BEHAVIORAL AND BRAIN SCIENCES (1998) **21,** 101–148
Printed in the United States of America

Theory of mind in nonhuman primates

C. M. Heyes
Department of Psychology, University College London, London WC1E 6BT, United Kingdom
Electronic mail: *c.heyes@ucl.ac.uk*

Abstract: Since the *BBS* article in which Premack and Woodruff (1978) asked "Does the chimpanzee have a theory of mind?," it has been repeatedly claimed that there is observational and experimental evidence that apes have mental state concepts, such as "want" and "know." Unlike research on the development of theory of mind in childhood, however, no substantial progress has been made through this work with nonhuman primates. A survey of empirical studies of imitation, self-recognition, social relationships, deception, role-taking, and perspective-taking suggests that in every case where nonhuman primate behavior has been interpreted as a sign of theory of mind, it could instead have occurred by chance or as a product of nonmentalistic processes such as associative learning or inferences based on nonmental categories. Arguments to the effect that, in spite of this, the theory of mind hypothesis should be accepted because it is more parsimonious than alternatives or because it is supported by convergent evidence are not compelling. Such arguments are based on unsupportable assumptions about the role of parsimony in science and either ignore the requirement that convergent evidence proceed from independent assumptions, or fail to show that it supports the theory of mind hypothesis over nonmentalist alternatives. Progress in research on theory of mind requires experimental procedures that can distinguish the theory of mind hypothesis from nonmentalist alternatives. A procedure that may have this potential is proposed. It uses conditional discrimination training and transfer tests to determine whether chimpanzees have the concept "see." Commentators are invited to identify flaws in the procedure and to suggest alternatives.

Keywords: apes; associative learning; concepts; convergence; deception; evolution of intelligence; folk psychology; imitation; mental state attribution; monkeys; parsimony; perspective-taking; primates; role-taking; self-recognition; social cognition; social intelligence; theory of mind

1. Premack and Woodruff's question

Premack and Woodruff (1978) asked "Does the chimpanzee have a theory of mind?" Since it was posed, 20 years ago, Premack and Woodruff's question has dominated the study of both social behavior in nonhuman primates (henceforward simply "primates") and cognitive development in children, but progress in the two fields has been markedly different. Developmentalists have established empirical methods to investigate children's understanding of mentality, and, forging links with philosophy of mind and philosophy of science, they have mustered the conceptual resources for disciplined dispute about the origins (innate module, convention, or testing), on-line control (simulation or inference), and epistemic status (stance, theory, or direct knowledge) of human folk psychology (e.g., Goldman 1993; Gopnik 1993; Gopnik & Wellman 1994). In contrast, those working with primates have continued to struggle with the basic question of whether *any* primate has *any* capacity to conceive of mental states.

Primatologists and other investigators of animal behavior use a variety of substitutes for the term "theory of mind," asking whether animals are capable of, for example, "Machiavellian intelligence" (Byrne & Whiten 1988; Whiten & Byrne 1988), "metarepresentation" (Whiten & Byrne 1991), "metacognition" (Povinelli 1993), "mind reading" (Krebs & Dawkins 1984; Whiten 1991), "mental state attribution" (Cheney & Seyfarth 1990a; 1990b; 1992), and "pan- or pongo-morphism" (Povinelli 1995). Some authors use these terms to refer to hypothetically distinct capacities (see Whiten 1994 and 1996b for discussion of terminology), but they usually function in research on social cognition in primates as synonyms. A researcher using the term "mental state attribution," for example, is no less likely than one using "theory of mind" to believe that law-like generalizations underlie mental state ascription.

In this target article, I assume that individuals have a theory of mind if they have mental state concepts such as "believe," "know," "want," and "see," and that individuals

Cecilia Heyes is a Reader in Psychology, and Associate of the ESRC Economic Learning and Social Evolution Research Centre, at University College London. Formerly a Harkness Fellow, and Research Fellow of Trinity Hall Cambridge, she has published a series of review papers on theory of mind and self-recognition in primates. Her principal research interests are social learning and imitation. She is currently writing a book on human imitation learning, and co-editing volumes on selectionist epistemology and the evolution of cognition. Dr Heyes is a member of the Association for the Study of Animal Behaviour, and was for nine years an Associate Editor of the *Quarterly Journal of Experimental Psychology.* She is member of the management committee of the Experimental Psychology Society.

© 1998 Cambridge University Press 0140-525X/98 $12.50

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

with such concepts use them to predict and explain behavior. Thus, an animal with a theory of mind believes that mental states play a causal role in generating behavior and infers the presence of mental states in others by observing their appearance and behavior under various circumstances. However, they do not identify mental states with behavior. For example, if chimpanzee Al has a theory of mind, he may judge chimpanzee Bert to be able to "see" a predator because it is daylight, Bert's eyes are open, and there is an uninterrupted line between Bert's eyes and the predator. But Al does not take seeing the predator to consist of these observable conditions. It is a further fact about Bert, inferred from these conditions, which explains why Bert runs away.[1]

In spite of nearly 20 years of research effort, there is still no convincing evidence of theory of mind in primates. We should stop asking Premack and Woodruff's question and considering the implications of a positive answer until we have designed procedures that have the potential to yield evidence favoring a theory of mind interpretation over other current candidates. Section 2 is a survey of the evidence of theory of mind from specific categories of behavior (imitation, self-recognition, social relationships, deception, role-taking, and perspective-taking), which argues for each study that the behavior reported could have occurred by chance or via nonmentalistic processes such as associative learning or inferences based on nonmental categories. In section 3, I argue that a theory of mind interpretation of the data reviewed in section 2 cannot be defended on the grounds of parsimony or convergence, and in section 4, I describe a test procedure that may be able to provide evidence of theory of mind in primates. Commentators are invited to identify flaws in this procedure and to devise alternatives.

2. Critique of evidence

The majority of those who have conducted empirical work on theory of mind in primates have claimed at one time or another that chimpanzees and possibly other apes, but not monkeys, have some components of a theory of mind (e.g., Byrne 1994; Cheney & Seyfarth 1990a; 1992; Gallup 1982; Jolly 1991; Povinelli 1993; Waal 1991; Whiten & Byrne 1991). The most commonly cited evidence in support of this view comes from studies of imitation, self-recognition, social relationships, deception, role-taking (or empathy), and perspective-taking. A sample of studies from each of these categories including the strongest and most influential is reviewed in sections 2.1–2.6.

In each of these six sections, two questions are addressed: (1) Competence: Is there reliable evidence that primates have the relevant behavioral capacity? (2) Validity: If present, would this behavioral capacity indicate theory of mind? For example, in the case of self-recognition, the competence question will be answered affirmatively if there is clear evidence that some primates are capable of using a mirror as a source of information about their bodies, and the evidence will be considered clear if there is no other at least equally plausible explanation for published observations of mirror-related behavior in primates. Similarly, the validity question will be answered affirmatively if there is no equally plausible nonmentalistic alternative to the hypothesis that mirror-guided body inspection requires or involves a self-concept. More generally, the competence question attempts to establish which environmental cues primates use to guide their behavior, and the validity question inquires about the psychological processes that lead them to use these cues rather than others.

The theory of mind hypothesis (or, more accurately, hypotheses) consists primarily of claims about what primates know or believe, about the *content* of their representations. Their distinctive, unifying feature is that they assert that primates categorize and think about themselves and others in terms of mental states. Consequently, the distinctive, unifying feature of nonmentalistic alternative hypotheses is that they do *not* assume that primates represent mental states. They assume instead that primates respond to or categorize and think about themselves and others in terms of observable properties of appearance and behavior. Behaviorism and learning theory are rich sources of nonmentalistic hypotheses, with those derived from behaviorism assuming no representation at all, and those derived from contemporary learning theory assuming, for the most part, some kind of imaginal, nonsymbolic representation (Dickinson 1980). However, denial that primates are capable of representation, or of abstract or symbolic representation, is not a necessary feature of a nonmentalistic hypothesis. Such a hypothesis might, for example, assume that primates are sensitive to whether a conspecific is "upright" or "supine" (see sect. 2.5 below), and that these are abstract or symbolically represented concepts, derived and applied through inference processes.

Consequently, it may be misleading to portray the debate about theory of mind in primates as a battle between the theory of mind hypothesis and "traditional learning theory" (e.g., Povinelli & Eddy 1996). The theory of mind hypothesis is primarily a claim about what is known or represented, whereas learning theory's most distinctive claims are about how knowledge is acquired (Dickinson 1980). Similarly, it may be counterproductive, although amusing, to think of animals that lack a theory of mind as "behaviorists" (e.g., Premack & Woodruff 1978). In one potentially confusing respect, scientific or philosophical behaviorists have a theory of mind just as surely as a scholar who believes that mental states cause behavior: they actively seek to explain, using argument and evidence, the nature and origins of behavior, including human use of mental state terms. In contrast, nonmentalistic alternatives to theory of mind hypotheses typically claim that primates "just do it"; they respond to observable cues, categorize them, form associations between them or make inferences about them, but they never ask themselves why, or whether other animals do the same thing. Thus, according to nonmentalistic hypotheses, primates are not psychologists, or indeed theorists, of any stripe.

Section 2.7 summarizes my answers to the competence and validity questions for each of the six types of behavior discussed. This may be used as a guide for selective reading by those who are not interested in detailed evaluation of evidence of competence when the behavioral capacity in question may not be a valid indicator of theory of mind.

2.1. Imitation

Motor imitation, the spontaneous reproduction of novel acts yielding disparate sensory inputs when observed and executed, has long been regarded as a potential sign of

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

higher intelligence in nonhuman animals (e.g., Thorndike 1898). It is relevant to theory of mind because it is thought to involve the ascription of purposes or goals by the imitator to the model (e.g., Tomasello & Call 1994; Tomasello 1996; Whiten & Byrne 1991). However, after nearly 100 years of research, there is still no unequivocal evidence of motor imitation in any primate species and even if there were, it would not imply the possession of mental state concepts.

Under uncontrolled and semicontrolled conditions the occurrence of imitation in monkeys (Beck 1976; Hauser 1988; Nishida 1986; Westergaard 1988), orangutans (Russon & Galdikas 1993), and chimpanzees (Goodall 1986; Mignault 1985; Sumita et al. 1985; Terrace et al. 1979; Waal 1982) has been inferred from the performance of a complex, novel, and previously observed act by a single animal or a succession of animals within a group. Even if one disregards the problem of the reliability of these observational or anecdotal data, they are not compelling. In all cases, the observed behavior could have been acquired by a means other than imitation (e.g., instrumental learning), and in many cases there is evidence that it was so acquired (Adams-Curtis 1987; Fragaszy & Visalberghi 1989; 1990; Galef 1992; Tomasello et al. 1993; Visalberghi & Trinca 1989). For example, the habit of potato washing was supposed to have been transmitted through the population of Japanese macaques on Koshima Island through imitation (Nishida 1986). However, given the order in which members of the troop were observed engaging in this behavior (first a juvenile, Imo, then her playmates, then their mothers), it is possible that, rather than copying the actions of potato washers, naive animals followed or chased them into the water while holding a potato. Once in that position, the pursuing animal would only have to drop and then retrieve its potato, now sand-free and with a salty taste, to acquire the behavior (Galef 1992; Visalberghi & Fragaszy 1992).

Remarkably few experiments have been conducted on imitation in primates. Their results may indicate only "matched dependent behavior" (Miller & Dollard 1941), the use of a demonstrator's behavior as a discriminative stimulus for the same response by the observer, and "stimulus enhancement" (Galef 1988; Spence 1937) that observing action can influence the degree to which the observer attends to certain physical components of a problem situation. Hayes and Hayes (1952) gave Viki, a "home-raised" chimpanzee, a series of 70 "imitation set" tasks. Each task consisted of the experimenter saying "Do this," and then performing an action such as patting his head, clapping his hands, or operating a toy. Hayes and Hayes claimed that Viki imitated more than 50 items in the set, including 10 completely novel, arbitrary gestures, but this conclusion is not secure because the report on Viki's behavior provided no indication of either the method used to measure the similarity between the experimenter's and the chimpanzee's behavior, or the degree of similarity observed.

Custance et al. (1995) carefully replicated Hayes and Hayes's study with two juvenile chimpanzees and provided a full report of their methods and results. The latter showed that after being shaped to imitate 15 gestures on the command "Do this," the chimpanzees spontaneously reproduced 13 and 17, respectively, of a possible 48 "novel" gestures, actions distinct from those in the training set. This is probably the strongest evidence to date that, at least after training, the form or topography of a primate's action can be influenced by observing the same action by a demonstrator. However, even when they reproduced novel gestures, the chimpanzees may have been engaging in matched-dependent behavior (Miller & Dollard 1941), that is, using the demonstrator's behavior as a discriminative stimulus for the same or similar behavior, without knowing that their behavior was similar to that of the demonstrator. For example, both chimpanzees reproduced lip smacking without being explicitly trained to do so in this study. However, they had been reared by humans, and humans have a strong tendency to play mutual imitation games with infants in which the infant is rewarded with smiles and cuddles for reproducing behavior, especially facial expressions (Piaget 1962). Hence, we cannot rule out the possibility that Custance et al.'s chimpanzees had been inadvertently rewarded for imitative lip-smacking (or imitative performance of a lip movement sufficiently like smacking to be scored as such in this study) before the experiment began. As Custance et al. point out, the reproduction of other novel items in the series may have been due to generalization from initial training within the study. For example, successful reproduction of nose touching may have represented fortuitous generalization decrement from prior training to reproduce chin touching. If this was the case, and if the chimpanzees' reactions to nose touching had been sampled many times in the absence of reinforcement (adventitious or otherwise), then one would have expected to see a range of responses to nose touching, including throat and cheek touching.

Tomasello and his colleagues (Tomasello et al. 1987) did not find any evidence of imitation of rake use in chimpanzees, but they reported positive findings for "enculturated" chimpanzees (i.e., animals with an extensive training history) in a later experiment (Tomasello et al. 1993). In this study, enculturated chimpanzees, relatively naive chimpanzees, and young children observed the experimenter manipulating 16 objects in various ways and, after observing each action, were given access to the same object either immediately or after a 48-hour delay. When the test was given immediately and the results for all objects were combined, the enculturated chimpanzees were comparable to the children in their tendency to act on the same part of the object, and with the same effect, as the demonstrator. However, for many objects, resemblance between the demonstrator and the observer could have been coincidental or due to stimulus enhancement rather than imitation. For example, when presented with a paint brush, the chimpanzees may have squeezed it with one hand, not because they had observed the trainer executing this particular action in relation to the brush, but simply in an effort to grasp an object which had been made salient through contact with the demonstrator. Since, by definition, the enculturated chimpanzees had been subject to more training procedures in the past than the other chimpanzees, there had been more opportunity for their fear of such procedures to habituate, and, specifically, more time for them to learn that objects handled by humans are often associated with reward. Therefore, even if the experiment by Tomasello et al. (1993) tested interest in novel objects and stimulus enhancement rather than imitation, one would expect the performance of the enculturated animals to be superior.[2]

The paucity of evidence of imitation in primates indicates neither that they are unable to imitate nor that such evidence is impossible to obtain for nonhuman animals. Relatively unequivocal evidence of imitation in budgerigars

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

(Galef et al. 1986) and rats (Heyes & Dawson 1990; Heyes et al. 1992) has been found by comparing the behavior of naive subjects that have observed a conspecific acting on a single object in one of two distinctive ways (but see: Byrne & Tomasello 1995; Heyes 1996). Whiten et al. (1996; Whiten & Custance 1996) recently gave a similar "two-action test" to chimpanzees, with mixed results. They found that chimpanzees that had seen a person withdraw bolts from rings with a twisting action for food reward subsequently twisted the bolts more than chimpanzees that had seen the person push the bolts through the rings with a poking action. However, as the authors pointed out, in the absence of data from subjects that did not observe any action on the bolts prior to testing it is difficult to rule out the possibility that what the chimpanzees learned by observation was not how to perform the twisting or poking hand movement but that certain movements of the bolts (e.g., rotation followed by lateral displacement toward the actor) were followed by reward. This has been described as emulation learning (Tomasello 1996).

Thus, surprisingly, it is not clear whether apes or indeed any other nonhuman primates can "ape" (Tomasello 1996), whether they are competent imitators. Furthermore, a capacity to imitate is not a valid indicator of theory of mind. It has been claimed that imitation involves the observer representing the demonstrator's mental state, its point of view, or its beliefs and desires (e.g., Gallup 1982; Povinelli 1995), but the case is not compelling. As far as I am aware, there is no evidence that the development of imitation in childhood is related to success in conventional theory of mind tests, and simple task analysis suggests that an observer could imitate a demonstrator's action without any appreciation that the demonstrator has mental states. To reproduce a novel action without training or tuition it would seem to be essential for the observing animal to represent what the demonstrator did, but not what it thought or wanted. When the action is perceptually opaque – it yields different sensory inputs to an animal when that animal observes the action and when it executes the action (e.g., a facial expression) – imitation further implies that the imitator can represent actions in a cross-modal or sense-independent code (Meltzoff & Moore 1983). But even in these fascinating cases, mental state attribution is not implied and indeed the ascription of a theory of mind to the imitator does not help to resolve the mystery of how the imitator translates sensory input from the demonstrator's action into performance that resembles, from a third party perspective, that of the demonstrator (Heyes 1994a; 1994b; 1996).

2.2. Self-recognition

A series of experiments using a common procedure apparently shows that chimpanzees and orangutans, but not other primates, are capable of "self-recognition" (Gallup 1970) or "mirror-guided body inspection" (Heyes 1994c); they can use a mirror as a source of information about their own bodies (Cheney & Seyfarth 1990a; Gallup 1982; Jolly 1991; Povinelli 1987). This capacity has been said to imply the possession of a "self-concept" and the potential to imagine oneself as one is viewed by others (Gallup 1982; Povinelli 1987). I will argue that there is no reliable evidence that any nonhuman primates can use a mirror to derive information about their own bodies, and that even if there were, such a capacity would not indicate the possession of a self-concept or any other component of a theory of mind.

In the standard procedure (e.g., Gallup 1970), an animal with some experience of mirrors is anesthetized and marked on its head with an odorless, nonirritant dye; several hours later, the frequency with which the animal touches the marks on its head is measured first in the absence of a mirror and then with a mirror present. Chimpanzees and orangutans typically touch their head marks more when the mirror is present than when it is absent, while monkeys of various species and gorillas touch their marks with the same low frequency in both conditions (Calhoun & Thompson 1988; Gallup 1970; 1977; Gallup et al. 1971; Ledbetter & Basen 1982; Platt & Thompson 1985; Suarez & Gallup 1981).

There is an alternative to the standard interpretation of the chimpanzee and orangutan tendency to touch their marks more in the presence of the mirror than in its absence. In the mirror-present condition, the animals had longer to recover from anesthesia and may therefore have been more active generally than in the previous, mirror-absent condition. If they were more active generally, they had a higher probability of touching the marked areas of their heads by chance. Thus, chimpanzees and orangutans may touch their marks more when the mirror is present than when it is absent simply because at the mirror-present stage, they have had longer to recover from the anesthetic and are therefore more active generally (Heyes 1994c).

In Gallup's (1970) original experiment, two additional chimpanzees that had no prior exposure to mirrors were anesthetized, marked, and observed in the presence of the mirror on recovery. They did not make any mark-directed responses, but that does not mean that the other, mirror-preexposed animals must have been using the mirror to detect their marks. Chimpanzees typically exhibit social behavior on initial exposure to a mirror, and it is therefore likely that the control animals were too busy responding socially to their mirror image to engage in the normal grooming behavior that had, by chance, given rise to mark-touching in the experimental subjects.

According to this anesthetic artefact hypothesis, which is also consistent with the results of mark tests that vary from the standard procedure (Anderson 1983; Anderson & Roeder 1989; Eglash & Snowdon 1983; Gallup & Suarez 1991; Lin et al. 1992; Robert 1986; Suarez & Gallup 1986b; see Heyes 1994c and 1995b for reviews), species differences in mark test performance arise from the fact that chimpanzees spontaneously touch their faces with a higher frequency than either monkeys or gorillas (Dimond & Harries 1984; Gallup et al. 1995; Heyes 1995b).

The anasthetic artifact hypothesis would be less plausible if the effects of mirror insertion on face-touching were larger. In studies reported by Gallup and his associates (e.g., Gallup 1970; Gallup et al. 1971; Suarez & Gallup 1981) it is difficult to assess either the magnitude of the effect on individual animals or its statistical reliability, because the results are presented as two group total scores: the number of mark-touches made by all members of a group of animals in the mirror-present and mirror-absent conditions. The smallness of the effect, however, is apparent in data reported by other authors: Calhoun and Thompson (1988) found that, after failing to touch their marks at all during the mirror-absent period, each of two chimpanzees made just

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

two responses in the mirror-present condition. Thirty chimpanzees tested by Povinelli et al. (1993, Experiment 4) touched their marks, on average (±SD), 2.5 (±3.7) times in the absence of the mirror and 3.9 (±8.0) times in its presence. Swartz and Evans (1991) reported that only one of 11 chimpanzees touched its mark more in the mirror-present condition, and that, on average, 3.3 (±3.7) touches occurred while the mirror was absent, and 2.9 (±7.19) when it was present. In all three of these experiments, the mirror-present and mirror-absent periods were each of 30 minutes duration. Thus, it would not be necessary for an anesthetic recovery gradient to be improbably steep, or especially uniform across animals, to account for the mark-touching effects typically observed.

It has been suggested that the anesthetic artifact hypothesis is inconsistent with the immediacy of the effects of mirror insertion on mark touching (Gallup et al. 1995). However, I cannot find any published, quantitative data showing that mark touching is more frequent at the beginning of the mirror-present period than at its end, or that the contrast between the mirror-absent and mirror-present periods is greatest when the terminal portion of the former is compared with the initial portion of the latter. Furthermore, if such data were available, they would be equally consistent with the hypothesis that the chimpanzees use their mirror images to detect their marks, and with the hypothesis that mirror introduction elevates arousal and thereby produces an increase in the frequency of a range of behavior patterns.

It is surprising that a straightforward mark test procedure that could disprove the anesthetic artifact hypothesis has not been implemented. The procedure in question would compare the frequencies with which chimpanzees touch the marked and corresponding unmarked areas of their faces, in mirror-present and mirror-absent conditions (see Heyes 1995b for a more complete design). If it showed that chimpanzees touch the marked areas more than the unmarked areas in the mirror-present condition but not in the mirror-absent condition, then there would be reason to believe that chimpanzees can detect marks on their heads using a mirror.

However, even if there were evidence that certain primates have this capability, it would not imply the possession of a "self-concept" or the potential to imagine oneself as one is viewed by others (i.e., theory of mind; Gallup 1982; Povinelli 1987). Simple task analysis suggests that to use a mirror as a source of information about its body an animal must be able to distinguish, across a fairly broad range, sensory inputs resulting from the physical state and operations of its own body from sensory inputs originating elsewhere. If the animal could not do this, if it lacked what might be described loosely as a "body concept," then presumably it could not learn that when it is standing in front of a mirror, inputs from the mirror correlate with inputs from its body. However, a "body concept" does not relate to a mental category, and, since it is equally necessary for mirror-guided body inspection and for collision-free locomotion, the former no more implies possession of such a concept than does the latter (Heyes 1994c).

A demonstration that the humble pigeon can learn to use a mirror to detect paper dots attached to its feathers (Epstein et al. 1981) makes it easier to appreciate that mirror-guided body inspection may not imply the use of mental state concepts (but see Gallup 1983 for objections to Epstein et al.'s interpretation of their results). More direct evidence of a dissociation between the two is provided by studies of autistic children who, although apparently incapable of ascribing beliefs to others, have been reported to begin using a mirror to inspect their bodies at the same age as normal children (Ungerer 1989).

2.3. Social relationships

There is a substantial body of evidence suggesting that the social behavior of primates is not affected only by concurrent events and the outcomes of previous, active engagements between the present interactants and third parties. The behavior of animal A in relation to animal B also may be affected by A's prior observations of B in relation to one or a number of other conspecifics, C, D, and so on. Evidence of this kind (reviewed in Cheney & Seyfarth 1990a) has been derived from observational and experimental studies of chimpanzees, baboons, and various macaques. For example, adult male chimpanzees are more likely to disrupt (through interposition, aggression, or a threat display) social interactions between pairs of high-ranking conspecifics than between pairs of mixed or low rank (Waal 1982).

Studies of this kind show that the social behavior of animals from a broad range of primate species is sensitive to what human observers naturally describe as "social relationships" among conspecifics. It has been said, in addition, to show that primates have knowledge of social relationships (Cheney & Seyfarth 1990a; Kummer et al. 1990; Waal 1991), and this seems entirely appropriate when the term "knowledge" is used in a very general sense and social relationships are understood to be observable properties. If, on the other hand, knowledge of social relationships is taken to involve the attribution to conspecifics of knowledge about their social interactants or dispositional mental states such as loyalty, dislike, or affection, and to be acquired by a means other than associative learning (Cheney & Seyfarth 1990a; 1992; Dasser 1988; Waal 1991), then the evidence to date does not support the conclusion that primates know about social relationships.

Two studies will illustrate the plausibility of simple associative accounts of sensitivity to social relationships. In the first (Cheney & Seyfarth 1980), free-ranging vervet monkeys heard the scream of an absent juvenile from a concealed loudspeaker. The adult female monkeys in the group typically responded to the sound of the juvenile's cry by looking at the juvenile's mother before the mother had responded to the cry herself. In so doing they displayed sensitivity to or knowledge of the mother–offspring relationship. But, as the authors recognized, this could have resulted from earlier exposure to a contingency between the cries of a particular juvenile and a vigorous behavioral reaction from a particular adult female (Cheney et al. 1986).

In the second study (Stammbach 1988), one subordinate member of each of a number of groups of longtailed monkeys was trained to obtain preferred food for the group by manipulating three levers. The other monkeys did not acquire the skill themselves, but those that received the most food as a result of the trained animals' activities began to follow them to the lever apparatus and spent an increasing amount of time sitting beside and grooming the trained animals, even when the apparatus was not in operation. The untrained monkeys may have behaved in this way because they attributed to the trained individuals superior knowl-

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

edge of the workings of the lever apparatus, and wanted to develop friendly relations with them in the hope of gaining more food (Kummer et al. 1990; Stammbach 1988). However, the results of an experiment with rats show that, rather than attributing superior knowledge, each untrained monkey may have learned an association between the trained animal in their group and receipt of preferred food. In this study (Timberlake & Grant 1975), rats acquired affiliative social responding to a conspecific that was fastened to a trolley and wheeled into an operant chamber as a signal for the delivery of food.

2.4. Role-taking

In the experiments that gave rise to the suggestion that chimpanzees have a theory of mind (Premack & Woodruff 1978), a "language-trained" chimpanzee, Sarah, was shown videotapes depicting human actors confronting problems of various kinds (e.g., trying to reach inaccessible food, to escape from a locked cage, and to cope with malfunctioning equipment). The final image of each videotape sequence was put on hold, and Sarah was offered a choice of two photographs to place beside the video monitor. Both of these represented the actor in the problem situation, but only one of them showed the actor taking a course of action that would solve the problem. Sarah consistently chose the photographs representing problem solutions, and this was interpreted as evidence that she attributed mental states to the actor (Premack & Woodruff 1978; see Premack 1983; 1988 for reservations about this conclusion). It was argued that if Sarah did not ascribe beliefs and desires to the actor then she would see the video as an undifferentiated sequence of events rather than a problem.

Close examination of the published reports of the videotape experiments (Premack & Premack 1982; Premack & Woodruff 1978) suggests that for any given problem Sarah could have responded on the basis of familiarity, physical matching, and/or formerly learned associations. For example, when the actor was trying to reach food that was horizontally out of reach, matching could have been responsible for Sarah's success because a horizontal stick was prominent in both the final frame of the videotape and the photograph depicting a solution. Similarly, when the actor was shivering and looking wryly at a broken heater, Sarah may have selected the photograph of a burning roll of paper rather than an unlit or spent wick because she associated the heater with the red-orange color of fire. Taken together, however, the results of Premack and Woodruff's videotape experiments are not subject to a single, straightforward nonmentalistic interpretation, and in this respect they are apparently unique in the literature on theory of mind in primates. Thus, according to this standard, no advance has been made on the original studies of theory of mind in primates.

Premack and Dasser (1991) have devised a method of finding out whether children use theory of mind rather than a matching or contiguity principle to solve videotape problems of the kind used by Premack and Woodruff (1978). This method, however, has not been applied to nonhuman primates, and the results of other experiments on role-taking in chimpanzees (Povinelli et al. 1992a; 1992b) are unfortunately no less ambiguous than those of Premack and Woodruff. In one of these other experiments (Povinelli et al.1992a), four chimpanzees were initially trained either to choose from an array of containers the one to which an experimenter was pointing (cue detection task), or to observe food being placed in one of the containers and then to point at the baited receptacle (cue provision task). Once criterion performance had been achieved on the initial problem, each chimpanzee was confronted with the other problem, and for three of the four animals this switch did not result in a significant decline in choice accuracy.

This result was tentatively interpreted as evidence of "cognitive empathy" or "role taking . . . the ability to adopt the viewpoint of another individual" (Povinelli et al. 1992a). This interpretation rests on two tenuous assumptions: (1) Training on the first task facilitated performance on the second, and, (2) this facilitation was due to the chimpanzees having the opportunity, during the first task, to see the problem from an interactant's perspective. The former assumption is unsupported because the results failed to show that each problem was learned faster when it was presented second than when it was presented first. Consequently, it is possible that the chimpanzees' fairly high rate of learning in each task was independently influenced by their pretraining and experience outside the experimental situation. The chimpanzees had learned to pull the levers to obtain food during pretraining, and they commonly encountered and exhibited pointing behavior in their day-to-day laboratory lives.[3]

If the results of the chimpanzee experiment (Povinelli et al. 1992a) 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 yet it is not clear which manipulations, if any, could provide unambiguous evidence that the opportunity for mental state attribution was responsible (Heyes 1993).

2.5. Deception

When applied to animal behavior, the term deception is often used in a functional sense (Krebs & Dawkins 1984) to refer to the provision by one animal, through production or suppression of behavior, of a cue that is likely to lead another to make an incorrect or maladaptive response. A mass of observational and anecdotal data leave no doubt that a broad variety of primate and nonprimate species (for excellent reviews see Cheney & Seyfarth 1991; Krebs & Dawkins 1984; Whiten & Byrne 1988) are capable of deception thus defined. However, the claim that theory of mind underlies this capacity in primates, that they sometimes act with the intention of producing or sustaining a state of ignorance or false belief in another animal, has little support. The evidence is almost exclusively anecdotal (Cheney & Seyfarth 1991; Whiten & Byrne 1988), and the behavior described in each anecdote is subject to one or more alternative interpretations.

