*, **6**, 343–390.

Dennis, W. 1929. The sensory control of the white rat in the maze habit. *J. Genet. Psychol.*, **36**, 59–89.

De Waal, F. 1982. *Chimpanzee Politics*. London: Jonathan Cape.

De Waal, F. 1986. Deception in the natural communication of chimpanzees. In: *Deception: Perspectives on Human and Nonhuman Deceit* (Ed. by R. W. Mitchell & N. S. Thompson), pp. 221–224. New York: State University of New York Press.

Dickinson, A. 1988. Intentionality in animal conditioning. In: *Thought Without Language* (Ed. by L. Weiskrantz), pp. 306–325. Oxford: Oxford University Press.

Dickinson, A. & Dawson, G. R. 1988. Motivational control of instrumental performance: the role of prior experience of the reinforcer. *Q. J. exp. Psychol.*, **40B**, 113–134.

Dickinson, A. & Dawson, G. R. 1989. Incentive learning and the motivational control of instrumental performance. *Q. J. exp. Psychol.*, **41B**, 99–112.

Frith, U. 1989. Autism and theory of mind. In: *Diagnosis and Treatment of Autism* (Ed. by C. Gillberg), pp. 87–102. New York: Plenum Press.

Galef, B. G. 1988. Imitation in animals: history, definition, and interpretation of data from the psychological laboratory. In: *Social Learning: Psychological and Biological Perspectives* (Ed. by T. R. Zentall & B. G. Galef, Jr), pp. 3–28. Hillsdale, New Jersey: Erlbaum.

Galef, B. G. 1992. The question of animal culture. *Human Nat.*, **3**, 157–178.

Gallup, G. G. 1982. Self-awareness and the emergence of mind in primates. *Am. J. Primatol.*, **2**, 237–248.

Goodall, J. 1971. *In the Shadow of Man*. London: Collins.

Goodall, J. 1986. *The Chimpanzees of the Gombe: Patterns of Behavior*. Cambridge, Massachusetts: Harvard University Press.

Heyes, C. M. In press a. Imitation, cognition and culture. *Anim. Behav.*

Heyes, C. M. In press b. Reflections on self-recognition in primates. *Anim. Behav.*

Heyes, C. M. & Dickinson, A. 1990. The intentionality of animal action. *Mind and Language*, **5**, 87–104.

Herrnstein, R. J., Loveland, D. H. & Cable, C. 1976. Natural concepts in pigeons. *J. exp. Psychol.: Anim. Behav. Proc.*, **2**, 285–311.

Humphrey, N. K. 1976. The social function of intellect. In: *Growing Points in Ethology* (Ed. by P. P. G. Bateson & R. A. Hinde), pp. 303–317. Cambridge: Cambridge University Press.

Jolly, A. 1985. *The Evolution of Primate Behaviour*. 2nd edn. New York: Macmillan.

Kaye, H., Gambini, B. & Mackintosh, N. J. 1988. A dissociation between one-trial overshadowing and the effect of a distracter on habituation. *Q. J. exp. Psychol.*, **40B**, 31–47.

Kohler, W. 1925. *The Mentality of Apes*. London: Routledge & Kegan Paul.

Krebs, J. R. & Dawkins, R. 1984. Animal signals: mind reading and manipulation. In: *Behavioural Ecology* (Ed. by J. R. Krebs & N. B. Davies), pp. 380–402. Oxford: Blackwell Scientific Publications.

Kummer, H., Dasser, V. & Hoyningen-Huene, P. 1990. Exploring primate social cognition: some critical remarks. *Behaviour*, **112**, 84–98.

Laudan, L. 1971. William Whewell on the consilience of inductions. *The Monist*, **55**, 368–391.

Leslie, A. M. 1987. Pretence and representation: the origins of a 'theory of mind'. *Psychol. Rev.*, **94**, 412–426.

Levins, R. 1966. The strategy of model building in population biology. *Am. Scient.*, **54**, 421–431.

Macfarlane, D. A. 1930. The role of kinesthesis in maze learning. *Univ. Calif. Publ. Psychol.*, **4**, 277–305.

Maki, W. S. 1979. Pigeon's short-term memories for surprising vs. expected reinforcement and nonreinforcement. *Anim. Learn. Behav.*, **7**, 31–37.

Maki, W. S. & Hegvik, D. K. 1980. Directed forgetting in pigeons. *Anim. Learn. Behav.*, **8**, 567–574.

Menzel, E. W. 1974. A group of young chimpanzees in a one-acre field. In: *Behaviour of Nonhuman Primates, Vol. 5* (Ed. by A. M. Schrier & F. Stollnitz), pp. 171–189. New York: Academic Press.

Mitchell, R. W. 1986. A framework for discussing deception. In: *Deception: Perspectives on Human and Nonhuman Deceit* (Ed. by R. W. Mitchell & N. S. Thompson), pp. 3–40. New York: State University of New York Press.

Morris, R. G. M. 1981. Spatial localization does not require the presence of a local cue. *Learn. Motiv.*, **12**, 239–260.

Peirce, C. S. 1936. Some consequences of four incapacities. In: *Collected Papers of Charles Saunders Peirce, Vol. 5* (Ed. by C. Hartshorne & P. Weiss). Cambridge, Massachusetts: Harvard University Press. Originally published in 1868.

Povinelli, D. J., Nelson, K. E. & Boysen, S. T. 1990. Inferences about guessing and knowing by chimpanzees. *J. comp. Psychol.*, **104**, 203–210.

Povinelli, D. J., Nelson, K. E. & Boysen, S. T. 1992. Comprehension of role reversal in chimpanzees: evidence of empathy? *Anim. Behav.*, **43**, 633–640.

Povinelli, D. J., Parks, K. A. & Novak, M. A. 1991. Do rhesus monkeys (*Macaca mulata*) attribute knowledge and ignorance to others? *J. comp. Psychol.*, **105**, 318–325.

Premack, D. & Premack, A. J. 1983. *The Mind of an Ape*. New York: Norton.

Premack, D. & Woodruff, G. 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.*, **4**, 515–526.

Suzuki, S., Augerinos, G. & Black, A. H. 1980. Stimulus control of spatial behavior on the eight-arm maze in rats. *Learn. Motiv.*, **11**, 1–18.

Visalberghi, E. & Fragaszy, D. M. 1990. Food-washing behaviour in tufted capuchin monkeys, *Cebus apella*, and crabeating macaques, *Macaca fascicularis*. *Anim. Behav.*, **40**, 829–836.

Whiten, A. & Byrne, R. W. 1988. Tactical deception in primates. *Behav. Brain Sci.*, **11**, 233–273.

Whiten, A. & Byrne, R. W. 1991. The emergence of metarepresentation in human ontogeny and primate phylogeny. In: *Natural Theories of Mind: Evolution, Development and Simulation of Everyday Mindreading* (Ed. by A. Whiten), pp. 267–281. Oxford: Basil Blackwell.

Wimsatt, W. C. 1981. Robustness, reliability and overdetermination. In: *Scientific Enquiry and the Social Sciences* (Ed. by M. B. Brewer & B. E. Collins), pp. 124–163. San Francisco: Jossey-Bass.

Woodruff, G. & Premack, D. 1979. Intentional communication in the chimpanzee: the development of deception. *Cognition*, **7**, 333–362.