Many anecdotal reports of deceptive behavior invite several alternative interpretations: that the behavior occurred (1) by chance, (2) as a result of associative learning, or (3) as a product of inferences about observable features of the situation rather than mental states (Heyes 1993; Kummer et al. 1990; Premack 1988). For example, "One of

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

the female baboons 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" (observation by Strum, cited as personal communication in Jolly 1985).

The female baboon may have intended to deceive the male about her intentions, but it may also have been no more than a coincidence that she began grooming the male when he was holding the carcass, and made a grab for the carcass when he was lolling back. Even if it did not occur by chance, the female's behavior may have been acquired through associative learning. For example, she may have snatched the carcass when the male was lolling back 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 could have acquired an association between snatching food and reward that was activated by the sight of a supine animal.

Even if observational studies of deceptive behavior could show that it was acquired through an inferential process rather than associative learning there would remain the possibility that the behavior was based on reasoning about observable features of the situation, or nonmental categories, rather than mental state concepts. Thus, the female baboon may have inferred from her experience of conspecific behavior that it is relatively safe to snatch food when the other animal is lying back, but she need not have regarded posture as an indicator of mental state.

The results of the only experimental investigation of intentional deception in primates (Woodruff & Premack 1979) are also equivocal. At the beginning of each trial in this study, a chimpanzee was allowed to observe food being placed in one of several inaccessible containers and then a human trainer dressed in green ("cooperative" trainer) or white ("competitive" trainer) entered the room and searched one of the containers. The trainer had been instructed to choose the container that the chimpanzee appeared to indicate through pointing, looking, or body orientation. When the cooperative trainer found food, he gave it to the chimpanzee, but the chimpanzee was rewarded on competitive trainer trials only if the trainer chose the incorrect container. After 120 trials, each of the four chimpanzees tested showed a reliable tendency to indicate the baited container in the presence of the cooperative trainer, and an empty container in the presence of the competitive trainer. Thus, the chimpanzees' behavior toward the competitive trainer was deceptive, in the functional sense, but the process underlying this behavior is not clear. The animals may have intended to induce in the competitive trainer a false belief about the location of food, or they may have learned, through association or otherwise, that indicating the baited container in the presence of a trainer wearing green led to nonreward (Dennett 1983; Heyes 1993).

2.6. Perspective-taking

2.6.1. Seeing and knowing. It is a fundamental tenet of human folk psychology that seeing is believing. When individuals have had visual access to a state or event X, they are likely to know about X, but without that visual access, they are likely to be ignorant with respect to X. Consequently, if nonhuman animals were spontaneously to behave in a different way toward individuals when they have and have not had visual access to an event, and if this behavior were akin to what a human would do when they took another to be either knowledgeable or ignorant with respect to that event, there would be a strong prima facie case for mental state attribution by the animal. Several experiments on "perspective-taking" in primates (Cheney & Seyfarth 1990b; Povinelli et al. 1990; 1991; Premack 1988) have been based on this kind of reasoning.

Two studies of perspective-taking in monkeys (Cheney & Seyfarth 1990b; Povinelli et al. 1991) and chimpanzees (Premack 1988) reported failure to find evidence that the subjects understood the relationship between seeing and knowing, or had the concept of "see." In the remaining study (Povinelli et al. 1990), chimpanzees were tested in a two-stage procedure. At the beginning of each trial in the first discrimination training stage, 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 so 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.

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, and the second stage of the procedure was designed to find out whether this discrimination was based on the trainers' visual access to the baiting operation. In each trial of this transfer 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. As before, the chimpanzees were rewarded if they selected the container indicated by the Knower. For each chimpanzee, mean choice accuracy in the final 50 trials of stage 1 was comparable with that in the 30 trials of stage 2, and this transfer performance was taken to indicate that the chimpanzees were "modelling the visual perspectives of others" (Povinelli et al. 1990). However, performance at the beginning of the transfer test was at chance level (Povinelli 1994), suggesting instead that the animals learned a new discrimination, between bagged and nonbagged trainers, during the test period. Povinelli and his colleagues have subsequently acknowledged that their experiments using the knower versus guesser procedure do not provide compelling evidence that chimpanzees understand or postulate a relationship between seeing and knowing (Heyes 1994d; Povinelli 1994).

2.6.2. Seeing and attending. Povinelli and Eddy (1996) recently published a series of experiments using simple discrimination procedures rather than conditional discrimination training followed by a transfer test, as in the knower versus guesser experiments. In their view, these experiments addressed the question of whether chimpanzees

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

understand "the attentional significance of seeing," "the mental connection engendered by visual perception" (Povinelli & Eddy 1996), and their procedures represented a methodological advance because they "allow for a very sensitive diagnosis" of whether animals' behavior is guided by elements of a theory of mind or by processes described by "traditional learning theory."

In this series of experiments (Povinelli & Eddy 1996), groups of 6 to 7 chimpanzees aged 5 to 6 years were each repeatedly presented with two trainers whose appearance differed in one of a variety of ways; the animals were rewarded with food for making a begging gesture in front of one of the trainers. For example, in one treatment condition one trainer was facing the subject (S+) while the other stood with his back turned (S−); in another condition one trainer wore a blindfold around the eyes (S−) while the other wore a blindfold around the mouth (S+). In every condition, the chimpanzees were rewarded if they gestured to the trainer that a human adult would judge to be able to see the subject (marked S+ in the foregoing examples).

Several findings from these experiments led Povinelli & Eddy (1996) to conclude that young chimpanzees probably do not understand the relationship between seeing and attending: (1) In the three conditions in which the sight of one trainer was occluded by an object (bucket, blindfold, and screen), the chimpanzees showed no "immediate disposition" to gesture to the other person. That is, in early training under these conditions they did not show a preference for the person without occluded vision. (2) When the two trainers differed on four out of five "naturalistic" dimensions, the chimpanzees did not show a preference for the S+ trainer at any point in the course of the experiments. Thus, the animals showed a preference for a person facing them over a person with his head and back turned. However, they did not gesture more to a trainer looking back over his shoulder than to one with both head and back turned, to a trainer with hands over his cheeks rather than his eyes, to someone with eyes open rather than closed, or to a person looking directly at the subject rather a person with eyes averted. (3) The subjects' performance "showed a learning curve from Experiment 1 to Experiment 13." For example, in early experiments, the animals did not gesture more to a trainer holding a screen on his shoulder than to a trainer holding the screen in front of his face, but later they performed above chance on this discrimination. (4) In the "attending versus distracted" treatment condition, one trainer looked directly at the subject (S+), while the other looked up and to the side (S−). On these trials, the chimpanzees often turned their heads in the direction of the S− trainer's gaze, a behavior that is regarded by some developmentalists as indicating understanding of the seeing-attention relationship; but in spite of this the chimpanzees gestured at random to the two trainers.

These results provide no encouragement for the view that young chimpanzees understand anything about "seeing," but neither do they constitute compelling negative evidence; they should not persuade us that young chimpanzees do not understand "seeing." One would expect animals with the concept "see" to be capable of using the visibility of the trainer's eyes, not merely his face or the front of his body, as a discriminative cue for begging. This capacity would not necessarily become apparent on the first trial of a laboratory test, however, nor indeed at any point in the set of trials given in Povinelli and Eddy's study. Even if the chimpanzees had the concept "see" before the experiment began, it could take them some time to become convinced that it was the basis on which they were required to discriminate in this particular set of problems. Furthermore, since eyes visible versus invisible, and eyes direct versus averted, are perceptually fine discriminations, the chimpanzees may have neglected to try hard on those trials, opting instead to collect their rewards during the easier trials in which the difficult ones were embedded.

Could the procedures used by Povinelli and Eddy (1996) have provided positive evidence of theory of mind? The experiments were presented as if certain outcomes would have supported a theory of mind interpretation over a nonmentalistic account or, more narrowly, a learning theoretic explanation. If this were true, these procedures would represent a major methodological advance because, as I have argued above (see also Heyes 1993), no other methods used to date in research on theory of mind in nonhuman primates have succeeded in doing this. Unfortunately, however, Povinelli & Eddy's procedures cannot do it either. Simple discrimination techniques of the kind they used can tell us which observable cues chimpanzees use when deciding whom to approach for food, but they cannot tell us *why* the chimpanzees use those cues; whether certain cues are important to them because, within the chimpanzees' theory of mind, those cues indicate "seeing," "attention," or "knowledge."

Imagine, for example, that Povinelli and Eddy had found that all of their chimpanzees immediately showed perfect discrimination on the basis of the visibility of the trainer's eyes. Thus, from the very first trial, the chimpanzees not only preferred a trainer with a bucket on his shoulder to one with a bucket over his head, but also preferred a person with his eyes open over one with his eyes closed, and even preferred a trainer looking directly at the subject (irises visible as circles), over a trainer with their eyes averted (irises visible as ellipses). By hypothesis, the data would indicate unambiguously that chimpanzees use eyes as a discriminative stimulus when deciding which of two trainers to approach for food. Even these data would be equally compatible with a theory of mind and a nonmentalistic explanation, or, as Povinelli and Eddy put it, with a "mentalist" and a "behaviorist" hypothesis. A theory of mind account would say that chimpanzees use eyes as a discriminative stimulus because they understand that an individual whose irises are visible as circles can "see" them, and that seeing is a mental state linked to attention or knowledge. A nonmentalistic account would say that the chimpanzees just do it; they have a learned or unlearned tendency to beg from people with visible eyes, and while the chimpanzees may even know that begging from people with visible eyes is more likely to lead to reward, they do not explain this contingency to themselves in mental terms or in any other way.

Note that the essential difference between the theory of mind hypothesis and the nonmentalistic hypothesis does not relate to whether the use of eyes as discriminative stimuli was learned or unlearned. If the chimpanzees in Povinelli and Eddy's experiments had shown perfect performance from the first trial, both mentalistic and nonmentalistic accounts could have attributed this to preexperimental learning or to an innate disposition. (Even "traditional learning theory" does not claim that all behavior

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

is learned.) The difference is that a theory of mind hypothesis would say that it was an understanding of the seeing-attending or seeing-knowing relationship, as well as a tendency to use eyes as discriminative stimuli, that was present before the experiment began. Similarly, improvement in performance over recorded trials could be attributed on both mentalistic and nonmentalistic accounts to learning during the experiments, or to the gradual unmasking of some preexisting tendency. Thus, a theory of mind hypothesis might say that the chimpanzees learned about the seeing-attending relationship in the course of the experiments, or that they already knew about that relationship but needed to discover its task relevance or to learn some new cues instantiating the seeing relation. A nonmentalistic hypothesis might say that the animals learned through the experiment to use eyes as discriminative stimuli, or that a preexisting tendency to do this only became apparent when the animals had become fully accustomed to all aspects of the testing procedure. In this example, and in the search for evidence of theory of mind in nonhumans more generally, the crucial difference between mentalistic and nonmentalistic hypotheses lies in their claims about "what is known," not about whether or how knowledge is acquired.

In view of their discouraging findings with 5- and 6-year-old chimpanzees, Povinelli and Eddy (1996) recommended that older chimpanzees be tested for theory of mind competence. This is a useful suggestion, but, if there is to be any chance of finding positive evidence of theory of mind, different test procedures must be found.

2.7. Summary

Research on imitation and mirror-guided body inspection (sects. 2.1 and 2.2 above) has not shown unequivocally that any primate has these behavioral capacities, and they could, in any event, be the products of associative learning and inferences involving nonmental categories. Thus, for imitation and self-recognition, the answers to both competence and validity questions are negative.

There can be little doubt that the members of many primate and nonprimate species exhibit sensitivity to social relationships and behavior that functions to deceive other animals (sects. 2.3 and 2.5 above); hence the answer to the competence question is affirmative for both social relationships and deception. However, in every case the relevant behavior could be based on one or a number of nonmentalistic psychological processes, and therefore these behavioral capacities are not valid indicators of theory of mind.

The position with respect to role-taking and perspective-taking is more complicated. Premack and Woodruff's (1978) research on role-taking (sect. 2.4) provided the first and arguably the strongest evidence to date of theory of mind in a nonhuman primate (Premack & Woodruff 1978). It showed that a chimpanzee was capable of matching problem-solution images; she had this behavioral competence, and it is difficult, but not impossible, to query the validity of this competence as an indicator of theory of mind. In contrast, Povinelli et al. (1992a) did not show that cue detection training facilitates chimpanzees' performance in a cue provision task, or vice versa, and even if such an effect had been demonstrated, it would not necessarily indicate theory of mind. Therefore, for the cue detection/provision task studies on role-taking, the answers to both competence and validity questions are negative.

The knower-guesser procedure used by Povinelli and his associates to investigate perspective-taking (sect. 2.6; Povinelli et al. 1990) involved a transfer test procedure with considerable potential. It could, I will argue below (sect. 4), provide evidence of behavioral competence validly indicating that primates have the concept "see." However, as yet, the answers to the competence and validity questions are negative for all perspective-taking studies. Neither the knower-guesser procedure nor the simple discrimination tests used by Povinelli and Eddy (1996) have shown that primates use the visibility of interactants' eyes to decide whom to approach for food; and such evidence would not be sufficient to implicate possession of the concept "see."

3. Procrastination

Progress in answering Premack and Woodruff's question requires experimental designs and test procedures that can distinguish the theory of mind hypothesis from nonmentalistic accounts of primate behavior. This requirement has been explicitly acknowledged by a few researchers (e.g., Povinelli & Eddy 1996; Premack 1988). However, the primacy of the need has been obscured and attempts to meet it may have been retarded by various attempts to show that data of the kind surveyed in section 2 either favor a theory of mind hypothesis outright or at least provide "suggestive" evidence of theory of mind. These arguments typically concede that each item of putative evidence for theory of mind in primates is susceptible to alternative interpretations, and an appeal is made to parsimony or convergent evidence to break the tie. Five arguments of this kind (two appealing to parsimony and three to convergence) are evaluated in this section.

3.1. Parsimony

3.1.1. Simpler for them. In their seminal paper, Premack and Woodruff (1978) suggested that "the ape could only be a mentalist . . . he is not intelligent enough to be a behaviorist." This raises the possibility that the application of theory of mind (or "mentalism") requires less intelligence of an ape than alternative "behaviorist" methods of predicting behavior, and therefore, by appealing to Lloyd Morgan's Canon or a similar principle of parsimony, one could justify preferring a theory of mind interpretation of behavior over an alternative when both are consistent with the data.

There are two problems with this argument in favor of the theory of mind hypothesis. First, there is no good reason to suppose that the acquisition and use of a theory of mind requires less intelligence, or is in any sense "simpler," for an animal than the acquisition and use of an alternative basis for predicting social behavior. Neither intelligence nor simplicity has been defined or measured in a way that would allow a reasonable comparison to be made. Premack and Woodruff pumped the intuition (Dennett 1980) that an alternative to theory of mind would require more intelligence by dubbing it "behaviorist," and thereby suggesting that the animal would have to master the contents of the *Journal of the Experimental Analysis of Behavior.* However, if one resists this sort of intuition, it is clear that, although a more consistent analogy would portray chimpanzees that lack a theory of mind as "associationists" or "cognitivists" rather than "behaviorists," all of these characterizations are misleading because alternatives to the theory of mind

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

hypothesis do not assume that chimpanzees and other animals know anything about the processes that they use to predict social behavior. Only the theory of mind hypothesis takes chimpanzees to be students of their own psychology. It claims that mental states such as wanting and believing control behavior, and that knowledge of such states – mental state concepts – is used in social interaction. In contrast, alternatives to the theory of mind hypothesis postulate just one layer of processes or representations that generate behavior in social contexts and elsewhere.

Second, even if theory of mind were demonstrably less demanding of intelligence or simpler than the alternatives (or vice versa) this would not be sufficient to justify preference for one account over another. The view that preference for more parsimonious explanations can be justified by appeal to a general ontological assumption such as the uniformity of nature (Hume 1748/1948), has been broadly rejected by philosophers of science (e.g., Boyd 1985; Sober 1988). Therefore, in addition to showing that theory of mind would be simpler than the alternatives, it would be necessary to argue that in the case of primate social behavior, in this particular corner of nature, a simpler process is more likely to be in operation than a more complex one (Sober 1988).

3.1.2. Simpler for us. Dennett (1983; 1989) has argued that taking "the intentional stance" toward animals, characterizing their behavior in terms of the actor's intentional states, can have practical advantages. He claimed that for field ethologists observing animals in their natural environments, the intentional stance is easier to use than the languages of behaviorism or information processing, and that by happy coincidence intentional descriptions of animal behavior provide important clues for the cognitive scientists whose job it is to explain that behavior by modeling the information processing systems that are really in control.

As far as I am aware, no one actively engaged in research on theory of mind in primates has explicitly claimed, with or without reference to Dennett, that theory of mind explanations should be preferred to nonmentalistic alternatives because the former are simpler for (some) people to understand. However, the "simpler for them" argument is commonly advanced and yet weak (see sect. 3.1.1), raising the possibility that researchers are implicitly assuming that theory of mind is simpler for primates to use because theory of mind hypotheses are often simpler for us to understand. Accordingly, it is worth reflecting on the "simpler for us" argument.

The first thing to note is that Dennett's arguments cannot (and were not designed to) justify a preference for the theory of mind hypothesis over nonmentalistic accounts of the kind of evidence reviewed in section 2 (Heyes 1987). On the contrary, they imply that, although it is legitimate for field ethologists to speak and write about animals as if they had mental states and mental state concepts, the broader research community should seek, and indeed prefer as explanations, theories that do not make reference to such states and concepts.

Leaving aside Dennett's more subtle position, it might be argued that if the theory of mind hypothesis is simpler for us to comprehend than alternative accounts of primate social behavior, this would be sufficient reason to prefer it over nonmentalistic accounts. This argument assumes that the principle of parsimony or simplicity is "purely methodological" (Sober 1988); that, regardless of whether we can justifiably assume that nature is simple, it is rational to prefer simple theories (e.g., Strawson 1952).

Even if one accepts that the principle of parsimony is purely methodological (and Sober 1988, gives compelling reasons not to accept this), there is a problem with the argument that because it is simpler to comprehend the theory of mind hypothesis should be accepted instead of nonmentalistic accounts of the current data on social behavior in primates. It is not clear that the theory of mind hypothesis is simpler in a way that should carry any weight. For some people, for example, who are unfamiliar with associative learning theory and cognitive psychology, it may be easier to understand and apply. However, this does not seem to be the kind of simplicity that was at issue in the historical episodes that led to parsimony being viewed as a methodological principle (e.g., Reichenbach 1951; Sober 1988). For example, it is unlikely to have been a user-relative conception of simplicity – a dimension defined by individual scientists' professional and educational backgrounds – that guided Einstein's reasoning to the special and general theories of relativity.

3.2. Convergence

3.2.1. More is better. Much of the putative evidence of theory of mind in primates is anecdotal; it consists of reports of single occurrences of a behavior, under uncontrolled conditions, made by isolated observers, or groups of observers who share a theoretical base. The profound weakness of this kind of evidence has been demonstrated repeatedly (e.g., Kummer et al. 1990; Premack 1988), and yet anecdotes continue to be published and treated as persuasive. In most cases, this is done without commentary or defense and, to their credit, Whiten and Byrne (1988) stressed that anecdotes are a prelude, not a substitute, for more systematic research and offered a rationale for their collection of anecdotes about deceptive behavior in primates. They suggested that a collection of anecdotes relating to the same category of behavior will constitute evidence of theory of mind provided that (1) the reports come from independent observers, and (2) each provides evidence that the act involved the agent representing the viewpoint or beliefs of others.

Whiten and Byrne's second criterion seems to be self-defeating. Their "multiple records" approach is designed to compensate for the fact that single anecdotes cannot provide evidence of theory of mind, and yet their second criterion requires each anecdote in a collection to provide such evidence for the ensemble to be persuasive. Nor does combining the second criterion with the first offer an escape from this circularity. Consider the hypothetical example of three animals seen by independent observers (criterion 1) snatching food that was previously available to a conspecific. The first, like the baboon reported in Jolly (1985, see sect. 2.4), grooms the conspecific and snatches when it is 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. Each observer might feel inclined to attribute the state of "intending to deceive with intimate behaviour" to the animal observed (Whiten & Byrne 1988), but the potential to attract the same mental state attribution from the human

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

observers might be all that the three animals have in common with regard to mental state concepts. Even if we could be sure that none of them had simply been lucky and that all of them had acquired the behavior through some inferential process, the possibility would remain that the animals learned to snatch from supine, sexually excited, and departing individuals, respectively.

This example illustrates that "the plural of anecdote is not data" (Bernstein 1988), but the point can also be generalized: the mere accumulation of data, whether anecdotal, observational, experimental, or a mixture of the three, does not necessarily provide convergent evidence. The literature reviewed in section 2 shows that in a range of social interactions (e.g., competitive and cooperative; dyads, triads, and larger groups; same and different gender, status, age, and species; in relation to feeding, grooming, mating, and mothering), the behavior of many individual apes has been interpreted as a manifestation of theory of mind. But to make the case for the theory of mind hypothesis more compelling on the grounds of convergence, one would need to show not merely that it can be applied to diverse phenomena but that for each of a range of phenomena it provides a better explanation than alternative, nonmentalistic hypotheses.

3.2.2. Apes can and monkeys can't. Humans have a theory of mind; nonhuman apes are more closely related to humans than are monkeys; and according to one school of thought closely related taxa are more likely than groups with a more distant common ancestor to have the same cognitive capacities. Therefore, one might argue, if nonhuman apes perform better than monkeys on tests designed to assess theory of mind, then, all other things being equal, the difference between the two groups provides convergent evidence that the apes' successful performance on the tests is a product of theory of mind rather than nonmentalistic thinking.

Unlike "more is better," this is a potentially sound convergence argument. However, it does not succeed in breaking the current deadlock between the theory of mind hypothesis and nonmentalistic accounts of primate behavior because in tests where apes have fared better than monkeys all other things have not been equal. For example, Gallup and his colleagues (e.g., Gallup 1970; Gallup et al. 1971; Suarez & Gallup 1981) have found that chimpanzees and orangutans pass, but various species of monkey fail, the mark test of mirror self-recognition. This could be owing not to the presence of a self-concept in apes and a lack of the same in monkeys but to the fact that apes spontaneously touch their faces more often than do monkeys (Dimond & Harries 1984; Gallup et al. 1995; Heyes 1994c; 1995b; 1995c; see sect. 2.2 above). Similarly, using the task in which subjects must choose a container indicated by one of two people, the Knower or the Guesser, Povinelli et al. (1990; 1991) found that chimpanzees did – and rhesus monkeys did not – learn to choose reliably the container indicated by the Knower. But this may not reflect a difference between the two groups in the capacity to model the visual perspectives of others, or to appreciate that seeing leads to knowing. Rather, it may have occurred because in the monkey experiment but not in the chimpanzee experiment the Knower moved around the room after baiting and before the subject had its choice. Thus, it would have been more difficult for the monkeys to remember on any given trial which trainer had been present during the baiting.

To be effective, an argument from ape-monkey contrast to the conclusion that apes have a theory of mind would need to show that the contrast in performance could not plausibly be ascribed to differences in task demands, sensory or motor functioning, or central processes not specifically related to theory of mind (e.g., working memory). As the foregoing examples illustrate, this has not been achieved, even in those rare and admirable experiments that have compared monkeys and apes using common procedures.

3.2.3. Chimps are like children. Another potentially strong but currently ineffective convergence argument is the following: the performance of chimpanzees (and/or other nonhuman apes) on theory of mind tasks is likely to reflect the use of a theory of mind rather than nonmentalistic processing, because the chimpanzees' performance resembles that of children in similar circumstances and there is independent evidence, often from verbal measures, that the children's behavior is based on a theory of mind. Current evidence does not support this argument, however, because, in the very few studies that have compared the behavior of chimpanzees and children under similar circumstances, the resemblance between the two or the independent evidence that the children were using theory of mind is weak.

Experiments on imitation (Tomasello et al. 1993) and self-recognition (Povinelli et al. 1993) provide examples of the first problem: poor resemblance between chimpanzees and children. Tomasello et al. (1993) found that in terms of their tendency to duplicate a model's actions on objects, "enculturated" chimpanzees were more like children than were nonenculturated chimpanzees. Although the children imitated fewer actions at a delayed test than at an immediate test, the enculturated chimpanzees showed the reverse pattern of performance.

Povinelli et al. (1993) reported that the mirror self-recognition behavior of chimpanzees and children is alike merely in that each shows a developmental trend, yet even this very general resemblance was not confirmed by the results. Reanalysis of the data from this study[4] (Heyes 1995b) showed that older chimpanzees were no more likely than younger ones to pass the mark test of self-recognition; and although 8- to 15-year-old chimpanzees showed more self-directed behavior in the presence of mirrors than 1- to 5- year-olds, the frequency of this behavior declined sharply between ages 15 and 39. The latter finding suggests either that, unlike humans, (1) chimpanzees typically acquire a self-concept as children and then promptly lose it on reaching adulthood, or (2) that self-directed behavior in the presence of mirrors is not a valid measure of self-conception.

In a study of perspective-taking, Povinelli's group (Povinelli et al. 1990; Povinelli & deBlois 1992) sought and found a more precise resemblance between chimpanzees and children, but in this example there was no compelling evidence that the children's behavior was guided by a theory of mind. Povinelli & deBlois (1992) found that 4-year-old children were more successful than 3-year-olds on a task similar to the Knower versus Guesser discrimination problem previously given to chimpanzees (see sect. 2.6; Povinelli et al. 1990). This does not, however, indicate that the chimpanzees' success on the problem was based on an understanding of the relationship between seeing and knowing, because the children who consistently chose the

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Knower were no more likely than the unsuccessful children to answer correctly a question about what the Guesser could see when they had left the room.

3.3. Conclusion

Each of the foregoing parsimony and convergence arguments could be put into reverse to motivate acceptance of nonmentalistic accounts of the data reviewed in section 2. Thus, it could be argued that theory of mind would require more intelligence of primates because it involves more than one layer or level of representations (sect. 3.1.1), and that nonmentalistic accounts are simpler from the investigator's perspective because they proceed from clearly specified assumptions rather than a largely implicit folk theory (sect. 3.1.2). Similarly, appealing to the "more is better" principle (sect.3.2.1), one could point to all of the nonsocial behavior of people and animals that can be explained by nonmentalistic theories; and, countering the argument from ape-monkey contrast (sect. 3.2.2), one could draw attention to the nonprimate species (including rodents, birds, and arthropods) that exhibit the kind of behavior interpreted as evidence of theory of mind when it appears in primates. Finally, one might note that, when direct comparisons have been made, it has turned out that in important respects chimps are *not* like children.

All of these arguments could be made at least as plausible as their counterparts in the existing literature on theory of mind in primates, but, in my view, it would be a mistake to pursue this option. To answer Premack and Woodruff's question, we need more strong experiments, not more weak arguments.

4. Proposals

4.1. Methods and questions

I have argued in sections 2 and 3 that research to date on theory of mind in primates does not show that they have such a theory. I also believe that it does not indicate that primates lack a theory of mind, or that Premack and Woodruff's question is unanswerable. There may be circumstances in which repeated failure to find evidence confirming a hypothesis can be interpreted rationally as a sign that the hypothesis is false, and it is conceivable that theory of mind and nonmentalistic accounts of primate social behavior are observationally equivalent. However, both of these negative conclusions would be premature because very few deliberate, potentially effective attempts have been made to test the theory of mind hypothesis against nonmentalistic alternatives. Research on imitation (see sect. 2.1) and self- recognition (sect. 2.2) has been used opportunistically to support the theory of mind hypothesis, most having been conducted to address other questions; and the vast majority of studies of social relationships (sect. 2.3) and deception (sect. 2.4) have used observational or anecdotal methods that lack the potential to distinguish the theories because they provide no information about the animals' histories (Heyes 1993). Just a handful of studies – of deception (Woodruff & Premack 1979), role-taking (Premack & Woodruff 1978; Povinelli et al. 1992a), and perspective-taking (Povinelli et al. 1990) – have been designed to pit the theory of mind hypothesis against an alternative while using a potentially reliable method to do so. Further empirical studies of theory of mind in primates are accordingly needed and warranted, but which methods should they use, and what kind of behavior should they examine?

The foregoing analysis (sects. 2 and 3) yields six principal recommendations for future research on theory of mind in primates, none of which is entirely original.

(1) Studies should be designed to distinguish the theory of mind hypothesis from nonmentalistic accounts of social behavior in primates. There is little point in reporting any more observations that are consistent with both kinds of account, or conducting experiments for which they would both predict the same outcome.

(2) It should be recognized that alternatives to theory of mind hypotheses are not necessarily "behaviorist" or derived from learning theory. The social behavior of primates may be based on abstract, symbolic representations of nonmental categories.

(3) Whether they are field- or laboratory-based, studies of theory of mind should involve experimental manipulation. Certain experimental methods (e.g., Povinelli et al. 1990; Premack & Woodruff 1978; Woodruff & Premack 1979) have come closer than any observational study to providing evidence of theory of mind in primates, and, although there are plans in place to increase the effectiveness of these methods (see Premack & Dasser 1991, and sect. 4.2 below), it is not clear how any observational study could distinguish the theory of mind hypothesis from its nonmentalistic alternatives.

(4) Investigations of role-taking, deception, and perspective-taking are more likely than research on imitation, self-recognition, and social relationships to tell us whether nonhuman primates have a theory of mind. The problems with attempts to demonstrate imitation and mirror-guided body inspection in primates are not intractable (for potential experimental designs see Heyes 1994c; 1995b), but there is little reason to suppose that mental state concepts are involved in imitation, self-recognition, and the kind of behavior examined under the heading of "social relationships" (sect. 2.3).

(5) Experiments that use a common procedure to compare the behavior of monkeys, nonhuman apes, and children (or adults) are more likely to yield compelling evidence of theory of mind in apes than studies of apes alone.

(6) The knower-guesser procedure used by Povinelli et al. (1990; see sect. 2.6 above) to investigate perspective-taking is particularly promising. This "triangulation" method (Campbell 1953; Heyes 1993) consists of conditional discrimination training followed by transfer tests, and its power lies in the fact that it requires animals to distinguish one mental state, X (e.g., knowing where food is hidden), from another, Y (e.g., not knowing where food is hidden), in two or more situations that 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 behavior, 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 behavior 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

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

respectively, then one has evidence of the application of mental state concepts 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 fallible, but in a different way.

More generally, it would be desirable for researchers with different expertise and theoretical commitments to collaborate in planning studies of theory of mind in primates. Combining commitment to the theory of mind hypothesis with skepticism and skills in experimental design with knowledge of the habits and natural history of primates would guard against confirmation bias, and would maximize our chances of developing procedures that are both practicable and potentially effective in testing the theory of mind hypothesis against nonmentalistic alternatives. To make this implementation of the "fishscale model of omniscience" (Campbell 1969) more than simply a pious wish, I describe below a test procedure that looks to me as if it could yield evidence of perspective-taking in primates. *BBS* commentators are invited to say what is wrong with it and how it could be improved or replaced by a potentially more effective method.

4.2. A potential study of perspective-taking

Initially, adult chimpanzees would be tested for perspective-taking using a version of the triangulation procedure developed by Povinelli et al. (1990; see sect. 2.6 above). Departures from this procedure would include (1) the presentation of nonreinforced probe trials rather than a new discrimination problem, when the initial discrimination has been learned; (2) use of trainers wearing opaque or translucent goggles, rather than a bag-on-head manipulation, for transfer trials; and (3) introduction of a pretraining phase in which the subjects are exposed to opaque and translucent goggles with distinctively colored rims. The first of these would ensure that successful "transfer" performance could not be due to learning of a new discrimination (see sect. 2.6 above) and, in combination, the latter two features of the experiment would make it unlikely that the animals could solve the problem using an observable cue, such as "eye-object line" (Heyes 1994d) – that is, by choosing the trainer for whom there is or was an unobstructed, notional straight line between their eyes and the baiting event. Preexposing subjects to the goggles would allow them, if they have the concept "see," to discover that one pair of goggles permits the wearer to see, while the other pair does not. If they subsequently prefer to take their cue from a trainer wearing translucent rather than opaque goggles, and if the only observable indication of which goggles the trainer is wearing is an arbitrary one (i.e., rim color) then it would seem that the subjects' preference for a person wearing translucent goggles could only be due to their attributing sight of the baiting event to that trainer. Use of goggles in a similar context was recommended by Gallup (1985; 1988) and Nicholas Humphrey (personal communication), and goggles were used by Novey (1975) in a study of infants. Cheney and Seyfarth (1990a) also used a similar manipulation in an experiment with monkeys.

In more detail, the procedure would be as follows.

(1) *Pretraining.* The chimpanzees would be trained, if necessary with food reward, to cover their eyes with two pairs of goggles. The two pairs would have rims of different colors, say red and blue. For half of the animals, the red-rimmed goggles would be opaque and the blue-rimmed translucent, while the other half would have the reverse assignment. Neither at pretraining nor at any other time will the chimpanzees see another person or animal wearing goggles. Furthermore, the opaque and translucent versions should be discriminable at a distance, that is, when worn by another individual, only in terms of their rim color. To check that this is the case, an attempt would be made to train chimpanzees that are not taking part in the main experiment on a simple discrimination between a trainer wearing opaque and translucent goggles with rims of the same color.

If it was found during pretraining that chimpanzees are highly resistant to putting goggles over their eyes, or that any aversion to the opaque goggles does not habituate in the course of pretraining (a possibility raised by Perner 1991), or that willingness to wear the two sets of goggles cannot be equalized by appropriate distribution of rewards, then one-way and two-way silvered screens, with distinctively colored frames, could be used in place of opaque and translucent goggles.

(2) *Training.* Using an apparatus and procedure like those of Povinelli et al. (1990), each chimpanzee would be presented on each trial with four containers and two trainers. One of the trainers would leave the room while a third person baited one of the containers; then each trainer would point at a container, and the chimpanzee would be rewarded for selecting the container indicated by the trainer who had been present during baiting.

(3) *Transfer.* When the animals had reached criterion on the training problem, trials of the kind used in training would be interspersed with occasional probe trials, in which both trainers would remain in the room and put on goggles during baiting. The Knower would put on translucent goggles, and the Guesser would wear opaque goggles. The subjects would never, or always, be rewarded on probe trials, regardless of the container they chose. The important point is that they would not be rewarded consistently for choosing either the Knower or the Guesser. If chimpanzees have the concept "see," then on probe trials one would expect them to choose the Knower, wearing translucent goggles, more often than the Guesser, wearing opaque goggles.

If, in the foregoing experiment, chimpanzees did not show a preference for the Knower over the Guesser, it may be worth running a variant that would contain fewer irrelevant cues or distracters, would make less demand on subjects' working memory, and would not rely on test trials in which the subjects' motivation is uncertain because responding is not differentially reinforced. This variant would begin with the same pretraining and would subsequently involve a successive, rather than a simultaneous, discrimination problem, using rate of learning rather than performance under nondifferential reinforcement as a measure. Thus, at the beginning of each trial in the training phase, a chimpanzee and a human trainer would face one another in a modified Wisconsin General Test Apparatus containing two covered food wells. The trainer would then either look intently at the food wells as one of them was baited by a third party (front trials) or turn, so that during baiting the chimpanzee and the food wells were behind the trainer's back (back trials). A screen between the wells and the chimpanzee would allow the latter to see the trainer and that baiting was occurring, but not where the food was

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

placed. After baiting, the trainer would face the subject and indicate one of the wells by placing his hand on it and the chimpanzee would be free to choose one well to search for food. On front trials, the trainer would point at the baited well and on back trials he would point at the other well. When the subjects had learned to select the well indicated by the trainer on front trials, and the other well on back trials, the transfer phase would begin, in which the trainer would wear translucent or opaque goggles. For half of the subjects, the trainer would indicate the baited well on translucent trials and the empty well on opaque trials (Group Direct) and for the other half, the trainer would indicate the baited well on opaque trials and the empty well on translucent trials (Group Reverse). If chimpanzees have the concept "see," one would expect Group Direct to learn faster than Group Reverse in the transfer phase. That is, Group Direct should learn to choose the well indicated by the trainer on translucent trials and the other well on opaque trials faster than Group Reverse learns to choose the well indicated on opaque trials and the other well on translucent trials.

The logic of both experimental designs requires training on only one discrimination problem before the transfer phase. In practice, however, it might be advisable to train the chimpanzees before transfer on more than one pair of stimuli instantiating the see versus cannot see distinction. This would help to ensure that if the chimpanzees have or can acquire the concept "see," they know by the time the transfer phase starts that it is relevant to the tasks in hand.

If either of these experiments had the predicted outcome, it would be desirable to repeat it using children as subjects. Each child would be tested using the same basic procedure as the chimpanzees but would also be given another test, preferably one that had already been validated as a measure of the theory of mind competence in question. Correlation between performance on the two tests would constitute convergent evidence that first measured some aspect of theory of mind and would encourage its use with other nonhuman species, including monkeys.

It would be very surprising indeed if these experimental proposals turned out to be easy to implement and did not contain any logical flaws. Research on theory of mind in primates would have made more progress in the last 20 years if single, crucial experiments were a possibility and if an effective research strategy were easy to formulate. However, I hope the proposals will contribute, after modification and refinement through open peer commentary, to the development of an effective experimental program.

4.3. On killing joy

In one of his inspired baptisms, Dennett (1983) gave the name "killjoy hypotheses" to explanations of behavior that eschew ascription of higher order intentionality or theory of mind to animals. Plenty of killjoy hypotheses have been discussed in this target article, and they will, as Dennett recognized, provoke a negative reaction in many readers. The idea that primates have a theory of mind is important and intriguing, and a great deal of careful labor has been devoted to its investigation. Therefore, it can be disappointing and irritating to be reminded that there are other, less exciting explanations for the reported data, especially when the recognition of these other possibilities requires close examination of methodology. It can seem as if elegantly bold ideas are meeting carpingly narrow objections, and in such a contest our instincts, or at least my instincts, are not to shout for the methodologists. But it is precisely because Premack and Woodruff's question is important and intriguing that it warrants a reliable answer; and without some sober reflection, acknowledging the limitations of current research, we may never know whether nonhuman primates have a theory of mind.

ACKNOWLEDGMENTS

I am grateful to Anthony Dickinson, Nicholas Mackintosh, Euan MacPhail, Henry Plotkin, Phil Reed, Elliott Sober, Andrew Whiten, and several anonymous referees for their comments on earlier drafts, and to Linnda Caporael, Richard Darby, Dorothy Einon, Christa Foster, Mark Gardner, Tim German, Chris Mitchell, Tristan Nokes, Kate Plaisted, Elizabeth Ray, and David Shanks for many useful and enjoyable discussions of the content. I owe the idea for this paper, and a great deal more, to Donald T. Campbell.

This research was supported by a grant from the UK Biotechnology and Biological Sciences Research Council.

NOTES

1. This target article adopts a realist position on mental states. It assumes that for most adult humans mental states and mental state concepts play a causal role in the generation of behavior, and it asks whether there is evidence that this is also true of any nonhuman primates. From a behaviorist perspective, the question "Do nonhuman primates have a theory of mind?" may be either incoherent or a question about whether human observers are willing to describe the behavior of nonhuman primates using certain mental terms. In either case, detailed analysis of the evidence of the kind presented here is otiose; the question is unanswerable, or the answer, apparent in common experience, is an emphatic "yes." People spontaneously speak, not only of other primates, but of nearly all other living things, as if they had mental states and a theory of mind.

2. Tomasello and his colleagues have advanced the interesting and more general thesis that, as a result of their extensive interaction with humans, enculturated apes engage in forms of social cognition beyond the capabilities of wild monkeys and apes (e.g., Tomasello 1996; Tomasello & Call 1994; Tomasello et al. 1993). This thesis is not a focus of the present discussion because, although Tomasello et al. claim that the behavior of enculturated apes is "intentional," they apparently mean by this that it is directed toward some purpose and involves thought of some kind, not, more specifically, that it implies theory of mind or the attribution of mental states.

3. The potential significance of pointing is indicated by evidence that rhesus monkeys, which do not normally show pointing behavior, did not immediately succeed on their second problem when switched from cue provision to cue detection, or vice versa (Mason & Hollis 1962; Povinelli et al. 1992b). Hess et al. (1993) showed that a rhesus monkey, Scarlet, who does point, fared no better than her conspecifics when switched from cue provision to detection. However, as Hess et al. acknowledged, since Scarlet is a single animal who may not point as much as the average chimpanzee, these data do not rule out the possibility that chimpanzees' performance on both tasks is facilitated by a preexisting habit of pointing.

4. I am grateful to Daniel Povinelli for supplying, immediately and in full, additional data from the studies reported by Povinelli et al. (1993).

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Imitation and mirror self-recognition may be developmental precursors to theory of mind in human and nonhuman primates

Kim A. Bard

Department of Psychiatry and Behavioral Sciences, Human and Behavioral Genetics Laboratory, Emory University School of Medicine, Atlanta, GA 30322. **kbard@emory.edu**

Abstract: Heyes argues that nonhuman primates are unable to imitate, recognize themselves in mirrors, and take another's perspective, and that none of these capabilities are evidence for theory of mind. First, her evaluation of the evidence, especially for imitation and mirror self-recognition, is inaccurate. Second, she neglects to address the important developmental evidence that these capabilities are necessary precursors in the development of theory of mind.

The development of imitation in humans proceeds from early imitation of facial acts and expressions and simple finger or hand movements in newborns through to imitation of complex actions on objects and delayed imitation in 2-year-olds (e.g., Meltzoff 1996). It would be extremely inaccurate to state that these imitative acts are not reinforced explicitly and repeatedly by parents throughout the infants' first 2 years. Along these lines, Heyes misses the point of the "enculturated" versus laboratory reared experiments; enculturated chimpanzees develop within a rich social-communicative environment where they are treated consistently as intentional communicators and thus are able to enter rather easily into some aspects of human communicative systems, such as taking turns. In these cases, chimpanzees understand what is expected in the imitation game through years of playing it with human partners. In contrast, in the Custance et al. (1995) study, we had to begin by "teaching" the chimpanzees the rules inherent in the imitation game. We designed interactions through which the chimpanzees, Scott and Katrina, learned the important aspects of the imitation game; the chimpanzees had to watch the model act, waiting until the action was over before taking a turn; in addition they had to perform the explicit task, which was to copy the actions they saw the model perform upon the verbal prompt "Do this"!

We argue that social experiences in addition to mental representational ability are necessary for higher level imitation (Bard & Russell, in press; Custance & Bard 1994). Many argue that it is through the intersubjective interchange involved in imitative interactions that children develop the generalized ability to imitate what they see others do, to imitate the *goal* of the modeled action (for the definition of emulation: Tomasello et al. 1987) and eventually to imitate the models' intentions (e.g., Kugiumutzakis, in press; Trevarthen, in press) and "mind-read" (Whiten 1996).

Heyes's argument that mirror-guided mark-directed behavior is an artefact of anaesthesia is erroneous, as there are many studies that did not use anaesthesia and have found mirror-guided mark-directed behavior in chimpanzees, orangutans, and in some gorillas (Hyatt & Hopkins 1994; Lin et al. 1992; Miles 1994; Patterson & Cohn 1994). Moreover, she appears to have misunderstood the meaning of "mirror-guided" which is essential to the definitions of both self-directed behavior and mark-directed behaviors indicative of self-recognition. Mirror-guided means that the subject receives constant feedback as they are behaving, as a consequence of watching the mirror image of their concurrent movements. The idea that mirror-guided means "looking in the direction of the mirror" (Heyes, 1994, p. 912) is simply incorrect: behavioral definitions clearly differentiate between looking at the mirror as an object and looking at the mirror image.

Looking at the mirror image is further differentiated into whether the mirror image is treated as a social partner or as a reflection of the self (indicated as mirror-guided self-directed behaviors or mirror-guided mark-directed behaviors). My own research in this area has been designed to document the age at which chimpanzees develop mirror-self-recognition (Lin et al. 1992). In a recent follow-up study (Bard 1997), the age at which chimpanzees demonstrate mirror self-recognition has been narrowed to between 28 and 30 months (one of five 28-month-olds touched the mark while looking at their mirror image whereas both 30-month-olds in both studies engaged in mirror-guided mark-directed behavior). Contingent behavior is a special case of not fully formed self-recognition: the subjects watch themselves act or explore the one-to-one correspondence between their own movements and the simultaneous movements of the mirror image. In the follow-up study, one 24-month-old engaged in contingent behaviors as her most complex behavior: she was reaching her hands up over her head while sitting in the lap of one of her favorite caregivers and facing the mirror. As she was moving, her attention was caught by the movement of the mirror image. Looking at the mirror image, she paused in her upward movement, and moved her right arm horizontally very slowly back and forth while glancing from the mirror image of her hand back and forth to her own hand. After a minute she resumed her play initiation with the caregiver. Some argue that any mirror-guided *self-directed* behavior is indicative of mirror self-recognition as it reflects the ability to direct behavior to the self as a result of looking at the mirror image. It is clear, however, that contingent behavior is indicative not of self-recognition but the beginning of understanding that the image moves when the self moves (Brooks-Gunn & Lewis 1984).

It is important to note that the inter-observer reliability was determined to be high in both studies, which verifies that several independent observers could agree on coding these subtle and complex behaviors. In Lin et al. (1992) the interobserver percent agreement was 85% and Cohen's kappa (a statistic that corrects for agreement by chance alone) was .76. In the follow-up study with multiple observers, the percent agreement ranged from 88% to 95%, and Cohen's kappa from .73 to .85. These scores are considered good to excellent (Bakeman & Gottman 1986) and document that multiple trained observers were coding the chimpanzees' behavior the same way and consistently over time.

A mental representation of the self, indicated by mirror self-recognition, is a developmental precursor to a reflective self-awareness (e.g., Mitchell 1993). Higher level abilities such as empathy, self-concept, and/or deception are predicated on this reflective self-awareness. Thus, self-recognition is not indicative of theory of mind but it is likely that theory of mind cannot exist without a foundation of self-recognition. Similarly, imitation is a necessary precursor to theory of mind. Imitation reflects intersubjective knowledge that forms the basis of social, emotional, and communicative exchanges. Delayed imitation (holding in mind the mental representation of others' actions) is one way that children act as they see others act and it may also be important in the development of empathy and perspective-taking, leading to mental attributions of thoughts, beliefs, and desires of others that may differ from one's own.

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Why not ask "Does the chimpanzee have a soul?"

William M. Baum

Department of Psychology, University of New Hampshire, Durham, NH 03824-3567. **wm.baum@unh.edu**

Abstract: The question, "Does the chimpanzee have a theory of mind?" is logically identical to the question, "Does the chimpanzee have a soul?" It is a peculiarity of our culture that we talk about anyone having a mind, and such talk is unhelpful for a science of behavior. The label "killjoy hypothesis" is an ad hominem attack.

Heyes offers a critical review of the research on theory of mind in primates and, after much incisive reasoning, leaps to the incorrect conclusion that refining method would resolve the issue.

The problem is not empirical but logical. Premack and Woodruff's (1978) question "Does the chimpanzee have a theory of mind?" is neither intriguing nor important. It only begs the question of whether it is helpful for a science of behavior to use concepts like "theory of mind" at all. If it were helpful, then it would make sense to discuss what the criteria would be for deciding the question, but first one must decide whether such a concept advances our understanding of behavior at all.

If Heyes were to apply the same critical thinking to comparable research with children or, if there is any, to adult humans, we might profitably wonder whether any of those results require the notion of "theory of mind." Children learn, as chimpanzees must learn, on the basis of observable cues. Unless you believe in ESP, there is no other way. The author's discussion seems to take it for granted that adult humans have a "theory of mind," but does not entertain the possibility that what people in our culture learn is, when presented with certain observable cues, to talk about what another person sees, knows, thinks, wants, and so on. Such talk is learned because it is reinforced in our culture. Such talk is just more behavior to be explained.

Other cultures or world-views make no assumption that humans, let alone chimpanzees, have a mind, let alone a theory of mind. Whorf (1956) explains, for example, that the Hopi have no such terminology in their talk about events or behavior. Talking about an inner mind and an outer world in no way makes these things real. Such talk is only a peculiarity of our culture that has no more use for understanding behavior than the terms "sunrise" and "sunset" have for understanding the mechanics of the solar system. The most one might hope to learn from the Heyes's experiment or any other such experiment is what cues are required before someone in our culture with training in cognitive psychology will start to talk about "theory of mind" in a chimpanzee.

Three hundred years ago it seemed intriguing and important to ask whether animals had souls. Premack and Woodruff's question is little different from the question, "Does the chimpanzee have a soul?" How is it different? We could establish a set of observations which, if they occurred, would be the occasion for us to say that the chimpanzee had a soul. If it can be taught to comfort another chimpanzee that is in distress, or if it can be taught to genuflect in front of a statue of Jesus, then we might say the chimpanzee has a soul.

The real question, however, is why should we care? What is gained by such an exercise? There could be debates if experimenters failed to demonstrate that the chimpanzee has a soul. Perhaps the methods were inadequate, and with further training or just the right training, the required behavior would appear. There would be some who, following the theological equivalent of Dennett (1987) would regard explanations of behavior that omitted the soul, as "killjoy hypotheses." Perhaps the killjoys who propose such explanations should be burned as heretics or at least ostracized by the community. If the debate were about the soul we might be more able to see that the label "killjoy" is just an ad hominem attack.

So much easier to attack straw men

Richard W. Byrne

Scottish Primate Research Group, School of Psychology, University of St. Andrews, St. Andrews, Fife KY16 9JU, Scotland. **rwb;@st-and.ac.uk psych.st-and.ac.uk.8080/people/lect/rwb.html**

Abstract: Rather than dealing with the important issues in the interpretation of behavioural data, Heyes seems only to reiterate lessons well-learned before she first reviewed the topic of primate deception. She also appears to misrepresent a series of published analyses. Despite her emphatic denials, the commonsense view is the best: informed observations and experiments can both provide evidence of theory of mind.

Isn't it funny, that when you read a story in the newspapers that you know something about, it's always wrong? Yet all the *other* stories are so good that you keep buying the paper. I had the same paradoxical feeling reading Heyes's attempt to discredit 20 years' study of primate mental capacity. Of the six areas she targets, I know most about the work on deception.

In Heyes's version, primatologists have been very naive. They mainly collect anecdotes of deception, and pile them up – as if a lot of unreliable accounts added up to solid data. And worse, they don't seem to understand that when animals learn in mechanical ways, well understood by experimental psychologists who study the laboratory rat, the results can nevertheless be impressive – just as if the animals had human understanding. Primate researchers therefore unscientifically argue for mentalistic explanations when the evidence is more parsimoniously explained by nonmentalistic means. They just don't seem to realize that "whether they are field- or laboratory-based, studies of theory of mind should involve experimental manipulation" (sect. 4.1). If only they had seen the wisdom of this years ago, they might have got somewhere.

To someone involved more directly with the data, Heyes's description is not so much a caricature as wholly unrecognizable. To convince readers that primate researchers do not understand the power of nonmentalistic learning, Heyes uses an account of a baboon who lulls another by grooming and then grabs its food. Many readers of this or a previous article that used the same example (Heyes 1993) might imagine that the naive primatologists are being rebuked for one of their past mistakes. Yet at no point has anyone claimed this particular record to *be* evidence of theory of mind. Interpretations have varied as to whether it showed functional deception, "deliberate misdirection" (Jolly, 1985, p. 412), or "perhaps a coincidence and not directed to the goal of obtaining a profit" (Byrne & Whiten, 1990, p. 37). Not only was a mentalistic interpretation not claimed, but the opposite was explicitly laid out:

> There is an alternative explanation for this class of behaviour, one that does not impute a deceptive intention and will often be hard to refute: when the most preferred course of behaviour is thwarted by the "target," execution switches to the next most preferred activity (grooming, in the above example) which coincidentally distracts the target and allows the "agent" to switch back to his most preferred activity. (Whiten & Byrne, 1988, p. 238)

The suggestion that those who observe primates are unaware of nonmentalist alternatives is hardly justified. Consider just a few of many published warnings: "Since, even in our limited data set, developmental explanations could range from shaping by reinforcement to 'intelligent' switching of solution methods between different problems, it becomes important to establish the antecedents of intimate tactical deception by detailed longitudinal study" (Byrne & Whiten, 1985, p. 672); "But may not this be simply due to the fact that chance experience had led to a situation through which a hobbling gait had acquired the meaning of more petting and attention than usual?" (Mitchell, 1986, p. 25, quoting Morgan's [1900] rejoinder to Romanes's [1883] story of a dog's deceptive limp); "Very often, observers had not collected precisely the right background information to distinguish between the trial-and-error type of explanation and the cognitive, intentional one" (Byrne & Whiten, 1987, p. 56); "Throughout the target article we

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

implicitly assumed the presence of what seems dictated by conventional wisdom, namely, first-order (or even zero-order) intentionality, acquired by conditioning; only in a very few cases did we claim there was second-order intentionality. We have spelled out conditioning explanations for some tactical deceptions elsewhere (*ibid.*), but *BBS* space limits did not allow these relatively familiar arguments to be detailed" (Byrne & Whiten, 1988, p. 267).

Demolishing her straw man, set up with a single account of a monkey's trick, makes it possible to ignore entirely a serious analysis of observational data as potential evidence of theory of mind. This programme did not pile up anecdotes – as if one could salvage poor quality data with sheer quantity. Rather, the putative evidence was examined sceptically, and only 18 records out of a corpus of 253 were suggested as cases where the evidence "implies that the primate can represent the mental states of others" (Byrne & Whiten 1990). Even these rather compelling records could each be explained without invoking theory of mind, as we demonstrated (Byrne & Whiten 1991) – if plausibility were set aside. But how often *should* one prefer an implausible, complicated, nonmentalistic account, over a simple mentalistic one? Once, certainly; but would Ockham's Razor permit this 18 times? And should we ignore the fact that those cases were concentrated in a *single* taxon of primates (great apes) which overall contributed fewer records of deception than monkeys? Not agreeing with Heyes's principle that only experiments can give evidence of theory of mind, we argued that "the approach of filtering out any evidence which can be explained without allowing animals to attribute intentions should continue to be used" (Byrne & Whiten, 1991, p. 140). Confronted by the clear and simple pattern thus revealed, we concluded that "great apes demonstrate an understanding of deception for which we have no good evidence in monkeys" (Byrne & Whiten, 1992, p. 624).

Most compelling were observations in which an individual used an apparently novel procedure to deceive a conspecific, rather than some act within the species' known repertoire. The point is that novel deceptive tactics can only be computed by individuals who already represent the mental states of others, even if animals "reason" (itself a strong claim, which I am surprised to find Heyes treat as null hypothesis). Accurate delimitation of repertoire is not simple, and there is not yet enough observational data to be entirely sure that great apes understand deceit, but to ignore the best – as Heyes does – hardly seems prudent.

Contrary to Heyes's suggestion, there is no sound evidence against the commonsense proposition that both observational and experimental evidence will be useful in deciding whether any animal has a theory of mind. To date, what is striking is just how many experiments conceived with the aim of settling the issue have singularly failed to do so. In contrast, the proper scientific analysis of observational data has been sadly lacking in many areas: we need better ethology, not just more experiments.

Seeing is not believing

Gergely Csibra

Medical Research Council, Cognitive Development Unit, London WC1H 0BT, United Kingdom.

g.csibra@cdu.ucl.ac.uk **cdu.ucl.ac.uk/gergely/home.htm**

Abstract: Heyes's proposed study for testing whether chimpanzees have a theory of mind is (1) too strong because it requires that the animals apply mental concepts to the interpretation of both their own experiences and the behaviours of others, and (2) too weak because dispositional rather than representational understanding of "seeing" is sufficient to pass it.

I cannot disagree with Heyes in her conclusion that no existing study has convincingly demonstrated either the presence or the absence of a theory of mind in apes. If she had been as rigorous in defining what she means by "theory of mind" as she was in surveying the methodological and theoretical validity of the animal studies claiming to provide evidence on this issue, she would have been in a better position to avoid some flaws in her proposed study. I shall focus on two aspects of her proposal: one of them may require too much from the animal, hence making a potential negative result inconclusive; the other, in contrast, enables the animal to pass the test without using real mental concepts, hence questioning the validity of a potential positive outcome.

In proposing to introduce a pretraining phase in which the chimpanzees would wear the goggles to be used later in the training and transfer phases, Heyes relies on the tacit assumptions that (1) if chimpanzees have a theory of mind, they must apply mental concepts not only in explaining the behaviour of others but in interpreting their own experiences as well; that (2) chimpanzees can recognize the equivalence of the situations in which either they or the trainers wear the goggles; and that (3) they are capable of transferring "first person" conceptual knowledge to "third person" interpretation of the experiences of others. These are very strong assumptions indeed, especially in view of the fact that Heyes's review casts doubt on whether chimpanzees have a self-concept at all (see target article, sect. 2.2) and whether they can map actions between themselves and others, as required, for example, for imitation (see target article, sect. 2.1). But does having a theory of mind really imply the validity of these assumptions?

Note that the question is not whether human beings apply their theory of mind more or less equally to others versus themselves (they do), and not even whether the first person viewpoint is essential in the development and/or the functioning of human theory of mind (as simulationist theorists suggest, see, for example, Barresi & Moore 1996; Goldman 1993). The real question is whether it is conceivable in principle that a creature applies representational mental concepts (beliefs, desires, etc.), with all the computational requirements they imply, in predicting and explaining other creatures' behaviour but fails to appeal to them when it interprets its own experiences.[1] If the answer to this question is, as I believe, positive, then this hypothetical creature, however sophisticated a theory of mind it may use to interpret the trainers' actions, would certainly fail Heyes's test, since it would not have any relevant information about the implications of someone's wearing one or the other pair of goggles. One could correct this problem by replacing the "first person" pretraining phase by a "third person" one, so that the chimpanzees, by observing the trainers' behaviour when wearing the goggles, could learn about their effects, and would possibly interpret them in mental terms. This modification would enhance the validity of a possible negative result but would not solve the other, thornier problem of the proposal.

The central aim of Heyes's proposed study is to demonstrate that chimpanzees have a concept of "seeing." The logic of the study is the following: If (1) the animals can discriminate circumstances that physically allow the establishment of a "seeing" relation between (i) a person and (ii) a relevant aspect of the situation from other circumstances that do not allow this, and (2) they can make use of this discrimination in choosing the person whose behaviour is more likely to indicate the location of the reward, then animals are said to use a theory of mind. But what aspect of the experiment guarantees that it is mental concepts that mediate between the chimpanzees' understanding of the seeing relation and their expected tendency to choose the person who has been exposed to this relation? In other words, how can we ascertain that chimpanzees conceive the seeing relation as a cause of epistemic (representational) rather than behavioural (dispositional) changes in the trainer who saw the relevant event? Focusing on a minor methodological problem in Heyes's proposal will help make this distinction clearer.

In the second, easier variant of her proposed study, Heyes would train the animals in two kinds of trials: the trainer is either looking at the baiting event (front trials) or it is occurring behind him (back trials). After baiting, "on front trials, the trainer would point at the baited well and on back trials he would point at the

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

other well." What would a chimpanzee with a theory of mind learn from these responses? My theory of mind would suggest that the trainer is perfectly informed in both kinds of trials; how else could he point to the empty location in all the back trials? If he had no means other than visual to gather information about the location of the bait, then I would expect random guesses and, accordingly, correct direction in half of the back trials. But, instead, he appears to help me to the reward in all the front trials and to deceive me in all the back trials! In fact, what I would learn in this situation is that I should avoid using my theory of mind to predict the validity of his pointing action on the basis of his computable epistemic states, and that I should rather use the observed "seeing" relation between the trainer and the baiting as an indication of his disposition to point to the baited versus the empty location. Note that successful (i.e., rewarding) interpretations of the trainer's behaviour in this situation as well as successful transfer of this knowledge to the condition where the trainer wears the goggles do not depend on applying any mental (representational) concept. All I need is an understanding of a particular physical relation between the other person's eyes and certain parts of the environment (similarly to the understanding of the physical relation between a camera and the objects it can "see") and an associative link between the presence or absence of this relation and the behavioural dispositions of the person.

Now, if I could do reasonably well without mental concepts in this situation where I actually had to suppress my theory of mind because of a minor problem in the experimental design, what makes it necessary to use mental concepts in the first version of the study, which is free of this problem? Nothing at all. Although I may use a theory of mind in this version, were I lacking it, I could still get the rewards as long as I understood the physical conditions for "seeing." Seeing is a mental concept if, and only if, it refers to an epistemic relation between a mind and an object/event that is established in a particular (visual) way; but it is not a mental concept when it refers only to the physical relations that may or may not give rise to the epistemic relation. Accordingly, demonstrating that animals can understand such a physical relation and can use it as a discriminatory cue to predict the usability of people's behaviour is not sufficient evidence for applying mental concepts. What is needed in addition is to demonstrate that the animals conceive the result of seeing as a representational rather than a dispositional state.

Twenty years ago, three philosophers – Bennett (1978), Dennett (1978a), and Harman (1978) – independently proposed different versions of the same basic idea for testing whether Premack & Woodruff's chimpanzee had a representational theory of mind. Since this method was applied to test children's theory of mind (Wimmer & Perner 1983), a whole industry of testing false belief-understanding has grown out of these proposals. Is it really impossible to adapt this method to chimpanzees?

NOTE

1. A similar point was advanced, and remained unanswered, by L. H. Davis (1978) in his commentary to Premack & Woodruff's (1978) original *BBS* target article.

Apes ape!

Deborah Custance

School of Psychology, University of St. Andrews, St. Andrews, KY16 9JU, Scotland. **dcustance.gold.ac.uk**

Abstract: Heyes's claim that the only unequivocal evidence of motor imitation comes from rats and budgerigars is contested. It is suggested that the rats' behavior can be explained by emulation and the budgerigars' by response facilitation. Behavioral matching in chimpanzees (Custance et al. 1995; Whiten et al. 1996) is reconsidered and interpreted in terms of imitation.

I would like to respond directly to some of the criticisms Heyes makes of two experiments my collaborators and I conducted on imitation in chimpanzees (Custance et al. 1995). It is not that I disagree with Heyes's view that "a capacity to imitate is not a valid indicator of theory of mind" (sect. 2.1). Instead I wish to contest Heyes's claim that the only unequivocal evidence for imitation comes from rats and budgerigars.

Heyes et al. (1992) found that rats compensated for a 180° difference in perspective and pushed a joystick in the same direction as a conspecific demonstrator (either to the left or right). However, the rats' behavior can be explained in terms of emulation combined with stimulus enhancement. The rats could have compensated for the difference in perspective in terms of the movement of the bar (emulation) rather than the behavior of the demonstrator (imitation). Although it is true that the rats in a control condition did not respond when the joystick moved automatically, independent of a demonstrator, it could have been that the movement of the joystick was not sufficiently salient unless a conspecific was touching it (see also Byrne & Tomasello 1995).

Heyes also argues that Galef et al. (1986) provide evidence of imitation in budgerigars. Two groups of budgerigars differentially reproduced alternative methods of pushing a lever: with either the beak or the foot. Both foot and beak grasping are species-typical behaviors in budgerigars. Hence their behavior can be explained in terms of a response facilitation (the reproduction of actions already within the subject's behavioral repertoire – Byrne 1994).

Heyes interprets the behavior of two chimpanzees tested by Custance et al. (1995) in terms of matched dependent learning (MDL) – where the demonstrated behavior acts as discriminative stimuli for a matching response. MDL is a good way of characterizing the 15 *taught* actions. The chimpanzees' behavior was shaped so that they would reproduce 15 different modeled acts after the verbal command, "Do this." However, MDL cannot explain how the chimpanzees were able to match *novel* demonstrated actions. Heyes suggests that they could be responding, by means of generalization, with one of the conditioned responses, then elaborating on that in a random fashion, and that by chance these random elaborations matched the novel demonstration more closely. It is true that many of the chimpanzees' behaviors appeared to be novel elaborations on taught actions. However, there are very many possible variations or elaborations that the subjects could have made on each of the taught acts. If these elaborations were random, as Heyes suggests, we would expect that with each subsequent response the chimpanzees' behavior would be just as likely (if not more likely) to diverge in similarity from the modeled act than to converge with it. The pattern of the chimpanzees' responses does not resemble that predicted by Heyes's hypothesis. If one studies the detailed descriptions of each of the chimpanzees' responses described by Custance (1994), one can see that they tended to improve consistently across the 3 to 4 demonstrations of the novel act rather than showing a random pattern.

There is a small subset of the chimpanzees' responses which were very different from any of the taught items. Heyes explains away "lip smacking" in terms of previous unconscious conditioning by the chimpanzees' human caretakers. Although I believe such an explanation to be highly unlikely, it is difficult to categorically prove otherwise. However, there are other examples of the chimpanzees' behavior which are much more difficult to explain away. For example, both chimpanzees accurately matched the novel demonstrated act of "touch back of head." One chimpanzee, Katrina, also accurately matched the novel demonstrated act of "peekaboo," where the hands were held up flat in front of the face and moved apart and together again on the lateral plane. It is implausible that the chimpanzees had been inadvertently rewarded by caretakers in the past for matching these behaviors and both "touch back of head" and "peekaboo" were very different from any of the taught actions.

Heyes also interprets the matching of chimpanzees to a "twist-pulling" action used to remove a pair of plastic rods from tubular

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

brass holders on an artificial fruit in terms of emulation (Whiten et al. 1996). She argues that what the chimpanzees could have learned was not the twist-pull technique used by the demonstrator, but rather that the rods move in a certain direction and rotate. It seems highly implausible that the chimpanzees inferred the twisting action from watching the rotation of the rods alone. It was very difficult for an observer to see that the rods rotated independent of the demonstrator's behavior. The front end of the rod was almost completely obscured from view by the grasping hand of the demonstrator. Only one inch of the far end of each rod was visible as it protruded from the tubular brass holder. Each rod was twisted and pulled four times during demonstration. After one or two twist-pulls the far end of the rod was no longer visible and the front end was still obscured by the demonstrator's hand. Therefore the most salient aspect of the "twist-pull" demonstration was not the rotation of the rod itself, but the demonstrator's behavior, and the chimpanzees' matching response is best interpreted in terms of imitation rather than emulation.

In conclusion, although the chimpanzees' responses in both experiments could not be considered perfect imitations, they do provide more convincing evidence of capacity for visual-motor cross-modal matching than either rat or budgerigar data do.

Mirrors and radical behaviorism: Reflections on C. M. Heyes

Gordon G. Gallup, Jr.

Department of Psychology, State University of New York at Albany, Albany, NY 12222

Abstract: Heyes's attempt to reinterpret research on primate cognition from the standpoint of radical behaviorism is strong on dialogue and debate but weak on evidence. Recent evidence concerning self-recognition, for example, shows that her arguments about differential recovery from anesthetization and species differences in face touching as alternative accounts of the behavior of primates in the presence of mirrors) are invalid.

Heyes's critique culminates an extended attempt over the past several years to apply radical behaviorism to a variety of different phenomena that have been taken as evidence of mental/cognitive processes in primates. Heyes has written at least two book chapters and seven journal articles in an effort to reinterpret these phenomena from the standpoint of radical behaviorism. Yet despite all the scarce archival space devoted to her arguments, many of which have no basis in fact, she has yet to collect a shred of experimental evidence of her own on primates that bear on any of these issues. Moreover, many of the research designs (including the one in her target article), in which she suggests a means of providing more definitive tests of mentalistic phenomena, are not, in fact, her own.

Heyes's arguments concerning self-recognition are a case in point. Heyes contends (in the absence of any evidence) that the reason chimpanzees and orangutans touch facial marks (applied under the effects of anesthesia) more frequently later in the presence of mirrors is not that they recognize themselves but that there are higher ambient levels of spontaneous face touching because of continued recovery from the effects of anesthetization. Heyes also attempts to dismiss the well-documented species differences in mirror self-recognition on the grounds that they are confounded by species differences in the incidence of face-touching behavior. It should be noted, however, that there is reason to question whether such differences even exist (see Suarez & Gallup 1986a), and if they do, unlike the data on self-recognition, they are most certainly relative rather than absolute. Heyes also ignores the fact that there are clear instances in which chimpanzees and orangutans use mirrors to inspect, manipulate, and explore facial as well as nonfacial parts of themselves that can only be seen in mirrors (e.g., Eddy et al. 1996). No one has ever reported compelling, replicable instances of comparable mirror-mediated behavior in monkeys.

Heyes expresses surprise that no one has attempted to compare the rate at which chimpanzees touch marked and unmarked portions of the face in the presence and absence of mirrors as a test of her anesthetization hypothesis, and she cites herself as the source of this idea. However, it is essential to note that not only should we be credited with devising this strategy in the first place (Gallup et al. 1995), but we have now applied this technique to a number of chimpanzees (Povinelli et al. 1997). In contrast to what Heyes's model would predict, when chimpanzees were tested for self-recognition, both the frequency and duration of touches to marked and unmarked portions of the face were the same in the absence of the mirror, but shifted almost exclusively toward marked areas when the mirror was made available. This shift toward investigating marked portions of the face was conditional and immediate upon exposure to the mirror and declined thereafter. Moreover, there was no overall increase in face touching from the control to test period, as would have been predicted by Heyes's recovery from anesthetization model. Not only do these results falsify Heyes's hypothesis, but her whole house of cards concerning species differences in self-recognition comes tumbling down as a consequence, since these results render face-touching behavior per se an inadequate explanation for the results of mark tests.

Heyes contends that in contrast to behavioral primatologists, developmental psychologists have established strong empirical methods for investigating mentalistic phenomena in children. However, research on self-recognition in nonhuman primates has typically been far more rigorous and more sophisticated than comparable work on human children (e.g., Gallup 1994).

Finally, it should be noted that Heyes erroneously credits Gallup (1982) and Povinelli (1995) with the claim that imitation requires mental state attribution. She also incorrectly credits Cheney and Seyfarth (1990a) and Jolly (1991) with experiments showing that chimpanzees and orangutans can but other primates cannot correctly decipher mirrored information about themselves.

Assessing theory of mind with nonverbal procedures: Problems with training methods and an alternative "key" procedure

Juan Carlos Gómez

School of Psychology, University of St. Andrews, St. Andrews KY16 9JU, United Kingdom. **jg5@st-andrews.ac.uk**

Abstract: This commentary criticizes nonverbal methods of assessing theory-of-mind on the basis of prior training of the critical response because they would encourage simple, nonmentalistic, associative solutions even in subjects with mentalistic capacities. I propose instead a new experimental paradigm based upon the use of spontaneous responses in less artificial situations. This method has already provided positive evidence of some level of ToM understanding in nonhuman primates.

Problems with training experiments. The procedures discussed by Heyes for assessing theory of mind (ToM) in nonhuman primates are based on prolonged training of the subjects in artificial tasks. I wish to argue that this is a mistaken approach. In procedures based on previous training of the critical response, the subjects may learn associations that were not foreseen by the experimenters. For example, in the goggles experiment Heyes assumes that the cues of "seeing" versus "not seeing" will be those provided by the eyes. However, head and body orientation also provide important cues about who is seeing what. The subjects could learn to respond to these grosser cues (or to the whole Gestalt head-body-eyes-oriented-to-target). Thus, during the first phase (goggles experiment, second version; sect. 4.2, para. 6), the trainer is supposed to "look intently at the food wells" during

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

baiting. The topography of "looking intently" is not described, but presumably it would involve keeping some head/body orientation (with open eyes) towards the food wells. Now, if these extra-ocular cues of "looking intently" are kept in the probe trials with the goggles, the chimpanzees may be responding to them: although the eyes are no longer visible, the rest of the Gestalt components may be enough to evoke the response. If the extra-ocular cues were eliminated in the probes, subjects may start responding at random because they have gotten used to the mechanical routine of responding to the grosser cues, or alternatively, they could (mentalistically) interpret the absence of body/head orientation in the experimenter with the translucent goggles as evidence of inattention.

Any attempts to overcome these problems by eliminating the extra-ocular cues during the training phase would lead to more training and increasingly bizarre situations for which the chimpanzees would have to learn specific (nonmentalistic) associations different from the (perhaps mentalistic) strategies they would use in natural situations. The possibility that in experiments based upon previous training subjects are trained *away* from mentalistic strategies would cause an underestimation of their ToM abilities. For these reasons, I suggest that a good ToM test should rely upon "natural" situations that do not require training of the critical responses. Chimps should not be taught who knows and who ignores something: if they are mentalists, they should recognize this from the first experimental trial. Children are not trained in preliminary versions of the Sally/Ann or Smarties tests. They pass or fail if they give the correct or incorrect answer at once. What we need for the apes is tests in which the target response occurs as a spontaneous reaction of the animal to the experimental manipulation. Let me illustrate this with an experimental paradigm that has already been empirically tested.

A "key" experiment of ToM. The subject sits in a cage. In front of the cage there are two boxes locked with padlocks. The keys to the padlocks are kept in a different container. This scenario is installed by the participants in full view of the subject. A "Caterer" enters the room, takes the keys, opens one padlock, baits the box, closes the padlock, returns the keys to the container, and exits. Some seconds later, the "Giver" enters, sits in front of the boxes and "asks" the subject where the food is (or simply waits for the subject to make a request). When the subject points to one box, the Giver fetches the keys, opens the padlock, gives the subject the food, closes the padlock, returns the keys, and exits. This is repeated several times.

Next, an experimental trial is introduced: the Provider, after baiting the box, takes the keys to a hiding place within the room and then leaves. If subjects understand the mental state of "ignorance," then when the Giver comes in they will point not only to the food, but also to the keys. At least two more experimental trials are run, interspersed with ordinary trials. Subjects who have an understanding of the mental state of ignorance would have to pass all three experimental trials at once (or a substantial proportion of whatever number of trials is conducted).

Control trials must also be performed, of course. The most important is the following: the relocation of the keys occurs in front of the Giver or the Giver himself does it. In these cases the subject should not point to the keys, even if they are in an unusual location. The beauty of this procedure is that it allows for these and other controls (Gómez & Teixidor, in preparation) and for different kinds of experimental trials that would make it possible to analyze in detail what counts for the subject as causing knowledge or ignorance.

The feasibility of the procedure has already been tested with a nonhuman primate. Dona, a female orangutan, failed six experimental trails like the ones just described, hence not showing evidence of ToM. However, in a subsequent run of the experiment with the same subject, in which the key was relocated by a Stranger who entered the room after the Provider's exit, the orangutan responded correctly in all 7 experimental trials. This performance was suggestive of mentalistic understanding. However, since she had witnessed the human looking for the keys in the wrong location during the six unsuccessful trials, her good performance could be affected by some learning.[1] But what counts here is that the *procedure* is feasible and serves as a more promising nonverbal method to investigate ToM. This is further highlighted by the recent results obtained by Whiten (in press) applying this paradigm to a "linguistic" chimpanzee who passed the test from the very beginning.[2]

I suggest that this nonverbal situation is a true test of mentalism, one that is flexible enough to allow for the introduction of modifications and controls to eliminate any potential problems without changing the basic paradigm. It illustrates the alternative approach of not training the "key" response, but leaving it to be spontaneously produced by the subject.

ACKNOWLEDGMENT

Commentary was written under DGICTY, Grant PB95-0377.

NOTES

1. A partial report and discussion of this application of the procedure can be found in Gómez (1996). An extensive description and discussion of both the procedure and its application to the orangutan (Gómez & Teixidor, in preparation) can be obtained from the author.

2. Making use of the flexibility of this "key" paradigm, Whiten modified the critical trial by having the Giver bait the box himself.

The prior question: Do *human* primates have a theory of mind?

Robert M. Gordon

Department of Philosophy, University of Missouri-St. Louis, St. Louis, MO 63121. **srmgord@umslvma.umsl.edu**
www.umsl.edu/~philo/vitaes/gordon.html

Abstract: Given Heyes's construal of "theory of mind," there is still no convincing evidence of theory of mind in *human* primates, much less *non*human. Rather than making unfounded assumptions about what underlies human social competence, one should ask what mechanisms other primates have and then inquire whether more sophisticated elaborations of those might not account for much of human competence.

Heyes concludes that "there is still no convincing evidence of theory of mind in [nonhuman] primates." This should have been a foregone conclusion, given the author's intellectualistic construal of "theory of mind": as demanding that primates *believe* that mental states play a causal role in generating behavior, *infer* the presence of mental states in others from their behavior, and *apply law-like generalizations* to predict and explain behavior. On this construal, one familiar with the recent literature ought to conclude that there is still no convincing evidence of theory of mind in *human* primates (Carruthers & Smith 1996; Davies & Stone 1995a; 1995b). Were this the only "mentalistic" option available to account for human competence in the anticipation, explanation, and social coordination of behavior, a good case could be made for a purely nonmentalistic account of the social behavior of human primates. Small wonder, then, that the social behavior of monkeys and apes, sophisticated as it is, has not been shown convincingly to spring from a theory of mind.

In its broadest sense, the term "theory of mind" is used, especially in developmental psychology, to designate the resources, whatever they may be, that human beings routinely call on in the anticipation, explanation, and social coordination of behavior. Thus a goal of much recent work on the development of theory of mind in children has been to determine whether children are in fact developing a theory (in a narrow sense of the term) or enhancing a capacity to simulate others in imagination (Gopnik & Wellman 1992; Harris 1992; Perner 1996). Using the term "theory of mind" in this broadest sense, to ask, "Do nonhuman primates have a theory of mind?" would be to ask, "Do they have what *we* have?" The latter question – with the qualification, "if

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

only in a rudimentary form, with a lower demand on higher cortical function" – seems to me to be a much more promising topic for research in comparative primatology than the heavily loaded either–or question Heyes seems to be asking: "Do they have a theory of mental states, with law-like generalizations, inferences from behavior to mental states, and so forth; or do they have only nonmentalistic resources?"

The thesis that if it's a mentalistic resource then it's a theory in a full-blooded sense is just one of the important undefended assumptions in Heyes's target article. Another is that our own success in psyching out our conspecifics is chiefly due to our capacity to attribute mental states to them, whether a theory underlies this capacity or not. Even this is highly questionable. For one thing, a very important part of our social behavior – our emotional responses to ephemeral shifts in another's vocal and facial expression – seems chiefly to rely on fast processing that does not await causal analysis, for example, the determination that the expression stems from anger, from fear, or from some other emotion. Another important bit of social behavior is that of turning one's eyes to triangulate with another's line of gaze, alternating one's glance between the other's eyes and the surrounding scene to confirm convergence on the same target; and the related tendency to protodeclarative pointing. Gaze-following behavior has not been shown to depend on a prior capacity to attribute mental states; yet, because it typically calls one's attention to what they are gazing at and responding to, it is clearly an important contributor to our capacity to anticipate and explain the behavior of others. Role-taking, too, has not been shown to require a prior, independent capacity to attribute mental states, to oneself or to others, much less a theory of mind in a robust sense; yet, it may be, as simulation theorists claim, an extremely important contributor to our ability to anticipate and explain the behavior of others.

Much of the behavior mentioned in the foregoing paragraph has been observed in nonhuman primates. That already suggests a partially positive answer – "Yes, at least in some degree" – to the question, "Do they have what we have, if only in a rudimentary form?" One of the morals to be drawn from Heyes's mistakes is that we should not interpret this question unilaterally: as, "We know what we have, now let's ask how much of it they have." Rather, one should independently ask what mechanisms *they* have that enable them to anticipate and explain behavior, and then inquire whether more sophisticated elaborations of these mechanisms, made possible by greater cortical power, could not account for much or even all human competence in the anticipation, explanation, and social coordination of behavior.

Theory of mind in nonhuman primates: A question of language?

Colin Gray and Phil Russell

Department of Psychology, University of Aberdeen, Aberdeen AB24 2UB, Scotland.

c.gray@abdn.ac.uk **www.psyc.abdn.ac.uk/dept/staff/gray.html**

Abstract: Two substantive comments are made. The first is methodological, and concerns Heyes's proposals for a critical test for theory of mind. The second is theoretical, and concerns the appropriateness of asking questions about theory of mind in nonhuman primates. Although Heyes warns against the apparent simplicity of the theory of mind hypothesis, she underplays the linguistic implications.

In subjecting to critical scrutiny the claim that some apes can "impute mental states to themselves and others" (Premack & Woodruff 1978, p. 515), Heyes has rendered researchers in this area a very considerable service. She rightly draws attention to and refutes some questionable arguments which have been used to bolster the case for theory of mind in nonhuman primates (sect. 3). She also shows the weakness of much of the evidence adduced in favour of the hypothesis (sect. 2).

Heyes's review of the evidence for theory of mind in six different areas (sect. 2) is admirable. She argues, in general, that an experimental approach is the only one capable of adequately pitting the theory of mind hypotheses against the alternatives and, in particular, that the paradigm followed by Povinelli et al. (1990) is the one most likely to resolve the issue of whether apes have theory of mind. The strength of such a triangulation strategy, she argues, lies in the fact that in the transfer discrimination, as many as possible of the stimuli that could have falsely cued the first discrimination are now uncorrelated with the alternatives in the new choice (sect. 4.1, para. 6).

But theory is one thing, practice quite another. Even leaving aside the general issue of whether any psychological experiment ever leaves absolutely no uncertainty, there may be some specific problems with the procedures suggested by Heyes. Following some general recommendations (sect. 4.1), Heyes offers (sect. 4.2) two scenarios, the first being a modified Povinelli experiment, the second taking a successive discrimination approach. Although both approaches embody what appear to be substantial improvements upon previous paradigms, it remains to be seen how feasible they really are. Heyes herself questions the compliance of the apes to the wearing of opaque goggles. This problem might well persist even if the appearance of the lenses of the two kinds of goggles were made more similar. Moreover, the experience of Povinelli et al. (1990) with colour-cueing did not indicate that the apes were receptive to such cues. The probe trials may overcome the problem of the animals learning another discrimination afresh; but they may make the training phase too confusing for them. In the successive discrimination scenario, the trainer's turning away introduces the same confounding variable of facial continuity that was present in Povinelli et al. (1990). And in a different way, the successive discrimination may also impose too great a load on working memory.

It is to be hoped that further research along the lines suggested by Heyes will prove fruitful, but in view of the potential problems with the Povinelli paradigm, it may be wise to keep the experimental portfolio broad enough to include some of the simpler methods Heyes has also reviewed.

Turning from methodology, we must now consider the more fundamental question of whether it is even appropriate to seek evidence for theory of mind in nonhuman primates. Arguably, the imputation of mental states to others in order to explain and predict their behaviour is not only mentalistic, but also *linguistic;* indeed, it is difficult to conceive of reflection upon mental states without a carrying language of some kind. Is it, then, sensible to ask a question that supposes language in the ape?

Heyes recommends that we use a methodology with which theory of mind could be demonstrated in children as well as in apes (sect. 4.1, para. 3). This, however, begs the question of whether such a paradigm could ever be found. It may be more than coincidence that whenever the same test of theory of mind has been used with both apes and children, the latter have given no evidence of using theory of mind (sect. 3.2.3). Although other evidence for theory of mind, even in very young children, is strong, it derives from tests that could not be used with apes, because of the linguistic component.

The link between the acquisition of language and the acquisition of theory of mind, though imperfectly understood, is a strong one. Peterson and Siegal (1995) found that a standard theory of mind test (passed by most hearing four-year-olds) was failed by much older prelingually deafened schoolchildren (of normal nonverbal intelligence) who had been raised in a spoken language environment. Deaf children are characteristically slow to acquire language – unless their parents are deaf signers. The difficulty the deaf (but otherwise healthy) child experiences with theory of mind tasks underlines the theoretical implications and difficulties of applying the theory of mind hypothesis to nonhuman primates.

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Anecdotes, omniscience, and associative learning in examining the theory of mind

Steven M. Green, David L. Wilson, and Siân Evans
Department of Biology, University of Miami, Coral Gables, FL 33124-0421.
sgreen@umiami.ir.miami.edu **fig.cox.miami.edu**

Abstract: We suggest that anecdotes have evidentiary value in interpreting nonhuman primate behavior. We also believe that any outcome from the experiments proposed by Heyes can be interpreted as a product of previous experience with trainers or as associative learning using the experimental cues. No potential outcome is clearcut evidence for or against the theory of mind proposition.

1. Anecdotes and ethology. The issue of whether some animals have a theory of mind has by no means "dominated the study of . . . social behavior in nonhuman primates" (sect. 1, para. 1). It has hardly caused a ripple of interest among most primate field workers. Ethologists implicitly accept it, as Heyes indicates (sect. 2, para. 1), much as we accept such a theory for people. The reasons are clear: our daily observations ("anecdotes") make it the most plausible interpretation of many kinds of behavior; it offers the clearest explanation for many problem-solving actions. Imputing a theory of mind to other beings is a product of everyday experience with other primates, whether human or nonhuman. Is the view that some nonhuman primate behavior is best explained by a theory of mind any less valid than such a view of people, one based on similar evidence, namely extensive experience?

In neither case do we have a critical experimental examination yielding only a single possible explanation. Beyond personal experience (of one's own mind) and the false-belief test experimental paradigm, the main evidence favoring a theory of mind in people is verbal explanations of mental processes that confirm our introspections. For nonhuman primates, the only possible parallels are experiments (sect. 2 below) and narrative descriptions of behavior that can be analyzed and interpreted.

Anecdotal descriptions have proved to be illuminating in chimpanzee self-recognition experiments (Gallup 1970) and anecdotes collected in natural settings can be similarly so, particularly with detailed knowledge of social context acquired in field studies. They may be the only means of recording rare events or complex interactions critical to understanding the behavior of primates. To ignore these events would impoverish our ability to investigate theory of mind. Recorded objectively, these are data concerning social interactions, problem solving, and so forth, that are subject to analysis and interpretation just as experimental variables are. As has been the case for human beings, a careful examination of observations accumulated from field and lab, both planned observations and experiments or serendipitous events, may provide more compelling evidence than relying on any single avenue in addressing whether nonhuman primate behavior is best explained by non-mentalistic hypotheses or by having a theory of mind.

2. Interpretations of experimental results. Let us assume that Heyes's experiment produces results that lead her to conclude the subjects have a concept of "see." This is not clear evidence that they possess a theory of mind because there are many alternatives, including explanations based on associative learning or that the subjects have the concept, however derived, that the Knower has superior knowledge.

Even results of the false-belief test on children (e.g., Baron-Cohen et al. 1985) can be explained by associative learning. Subjects are asked to indicate the belief of two dolls, observing (witness) and nonobserving (absent), when a token (marble, candy, etc.) is moved from an initial location "seen" by both to another. A theory of mind is imputed if subjects, queried as to where each doll would look for the token, point to the container where each doll last "saw" the token, thus demonstrating that the nonobserving doll holds a false belief and putatively illustrating that subjects have the concept of "see." An alternate explanation is that subjects formed an association among three objects: nonobserving doll, token, and initial location containing the token. When asked a question that references two of these (non-observer and token, e.g., "Where will Sally look for her marble?"), the subject points to the specific container with which these two were last observed together. This outcome demonstrates a learned association. A theory of mind offers no better explanation and the classic test is therefore ambiguous in the absence of information explaining the basis for the choice.

A confirmatory result of Heyes's experiment could also be explained as demonstrating that her subjects associate aversive or nonaversive stimuli with salient or nonsalient pointing cues. ("One trainer has aversive [opaque] goggles – I don't want anything more to do with him" whereas "less aversive [translucent] goggles – less imperative to avoid", i.e., simple associative learning.)

Nonconfirmatory results might occur even if subjects do have a "theory of mind." Subjects may believe that, however unlikely, the Knower is deceitful or the Guesser omniscient, having special knowledge (ESP, clairvoyance, etc.). Well-studied chimps have often seen people who have acquired "mysterious" knowledge (e.g., correct container) without obvious means. Heyes's variant of the proposed experiment – trials with the trainer's back toward the food wells when baited, and the trainer then pointing to the nonbaited well – would exemplify this mysterious omniscience. From the chimp's perspective, how did the trainer learn which well was empty? If the trainer is ignorant, pointing to the two wells should be equally likely. Chimps may have acquired expectations about human knowledge that are part of their "theory" of human minds and that will result in experimental performance belying the underlying mental processes.

Thus, confirmatory results can be explained in many ways and nonconfirmatory results do not demonstrate that chimps lack a theory of mind. As such the proposed experiments are not a critical test. More generally, it may be impossible to design a test of a falsifiable hypothesis on the issue of theory of mind in primates. Even the false-belief test in humans which scores pointing, depends on a verbal query. It is difficult to conceive of an unambiguous belief test that does not require language as in the explanation of choice suggested above for the false-belief test on children.

To examine the issue in the same fashion in both human and nonhuman primates, we must either (1) accept all varieties of evidence (as suggested above) or (2) use a test with results not explicable by associative learning or other simple nonmentalistic hypotheses and that can be conducted in parallel fashion in both taxa, perhaps requiring the participation of language-trained apes.

Theory of mind in young human primates: Does Heyes's task measure it?

Deepthi Kamawar and David R. Olson
Centre for Applied Cognitive Science, Ontario Institute for Studies in Education, University of Toronto, Toronto, Ontario, Canada M5S 1V6.
dolson@oise.utoronto.ca

Abstract: Three- to six-year-olds were given Heyes's proposed task and theory of mind tasks. Although they correlated, Heyes's was harder; only 50% of participants with a theory of mind reached a criterion of 75% correct. Because of the complex series of inferences involved in Heyes's task, it is possible that one could have a theory of mind and fail Heyes's version.

We attribute a theory of mind to a child or animal if they: (1) believe that mental states causally explain actions and (2) appropriately attribute those states to themselves or others when appropriate information conditions obtain. Heyes proposes a task with these critical inferential steps: (i) I cannot see through opaque glasses so neither can another; (ii) because one cannot see one cannot *know;* and (iii) because one does not know one cannot indicate information. "Knowing" is the critical mental state and the critical understanding is that seeing causes knowing.

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

In order to get some insight into the demands of Heyes's task, we have presented an analogous task to 24 three- to six-year-olds along with standard theory of mind (ToM) tasks: appearance-reality, change of location, and unexpected contents. Our commentary focuses on the relation between Heyes's and standard ToM tasks.

To make Heyes's task suitable for children, some changes were made: (a) the task could not be nonverbal, but information was kept to a minimum and used no mental verbs; (b) stuffed animals acted as guesser/knower; and (c) mirrored sunglasses were used instead of goggles – one pair made opaque from the inside (the colour of the opaque glasses was counterbalanced). The activity was introduced as a game in which a sticker would be hidden in one of four boxes and then animals ("Bear" and "Lion") would tell the child where to look. If the child found the sticker, she kept it. Then the glasses were introduced ("Here is the red/black pair") and the children wore them for about 30 seconds. The game would begin: one animal would be on the experimenter's side of the screen observing the boxes while the other was "all covered up" by being placed in a fabric bag on the child's side. A sticker would be hidden, the screen removed, and the hidden animal brought back. Each animal would indicate a box and the child would choose one. Both the boxes named and the knower were counterbalanced. After six trials, the glasses were reintroduced, one pair at a time, and the children tried them on again ("Remember the red/black pair?"). Immediately after the child removed them, they were placed on an animal. This was done to minimize concerns about the children not being able to remember which pair was opaque. Four Heyes trials were then presented, interspersed with knower/guesser trials; the Heyes trials were counterbalanced so that half the children were never rewarded and the rest were always rewarded. The order to the two types of tasks (Heyes and ToM) was counterbalanced.

Results. First, as is well established in the literature, the ToM tasks were highly related to age ($r\ (df = 24) = .63$, $p < .01$), with the major break occurring at age 4. Performance on the pretraining task in which participants were rewarded for selecting the option indicated by the knower was also related to age ($r\ (df = 24) = .48$, $p < .02$) and the correlation with the ToM tasks was not significant. Performance on Heyes's task was related to age ($r\ (df = 24) = .44$, $p < .05$) and the largest shift occurred at age 4, when one-half of the participants passed Heyes's task (3 or more correct out of 4) and none of the 3-year-olds did. That proportion remained at about 50% at ages 5 and 6 as well. There is a barely significant correlation between ToM and Heyes's tasks ($r\ (df = 24) = .40$, $p = .05$) which does not remain significant if age is removed as a factor. Nonetheless, it does appear that there is some relation between grasping "who knows" in Heyes's task and the standard ToM tasks, although only about half of the children who succeeded on ToM tasks (85% or more) succeeded on Heyes's task. Heyes's task is by and large more difficult, presumably because of the long and diverting series of inferences required in addition to simply grasping the relation between seeing and knowing. It seems fair to infer that one could have a ToM (as assessed by standard tests) and still fail Heyes's version.

Interviews with individual children were striking. While they acknowledged that "Bear can't see" when wearing opaque glasses, several of the youngest nonetheless pointed to the box indicated by Bear. Successful children, exemplified by one 4-year-11-month-old, said, pointing to Lion, "He can't see" and chose Bear's box. On the next trial, he said, "He [Bear] can't see. When [I] wear them, I can't see" and pointed to Lion's box. While all the children talk about "seeing," none of them refer to the mental state of "knowing."

Children did not immediately infer the opacity of the glasses from the colour of the frames. We had to let the children try the glasses again before putting them on the observers. Merely discriminating the colour of the frames (as Heyes proposes) is not enough. Unless the connection is made, the crucial inferences cannot be drawn.

A final comment on the methodology. The critical issue in theory of mind is the ascription of mental states such as knowing and believing. To study "knowing," one examines whether subjects understand informational causation (Wimmer & Weichbold 1994); to study "believing" one studies deception. [See Whiten & Byrne: "Tactical Deception in Primates" *BBS* 11(2) 1988.] But one may be able to understand "seeing" without inferentially connecting it to knowing, and one may be able to understand how to mislead without understanding that the misled actually holds a false belief (Peskin 1996). For this reason, it is appealing to developmentalists to divorce the conceptual structure – the implicational relations holding among a set of linguistically coded concepts – from the behavior which may be characterized in terms of that conceptual structure (cf. Dennett 1978b). Only language-using humans can be expected to construct the former and then primarily for explanatory purposes. Many practical activities, including social activities, can be carried out without the conceptual underpinnings of a theory.

Given the results of this study, it seems likely that Heyes's task measures more than a theory of mind; one can pass standard false belief tasks and still fail Heyes's task. Hence, this task might not be a suitable theory of mind measure for nonhuman primates; it demands too much of them.

Having a concept "see" does not imply attribution of knowledge: Some general considerations in measuring "theories of mind"

David A. Leavens

Department of Psychology, University of Georgia, Athens, GA 30602; and Division of Psychobiology, Yerkes Regional Primate Research Center, Emory University, Atlanta, GA 30322. **dleavens@uga.cc.uga.edu**

Abstract: That organisms have a concept "see" does not necessarily entail that they attribute knowledge to others or predict others' behaviors on the basis of inferred mental states. An alternative experimental protocol is proposed in which accurate prediction of the location of an experimenters' impending appearance is contingent upon subjects' attribution of knowledge to the experimenter.

Heyes correctly notes that there is no necessary relationship between behavioral competence in experimental tasks designed to measure any of (a) imitative behaviors, mirror-guided behaviors, differential social responses, deceptive behaviors, role-taking behaviors, or perspective-taking behaviors, and (b) mental state attribution ("theory of mind") in nonhuman animals. Yet, Heyes insists that behavioral competence in a task designed to measure the presence of a concept "see" is a valid measure of a theory of mind, because the possession by an organism of a concept "see" implies both (a) the attribution of mental states to others and (b) a belief that these mental states are causal in the behavior of others (sect. 1, para. 3).

Heyes's experimental design(s) require chimpanzees to discriminate experimenters who can see the baiting of a food-well from experimenters who cannot see this baiting. Hence, Heyes's triangulation method will not tell us whether chimpanzees have "theories of mind," but only whether chimpanzees discriminate second party "seeing" and "not-seeing." If the subjects exhibit better than chance performance after being given multiple, successive sets of discriminanda for the distinction between "seeing" and "not-seeing" (sect. 4.2, para. 7), then we might validly conclude that the chimpanzees "have" the concept "see" (i.e., that concept formation has occurred).

That chimpanzees without special training exhibit a discrimination between "seeing" and "not-seeing" in others, including both conspecifics and humans, is a robust finding in recent experimental and observational studies on audience effects on chimpanzees'

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

communicative behaviors; these studies have established that language-naive chimpanzees are extremely sensitive to at least some of the behavioral concomitants of visual attention in both humans and conspecifics (e.g., Leavens et al. 1996; Povinelli & Eddy 1996, experiment 1, condition C; Tomasello et al. 1985). Even in very unusual experimental contexts, filled with strange objects and unnaturally static postures by experimenters, juvenile chimpanzees rapidly learn to discriminate between states of "seeing" and "not-seeing" in second parties (Povinelli & Eddy 1996, experiments 10, 11, and 13). Because these discriminations cannot uniquely implicate the attribution of knowledge to others by chimpanzees (Leavens et al. 1996), Heyes incorrectly subsumes "seeing" into the general category "mental state concepts" (sect. 1, para. 3).

In human developmental research, both the attribution of mental states to others and the use of these attributions to explain and predict behavior are cardinal defining features of "theories of mind" (e.g., Perner et al. 1987). An experimental demonstration of "theories of mind" in nonhuman organisms will therefore require more than merely a differential response to others' abilities to see aspects of stimulus arrays. Such an experiment would demonstrate chimpanzees' prediction of others' behaviors that can only be plausibly attributed to these chimpanzees' attributions of knowledge states to second or third parties.

With respect to the experimental design(s) proposed by Heyes, two questions need to be addressed. First, would postive, early transfer performance on any version of Heyes's experimental design demonstrate that the chimpanzees are attributing knowledge of food location to an experimenter? The dependent variable in Heyes's experimental designs is food-well choice. Because this dependent variable is used to measure *both* subjects' discriminations of second party visual access and subjects' attribution of knowledge to second parties, it is not possible to distinguish the interpretation (a) that the subject understands that second party "knowledge" follows second party "seeing" (mentalistic) from the interpretation (b) that the subject discriminates second party "seeing" from second party "not-seeing" (nonmentalistic).

The second question is, Would high performance on transfer in Heyes's design constitute evidence for subjects using these attributions to predict or explain experimenters' behavior? It is clear that Heyes's experimental design does not require the subjects to predict the behavior of one or more experimenters on the basis of the experimenter's knowledge about food location. Because space limitations prohibit a fully explicated alternative experiment, only some general recommendations are offered here.

First, a theory of mind experiment will necessarily require subjects who can distinguish "seeing" from "knowing"; because this discrimination is linguistically based in human studies ("Where did John see the chocolate?" vs. "Where does John think the chocolate is?"), this requirement poses special difficulties for the study of knowledge attribution in nonhumans (cf. Nelson 1996). One solution might be to permit our subjects to observe others in multiple conditions, receiving information in more than one sensory modality (e.g., visually and in the auditory domain); this will uniquely implicate responses based on others' knowledge over responses based on modality – specific, discriminative features of a stimulus array.

Second, subjects should observe experimenters experiencing both true and false information (not merely information vs. no information, as in Heyes's design), otherwise an essential ingredient of theory of mind measurement in humans will be lost (e.g., Perner et al. 1987). Third, subjects should be required to *predict* the behavior of experimenters based on experimenters' inferred knowledge states. In the absence of language, this is especially difficult to measure, but manipulating (a) knowledge states of an experimenter and (b) spatial locations of experimenter arrival and measuring anticipatory responses of the chimpanzees (location, visual orientation, etc.) would create the potential of an outcome in which subjects' responses could be attributed to their attributions of knowledge to experimenters. Correct anticipations of the locations of experimenters' arrival would depend on correct attributions of knowledge to the experimenters.

Fourth, in a recent study on gestural communication in a sample of over a hundred chimpanzees (Leavens & Hopkins, in press), juveniles (3–7 yrs.) exhibited a striking decrement in their propensity to communicate with adult, human experimenters, compared to all older chimpanzees (8–56 yrs.). I would accordingly recommend working only with adolescents and older chimpanzees (cf. Heyes's sect. 2.6.2., para. 8), unless juveniles are raised in more intimate association with adult humans than is typical for most laboratories.

Finally, it is worth pointing out that humans typically require four or more years of intensive (albeit, incidental) training before exhibiting theories of mind in experimental contexts; hence we should not put too much emphasis on failures to find theories of (human)mind in animals who have had far less training on such tasks or far less exposure to communicative interactions with humans.

Attribution is more likely to be demonstrated in more natural contexts

M. D. Matheson, M. Cooper, J. Weeks, R. Thompson, and D. Fragaszy

Department of Psychology, University of Georgia, Athens, GA 30602.
cmspsy37@uga.cc.uga.edu

Abstract: We propose a naturalistic version of the "guesser–knower" paradigm in which the experimental subject has an opportunity to choose which individual to follow to a hidden food source. This design allows nonhumans to display the attribution of knowledge to another conspecific, rather than a human, in a naturalistic context (finding food), and it is readily adapted to different species.

Celia Heyes has pointed out methodological problems of previous work with nonhuman species utilizing the "guesser-knower" paradigm (Povinelli et al. 1990) to document the presence of "theory of mind." Positive outcomes in these studies, if found, would all be equally interpretable as governed by the subjects having acquired new, learned discriminations during purported transfer test phases. As Heyes points out, the design requirements to isolate the theory of mind interpretation as the single sufficient explanation are (1) that only the subject's previous experience can serve as the discriminative cue, and (2) that the subject experiences no reinforcement on test trials, and therefore no possibility of learning new discriminations.

However, the modified guesser-knower paradigm suggested by Heyes still suffers from problems that would make it unlikely to produce positive results, even if the subjects were able to use their own experience to predict what another individual knows (sees). These problems include: that subjects are asked (1) to make inferences about humans, rather than conspecifics, and (2) to perform rather arbitrary tasks, which are at odds with apes' usual behavior. Although these problems may not completely vitiate attempts to detect knowledge attribution in a nonhuman species, they could potentially mask it. It would accordingly be wise to devise a more species-neutral task that meets the requirement to restrict potential cues to the subject's own experience, as suggested by Heyes. A suggestion of such a task follows. Anticipating a positive outcome, it would also be useful if the task were sufficiently flexible to use with other species, to foster broader comparative inquiry. There is at present no reason in principle to exclude any species *a priori* from the search for "theory of mind." Our task has the advantage of ready adaptability to testing other species, and using sensory modalities other than vision as the source of knowledge. We present the task first in a format friendly to chimpanzees, and then suggest how it can be modified for another species (dogs). The task will be familiar to some readers as a variation of a classic study conducted by Emil Menzel with young chimpanzees (Menzel 1974).

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

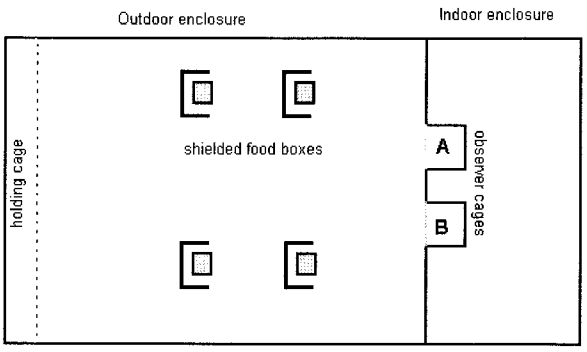


Figure 1 (Matheson et al.). Both observer cages have doors leading to the inside and outside enclosures. Observer cage A is equipped with two-way glass. Observer cage B is equipped with one-way glass, preventing the occupant from observing events in the outdoor enclosure. The holding cage is separated from the outdoor enclosure by a chain-link fence.

Apparatus. This experiment should be conducted at a location in which a colony of chimpanzees are housed in an indoor–outdoor enclosure (see Fig. 1). Food stations (bins with lids) are set up in the outdoor area, with blinds set up on one side of all food stations to block their vision from one end of the enclosure (say, the south end). The food stations are placed symmetrically around the center point of the cage. A holding cage is set up in the south end of the enclosure, so that an individual in the cage would be prevented by the blinds from seeing the food stations, but could see the exterior wall of the indoor enclosure.

In the indoor enclosure, two experimental cages are attached to the wall separating the indoor and outdoor enclosures, with glass windows looking out into the outdoor area. One cage has a two-way window; the other cage has a one-way window so that visual access to the outdoors is prevented, whereas visibility from outdoors into the indoor enclosure is unimpeded. The exterior and interior wall in which the glass panel is set is one color for the two-way window (say, red), and a different color for the one-way window (say, blue). These cages have doors into the outdoor enclosure, as well as the larger section of the indoor enclosure.

Training. The glass panels are visible from both the inside and the outside enclosures, so all chimpanzees have liberal incidental experience with the glass and the differently colored walls during their daily activities. During active training, an individual chimpanzee is placed in one of the two indoor experimental cages while the rest of the group is placed in the main section of the indoor enclosure. An opaque, chimpanzee-proof blind is placed around the experimental cage to isolate visually the individual in training from the rest of the group. While the chimpanzee is in the experimental cage, a researcher drives a cart containing food into the outdoor enclosure, stopping at and manipulating all food stations in a set order but baiting only one with a large quantity of highly desirable food. The chimpanzee trainee is then released into the outdoor enclosure and allowed to retrieve and consume the food, while the rest of the group remains inside. All adult chimpanzees are given several training trials in each of the indoor cages described above (one- or two-way glass) in which they experience each of the food bins as the single baited bin, until it is evident that the subjects quickly move to any of the baited stations when they have seen the baiting, and that they search actively until they find the food in any bin when they have not seen the baiting (i.e., when they had been locked in the experimental cage without visual access to the enclosure; the cage with the blue wall).

Testing. During testing, the full group is locked inside. Then, two "viewers" are moved to the indoor experimental cages, at this time in full view of the group. Next, a subject (who has just seen which individual went into the red cage, and which into the blue) is taken from the group to the holding enclosure at the south end of the outdoor cage. Blinds are set up inside the building between the remainder of the group and the two "viewers." Finally, to standardize the behavior of the two viewers, a sheet of one-way glass is inserted on the inside of the two-way glass cage so that neither "viewer" is actually able to see the outdoor enclosure. A wall panel of the appropriate color (blue, in this example) is placed around the inserted one-way glass on the inside wall, so that the color cue and visual access contingency matches the viewer's previous experience. Thus, none of the three subjects participating in the test (the two "viewers" and the experimental subject enclosed in the south end of the enclosure) can see where the food is hidden. However, the experimental subject moved to the south end of the enclosure sees one viewer move into the cage that does not allow a view of the enclosure; another viewer move into the cage that ordinarily does allow a view of the enclosure. Moreover, the experimental subject is able to see the "viewers" moving in the interior experimental cages. The exterior walls of both cages, which are visible to the test subject, are still marked (red and blue) as they were during training.

While the three individuals are in the cages as described, a researcher drives through the enclosure, stopping at and baiting *all* the food stations. When this has been completed, all three subjects are simultaneously released into the outdoor enclosure. The behavior of the test subject who has been locked in the outdoor area is recorded with respect to (1) which food station is approached first, and (2) whether or not either "viewer" approached that station first.

This procedure adapts the "guesser-knower" paradigm to a more naturalistic setting. Moreover, a greater number of subjects may be available in such a setting, and less training is required because the task itself (retrieving food) is one that is already in the chimpanzees' repertoire; it is also a less complicated procedure, with no paraphernalia on the animals. Previous work (Menzel 1974) has demonstrated that chimpanzees will follow an individual who has been shown the location of food, when they are released together, particularly if the "leader" has a history of sharing food with others. It will be important to have documented social relations in the group prior to testing, to avoid pairing subjects with nonsharing or intolerant "viewers."

The relevant prediction in this study is that the chimpanzee released from the south end of the outdoor enclosure will reliably follow the chimpanzee from the "viewing" cage with the red exterior wall (the one from which the baiting during the training period was visible) to the first station it visits more often than it will follow the other chimpanzee or search out the nearest station on its own. On the basis of previous experience, the subject would expect just one feeder to be baited; hence it should be strongly motivated to find the single baited feeder. Each subject would contribute one data point, although each could serve as a viewer after having been a subject.

Adapting this task to another species and another modality would not be difficult. For example, to test dogs, one could use the same design, or one could change the cues available to the "viewer" to auditory rather than visual ones. In the latter case, the dogs would be trained to go to one of the four food stations in response to the appropriate auditory cue. During training the dog in the interior kennel marked with one distinctive wall (both inside and outside) and fixed location would hear the word "one." Only a trip to the food station designated as "one" would be rewarded (and so forth for the other three stations). When placed in the other kennel (with the wall marked in a distinctively different pattern), the animals would not receive an auditory cue, and would have to search the bins to find the sole baited one. Neither indoor kennel would have visual access to the outdoor enclosure. The testing procedure would parallel that used for the chimpanzees. As with the chimpanzees, it would be important to document beforehand whether subject and "hearers" would co-feed at the same site amicably.

The predictions in this test are the same as for the visual tests: if the dog restrained in the outdoor kennel at the far end of the

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

enclosure can infer from its own experience with the kennels that are distinctively marked and in a fixed location which other dogs will hear the cue, it ought to follow that dog to a bin. Inference on the basis of vision should not be a litmus test for theory of mind.

We have another design to suggest as well, a variation of the video tape method developed by Premack and Woodruff (1978), one that we believe provides a clearer test of the hypothesis that chimpanzees can infer knowledge of another than Heyes's suggested design. Subjects are presented with video tapes in which a guesser and knower (roles alternated randomly) see food hidden, and then the "guesser" leaves the room, after which the location of the food is switched in the presence of the "knower" (the "Sally and Anne task" used by Wimmer and Perner 1983). The subject is trained using positive reinforcement techniques to pick a photograph at the conclusion of the video depicting the knower making the correct choice. On some training video tapes both the guesser and knower will stay in the room, or both will leave the room. When it has been established that the subject can choose which individual(s) should be pictured finding the food, and following the training noted below, a test condition is run presenting a novel configuration of events involving both the guesser and knower wearing distinctive goggles, and neither guesser or knower leave the room.

One control procedure is necessary to confirm that the subjects can discriminate the goggles on video tape. Nonparticipant subjects are required to discriminate the colors of the goggles in a simple two-choice setting, and are then shown humans on video tapes wearing the goggles, to confirm transfer of the discrimination to this new form of stimulus presentation. One additional training procedure is necessary as well, to ensure that the subjects experience the properties of the goggles. Subjects are given experience wearing opaque and translucent goggles of the same color as those used in the testing. After some experience with them, they are given a series of choice trials while wearing (in turn) opaque or translucent goggles. In the choice trials, two food items, one more preferred and the other less preferred, are presented on a tray, with alternating locations for the two food items on the tray. Subjects wearing opaque goggles are expected to make random choices, and subjects wearing translucent goggles are expected to choose the rare and preferred treat.

The experiment is conducted shortly after the conclusion of the goggle training and the attainment of criterion on the video-choice training. The subject is presented with the familiar training sequences interspersed with three novel ("probe") film sequences. The first probe shows the experimenter hiding food in cup A while the guesser and knower are present without goggles. The guesser puts on opaque goggles (red) and the knower puts on translucent (blue) goggles. The experimenter then switches the food from cup A to cup B. The guesser and knower take off their goggles and prepare to make their choice. The tape is stopped, and the subjects are presented with the same photographs as were used in the training films. In the second and third probes, both humans wear opaque (red) goggles or translucent (blue) goggles. If the subjects attribute knowledge to humans on the basis of vision, then a significant number of subjects should choose cup A for the wearer of the opaque goggles and cup B for the wearer of the translucent goggles.

These modifications to Heyes's proposal have three advantages. First, using only one probe trial per condition avoids the difficulty of explaining improved performance over repeated probes. Second, the subjects are asked where they believe the human thinks the food is hidden. The additional step of asking where the subjects think the food is hidden is not required. Third, allowing for two guessers and two knowers on the same trial yields three measures from the same subject, rather than a single measure. If the subject infers the human's knowledge state on the basis of its own experience with the goggles, it should choose as effectively with two knowers and two guessers as with a single knower and guesser.

In closing, we suggest that we would do best to adapt our tasks to the subjects' concerns in an ecologically and socially valid context to provide convincing evidence of knowledge attribution. Failures to demonstrate the phenomenon of interest in contrived situations (such as with goggles) would not be convincing negative absence, although positive evidence would be convincing. We think that the ecologically valid design listed first is more likely to demonstrate the phenomenon, should it exist, and it allows testing to be extended to other species more readily than the contrived situations involving goggles and videotapes.

Methodologies, not method, for primate theory of mind

H. Lyn Miles and Warren P. Roberts

Department of Sociology and Anthropology, University of Tennessee, Chattanooga, TN 37403 and Department of Anthropology, University of Georgia, Athens, GA 30602

lmiles@cecasun.utc.edu www.utc.edu/~lmiles/ wrobertsjr@juno.com

Abstract: Heyes correctly points out some problems in primate theory of mind, but lacks a critical approach to children's theory of mind, and at times implies meta-awareness when discussing theory of mind. Also, in selecting pure experimental designs, she ignores its limitations, as well as the merits, and at times the necessity, of other methodologies.

Heyes is perceptive in her analysis of some methodological problems in primate theory of mind research, but her lack of a critical approach to children's theory of mind, and her acceptance of research with children as universally productive and disciplined, is perplexing. Some researchers are now questioning whether children have a consistent theory of mind (Bjorklund 1995). At what age should we accept the emergence of theory of mind in Western cultures? Which domains and paradigms are viewed as clear evidence for theory of mind? Should these be prediction of object motion, social gesturing, competent reading of emotional signals, or merely language mediated false belief paradigms (Frye & Moore 1991)?

Unfortunately, Heyes also presents an all too familiar double standard in ape and child research. For example, in her criticism of Custance et al. (1995), she suggests that the human-reared apes had received inadvertent rewards for imitative lip-smacking, but she apparently does not apply this critique equally to studies of children, despite citing Piaget's (1962) observation that human caregivers socially reward infants during imitative games.

Assigning priority of pure experimental design over normalcy of populations is unhelpful, and allows odd conceptual pairings to emerge. For example, Heyes's criticism of self-recognition uses studies of autistic children as supporting evidence, but mirror self-recognition and mental state dissociation in abnormal populations says little about the interdigitation of the two in normal populations. Lesions can dissociate sight from awareness of visual perception (Flanagan 1992) but this does not suggest that normal individuals are not aware of what they see, nor that awareness and perceptual systems did not coevolve.

It is not always clear when researchers are discussing theory of mind where a child or an ape might act on knowledge of others' beliefs, or third and fourth-order meta-awareness where they might be further able to focus attention on the knowledge itself and to know that they know. Heyes seems to claim the former when she reports that theory of mind is about the "*content*" of their representations, but suggests the latter when she states, "an animal with a theory of mind believes that mental states play a causal role in generating behavior" (sect. 1).

In addition to Heyes's approach, there are other approaches to primate cognition, including description, structured and systematic observations, and developmental anecdotes. [See Whiten & Byrne: "Tactical Deception in Primates" *BBS* 11(2) 1988.] These can consider the role of context and can include more naturalistic conditions (Mitchell et al. 1997) whereas an experimental method has the limitation of sometimes restrictive and impoverished

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

conditions that are so confining as to distort the phenomena studied and to miss their critical aspects.

A restrictive approach can actually subvert symbolic processes; for example, it is impossible to study representational ability in apes without providing a rich ecological, ethological, and symbolic context of enculturation (Miles 1997; Savage-Rumbaugh & Lewin 1994). Researchers may need to act "as if" as a methodological requirement in order to determine whether animals have mental abilities. In fact, rich symbolic and social contexts and rewards may be crucial for understanding human biosocial and mental development (Shore 1996), as well as being a critical component of a normal rearing history for all hominoids. Social rewards in closely interactive units are certainly the norm in the development of socially competent chimpanzees (Goodall 1986).

Most important, to argue that only one method can allow us to conclude that children or animals do or do not have a theory of mind leads to other surprisingly narrow-minded analyses. For example, Heyes questions that the home raised chimpanzee, Viki, could imitate (Hayes & Hayes 1952) on the grounds that no method was reported in which the experimenter measured the degree of similarity of Viki's behavior. But Heyes fails to note that this was later provided by a replication of this study with the enculturated orangutan Chantek (Miles et al. 1996). Heyes also considers only mirror self-recognition, and ignores other measures of self-awareness, including personal pronouns and possession (Itakura 1994; Miles 1994; Patterson & Cohn 1994) as well as Menzel's (1974) experiment on chimpanzee deception. Finally, she ignores that all child studies are linguistically mediated either by virtue of the procedures used or by the prior enculturation of the children. Such adherence to a single perspective may hold some benefits, but overall it is myopic.

Primate theory of mind is a Turing test

Robert W. Mitchell[a] and James R. Anderson[b]

[a]*Department of Psychology, Eastern Kentucky University, Richmond, KY 40475.* **psymitch@acs.eku.edu;** [b]*Department of Psychology, University of Stirling, Stirling FK9 4LA, Scotland.* **j.r.anderson@stir.ac.uk**

Abstract: Heyes's literature review of deception, imitation, and self-recognition is inadequate, misleading, and erroneous. The anaesthetic artifact hypothesis of self-recognition is unsupported by the data she herself examines. Her proposed experiment is tantalizing, indicating that theory of mind is simply a Turing test.

We agree with several of Heyes's ideas, especially those which remind us of earlier arguments that self-recognition and bodily imitation need not imply theory of mind (Mitchell 1993), and that information about an animal's history is necessary to interpret its deceptions (Mitchell 1986; Morgan 1894). Our own research, replicating Woodruff & Premack's (1979) chimpanzee study with capuchin monkeys, supports the idea that communicative and deceptive pointing can derive from discrimination learning (Mitchell & Anderson 1997), though it raises the question of why macaques failed to learn deceptive pointing in a similar experimental setting (Blaschke & Ettlinger 1987). Perhaps a more accurate and thorough literature review than the one provided by Heyes would establish more common ground.

Heyes claims that all instances of imitation in nonhuman primates "could" or "may" have come about by chance or nonimitative means, even when the same animal produces a variety of responses supporting a generalized ability for the imitation of behavior. Instead of evidence, scenarios of possible or spurious reinforcements for each particular instance are offered. Such speculations are not evidence against an ability for imitation; even in humans, imitation develops from nonimitative processes and generalization of previously learned behaviors (Guillaume 1926/1971). Alternative explanations for single behaviors produced by rats and budgerigars are ignored, no evidence is provided that they can imitate a variety of actions, and their imitations appear unrelated to cross-modal imitations present in humans and some apes (Mitchell 1996; 1997). Relevant studies on primate imitation are ignored by Heyes: one experiment explicitly designed to shape imitative responses resulted in failure with a macaque (Mitchell & Anderson 1993), another in success with an orangutan (Miles et al. 1996). Contrary to Heyes's claims, Gallup (1982) and Povinelli (1987) never discuss imitation as evidence of theory of mind, and only two (not all four) chimpanzees in Woodruff & Premack's (1979) study pointed to an empty container with the competitor, and those on less than 80% of the last 60–94 trials, after about 190 trials each with competitor and cooperator.

In presenting the self-recognition literature, Heyes ignores evidence from apes of self-exploration of body-parts not visible without a mirror, studies using sham marking instead of anaesthesia, evidence of self-recognition in gorillas, and methodological flaws in studies of human infants (Parker et al. 1994). Contrary to her claims, studies using variants of the traditional mark test procedure do not support her anaesthesia artifact hypothesis, as these studies used monkeys (which never pass any form of the mark test) or very young great apes (see Gallup et al. 1995). Given that Heyes's hypothesis is intended to explain why chimpanzees pass the mark test, her inclusion of mark-directed touches by chimpanzees failing the mark test (in the means and standard deviations for the bimodal distribution created by combining both passing and failing animals) is baffling. For Swartz and Evans (1991) the one chimpanzee who passed the mark test (on all 3 trials) showed from 10–23 mark touches while looking at herself in the mirror, and from 1–7 mark touches in the control condition; for Povinelli et al. (1993), the 10 chimpanzees who passed the mark test touched the mark on average 11.2 times ($s = 10.8$) in the mirror condition, but only 1.6 times ($s = 2.5$) in the control condition (and including "spurious" mark-rubs does not change the difference between conditions). By contrast, those chimpanzees who failed the mark test showed infrequent mark touching, but more of it in the control condition than in the mirror condition, indicating that chimpanzees are not, generically, more likely to touch the mark by chance in the mirror condition than in the control. For Swartz and Evans, the 10 "failing" chimpanzees touched the mark infrequently in the mirror condition (21 tests with no touches, 5 with 1–4 touches), whereas most touched the mark in the control condition (11 tests with no touches, 15 with 1–12); for Povinelli et al., 18/20 of these chimpanzees failed to touch the mark at all in the mirror condition, but 12/20 touched it at least once (range 1–13) in the control condition (and including "spurious" mark-rubs just increases the latter range). For most chimpanzees, then, mark-touching frequency actually *decreased* as the time interval since anaesthesia increased.

Heyes argues that to detect itself via a mirror, an organism must distinguish its own bodily inputs from external inputs. The immediate relevance of this internal/external distinction to self-recognition remains unclear (Anderson & Gallup, in press; Mitchell 1996): human children make this distinction by 5 months of age (Watson 1994), yet do not show any signs of self-recognition until 10–19 months later. The logical conclusion from Heyes's argument is that all visually capable organisms should pass the mark test, yet they do not. Although she cites Epstein et al.'s (1981) "self-recognizing" pigeon experiment favorably, it has failed to replicate despite extensive efforts (Thompson & Contie 1994). Heyes's assertion that autistic children show self-recognition at the same age as normal children is surprising, given that the youngest autistic children tested are 3-year-olds (Mitchell 1997).

Heyes's proposed experiments raise some surprising issues about theory of mind and about how a chimpanzee might interpret human actions. They presume that laboratory chimpanzees believe that humans know where objects are located solely through vision; yet these chimpanzees presumably have numerous experiences in which humans appear to know the location of something without having seen it. Indeed, in the second (variant) test, the human who does not observe the baiting or wears opaque goggles

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

nevertheless knows enough to point to a container, and sometimes even to the correct one, which seems confusing. Greater frequency of correct solutions with Knower than with Guesser, or faster learning with Group Direct than with Group Reverse, do not seem adequate as evidence of theory of mind unless the correct solutions occurred from the start of the transfer trials. In fact, surely only a correct choice on the first transfer trial can count toward evidence of theory of mind, whether or not chimpanzees are consistently rewarded: reward would contaminate subsequent responses, and lack of reward could indicate error and therefore lead the chimpanzee to respond to other stimuli. Either way, simple task analysis indicates that the experiment she describes remains essentially a multiple discrimination learning task, which even macaques can perform (see Mitchell & Anderson 1997). Perhaps, in fact, we can never be sure whether an animal is responding to another based only on the other's behavior, or on mental-state inferences from that behavior. As Turing (1950, p. 446) argued, we may have to maintain the "polite convention that everyone thinks" until his or her behavior suggests otherwise.

ACKNOWLEDGMENT

We thank Karyl Swartz for her assistance.

Primate cognitive neuroscience: What are the useful questions?

A. Parker

Department of Experimental Psychology, University of Oxford, OX1 3UD, United Kingdom. **amanda.parker@psy.ox.ac.uk**

Abstract: Study of "theory of mind" in nonhuman primates is hampered both by the lack of rigorous methodology that Heyes stresses and by our lack of knowledge of the cognitive neuroscience of nonhuman primate conceptual structure. Recent advances in this field indicate that progress can be made by first asking simpler research questions.

In the two decades that have elapsed since Premack and Woodruff (1978) opened the debate about theory of mind in nonhuman primates (NHPs), a great deal has been learned about the development of social intelligence in human infants. Heyes attributes the contrasting lack of progress in our understanding of NHP theory of mind to ineffectual experimental methodologies and lack of clear theorizing. To some extent she may be right. But another interpretation of the research covered in her target article might be that, in the present state of knowledge about NHPs, we are asking the wrong questions, or trying to force comparisons that are simply inappropriate. This is particularly true as theories of human theory of mind development have become sophisticated very quickly.

One recent proposal is that theory of mind development in human infants depends on the development of "modules" for intentionality detection, gaze direction detection, and shared attention, leading to the development of a further theory of mind module (Baron-Cohen 1995). The importance of linguistic processing in this last stage should not be underestimated. A useful research question would therefore be, To what extent do these modules, or their precursors, occur in NHPs? Heyes's proposed experiment deals with one aspect of visual attention, the inference of another's knowledge of an event from a calculation of whether or not they have perceived it. Her proposed method is likely to produce valuable insights on this topic. At the present point, we need precise quantitative knowledge about the range of NHP conceptual abilities about other minds.

Abilities are likely to be tied to neuroanatomy. So we should also ask about the extent to which different species of primates have similar or different neural structures and connections, and to what extent this leads to similar abilities. Neuroanatomical study of the frontal lobes of humans and macaque monkeys has revealed that their basic architectonic plan is the same (Petrides & Pandya 1994). A recent comparison of cytoarchetectonic areas thought to be uniquely human with chimpanzee and macaque brains, likewise suggests that the basic organization of these areas is the same (Passingham 1997). Moving from anatomy to behavior, we have found that damage to structures known to be important in episodic memory in humans causes large deficits in object-in-place memory in monkeys (Parker & Gaffan 1997a; 1997b). Similarly, damage to perirhinal cortex will affect the monkey equivalent of semantic memory – conceptual knowledge about objects. (Parker & Gaffan 1997c). It is unlikely however, that autobiographical memory will develop without language (Nelson 1993), and it seems likely that autobiographical, rather than episodic, memory is the key feature of the human conception of the self and the consequent full development of complex representations of other minds.

Primates are able to understand the world, including their social world, because of their highly developed conceptual resources. An emergent property of the neural structure of the temporal and frontal lobes is that they store information as categories. In the temporal lobe, these structures represent objects, while the frontal lobe stores action, intention, and affect related schemata. Combining these two types of information enables primates to produce complex and subtle behaviors over a wide range of situations. At a certain point this conceptual structure may become elaborate enough to sustain a theory of mind. What are its basic building blocks? This is a more answerable question. One source of answers may lie in the way that object representations in the temporal lobe become integrated with appropriate strategies in the frontal lobe, and the effects of damage to these structures on this knowledge. Current experiments involve only abstract stimuli and arbitrary categories (Parker & Gaffan 1997d) and indicate that when interaction between object knowledge and action strategies is prevented, behavior is at chance levels. Future research with this paradigm will examine real categories of objects. It is a short step from here to the study of how monkeys apply strategies to categories of conspecifics.

To conclude, much of the available evidence points to language as being vitally important for the development of a complete human theory of mind. This does not mean that components of this ability are not present in NHPs, but it does mean that using anthropomorphic methods to search for them is inappropriate. We should be searching for the basic elements of social understanding in primates, and using rigorous experimental methodologies, as Heyes suggests.

To see or not to see, that is the question: Designing experiments to test perspective-taking in nonhumans

Irene M. Pepperberg

Department of Ecology and Evolutionary Biology and Department of Psychology, University of Arizona, Tucson, AZ 85721. **imp@biosci.arizona.edu**

Abstract: Heyes argues that we need alternative experiments to study those animal abilities generally considered to involve "theory of mind." The studies she proposes, however, have as many problems as those that she criticizes. Further interactions should exist among researchers examining these capacities before additional experiments are undertaken.

As a researcher who studies cognitive and communicative capacities of nonhumans (e.g., Pepperberg 1990; 1996), I both agree and disagree with Heyes's target article. I agree that questions concerning animal capacities such as self-awareness and perspective-taking are intriguing, but that much published research into these areas has either failed to show such abilities or made unfounded claims for these capacities because of problems in experimental design. I disagree with some specific criticisms that Heyes has offered; because many of my colleagues have previously engaged

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

her in debates on these issues (see, for example, Gallup et al. 1995), I will instead, confine my discussion to a critique of her proposed experiments.

I find Heyes's experiments (sect. 4.2) for determining "seeing" abilities (perspective-taking) confusing. Maybe I misunderstand her rationale, but I find that her experiments lack a certain internal – and occasionally external – validity. I'll take each point of confusion in sequence.

I do not completely understand her pretraining (sect. 4.2.1) paradigm. Heyes first argues for training chimpanzees to use opaque versus transparent goggles despite the likelihood of opaque goggles being aversive. Using Heyes's own "killjoy" approach, one can critique her experiment using the following arguments: given the proposed pretraining, a chimpanzee will associate the opaque goggles with an aversive situation (not being able to see anything, never mind the experimental apparatus), and will very likely learn to avoid every and any thing that is associated with such goggles, that is, any experimenter wearing them and that experimenter's actions. The chimpanzee might then pick the Knower (or Group Direct) not specifically because translucent goggles had something to do with "seeing" and opaque goggles specifically with "not seeing," but simply to avoid something aversive. The problems involved in avoiding aversive situations are also likely to permeate the one- versus two-way screen design. Imagine, for example, that during the pretraining sessions, a chimpanzee experiences the one- and two-way screens on a random basis; in both cases the experimenter hides a treat under one of two cups, the screens are removed, and the chimpanzee is then allowed to choose. Such a protocol would presumably provide the chimpanzee with the appropriate experience so that it could transfer this experience to humans who were placed behind these screens during the hiding of the treat. But it is quite possible that the chimpanzee would simply find the one-way situation, with its concomitant difficulties in finding the treat, to be aversive, and thus respond on that basis, rather than on the basis of understanding the underlying rules of perspective-taking. The same problem would arise in using such screens for a variant of the procedure involving the Wisconsin General Test Apparatus.

During transfer (sect. 4.2.3), Heyes claims that by using a change of procedure (goggles versus the movement of a trainer) as a probe and either consistently rewarding or not rewarding the chimpanzee during these probes, an experimenter could tell whether the chimpanzee had the concept "see": that is, Heyes expects that one can determine what the chimpanzee understands by examining whether it would choose the Knower *more often* than the Guesser. I disagree, because in such a situation *only the first trial* would be relevant: A chimpanzee would be highly sensitive to the altered condition of a probe and would not necessarily connect this experience with that of the nonprobe situations: a chimpanzee who chose wrongly on the first probe trial but was rewarded would consistently choose wrongly again given this altered condition; a chimpanzee who chose correctly on the first probe and was rewarded would likely consistently choose correctly again in this altered condition; similar situations would exist for unrewarded probes, although in such cases the animal might continue to switch back and forth in an attempt to receive a reward.

Last, I question Heyes's use of children for triangulation (and the triangulation process itself), given the possibility that similarities in observed (i.e., "surface") behavior may have little to do with similarities in the mechanisms underlying such behavior across species. Humans, for example, when given a collection of small numbers of different items (e.g., blue and red balls and blocks) and asked to quantify one subset (blue blocks versus blue balls or red blocks or balls) will count, rather than subitize (a perceptual mechanism generally used to enumerate small quantities [less than five] if identical items comprise the set; see Trick & Pylyshyn 1989). That humans count in this task is determined by examining both reaction times and accuracy and comparing the data to other tasks that require counting or allow subitizing. A parrot's data on this red/blue/ball/block task matches that of humans with respect to accuracy (Pepperberg 1994), but claims cannot be made that the bird can or cannot "count" based on the data; parrots' skills, unlike those of humans, may simply be such that larger numbers ($\geq$ five) can be subitized.

In sum, I do not think Heyes's proposed methodology is superior to other procedures. Her challenge to devise appropriate experiments, however, should be taken seriously. Specifically, given the resources currently devoted to such projects and the problems involved in their experimental design, I suggest that additional pre-experimental interactions among research groups, consisting of design critiques, would greatly improve the quality of work in this area.

Tactics in theory of mind research

Jesse E. Purdy[a] and Michael Domjan[b]

[a]*Department of Psychology, Southwestern University, Georgetown, TX 78626.* **purdy@ralph.southwestern.edu;** [b]*Department of Psychology, University of Texas at Austin, Austin, TX 78712.* **domjan@psy.utexas.edu**

Abstract: Progress in the "theory of mind" debate would be better served at this point by abandoning the search for a perfect "critical experiment" and developing an incremental research program based on a systematic theory of "theory of mind." Studies using the goggle procedure advocated by Heyes should dissociate the ability to see from possible behavioral artifacts of "blind" trainers.

The Heyes target article contributes to a large literature on what constitutes evidence of "theory of mind" in nonhuman primates. We offer the following comments with a bit of reluctance because our first reaction to the article was that progress in this area would be better served by empirical contributions rather than additional discussion in the absence of new data. It is difficult to determine *a priori* what would be the best way to demonstrate a phenomenon. Progress in behavioral science rarely comes from a decisive "critical experiment." Rather, progress emerges from incremental empirical efforts in which the phenomenon of interest is gradually isolated from other alternatives.

We are also a bit skeptical about attempts to prove or disprove something as potentially complicated as "theory of mind." A consideration of the history of other forms of complex cognition is instructive in formulating an investigative approach to "theory of mind." Studies of language learning in chimpanzees provide a good example. These investigations started out with the goal of yielding a "yes" or "no" answer (Kellogg 1968). However, language is no longer viewed as monolithic but as "consisting of a large number of component parts and interacting functions" (Rumbaugh et al. 1991, p. 145). Within this incremental framework it is no longer meaningful to ask whether chimpanzees can or cannot learn language. Rather, the meaningful questions concern identifying the components of linguistic skill, determining how those components are learned, discovering the order in which they have to be learned, and so on (Roitblat et al. 1993).

Following the example set by studies of language learning, it may be time to abandon the attempt to answer in a "yes" or "no" fashion whether animals have a theory of mind. Instead, it may be more profitable to devote effort to identifying what might be components of a theory of mind, how these components, might be learned, and the order in which these components are learned most efficiently. This requires a systematic theory of "theory of mind" in place of a list of various phenomena (imitation, self-recognition, role taking, deception, and perspective taking) that may have little in common.

Whether or not one favors the "critical experiment" approach, any experiment that is performed should minimize the impact of irrelevant factors. Heyes proposes to test the concept "seeing and knowing" by using a procedure modeled after Povinelli et al.

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

(1990). Chimpanzees are first conditioned to discriminate between opaque and translucent eye glasses on the basis of the color of the frames. They are also trained to choose a container pointed to by a "knowledgeable" trainer. After they have learned this task, the chimpanzees are presented with probe trials in which two trainers, one wearing opaque glasses and the other wearing translucent glasses, observe a third person bait one of four containers. Each trainer then points to one of the containers and the chimpanzees follow with their selection. Evidence for the concept "seeing and knowing" would be provided by the chimpanzees choosing the container pointed out by the "sighted" trainer more often than the container pointed out by the "blind" trainer.

With this procedure it is unlikely that the opaque glasses would be the only cue the chimpanzees might use to tell whether the trainer can see. While wearing opaque glasses, the trainer might move awkwardly, reach out slightly to avoid potential obstacles, point inaccurately or with hesitation, stumble slightly, or provide other nonverbal cues. Such cues could be better indicators of the trainer's inability to see than simply the color of the glasses. Chimps may avoid using information obtained from trainers who move awkwardly. This could result in the chimp choosing the "sighted" knowledgeable trainer without having to experience the opaque and translucent glasses and without knowing whether the trainer could see.

To eliminate this potential confound, the opaque and translucent glasses should be switched during one half of the probe trials, keeping the color of the glass fames constant. As a result, the trainer considered to be "blind" by the chimpanzees will in fact be able to see on some trials. Similarly, the trainer considered to be "sighted" by the chimpanzees will not be able to see on some trials. By properly counterbalancing the actual ability of the trainers to see against the presumed knowledge of the chimpanzees, one could determine whether the chimpanzees were responding on the basis of their presumed knowledge or on the basis of the differential behavior of sighted and "blinded" trainers.

We would also like to suggest that the reversal procedure that was proposed for the second experiment be used in the first experiment. As mentioned by Heyes, if chimpanzees do have an innate sense of "see," or if the chimpanzees have acquired such a sense, then it should be more difficult to teach them that a "blind" observer knows where food has been hidden than it would be to teach them that a "sighted" trainer knows where food has been hidden.

A third procedural change might be more difficult to implement, but is worth considering. A chimpanzee is likely to find it easier to figure out what its conspecifics "see and know" than what human trainers "see and know." Therefore, it would be interesting to replace the human trainers with chimpanzee trainers. Given how much difficulty we humans seem to be having determining what is going on in the head of a chimpanzee, chimpanzees may have similar difficulties determining what is going on inside the heads of human beings.

Seeing is not (necessarily) believing

Virginia Slaughter and Linda Mealey

School of Psychology, University of Queensland, Brisbane Australia 4072.
vps@psy.uq.edu.au lmealey@psy.uq.edu.ac

Abstract: We doubt that theory of mind can be sufficiently demonstrated without reliance on verbal tests. Where language is the major tool of social manipulation, an effective theory of mind must use language as an input. We suspect, therefore, that in this context, prelinguistic human and nonhuman minds are more alike than are human pre- and postlinguistic minds.

The task described by Heyes is an interesting one, however, given the failures of Povinelli and Eddy (1996), we doubt that it would be successful. Even if it were successful, could we, based on those results, conclude that chimpanzees have a theory of mind? We think not.

Success on Heyes's task would show that a chimpanzee knows something about the *process* of seeing, but not necessarily anything about the *content of the resultant mental state.* Heyes's goggle task is essentially a "Level 1," perspective-taking task (Flavell et al. 1981), which is distinguished from a more complex "Level 2" task. At Level 1, a child understands that another person can or cannot see something; at Level 2, a child also understands how something may look different to another person. It is only at Level 2 that children are attributed understanding of another person's mental content. Chimps who passed Heyes's task would demonstrate knowing that opaque goggles block sight and that a person wearing them cannot see, but we could not conclude that they necessarily represented any mental state of the goggle-wearer. Even if a chimpanzee who passed Heyes's test could make the inference that "the person exposed to the information (the translucent goggle-wearer) is the one I should follow," such a level of understanding would not necessarily require representing another's mental content (Barresi & Moore 1996; Flavell 1988; Perner 1991). Success on Heyes's task would provide evidence of a sophisticated and impressive intellect, but such success could be achieved just as well by a "radical behaviorist" chimp (Gallup 1996) or a "nomothetic psychologist" chimp (Mealey 1992) as by a chimp using a theory of mind. Since it is *the capacity to represent another's mental content as different from one's own* that has become the litmus test for a theory of mind (Dennett 1978; Fodor 1992), success on the goggle task would be consistent with, but not sufficient to demonstrate, the presence of a theory of mind.

Those investigating theory of mind in nonhuman primates face the same problems as researchers who study theory of mind in preverbal children (Cheney & Seyfarth 1992 and commentaries). In fact, the current state of the literature with regard to preverbal infant theory of mind is quite similar to the situation Heyes describes in her section 1 with regard to primate theory of mind; to paraphrase: "those working with [preverbal infants] have continued to struggle with the basic question of whether any [12-month-old] has any capacity to conceive of mental states." This debate is currently at the center of infant research, with some theorists arguing that gaze-following, social referencing, intentional communication, and other achievements of the child's first year constitute a first theory of mind, while others argue that all of those skills can be explained by appeals to simple learned contingencies (see Moore & Dunham 1995).

Heyes's optimistic description of the progress made in understanding the development of the child's theory of mind (sect. 1) refers primarily to work done on tests such as the false belief task, in which children explicitly state what they think another person believes, or sees, or wants. When a child says "he thinks it's Smarties, but I know it's a pencil in the box," there is no question that the child is representing another's mental content. However, as far as we know, no nonverbal test of theory of mind satisfies this criterion. Indeed, recent work with very young children and with autistic and deaf subjects suggests that a minimum linguistic capacity may be a necessary precursor not only for testing, but also for developing such representations (Baron-Cohen 1991; Budwig & Bamberg 1996; Gopnik 1993; Nelson 1996; Peterson & Siegal 1995).

The evolutionary advent of language as a tool of social manipulation may have resulted in selection pressures for a human theory of mind which were qualitatively different from the selection pressures leading to advanced social intelligence in chimpanzees (and other long-lived social animals). We suspect accordingly that pre- and post-linguistic minds, human or not, use quite different methods to achieve similar social ends.

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Simpler for evolution: Secondary representation in apes, children, and ancestors

Thomas Suddendorf

Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand. **t.suddendorf@auckland.ac.nz**

Abstract: Great apes show behavioral evidence for secondary representation similar to that of children of about two years of age. However, there is no convincing evidence for metarepresentation in apes. A good evolutionary interpretation should be parsimonious and must bring developmental and comparative data in accord. I propose a model based on the work of Perner (1991) and close by pointing out a logical flaw in Heyes's second proposed experiment.

Although Heyes begins her argument by acknowledging developmentalists' progress in the field, she ignores the empirical and theoretical advances that have been made. Yet these may hold the key to a sensible reconceptualization of the nonhuman primate data in an evolutionary framework. Such a framework can be based on Perner's analysis of children's developing understanding of the representational theory of mind.

The acid test for theory of mind in developmental psychology is the ability to attribute false beliefs (Wimmer & Perner 1983) because it implies an understanding that mental states are attitudes to representations of the world rather than to the real world. Children pass false-belief tasks by about age four. According to Perner (1991), this is due to children's emerging general capacity for metarepresentation (i.e., understanding representations *as* representations). With this ability children can also distinguish between appearance and reality and can understand representational change (e.g., Gopnik & Astington 1988). Other correlates include episodic memory (Perner & Ruffman 1995), divergent thinking (Suddendorf & Fletcher-Flinn 1997), imaginary object pantomime (Suddendorf et al. 1996) and a host of other skills that I categorize under the label *metamind* (Suddendorf, in press). Heyes is right that there is as yet no convincing evidence for a representational theory of mind in nonhuman primates, nor is the evidence convincing for any of the correlates of metarepresentation.

But children show a rudimentary consideration of mental states long before acquiring a metamind. By age two, they talk about mental states, engage in social pretence play, pass mirror self-recognition tasks, grasp synchronous imitation, and show empathic behaviour. Extrapolating from Perner (1991), one can argue that all these skills reflect children's new ability to form secondary representations (Suddendorf, in press). This is the ability to entertain and collate off-line mental models (e.g., about past, future, or imaginary situations) in addition to the primary reality model. This ability is also evident in two-year-olds' skill in understanding hidden displacement, interpreting pictures, and insightful problem solving. Correlations between mirror self-recognition, empathic behaviour, and synchronous imitation have been taken as support for the emergence of the ability to form secondary representations (Asendorpf et al. 1996; Bischof-Köhler 1989; Suddendorf, in press).

The facts that all these skills develop in tandem and that they all appear logically to require secondary representations, ought to be considered in comparative and evolutionary accounts of theory of mind. Great apes are capable of secondary representation! Their capacity for insight (e.g., Köhler 1927) shows that they can mentally compare a goal (secondary) situation with the present (primary) situation to figure out how to get from one to the other. In this light it is not very surprising that apes also display this skill in other realms such as understanding hidden displacement, pretence, mirror self-recognition, synchronous imitation, empathic behaviour, interpreting pictorial representation, and mental attribution of motivational states.

Great apes, in contrast to monkeys, show behavioral evidence very similar to that in two-year-olds in all these areas. The most parsimonious explanation of this similarity is that the same underlying mechanisms are involved. From an evolutionary perspective when all species of a superfamily (i.e., Hominoidea) share the same behavioral phenotype this suggests homology (i.e., a common ancestor already possessed that skill). There is no apparent reason to assume that convergent evolution produced different mechanisms to create the same skills in all five sister species. Parsimony here does not refer to affordances on part of the individual or the researcher (as Heyes discusses), but to the simplicity of an evolutionary account of the data.

Combining developmental and comparative data, then, it seems reasonable to attribute the capacity to form secondary representations to two-year-old children, great apes, and our common ancestor 15 million years ago. Although the ability to form secondary representations allows for limited attribution of mental states, it need not imply an understanding of representations *as* representations. Only by about age four do children develop a metamind. Since great apes have not yet demonstrated metarepresentation or any of its correlates, it has to be assumed that it evolved after our ancestors split from the line that led to modern chimpanzees. I have suggested that this occurred with *H. erectus* (dating from 1.8 mya) (Suddendorf, in press; Suddendorf & Corballis 1997). However, it can of course never be proven that apes do not have a representational theory of mind. Experiments, like the ones Heyes proposes, should continue.

Heyes's second experiment, however, has a logical flaw. A subject who actually has a theory of mind would have to wonder how the trainer on the "back trials" knows how to consistently choose the wrong well. Consistently getting it wrong implies as much knowledge about which well is baited as consistently getting it right. A clever subject would therefore abandon a mentalistic strategy (choose the trainer who sees the baiting because he knows where it is) for a behavioral one (choose the indicated well if the trainer faces the well and the other if he turns his back). In order to encourage the subject to adopt a mentalistic strategy training should be realistic. That is, the trainer in the "back trials" should be guessing about the food location, and therefore be correct on half the trials. It would also seem sensible to validate the test design with children *before* attempting the more difficult task of testing nonhuman primates. More collaboration between comparative and developmental research is desirable. An evolutionary perspective should be the link.

Precursors to theories of mind in nonhuman brains

Stephen F. Walker

Centre for Life Sciences, Birkbeck College, London WC1E 7HX United Kingdom. **s.walker@psychology.bbk.ac.uk**
www.psyca.bbk.ac.uk/staff/sfw/htm/

Abstract: Heyes is right that behavioural tests able to distinguish mentalistic from nonmentalistic alternatives should be sought, but the theoretical issue is less about the passing of behavioural tests than it is about the internal mechanisms which allow the passing of the tests. It may be helpful to try to assess the internal mechanisms directly by measuring brain activities.

The theory of "having a theory of mind" is loose and needs to be tied down, even with human subjects. There is overwhelmingly greater evidence for the development of mental state concepts in human infancy, however, and a certainty that such concepts function in human adults. There is also some support for the notion that there is a theory of mind "module" in the human case, or at least a network of links between the various separately measurable social skills (Fletcher et al. 1995; Karmiloff-Smith et al. 1995). As a particular case, contrary to the suggestion at the end of section 2.1, there is evidence from large samples of children supporting a very close correlation between mirror self-

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

recognition and sensitivity to imitation (Asendorpf & Baudonniere 1993; Asendorpf et al. 1996) plus a large literature relating imitation and other theory of mind tests (e.g., Azmitia & Hesser 1993; Loveland et al. 1994; Smith & Bryson 1994; Vonhofsten & Siddiqui 1993).

This literature contrasts starkly with the very limited corpus of positive behavioural evidence for mental state concepts corresponding to "want" and "know" even in chimpanzees, and the consensus is that social attribution and mirror self recognition are absent in other nonhuman primates and large-brained higher vertebrates (Povinelli 1989; 1993; Povinelli & Preuss 1995). There is little room for argument about whether support for the mentalistic theories requires behavioural evidence to distinguish mentalistic capacities from alternatives – we should surely presume that natural selection is not intelligent enough to be anything but behaviourist. Therefore if mentalistic capacities are the result of selection they must have arisen because they produce behavioural effects that increase the inclusive fitness of the individuals that have them. The only reservation I have about the study of perspective-taking proposed by Heyes is that the distinction between red-rimmed and blue-rimmed goggles seems rather remote from any naturalistic function that precursors to perspective-taking in wild chimpanzees might serve. One-way and two-way silvered screens are suggested as alternatives and these would seem better as a starting point, as more similar to completely opaque or partly opaque vegetation. However, in any such stringent test it seems quite probable, on the basis of reports published so far, that chimpanzees would fail.

In that case, should all discussion and experimentation on mentalistic cognitive processes in nonhuman primates cease? It is unlikely that it would, because although the experimental manipulations suggested by Heyes are helpful, there are implicit and sometimes explicit theoretical assumptions related to the "theory of mind" tests which are wider than any particular crucial test. One underlying assumption is that nonhuman primates have enlarged brains by general mammalian standards, and that the expansion of primate neocortex is to some degree related to sociality (Humphrey 1976). [Cf. Falk: "Brain Evolution in *Homo:* The 'Raditor' Theory *BBS* 13(2) 1990.] Data to test these assumptions are still fairly limited (Barton 1996), but presumably the large brain size of primates is not an accident, and a functional relation with sociality could hold even if primate mentalistic capacities were very severely limited by comparison with those demonstrated by the passing of false-belief tests by human 10-year-olds. Phylogenetic relatedness itself means that there is special value in examining the brain mechanisms of cognition in primates; and the details of brain functioning, rather than the behavioural effects of brain functioning, provide an independent source for investigating similarity between human and nonhuman mechanisms, even where human capacities may be qualitatively different from those of any other extant species. There is continuing interest in human and primate brain mechanisms for self-related aspects of visual attention, and visually controlled reaching and grasping (Graziano et al. 1994; Hietanen & Perrett 1996; Jeannerod et al. 1995; Johnson et al. 1996; Kertzman et al. 1997; Witte et al. 1996). There is already substantial evidence for a commonality in human and nonhuman primate brain mechanisms in at least some precursors to theory of mind tests.

In particular, the integration and transformation of visual information into representation of motor activity would be a prerequisite for imitation under most definitions. Cells in macaque temporal cortex recognize the direction of motion and view of the body, and a proportion of these continue to be selective when the information is limited to the movement of light patches attached to the points of limb articulation (Oram & Perrett 1994). Even more closely related to precursors of imitation of object manipulation, there are cells in parietal cortex which respond to objects according to type of manipulation (Murata et al. 1996). These provide input to an area *s* in premotor cortex of macaques where some cells discharge either when the animal performs a grasping action itself (even in the dark) or when it observes the human experimenter or another monkey perform the action (Rizzolatti et al. 1996). It has been proposed (not unreasonably) that these cells may be part of an observation/execution matching system (Gallese et al. 1996) which shows some degree of comparability between macaques and humans (Fadiga et al. 1995; Grafton et al. 1996).

This does not at all imply that macaque capabilities for the purposive imitation of grasping and gripping actions are even remotely equal to those of humans, and in a sense this sort of special purpose sensory-motor transformation system is the alternative to sweeping mentalistic accounts. There is also no reason to suppose that some degree of multi-modal transformation of information does not occur in the brains of rats (Chudler et al. 1995) and indeed even in the superior colliculus of lower vertebrates (Spreckelsen et al. 1995). But a promising direction for research on theory of mind in primates would be to study commonalities of brain mechanisms. If functional brain imaging studies of children, chimpanzees, and macaques in mirror self-recognition conditions becomes feasible and the patterns of activity observed in ostensibly self-recognizing children and chimpanzees are found to be similar, this would be evidence complementary to that obtained by purely behavioural controls, but something quite different is correlated with face touching in macaques.

Triangulation, intervening variables, and experience projection

Andrew Whiten

School of Psychology, University of St. Andrews, St. Andrews, KY16 9JU, Scotland. **a.whiten@st-andrews.ac.uk psych.st-and.ac.uk:800**

Abstract: I focus on the logic of the goggles experiment, which if it as watertight as Heyes argues, should clearly support ape theory of mind if positive, and clearly reject it if negative. This is not the case, since the experiment tests for only one kind of mindreading, "experience projection": but it is an excellent test for this, given adequate controls.

There are several substantive points of agreement between Heyes's analysis of the state of the art in primate theory of mind research and my own (Whiten 1997; also especially 1994 and 1996 as cited in the target article). Independently, both Heyes (1993) and Whiten (1993) recognised that the nature of mental state attribution is such that a particular kind of complexity will be necessary in its empirical identification: specifically, to the extent that mental states function as "intervening variables" for the mindreader (Fig. 1; Whiten 1993; 1996), the identification of mindreading will require the kind of evidence that Heyes (1993 and target article) called "triangulation." However, I think a further and different complexity is raised by the experiment with which the target article culminates: given that I also agree with Heyes that the "right" experiments are vital (Whiten in press a; in press b) I shall restrict myself to the interpretation of the "goggles" test that she proposes, trusting that the commentaries by my coworkers Byrne and Custance will deal with what seem to be vigorous misinterpretations by Heyes with respect to earlier publications on tactical deception, Machiavellian intelligence, and imitation.

If the goggles experiment is the watertight kind that Heyes advocates, passing means theory of mind and failing means no theory of mind. A pass would be very convincing. But that would also have been true of Premack and Woodruff's (1978) first attempt at a test of false belief attribution in chimpanzees: the problem was that when the ape failed, excuses could be suggested and the experimental hypothesis could not easily be rejected (Premack 1988, p. 178). One reason the goggles experiment is powerful if passed but ambiguous if failed relates to the now voluminous literature on the different ways in which mindreading could get done (e.g., Carruthers & Smith 1996; Davies & Stone

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

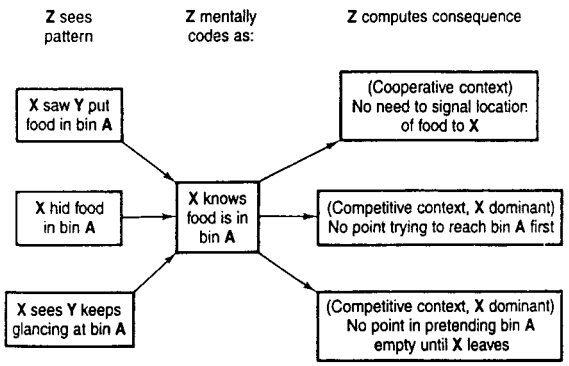


Figure 1 (Whiten). Here, a hypothetical primate, Z, reads the mental state of *knowing* in an individual X, coding this state on different occasions by recognising as an equivalence class a variety of circumstances like those shown on the left. Once X has been classed in this state, predictions appropriate to different circumstances such as those shown on the right can be generated. The state thus functions as an intervening variable, offering economy of representation (after Whiten 1993; 1996). Methodological implications for triangulation are discussed in the text.

1995). The goggles experiment, as outlined by Gallup and refined by Heyes, imaginatively tackles one of these ("experience projection" – see below): but it would not be a watertight test of mindreading per se, because a fail would be consistent with chimpanzee mindreading being done in one of several other ways (and as the publications just cited make clear, we are far from agreement over which kind of mindreading humans do!).

Two different ways of mindreading seem implicit in Heyes's paper and I shall try to make them explicit. I can best explain this by reference to the analysis that Heyes cites in which I distinguished four senses in which nonverbal mindreading might differ from "mere" behaviour-reading (Whiten 1996) – not a trivial distinction, because mindreading is not telepathy, but instead has to be done through observing others' behaviour and/or circumstances. Two of the four candidate differences I judged to be relatively weak and there is not space to discuss them here. Of the other two, one was that a mental state has the status of an "intervening variable" (see Fig. 1), so mindreading necessitates recognising such states. There is not space here to rehearse all the implications of this (see Whiten 1996, pp. 285–88), but the one which is highly relevant here concerns methodological identification of mindreading when conceived of in this way (Whiten 1996, p. 288). It requires testing for transfer in different conditions – what Heyes calls "triangulation." However, as she describes it, this requires testing a subject for recognition only of the left hand side of an intervening variable's "web," like that in Figure 1. In fact I would suggest that the methodological requirement is more severe than Heyes herself acknowledges: the right hand side needs testing also, to reject the hypothesis that the putative mindreader just knows simple rules for how a couple of antecedents on the left directly predict just one of the outcomes on the right. This means that identifying mindreading of this character is never going to be absolute, and probably not achievable through a one-off experiment, a point emphasised by Dennett (1996, p. 125). However, the power of the first strong test of false-belief attribution done on children recognised this and did test the "right side" of the web – not just the left, which Heyes considers – by examining an appreciation by the subject that the character "Maxi" could use his belief either to deceive his sibling or to help his grandpa (Whiten 1997; Wimmer & Perner 1983). So Heyes and I agree on the importance of triangulation, but Heyes's conception is too narrow, at least if the force of the intervening variables conception is accepted.

The other sense of nonverbal mindreading considered by Whiten (1996) was "experience projection," a version of what many theorists have discussed recently under the heading of "simulation" (Carruthers & Smith 1996; Davies & Stone 1975). In this, mindreaders operate essentially by using their own mind to compute how another individual is likely to behave under the circumstances at stake. Now it seems quite clear that the goggles test is a test of mindreading by experience projection, or simulation: the putative mindreader will succeed by projecting onto the person wearing the goggles their own prior experience of being able to see, or not see, as associated with the colour of the goggle frames concerned.

Thus the central point I am making is that although Heyes prefaces her presentation of the goggles experiment with a discussion of triangulation, the goggles experiment and triangulation are referring to importantly different methodologies and potentially quite different kinds of mindreading. If true, this is of considerable importance. The reason that a positive result on the goggles test would be so powerful is partly that it sidesteps the tricky triangulation issue (including the more complex operation I suggest is necessary given the shape of Fig. 1). As I've said before, I think experience projection is important to test, on both human and nonhuman primates (Whiten 1991b, p. 330). Conversely, however, it is important to recognise that a negative result, although counting against experience projection, would leave open the question of whether the subject mindreads by one of the several alternatives distinguished in the literature cited below.

What can we learn from the absence of evidence?

Thomas R. Zentall

Department of Psychology, University of Kentucky, Lexington, KY 40506.
zentall@pop.uky.edu

Abstract: Heyes discounts findings of imitation and self recognition in nonhuman primates based on flimsy speculation and then indicates that even positive findings would not provide evidence of theory of mind. Her proposed experiment is unlikely to work, however, because, even if the animals have a theory of mind, a number of assumptions, not directly related to theory of mind, must be made about their reasoning ability.

Heyes presents an important message that has sometimes been lost on "animal cognition" researchers. Mentalistic interpretations of data require a special level of support because the evidence is always indirect and the support is often by exclusion (i.e., once nonmentalistic accounts are ruled out, mentalistic accounts are all that remain).

To support her claim that theory of mind has not been shown in nonhuman primates, Heyes attacks a number of peripheral areas of research (imitation and self recognition). The logic of whether evidence of imitation and self recognition would support theory of mind in animals aside, her attack on the validity of the research itself seems over-critical and unsubstantiated.

On the one hand, Heyes accepts data from research with rats (Heyes & Dawson 1990) and budgerigars (Galef et al. 1986) involving the two-action method (in which a demonstrator acts on an object in one of two distinctive ways and the observer matches the response topography of the demonstrator; see also Zentall et al. 1996, with pigeons, and Akins & Zentall 1996, with Japanese quail). On the other hand, she discounts Custance et al.'s (1995) replication of Hayes and Hayes's (1952) demonstration of generalized imitation in which the concept "do this" was trained and found to apply successfully to novel gestures. According to Heyes, these novel imitations were *possibly* (1) trained prior to the experiment (e.g., imitation of lip smacking may have occurred during rearing by humans) or (2) generalized from original training (e.g., trained nose touching may have generalized to tested chin touching). But in the absence of supporting evidence, speculation about experience prior to the experiment is unconvincing.

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Furthermore, the notion that transfer from one behavior to another can be explained as generalization from training shows a cavalier disregard for the mechanisms which govern generalization.

Stimulus generalization occurs when a trained response is made in the presence of a stimulus different from the training stimulus. Following nose-touch training, the chin-touch test clearly represents a different stimulus, but according to stimulus generalization theory, it is the probability of the trained response that is in question (i.e., nose touching) rather than the probability of a new behavior (i.e., chin touching). Response generalization, on the other hand, is typically viewed as random variability about a target (reinforced) response and thus, such response variability should be uncorrelated with the change in stimulus. To say that the novel demonstration of chin touching results in a corresponding shift in response from nose touching to chin touching suggests either chance correspondence between stimulus change and response change (quite unlikely given that at the end of training of nose touching, chin touching must have been relatively unlikely to occur) or true imitative learning.

The second area in which Heyes may be too critical without supporting evidence is that of self recognition (e.g., Gallup 1970). To suggest that the within-subject control test without mirror may have resulted in fewer mark touches (relative to the mirror test performed immediately afterwards) because of the aftereffects of the anesthetic is without serious merit – especially considering Suarez and Gallup's (1981) delayed control test (4–5 hrs after marking).

The notion that imitation may involve theory of mind may derive from Piaget's (1945) view that imitation involves perspective taking, but given the number of species that appear to show imitative learning, perspective taking is unlikely to be the underlying mechanism. Furthermore, as Heyes notes, even if perspective taking is necessary to account for imitation, it does not imply a theory of mind. The same can be said of self recognition.

What then of Heyes's suggestion about modifying Povinelli et al.'s (1990) knower/guesser technique? I think this proposal may represent a nice "thought experiment," but one which may not work even if chimpanzees do have theory of mind. First, how could one ensure that the animals associate the color of the rims of the goggles with their light-transmitting characteristics? or that they associate the goggles on themselves with the same goggles on another? These are critical links in the chain without which theory of mind cannot be demonstrated, regardless of its validity. Second, might a differential emotional reaction to the wearing of the two pairs of goggles (e.g., fear produced by loss of sight when wearing the opaque goggles) generalize to the test context and result in avoidance of the opaque goggles on another animal (an experimental confound).

The "negative transfer" control procedure (for half the test animals the opaque goggles would be on the "knower") suggested by Heyes does provide a useful comparison to control for a variety of possible artifacts. But an additional comparison of results might be made with a potential "zero transfer" control procedure. Under the assumption that experience with the colored-rim goggles is necessary for the animals to understand the differences in their light-transmitting characteristics, how rapidly would a group of animals learn the knower/guesser task without any wearing experience with the two pairs of goggles on the part of the target animal?

The real problem with such experiments (as with all research on the learning capacities of animals), however, is that one can conclude little from the absence of evidence. Only positive evidence is interpretable.

The above comments notwithstanding, Heyes's cautions are well taken. Furthermore, the critical evaluation of the work of others can often lead to a more critical examination of one's own ideas by others. In this regard, Heyes is to be commended for taking this risk in the name of designing a better test of theory of mind in animals.

Author's Response

Liberalism, chauvinism, and experimental thought

C. M. Heyes

Department of Psychology, University College London, London WC1E 6BT, United Kingdom. **c.heyes@ucl.ac.uk**

Abstract: The target article argued that there is currently no reliable evidence of theory of mind in nonhuman primates and proposed research methods for future use in this field. Some commentators judged the research proposals to be too chauvinist (in danger of falsely denying that primates attribute mental states), but a majority judged them to be too liberal (in danger of falsely affirming theory of mind in primates). The most valuable comments from both camps exemplified "experimental thought," the obverse of "thought experiments," and recommended specific alterations and alternatives to the studies I proposed. This Response evaluates these recommendations and presents a revised version of the proposals that appear in the target article. Other valuable commentary cast doubt on the assumption that people have a theory of mind, aired the possibility that language may be a prerequisite for either possession or detection of a theory of mind, questioned the notion of critical experiments, and emphasized the distinction between attribution of sight and belief. In addition to commenting on these issues, I respond to objections to my interpretation of existing research on self-recognition, imitation, and deception.

R1. Experimental thought

The target article was an exercise in "experimental thought," the obverse of "thought experiments." Although thought experiments substitute for real empirical investigation, the purpose of experimental thought is to provoke and prepare effective empirical work. It can do this in two ways. First, thinking about existing experiments and other empirical studies can reveal that more are needed to answer a particular question. This was the purpose of the target article's review of the literature on theory of mind in primates. Second, thinking about potential experiments, particularly when it combines the expertise of many researchers, can lead to the implementation of experiments that have a good chance of answering the question at issue. This was the purpose of the methodological proposals in the target article and of subjecting them to Open Peer Commentary.

Several of the commentaries included incisive experimental thought of just the right kind. They identified weaknesses in the experiments I proposed and suggested either specific alterations and refinements to those procedures or complementary research strategies. One of these commentaries even reported the results of a pilot study using one of the procedures I recommended.

A second group of commentaries raised important general issues relevant to empirical enquiry about theory of mind in primates. They cast doubt on the assumption that people have a theory of mind, aired the possibility that language may be a prerequisite for either possession or detection of a theory of mind, emphasized the distinction between attribution of sight and of belief, questioned the notion of "crucial experiments," and suggested that the hypothesis that nonhuman primates have a theory of mind requires a special kind or level of evidential support.

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

The final group of commentaries were not so valuable. They found fault with specific sections of the literature review, but did not explain how these weaknesses, if genuine, would affect the big picture, on Premack and Woodruff's (1978) question. For example, those who insisted that nonhuman apes really *do* imitate and/or recognize themselves in mirrors did not provide any reason to believe that imitation and mirror self-recognition are related to theory of mind.

On another dimension, the commentaries fell into two groups: some took my approach to be too chauvinist, in danger of falsely denying that nonhuman primates have a theory of mind, and others thought it too liberal, in danger of falsely attributing theory of mind to these animals (Block 1978). A majority of commentators considered the target article to be too liberal, but those who found it too chauvinist were more stridently critical.

I will respond to all substantive comments and in roughly the order in which the relevant issues arose in the target article. Thus, R1 deals primarily with comments on section 1 of the target article, R2 with comments on section 2, and so on for R3 and R4. However, at the subsection level, this mapping between the target article and the response does not apply. For example, the contents of R4 are subdivided further and in different ways than the contents of section 4, reflecting my bias in this response toward summarizing and evaluating commentators' specific experimental proposals.

R1.1. Is Premack and Woodruff's question worth asking?

A number of commentators stated or implied that it is not worth trying to answer Premack and Woodruff's (1978) question empirically. For some, the problem is epistemological: even if there are nonhuman primates that have a theory of mind, we cannot know this. In **Mitchell & Anderson**'s case, this view seems to be born out of solipsistic doubt that we can know anything about what another person or animal thinks, but for **Gray & Russell, Green et al.** and **Slaughter & Mealey** the epistemological problem is more specific. They doubt that we can find out whether a person or animal has a theory of mind unless that individual uses language.

For other commentators, the problem is ontological: it is so unlikely that nonhuman primates have a theory of mind that it is not worth trying to find out whether they do. Like **Mitchell & Anderson** on the epistemological side, **Baum**'s ontological objection is very general. He views all reference to theory of mind, and presumably to mental states in general, as behavior to be explained in terms of reinforcement history, and therefore denies on logical grounds that any person or animal could have a theory of mind in any deeper sense than that of being a user of mental state terms. If this is true, then nonhuman primates obviously do not have a theory of mind, and human use of mental state terms is not explained by attributing to them a theory of mind or any other internal psychological properties.

Other more specific ontological objections also focus on language, but they assume that having a theory of mind is linked with, but does not consist of, using mental state terminology. Thus, **Slaughter & Mealey** suggest that language use provided the selection pressure for evolution of theory of mind in humans, and **Parker** presents the interesting hypothesis that language is necessary for autobiographical memory, and that this kind of memory is in turn necessary for self-conception and conception of other minds. **Kamawar & Olson** argue that language is necessary to understand implicational relations between mental states. In their view, such understanding is a hallmark of genuine theory of mind, and therefore whether or not nonhuman primates exhibit behavior that can be described, predicted, and explained using human theory of mind, these animals cannot possess such a theory themselves. In contrast, **Gordon** and **Green et al.** are inclined to think that humans do not have a theory of mind in a "full-blooded" sense (one that includes law-like generalizations and inferences to behavior from mental states), and therefore suspect that it is a waste of time to look for this sort of underpinning of social behavior in nonhuman species.

R1.1.1. Philosophical theories of mind.

For me, the most interesting feature of **Mitchell & Anderson**'s and **Baum**'s very general objections is their similarity. Mitchell & Anderson's Cartesian introspectionism takes mental states to be very real and essentially private, while Baum's radical behaviorism casts mental states as unhelpful fictions, but both parties conclude that empirical enquiry about mentality, and therefore about theory of mind in primates, is ultimately hopeless. Although **Gallup** has me down as a radical behaviorist, evidently I disagree with them both. Like most cognitive scientists, I adopt a functionalist view of mental states. I take them to be theoretical entities, characterized by their relationships with sensory inputs, motor outputs, and other (functionally defined) mental states, not by their subjective properties and therefore to afford empirical investigation. This philosophical (rather than "folk") theory of mind contrasts with Cartesian introspectionism and radical behaviorism, but it would be inappropriate for me to give my reasons for subscribing to functionalism rather than one of the other two since neither Mitchell & Anderson nor Baum justify their commitments. On this issue, we are all to some degree restating dogma, but at least Baum's restatement was lucid and entertaining.

R1.1.2. Language and theory of mind.

It may well be true that language is necessary for the detection of theory of mind (**Gray & Russell, Green et al., Slaughter & Mealey**). However, unless one assumes that having a theory of mind consists in the appropriate use of mental state terms, this is an empirical question, and the only way we can answer it is by trying to devise nonverbal tests. Perhaps a detailed survey of developmental research would reveal that this has been attempted with determination and ingenuity for children and failed. But the commentators did not offer or cite such a survey, and even if they had, it would be worth trying again with nonhuman primates because failure with children could be related to the relatively limited motor capacities of prelinguistic children or to nonlinguistic deficits in abnormal populations. As argued in the target article, very few attempts have been made to devise a valid test of theory of mind in nonhuman primates (as opposed to a test in which a positive outcome could equally well be explained in nonmentalistic terms); hence the lack of such a test for this group is not in itself an indication that language is necessary for theory of mind detection.

Turning to those who believe there is an ontological link between language and theory of mind (**Slaughter & Mealey, Parker, Kamawar & Olson**), I share their hunch that language is phylogenetically and ontogenetically necessary for mental state attribution, and therefore my pre-

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

diction would be that mentalistic processes will not be found in nonhuman primates. There are several reasons why, in spite of this, it is important to devise and conduct experiments to answer Premack and Woodruff's question.

First, the proposal that nonhuman primates have a theory of mind is not a flat earth hypothesis; it is entirely conceivable that they do, and there are many intelligent people who feel certain that the hypothesis is correct. Second, because there are many people who believe in the hypothesis, inside and outside the research community, it is likely that resources will continue to be used for research on theory of mind in primates, and they are better spent on experiments with the potential to answer Premack and Woodruff's question than on the production of additional ambiguous data. Finally, the costs of a false negative answer to their question would be high because compelling evidence that nonhuman primates attribute mental states would have important implications for our understanding of human folk psychology and the evolution of intelligence. For example, it would seem to favor innate module theory (e.g., Baron-Cohen 1995; Leslie 1991) and simulation theory (e.g., Morton 1980) over theory theory (e.g., Gopnik 1993), because while the former could ascribe the occurrence of theory of mind in human and nonhuman species to inheritance of a common computational device (innate module theory), or to common contents of introspection (simulation theory), under theory theory it would be a highly improbable coincidence. Theory theory emphasizes the role of language and social interaction in the development of theory of mind, but nonhuman primates do not have language and their social environments are very different from ours. Evidence that primates do and that other species do not attribute mental states would also support the "social function of intellect" hypothesis (Humphrey 1976) – the suggestion that it was the social environment, rather than the advantages of object manipulation skills, that was the primary source of selection pressure for the evolution of intelligence in the primate line.

R1.1.3. Do humans have a theory of mind? It is suggested by **Gordon** and **Green et al.** that humans do not have a theory of mind of the kind that I, and many primate researchers, are seeking in nonhuman species. In the case of Green et al., this suggestion follows from the claim that children's performance on the Sally/Anne false belief task can be explained in nonmentalistic terms, and specifically as being due to associative learning. The hypothesis they offer, however, is unrelated to any theory of associative learning that I have ever encountered, and fails to explain, among other things, why some children pass the test while others fail, and why those who pass give different answers to questions about where Sally will look for the hidden object, and about the true location of that object.

On the other hand, **Gordon** believes that humans use "mentalistic resources" to predict and explain the behavior of others, but doubts that those resources include law-like generalizations allowing inferences of mental state from behavior. This is a potentially important distinction, but Gordon does not draw it very clearly. What are these "mentalistic resources" that are both distinguishable, conceptually and empirically, from nonmentalistic processes, and fail to involve generalizations such as "[Conspecifics] act in such a way that will satisfy their desires if their beliefs are true" (Fodor 1992), and inferences such as "He grabbed the banana because he wanted to eat it"? Maybe they are the resources ascribed to humans by simulation theory. If so, there is no fundamental incompatibility between Gordon's view and that taken in the target article. Although I characterized individuals with a theory of mind in the style of theory theory (sect. 1), my criticisms of existing research on theory of mind in primates and proposals for future research do not depend on the validity of theory theory. For example, if nonhuman primates have a theory of mind by their respective lights, both theory theory and simulation theory would predict the same outcome for the "goggles" experiments I outlined. Indeed, **Whiten** argued that the task I recommend is particularly compatible with simulation theory because it encourages subjects to extrapolate from their own experience of the opaque and translucent goggles to that of others.

Kamawar & Olson mention another distinction that may be important, but which I find rather cryptic, between understanding the implicational relations holding among a set of linguistically coded concepts and exhibiting behavior that may be characterized in terms of that structure. They suggest that humans use the former only when they are formulating verbal explanations for behavior, which leads one to wonder whether they might not be conceiving of the understanding of implicational relations between mental state concepts in such a way that it consists of being able to offer verbal explanation of behavior, or, at least, could not be detected in any other way. In either case, I readily conceded that research of the kind I support could not tell us whether primates have a theory of mind in Kamawar & Olson's first sense. However, if it has the potential to show that human primates exhibit behavior that may be characterized *uniquely well* in terms of a conceptual structure involving mental states (i.e., more plausibly than in nonmentalistic terms), then it will be a major advance on the existing data, and would satisfy me that nonhuman primates have a theory of mind that is similar in important respects to that of humans.

R1.2. Relative strength of developmental and primate research on theory of mind. Finally, in this roundup of objections to section 1 of the target article, it should be noted that several commentators took exception to my claim that developmentalists have made much more progress in their research on theory of mind than primate researchers. **Gallup** and **Mitchell & Anderson** pointed out that there are methodological weaknesses in research on self-recognition in children. I agree entirely, and this would be contrary to my view only if I had claimed that developmental research is perfect. **Miles & Roberts** drew attention to the fact that developmentalists disagree among themselves about many issues, which was, in a sense, precisely what I identified as the hallmark of their progress. They are in a position to engage in "*disciplined* dispute" (sect. 1 of the target article, emphasis added); relative to primate researchers, their disparate views are clearly specified and subject to resolution by empirical means.

R2. Defences of existing research

Existing research in six fields was reviewed in the target article (imitation, self-recogniton, social relationships, deception, role-taking, and perspective-taking), and the commentaries included objections to the conclusions reached

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

in three of these fields (imitation, self-recognition, and deception). However, no commentator identified a study or studies and claimed that the results could not be plausibly explained in nonmentalistic terms.

Nearly all of the comments about imitation and self-recognition concerned competence rather than validity; they tried to show that primates are capable of imitation and self-recognition, but did not present arguments or evidence to suggest that imitation and self-recognition are valid indicators of theory of mind. Thus, the preoccupations of these commentators (principally **Bard, Custance, Gallup, Mitchell & Anderson, Zentall**) indirectly reveal that I made a strategic mistake in the target article. Since my main points about imitation and self-recognition concerned their validity as indicators of theory of mind, I should not have argued in addition that the evidence of primate competence is these areas is inadequate. Doing so created the opportunity for some commentators to ignore the main points and focus on competence, and now I feel duty bound to answer their objections, however irrelevant this exchange may be to the question of whether primates have a theory of mind. To assist the reader who does not want to come along for this ride, I respond in R2.1 to the few comments that were made about validity and devote R2.2 and R2.3 to questions of competence.

R2.1. Are imitation and self-recognition valid indicators of theory of mind? In the target article I denied that evidence of imitation and mirror self-recognition in nonhuman primates would imply mental state attribution, that is, that nonhuman primates need mental state concepts in order to imitate body movements or to use a mirror to derive information about their own bodies. No one directly challenged this position, but some commentators claimed that there is a link of some sort between imitation and mirror self-recognition (**Walker, Suddendorf**) and/or between imitation and self-recognition on the one hand and theory of mind on the other (**Bard, Walker, Miles & Roberts, Mitchell & Anderson**). I will respond to these comments simply in order to make it clear that they do not contradict my position or imply that imitation and self-recognition are valid indicators of theory of mind.

There can be little doubt that some of the cognitive operations involved in imitation and mirror self-recognition are the same. Unless the common operations consist of mental state attribution, however, or of processes that lead specifically and directly to the capacity for mental state attribution, they are irrelevant to the question of whether imitation and self-recognition are valid indicators of theory of mind. The research mentioned by **Walker, Mitchell & Anderson** and **Suddendorf** does not show an interesting, specific correlation between imitation and mirror self-recognition as they are defined in the primate literature, let alone that any correlation is due to mentalistic processing or direct precursors of mentalistic processing. For example, Asendorpf and Baudonniere (1993, cited by **Walker**) showed that infants who passed a mirror self-recognition test engaged in more "synchronic imitation" than those who failed the test, but their measures of self-recognition and imitation were not comparable to those used in nonhuman primates. The self-recognition test included verbal measures, and the imitation test would be regarded by comparative researchers as a measure of local enhancement. Furthermore, Asendorpf and Baudonniere's study did not control for mental age or developmental level, and therefore the children who passed the self-recognition test may have been superior to those who failed on a wide range of indices. Similarly, Mitchell & Anderson (1993) (cited by **Mitchell & Anderson**) showed that a single rhesus monkey, Rodrique, neither passed a mark test of mirror self-recognition nor generalized on the basis of training to scratch the same part of his body as the experimenter. If this counts as evidence that imitation and self-recognition are linked, then it also shows that these two are correlated with playing the stock market. Presumably Rodrique didn't do that either.

The same principles apply to purported links between imitation and/or self-recognition on the one hand and theory of mind on the other. There must be many necessary conditions for theory of mind development, including respiration and sensory function. Thus, even if imitation and/or self-recognition are among these necessary conditions, as suggested by **Bard,** they cannot be regarded as indicators of immanent or actual theory of mind unless we have good reason to believe that they lead directly and specifically to mental state attribution, and the commentators do not allude to such reason. For example, Smith and Bryson's (1994) literature review (cited by **Walker**) converges on the conclusion that autistic impairments in imitation and theory of mind both arise from core deficits in attention and perceptual integration, and therefore casts imitation and theory of mind problems, not as cause and effect, but as correlated consequences. Similarly, **Miles & Roberts** point out that self-recognition and theory of mind could be dissociated in autists but related in the normal population. Yes, that is conceivable, but because there is evidence that they are dissociated in autism (Dawson & McKissick 1984; Ferrari & Matthews 1983; Ungerer et al. 1981), it would take more than a plausible suggestion that they are linked in the normal population to make mirror self-recognition a valid means of diagnosing theory of mind in human or nonhuman primates.

Thus, the commentary process has not revealed, and I do not know of any good reason to believe that we can find out about theory of mind in primates by establishing whether they can imitate body movements or pass a mirror test of self-recognition. Indeed, two recently published studies (Kitchen et al. 1996; Hauser et al. 1995) provide further reason to doubt that mirror-guided body inspection is related to self-conception. Kitchen et al. (1996) took chimpanzees that had passed the mark test of self-recognition using a normal mirror and exposed them to convex, concave, and triptych mirrors. They did not find any evidence of surprise or alarm when the animals were thus confronted with distorted and multiple images of their own bodies, suggesting that the chimpanzees' interest in the mirror images was based on their action-contingent properties, and did not involve comparison between the mirror image and a mental representation of the viewer's own body, let alone a mental representation of its "self."

In an unusually well-controlled study, Hauser et al. (1995) found evidence that cotton-top tamarins pass a mark test of mirror self-recognition when the marking procedure consists of dying the tufts of distinctive species-typical white hair on the tops of the monkeys' heads. Researchers who believe that there are already valid tests of theory of mind in primates generally do not believe that New World monkeys pass these tests. Therefore, by their lights, this

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

evidence of self-recognition in monkeys suggests that mirror self-recognition is not an indicator of theory of mind.

R2.2. Imitation: Competence. There were three substantive challenges to my analysis of research on imitation, focusing on the issue of generalization, a study of an orangutan (Miles et al. 1996), and data from rats and budgerigars.

The first of these is the most interesting. **Zentall** and **Custance** argued that the imitation of novel acts reported by Custance, Whiten, and Bard (1995) was unlikely to be due to generalization from earlier training, and **Mitchell & Anderson** suggested that it would be equally significant whether or not it resulted from generalization. In response to Zentall, I should clarify that I was suggesting that apparent imitation of chin touching could have resulted from *stimulus* generalization, and assuming that changes in the stimulus (e.g., from nose to chin touching) can result in a change in the vigor, as well as the probability, of the trained response. Since chin touching does not require the arm and hand to be elevated as far as nose touching, the response to the new stimulus might therefore take the form of chin touching. However, I concede to Custance that if performance tended to improve across successive demonstrations of a single novel act, *and* if there was no adventitious reinforcement of imitative responding, then stimulus generalization is unlikely to have been responsible.

In contrast with **Mitchell & Anderson,** I believe that if stimulus generalization of the kind I propose were responsible for the chimpanzees' behavior in response to novel demonstrations, this behavior would be much less interesting because it would not imply that these animals understand that there is spatial correspondence between, for example, a trainer's chin and their own chin. An understanding of this correspondence would, in turn, be interesting because it would seem to derive from something more complex than a sensory matching process; perhaps from some kind of spatial analogical reasoning. (But let's not get carried away. Not only would it not require the attribution of mental states to recognize the similarity between my chin and yours, but it is not at all clear how mentalistic concepts could help me to solve this mapping problem.)

Thus, the commentaries have reinforced my view that Custance, Whiten, and Bard have provided "the strongest evidence to date that, at least after training, the form or topography of a primate's action can be influenced by observing the same action by a demonstrator" (sect. 2.1, target article). **Mitchell & Anderson** note that I did not mention a similar study of an orangutan (Miles et al. 1996). This is because I was focusing on the strongest and most widely cited putative evidence of theory of mind in primates, and this study does not fall into either category. For example, Miles et al. did not report any information about which actions were presented during training, and therefore all of the apparent imitations could have been matched dependent behavior.

I referred to experiments on imitation in rats and budgerigars in order to recommend two-action tests and bidirectional control procedures for research on imitation. Since both **Custance** and **Zentall** have published data from such procedures themselves, I assume that there is no quarrel between us about their value, and otherwise the debate about whether rats and budgerigars can imitate has little relevance to Premack and Woodruff's question. Therefore, I refer the reader again to Byrne and Tomasello (1995) and Heyes (1996) for discussion of these issues, and note in passing that, if there is a killjoy explanation for the rat bidirectional control data (e.g., Heyes et al. 1992), my hunch is that it will relate to scent cues.

R2.3. Self-recognition: Competence. I dearly wish that, as I write, the critical study cited by **Gallup** (Povinelli et al. 1997) had already been published. If, as Gallup claims, it really meets the requirements for a valid test of mirror self-recognition noted in the target article (and in greater detail in Heyes 1995b), then the question of whether chimpanzees are capable of mirror-guided body inspection has finally been settled, and the answer, as I among others anticipated (Heyes 1994c), is "yes." In this case, attention could be confined to the outstanding questions: Why has it taken so long? Why was it so widely believed that apes can recognize themselves in mirrors before the convincing evidence arrived? How do chimpanzees do it? The last of these is by far the most important, but the others are of interest from a science studies perspective. Unfortunately, since these data are not yet on public record, and a previous study that was claimed to have met similar requirements did not do so (reanalysis of Megan's data reported in Gallup et al. 1995; see Heyes 1995b for commentary), I should respond, albeit briefly, to comments about other data that have been upheld as evidence of self-recognition competence.

The methodology used in all four studies cited by **Bard** as providing evidence of self-recognition without anaesthesia was weaker yet than that of the studies reviewed in the target article. (**Mitchell & Anderson** also alluded to studies involving sham marking but did not provide any citations.) In three of them (Lin et al. 1992; Miles 1994; Patterson & Cohn 1994) the marking procedure was such that the animals could have detected either mark application or the marks themselves using tactile, olfactory, or direct, rather than mirror-mediated, visual cues. The fourth (Hyatt & Hopkins 1994) reports on the spontaneous behavior of common chimpanzees and bonobos in the presence of mirrors and does not compare it with their behavior in the absence of mirrors. Thus, it assumes, in common with Bard and with Gallup and Mitchell & Anderson when they draw attention to reports that apes use mirrors to manipulate otherwise invisible body parts, that researchers can tell through casual observation whether an animal is seeing and using its mirror image, or merely looking in the direction of the mirror. If it were so easy to distinguish looking and seeing, answering Premack and Woodruff's question would pose much less of a challenge. As it is, both "folk" and cognitive psychology define mental states such that they can be inferred from, but not observed in, behavior (see Heyes 1996; Mitchell 1996 for further discussion of this issue, and Heyes 1994c; 1995b for analysis of other self-recognition studies that have not used anesthesia).

Taking an alternative tack, **Zentall** suggests that more mark touching when a mirror is present than when it is absent could not be an artefact of recovery from anesthetic because Suarez and Gallup (1981) found this effect using a delayed control test. In fact, Suarez and Gallup (1981) allowed a typical interval, of 4 to 5 hours, to elapse between marking and testing and used a relatively high dose of ketamine HCl (10 mg/kg). Therefore, as in other studies using the standard mark test, it is entirely possible that any increase in mark touching in the mirror-present condition

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

was due to an increase in frequency of spontaneous face touching with anesthetic recovery.

Gallup says that there is reason to doubt Dimond and Harries' (1984) evidence of species differences in frequency of spontaneous face touching, and yet the only comparable subsequent study, co-authored by Gallup himself (Gallup et al. 1995), provided data confirming that chimpanzees touch their faces more than monkeys (Heyes 1995b, pp. 1540–41).

Mitchell & Anderson claim that the results of variants of Gallup's mark test are consistent with the hypothesis that this test measures self-recognition because they involved monkeys and very young apes. It is true that the failure of monkeys on these tests is consistent with both the self-recognition and anesthetic artefact hypotheses, but according to a straightforward reading of the former, the young apes should have passed. A recent study purporting to demonstrate that young chimpanzees are less likely to be capable of self-recognition (Povinelli et al. 1993), shows upon reanalysis that, if anything, young chimpanzees are more likely than their elders to pass the mark test (Heyes 1995b).

Mitchell & Anderson also wonder why, when calculating descriptive statistics, I included chimpanzees that did, and chimpanzees that did not touch their marks more in the mirror-present condition. I did so to indicate that, on average, chimpanzees subjected to the mark test do not touch their marks much more when the mirror is present than when it is absent, and therefore that there is only a small behavioral effect to be explained by the anesthetic artefact hypothesis, or indeed any other hypothesis. However, I am grateful to Mitchell & Anderson for emphasizing through their summary of the same data that there may be no effect at all to be explained. It may be that, on average, chimpanzees touch their marks with the same or a lower frequency in the second, mirror-present condition than in the first, mirror-absent condition. There is certainly no justification for assigning animals to separate groups for the purpose of analysis purely on the basis of whether they did or did not touch their marks more when the mirror was present. This procedure begs the question of what, if anything, mark touching is measuring. It simply assumes that whether or not an animal makes more touches in the presence of a mirror is not due to chance but to whether or not the animal is capable of self-recognition.

Another of **Mitchell & Anderson**'s comments about self-recognition likewise reflects a basic misconception, this time relating to the distinction between necessary and sufficient conditions. Had I claimed that the ability to distinguish response feedback from other sensory input is a sufficient condition for mark test success, then I should indeed expect a very broad range of creatures to pass the mark test. In fact, I suggested that this is a necessary condition, and not an especially important one, but that its existence may have contributed to the spurious plausibility of the claim that mirror-guided body inspection implies self-conception (see Heyes 1994c, pp. 917–18; 1996; Mitchell 1996 for further discussion).

R2.4. Deception. First, a detail about deception: **Mitchell & Anderson** suggested that two, not four, of Woodruff and Premack's (1979) chimpanzees deceived the competitive trainer and that they did so after 190, not 120, trials. In fact, Figure 1 in Woodruff and Premack's report clearly confirms that, between trials 96 and 120 of the production test, there was a reliable difference between the choice accuracy of cooperative and competitive trainers for each of the four chimpanzees.

Defending serendipitous reports of deception, rather than experimental data, **Byrne** suggests that I picked an easy target when I chose to illustrate the ambiguity of this kind of evidence using Jolly's baboon anecdote. I thought he would follow this up by identifying a more appropriate example, one that cannot be explained in nonmentalistic terms, but instead he concedes that all such reports are ambiguous, that is, explicable in both mentalistic and nonmentalistic terms, and claimed that primate researchers know very well that they are ambiguous. If this is true, and I'll take his word for it, then why do some researchers continue to collect and publish these data and treat them as persuasive?

R3. Parsimony and convergence

Byrne's answer to this question takes us into the realm of parsimony and convergence arguments. He says that primate behavior that can be explained in both mentalistic and nonmentalistic terms provides "potential evidence of theory of mind." By the same token, all behavior that can be described in mentalistic terms provides potential evidence of theory of mind, including that of a broad range of vertebrate and invertebrate species (Bitterman 1988), and possibly of my old and temperamental car. So, if this is the reason, or even part of the reason, for the enduring currency of ambiguous data, why is research effort focused on primates? Judging from his commentary, Byrne would reply that, at least in the case of deception, there are more reports about primates that can be more simply explained in mentalistic than nonmentalistic terms because the observed animal used an apparently novel procedure to deceive a conspecific. But this reply is unsatisfactory for three reasons. First, it does not explain how we are to distinguish simple hypotheses from more complex ones, or why we should prefer the former (see sect. 3.1 of target article). Second, field observation is the weakest method for detecting behavioral novelty. Since individual animals are observed only intermittently, and there is no control over their opportunity to engage in various forms of behavior, we cannot infer that the first occasion on which a behavior was witnessed was also the first occasion on which it occurred. Finally, since the anecdotal method does not, by definition, involve any systematic sampling procedures, it cannot provide reliable information about the distribution of behavior across taxa. Variance in the distribution of *reports* of certain forms of behavior, whether it is between crustacea and primates or monkeys and apes, does not provide this information because it is uncontrollably and immeasurably biased by our preconceptions about animal mentality.

Byrne takes the main thrust of my argument to be that nonmentalistic explanations for primate behavior should be preferred because they are more parsimonious. In fact, I argue that all appeals to simplicity of explanation are a hindrance in this area of research and carefully resist the temptation to make one myself.

R4. Proposals

The remainder of this Response relates to comments on my proposals for future research on theory of mind in nonhu-

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

man primates. (I refer to them as "my" proposals for brevity, but as the target article indicated, even before this commentary process began, they owed a great deal to the insights of Campbell, Cheney and Seyfarth, Humphrey, Povinelli, Premack and others, and by the end they will be the product of still further distributed cognition.) Section R4 is divided into five subsections, with the first concerning general methodological issues. The second and third deal with objections that my proposals were too chauvinist and too liberal, respectively. Sections R4.1 to R4.3 include assessment of all of the commentators' recommendations for alteration and refinement of my proposals. The fourth section summarizes those revisions, suggested by the commentary process, that I think should be adopted in future research, and in the final section I reflect on the value and feasibility of large-scale collaboration in planning experimental research.

R4.1. General methodological issues

R4.1.1. Why only experiments? Several commentators took exception to my view that experiments are what are now needed to address Premack and Woodruff's question, and bore witness to their belief in the power of observational or descriptive methods. However, only one commentary (**Green et al.**) gave a specific reason, identifying what it is that may be revealed by these other methods and not by experiments.

Green et al. suggested that field observations may be the only way of recording rare events or complex interactions that are critical to understanding the behavior of primates, and I basically agree with them. Field observation is indispensable for establishing what animals do in their natural environments and therefore important in formulating hypotheses about what their mental capabilities might be. However, it is largely powerless to test hypotheses about what animals think, to determine which of a number of possible psychological mechanisms is generating behavior. Therefore, it must give way to field or laboratory experiments once the hypotheses have been formulated, and in the case of primate social cognition we already have plenty of provocative anecdotes so the transition is overdue.

R4.1.2. Why training experiments? Gómez values an experimental approach but believes that laboratory training prevents animals from using mentalistic processes. I agree that in learning environments where multiple discriminative cues are available, animals may sometimes use a salient cue that is not necessarily the most conducive to mental state attribution. For example, if both an individual's head orientation and their eye orientation could guide responding, an animal might use head orientation, the more readily perceptible cue, and consequently miss an opportunity to learn about sight, or, if they already have the concept of sight, fail to notice that the problem in hand can be solved using that concept. What I do not understand is why Gómez believes that this will always happen when the learning environment has been arranged by a researcher, that is, in a training experiment, and not when it has arisen spontaneously. Experience in which eye orientation is the most reliable of a range of discriminative cues is likely to be very helpful, and perhaps necessary, for learning the concept of sight, and conducive to the realization that sight is a key variable in a given situation, but I see no reason why a training regime should not provide such experience. Consequently, I take the implication of Gómez's objection to be that, in the experiments I propose, chimpanzees should be given a range of different types of training trials before the goggles trials are introduced, and that across these training trials, eye visibility and orientation should be confounded with a variety of additional cues while always remaining a reliable indicator of correct responding. Pretransfer training on more than one discrimination problem was recommended in the target article, but I am grateful to Gomez for underlining its importance.

I also admired the specificity of **Gómez**'s alternative research proposals, but I am not optimistic about the experimental design he described. Like others based on a "trapping" method (Heyes 1993), it hides but does not eliminate the animals' opportunity to solve the problem through associative learning. Gómez suggests that if an animal in his task (1) points to the keys' location when they have been moved in the Giver's absence, and (2) does not point to the keys' location when they have been moved in the Giver's presence, then we have evidence that the animal can attribute ignorance to a human. (He refers to other control trials, but does not give any details.) My principal concerns about this interpretation of probe trial performance arise from uncertainty about what happens, and what is supposed to happen, in the original (training) trials; those in which the keys are not moved in the interval between baiting and the arrival of the Giver. Does the animal point at the key container on these trials, as well as at the food container? If so, it has the opportunity to learn that pointing at the key location leads to reward, and this may be responsible for pointing in the first type of probe trial. The absence of pointing in the second type of probe trial would not rule out this possibility because the second type of probe involves a substantial change in the procedure and lack of responding could be due to generalization decrement. On the other hand, if the animal does not point at the keys' location in the original trials, then the predicted probe trial performance would not make sense under a mentalistic interpretation. The Giver is not in the room when the keys are moved in either the original trials or the first type of probe trials, so why should the animal assume that the Giver is ignorant of the key location in one case and not in the other?

Thus, I stand by my recommendation that future experiments on theory of mind in primates involve explicit training, and indeed I agree with **Purdy & Domjan** and **Zentall** that the negative transfer goggles experiment, described second in the target article, which involves training in both stages, is the more promising of the two. The reason for this was highlighted by **Pepperberg** and **Mitchell & Anderson:** there is a danger that when subjects are given probe trials that are not differentially reinforced, that is, when they are not trained at the transfer phase, only their performance on the first trial will be informative because on subsequent trials they may perseverate or alternate. Zentall made the further, valuable suggestion that zero transfer control groups be added to the negative transfer experiment. The animals in these groups (ZERO-DIRECT and ZERO-REVERSE) would not be given the opportunity to wear goggles during pretraining, and their performance on the goggles discrimination task would provide a further check on whether the animals in the other groups are really using rim color as a cue in this task (see also Purdy & Domjan), and potentially provide information about

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

whether the experimental animals learn that transparent goggles afford seeing, that opaque goggles do not afford seeing, or both.

R4.1.3. Why not more "naturalistic" experiments? I have often wondered what people mean when they talk about "naturalistic" experiments, and I am still puzzled after reading the commentary by **Matheson et al.** I do not understand how experimental tasks can be both naturalistic for each species tested and species-neutral, and I cannot see what is natural about chimpanzees watching videos. Nonetheless, I am glad that Matheson et al. offered specific experimental proposals, and I see many merits in the procedures they recommended, especially the first. This experiment illustrates the use of screens rather than goggles to manipulate sight, a possibility mentioned only in passing in the target article, and allows investigation of whether animals attribute sight to conspecifics rather than people (see also **Purdy & Domjan**). It is also a substantial asset that Matheson et al.'s procedure could be adapted to test nonprimate species and modalities other than vision. My principal concern about the procedure arises at the test stage and possibly from its naturalism. Simultaneous release of the Knower, Guesser, and subject on transfer trials creates a competitive feeding situation, and I am worried that the sequel would be an uninterpretable melee. Furthermore, even if it were possible to record reliably which animal approached which station first, the procedure would be likely to yield less, not more, data per subject than the experiments I proposed. Each subject could yield a maximum of one data point, and this would be lost in many cases, for example, when the Guesser got to the food before the Knower approached a container, when the Knower was known by the subject not to be a reliable food-sharer, or when the subject was so quick on his feet or so aggressive that he was able to search all the containers before the Knower and Guesser got a look in.

The second procedure proposed by **Matheson et al.** is a video version of the first goggles experiment I proposed, and in comparison it places additional task-irrelevant demands on the subjects. My impression is that its resemblance to the Sally/Anne task is superficial because subjects cannot be interrogated independently about where the food is located and where the Guesser believes that the food is located. Therefore, the procedure is at best a test for the concept of seeing, and could not provide evidence of the attribution of false belief (see R4.3.2).

Matheson et al.'s second procedure also differs from my first in giving subjects two additional types of probe trials, in which both trainers are wearing the same kind of goggles, opaque or translucent. The problem with this innovation is that mentalistic and nonmentalistic hypotheses do not make specific or distinct predictions about performance on these trials. During training, subjects will be taught to select the trainer who saw, or was present during, the (most recent) baiting. On probe trials of the second and third type, neither or both trainers will see/be present. Therefore, both hypotheses would be consistent with random performance and lack of responding on the part of the chimpanzees, but neither would predict a systematic response bias.

R4.1.4. Critical experiments and special evidence. Like **Purdy & Domjan, Gray & Russell, Green et al.,** and **Whiten,** I do not believe in critical experiments. I accept that theory is always underdetermined by data, *in principle* there is never only one hypotheses that will fit any given data set (Quine 1951; 1969), and I assume that scientific advances are made incrementally, through ramification-extinction of plausible rival hypotheses (Campbell 1986; 1997). Given these conventional postpositivist assumptions about science, it follows that I see the purpose of the experiments I proposed to be taking one step forward in answering Premack and Woodruff's question by trying to obtain data from chimpanzees that can be plausibly explained by the hypothesis that these animals attribute sight, and not with reference to any current theory of associative learning or other nonmentalistic processing. I do *not* imagine that, in themselves, these experiments could tell us about other nonhuman primates or attribution of other mental states, and I would be surprised if positive results did not provoke the formulation of additional nonmentalistic hypotheses that would need to be tested in further experiments on attribution of sight in chimpanzees. Viewed in this way, the advance would be a small one even if the experiments achieved their purpose. However, it would not be trivial because no previous study of theory of mind in primates has had even this degree of discriminative power.

It also follows from my postpositivist assumptions that I do not regard mentalistic hypotheses as requiring a special level of support or unique methods of investigation. Unlike **Zentall,** I take all evidence to be indirect in the sense that all observation is to some degree theory-laden, and therefore I would not distinguish enquiry about behavior and about mental states according to whether the phenomena are directly observable. More specifically, and in contrast with **Whiten,** I take triangulation to be a research strategy which is indispensable in a broad range of scientific fields (Wimsatt 1981), and certainly throughout comparative psychology, not as a method that is specially or uniquely suitable for detecting what Whiten describes as triangulation processes in primates.

R4.2. Too chauvinist. I was delighted to find that **Kamawar & Olson** have piloted with children the first of the two goggles experiments I proposed. I was also pleased to see that the results provided evidence of the basic validity of the procedure by showing that the children's performance was related to age, and that there was a modest correlation between performance on the goggles task and on existing measures of theory of mind. However, their most important finding was that many children who passed the other tests failed the goggles test. Although it may be harder for children to attribute mental states to stuffed toys than for chimpanzees to attribute them to live humans, this result strongly suggests that, in its current form, the goggles test is too chauvinist or conservative; it carries a high risk of promoting false negative conclusions. Therefore, the next task is to identify what might be making the test too chauvinist.

R4.2.1. Demands on memory. Several commentators (**Kamawar & Olson, Gray & Russell, Zentall**) made the useful suggestion that the goggles test may be too chauvinist because it requires subjects to learn an association between the rim color and the light-transmitting properties of the goggles, and to remember this pair of associates (e.g., red-transparent, blue-opaque) over an extended period between pretraining and testing. The experiment by Kamawar & Olson suggested one way of reducing this irrele-

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

vantly demanding feature of the task, by allowing subjects to reexperience both pairs of goggles shortly before each test trial, and **Purdy & Domjan** wisely proposed that the rim colors should be switched on some probe trials to check whether the associations have been learned and remembered, and that rim color is being used as the discriminative cue.

R4.2.2. My mind and other minds. It was also suggested that subjects may fail the goggles test if they have a theory of mind but do not apply it equally to themselves and to others. This objection took several forms: **Zentall** said that chimpanzees may not associate the goggles on themselves with the goggles on another; **Whiten** suggested that they may treat the mental state of another as an intervening variable rather than using "experience projection"; and **Csibra** pointed out that it is conceivable that chimpanzees apply theory of mind to others but not to themselves.

I found **Csibra**'s explication of this point admirably clear, but I cannot agree with him that it makes the test worryingly chauvinist. As he points out himself, humans apparently apply their theory of mind equally to themselves and to others, and according to simulation theory, first person knowledge is ontogenetically and operationally primary. Therefore, while it is conceivable that chimpanzees have a theory of other minds but do not attribute mental states to themselves, this is a possibility, in principle, rather than a plausible hypothesis, and given the perennial problem of underdetermination (see R4.1.2), an experiment is worth doing if it "only" discriminates between hypotheses for which there is already some evidential support. (I would offer the same response to **Green et al.** and **Mitchell & Anderson**, who suggest that chimpanzees with a theory of mind could fail the goggles test if they believe in ESP.)

Although I am not unduly worried about the possibility that chimpanzees have a theory of other minds exclusively, I would like to thank **Csibra** for spotting an absolute howler of a mistake in my negative transfer experiment and for explaining *why* it is a mistake. (**Green et al., Mitchell & Anderson,** and **Suddendorf** queried the same point, but were not very clear about why it is a problem.) My mistake was to propose that the trainer wearing the opaque goggles for Group Direct and the trainer wearing the translucent goggles for Group Reverse should always point to the empty container, rather than at random. This would make the negative transfer experiment chauvinist because animals with the capacity to attribute sight could get the impression that the trainer who is unable to see nonetheless has full information about the location of the bait, and therefore conclude that seeing is irrelevant to the task. Thus, contrary to my recommendation in the target article, in the second phase of the negative transfer experiment, the trainer wearing opaque goggles for Group Direct and the trainer wearing transparent goggles for Group Reverse should point to the baited cup at random, and therefore, on average, 50% of trials on which those trainers are chosen by the chimpanzees will end in reward.

Given this change in the contingencies, it would be advisable to use a simultaneous rather than a successive discrimination procedure in both conditional discrimination and transfer phases of the negative transfer experiment. If only one trainer is present on each trial (successive procedure), subjects may choose the container indicated by the Guesser simply because there is no better response strategy available (A. Dickinson, personal communication; E. Ray, personal communication).

Whiten and **Zentall** may have been suggesting the opposite asymmetry, that is, that chimpanzees attribute mental states to themselves but not to others. This is more plausible because if, as simulation theory claims, first person knowledge of mental states is a developmental precursor in humans of the application of theory of mind to others, then chimpanzees may resemble a human ancestor species that had evolved introspective access to mental states without the realization or belief that others also have these states. In this case, however, I am pessimistic about the possibility of useful empirical enquiry. I suspect that there is no way of testing whether a nonlinguistic creature has a theory of mind that it applies only to itself, and therefore do not regard it is a serious weakness in the experiments I propose that they are unable to evaluate this hypothesis.

R4.3. Too liberal

R4.3.1. Fear of the dark. In the target article (sect. 4.2) I mentioned that chimpanzees may show an aversion to opaque goggles during pretraining and suggested ways in which this might be overcome before administration of the probe trials. Several commentators (e.g., **Csibra, Green et al., Pepperberg, Zentall**) spelled out why it would be important to overcome any aversion of this kind. If the chimpanzees dislike the opaque goggles, and this aversion becomes associated with their color, then on transfer trials they may avoid looking at the trainer wearing goggles of that color, and choose the container indicated by the wearer of transparent goggles by default. Thus, if subjects showed fear of the opaque goggles at the end of pretraining, the experiment would be too liberal, in danger of leading to the conclusion that chimpanzees can attribute sight when, in fact, their performance on probe trials was due to aversion to the rim color associated with opaque goggles.

This possibility presents a significant practical challenge, but there is no reason to suppose that it is insurmountable. As I suggested in the target article, animals could be rewarded during pretraining for putting on the opaque goggles, or screens could be used instead of goggles, and **Matheson et al.** provide a clear illustration of the latter strategy. Another option would be to use broad hollow tubes, painted red or blue, with a transparent or opaque screen inside. In this case, the opaque screen could consist of a kaleidoscope that the chimpanzees might enjoy looking at, but could not, of course, see through.

Matheson et al. proposed that pretraining should involve the subjects looking through opaque and transparent goggles in search of food, but I would be reluctant to pursue this option. By heightening the animals' motivation to see through the goggles, it may exacerbate any aversion to the opaque. Similarly, the animals should not observe trainers wearing goggles during pretraining (contra **Csibra**) because it would give them an opportunity to learn that the wearer of, for example, red-rimmed goggles is unresponsive or provides unhelpful cues, and therefore to show a bias on transfer trials based on rim color alone, that is, without any appreciation of the significance of rim color with respect to sight.

R4.3.2. Looking, seeing, and mental representation. In the target article, I described the goggles experiments as studies of "perspective-taking." A study labelled in this way

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

could be investigating whether the subjects understand (1) *looking* – whether they are sensitive to gaze direction in social interaction, (2) *seeing* – that vision establishes some kind of cognitive or mental connection between the subject and the object of perception, (3) *seeing is believing* – that vision gives rise to mental representation of an object, and that mental representations can be true or false and can vary over time and among viewers without change in the object itself.

I hope it will clear up some confusion when I say that I was and am claiming that the goggles experiments could provide evidence that chimpanzees understand seeing, the second item in this rough and ready typography, and I am grateful to **Csibra, Kamawar & Olson,** and **Slaughter & Mealey** for pointing out that I did not make this explicit in the target article. As their comments indicate, the experiments I proposed would be too liberal if positive results were interpreted to mean that chimpanzees have a full-blown understanding of mental representation. Establishing whether chimpanzees understand seeing may seem like a modest goal, especially to developmentalists who have long had false belief tests at their disposal, and to anyone who believes that it was shown years ago that primates are capable of intentional deception, self-conception, and the like. However, as my literature review showed, any study of social cognition in primates yielding results that are not explicable in nonmentalistic terms, and that thereby suggest even a low-level understanding of mentality in primates, would be a significant step forward.

Some commentators (**Leavens, Matheson et al.,** and perhaps **Mitchell & Anderson**) suggested that the goggles experiments would be too liberal because they mistakenly believe, not only that I was claiming that they test for (3), seeing is believing, but also that the experiments would demonstrate no more than (1), looking. In fact, positive results in the goggles experiments would reveal more than sensitivity to gaze direction, and therefore more than has previously been demonstrated in any experiment on perspective-taking in chimpanzees (e.g., Leavens et al. 1996; Povinelli & Eddy 1996a; 1996b), because the subjects would not be able to see the trainers' eyes on transfer trials.

R4.3.3. Evidence of absence. In another, more general sense, the goggles experiments may be too liberal if positive outcomes would provide evidence of the attribution of sight, while negative results would be uninterpretable (**Matheson et al., Suddendorf, Whiten, Zentall**). To some degree, an asymmetry of this kind is inevitable, and therefore cannot be regarded as a weakness in the goggles experiments specifically. Null results are always more difficult to interpret than those in which the experimental manipulation has a reliable effect, because a null result could be due to test insensitivity. However, I think that the goggles experiments could provide some evidence of the absence of a capacity to attribute sight, especially if negative results were followed by positive results on an analogous task in which chimpanzees were required to match functional rather than mental relations.[1]

For example, it may be possible to adapt the procedure used by Premack and Premack (1983) to show that a chimpanzee, Sarah, is capable of analogical reasoning. Thus, if the functional relationship to be matched were "opening," the food might be hidden in a locked box on training trials, and the animals would be rewarded for selecting the trainer holding a large key. Pretraining would consist of allowing the chimpanzees to discover through their own efforts that only one of two instruments (e.g., a brush and a can opener) will open a novel container (e.g., a tin can). Then, on transfer trials, the novel container would be presented in place of the locked box, and the chimpanzees would be allowed to choose between trainers each holding one of the two instruments used in pretraining. Provided that physical similarity and contiguity between the correct pairs of stimuli in the training and transfer trials were properly controlled, transfer trial preference for the trainer holding the instrument that will open the novel container would suggest analogical reasoning about functional relationships by the chimpanzees. This, in turn, would indicate that the basic test procedure is one which can detect complex mental processing in chimpanzees, and therefore make it less plausible that failure on the goggles task was due to test insensitivity.

The foregoing proposal is no more than an outline. It would need a good deal more thought before implementation, and even if it were conducted properly and provided positive results, contrasting with those of the goggles experiments, of course we could not conclude once and for all on this basis alone that chimpanzees cannot attribute sight, much less that they do not have a theory of mind. In this respect, positive and negative evidence is the same: it accrues gradually and in small steps (see R4.1.2).

A similar point applies to inferences from goggles experiments with monkeys and children. For example, if there was a strong correlation between children's performance on a goggles test and on other, verbal measures of theory of mind, it would support the inference that chimpanzees' success on the goggles test is due to attribution of sight, but, as **Pepperberg** points out, it certainly would not guarantee the truth of this proposition.

R4.4. Summary of revisions. As a result of the commentary process, I would revise the proposals made in section 4 of the target article as follows.

1. Try the negative transfer goggles experiment first (**Purdy & Domjan, Zentall,** R4.1.2).
2. Use a simultaneous rather than a successive discrimination procedure in both conditional discrimination and transfer phases of the negative transfer experiment (A. Dickinson, personal communication; E. Ray, personal communication, R4.2.2).
3. Give pretransfer training on more than one discrimination problem that can be solved using the concept of sight (**Gómez,** R4.1.2).
4. The trainer wearing opaque goggles for Group Direct and the trainer wearing translucent goggles for Group Reverse should indicate a container at random on transfer trials in the negative transfer experiment (**Csibra,** R4.2.2).
5. Give subjects the opportunity to reexperience both pairs of goggles shortly before transfer trials (**Kamawar & Olson,** R4.2.1).
6. Switch rim colors on some transfer trials (**Purdy & Domjan,** R4.2.1).
7. If a sufficient number of chimpanzees are available, add zero transfer control groups to the negative transfer experiment (**Zentall,** R4.1.2).
8. If subjects show a bias against the opaque goggles during pretraining, try not only screens but also viewing tubes in the place of goggles (R4.3.3).

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

9. In the event of negative results, follow up with an analogous task assessing subjects' understanding of functional rather than mental relations (U. Frith, personal communication; F. Happe, personal communication, R4.3.3).

The commentary process has also convinced me that it is well worth continuing attempts to develop tests for primate and nonprimate species assessing the attribution of sight to conspecifics (**Matheson et al.**), and that neuroanatomical approaches could be profitably pursued in parallel with the development of behavioral tests (**Parker, Walker**).

R4.5. Ethnocentrism of disciplines. In addition to tackling substantive issues, a few commentators sought to undermine my arguments by implying that they had the same ideas first, or by pointing out that I am not a primate researcher. The first strategy reflects a desire for personal recognition that may be, not just venial, but a powerful engine of scientific progress (Hull 1988a; 1988b). The second, although it emerges from equally pervasive human tendencies, is potentially much more destructive. It exemplifies "ethnocentrism of disciplines" (Campbell 1969), a variety of tribalism or ingroup partisanship, that encourages narrow specialization in science, and discourages pooling of resources among researchers with overlapping but diverse skills and knowledge bases. Fortunately, the bulk of the commentaries illustrate the value of resisting such ethnocentrism and, in line with the principal purpose of *Behavioral and Brain Sciences*, fostering disputatious communication across specialist boundaries. Specifically, the commentators' experimental thought has brought us yet closer to being able to answer Premack and Woodruff's challenging, 20-year-old question: Does the chimpanzee have a theory of mind?

NOTE

1. I am grateful to Uta Frith and Francesca Happe for suggesting the use of a task analogous to the goggles experiments to aid interpretation of negative results.

References

Letters *a* and *r* appearing before authors' initials refer to target article and response, respectively.

Adams-Curtis, L. E. (1987) Social context of manipulative behaviour in *Cebus apella*. *American Journal of Primatology* 12:325. [aCMH]

Akins, C. & Zentall, T. R. (1996) Imitative learning in male Japanese quail (*Coturnix japonica*) involving the two-action method. *Journal of Comparative Psychology* 110:316–20. [TRZ]

Anderson, J. R. (1983) Responses to mirror image stimulation and assessment of self-recognition in mirror- and peer-reared stumptail macaques. *Quarterly Journal of Experimental Psychology* 35(b):201–12. [aCMH]

Anderson, J. R. & Gallup, G. G., Jr. (in press) Self-recognition in nonhuman primates: Past and future challenges. In: *Brain, behavior, and cognition: Animal models and human studies*, ed. M. Haug & R. E. Whalen. American Psychological Association. [RWM]

Anderson, J. R. & Roeder, J. J. (1989) Responses of capuchin monkeys to different conditions of mirror-image stimulation. *Primates* 30:581–87. [aCMH]

Asendorpf, J. B. & Baudonniere, P. M. (1993) Self-awareness and other-awareness: I. Mirror self-recognition and synchronic imitation among unfamiliar peers. *Developmental Psychology* 29:88–95. [SFW, TRZ]

Asendorpf, J. B., Warkentin, V. & Baudonniere, P. M. (1996) Self-awareness and other-awareness: II. Mirror self-recognition, social contingency awareness, and synchronic imitation. *Developmental Psychology* 32:313–21. [TS, SFW]

Azmitia, M. & Hesser, J. (1993) Why siblings are important agents of cognitive-development: A comparison of siblings and peers. *Child Development* 64(2):430–44. [SFW]

Bakeman, R. & Gottman, J. (1986) *Observing interaction: An introduction to sequential analysis*. Cambridge University Press. [KAB]

Bard, K. A. (1997) New data and issues about self-recognition in chimpanzees. *American Journal of Primatology* 42:93 (abstract). [KAB]

Bard, K. A. & Russell, C. L. (in press) Evolutionary foundations of imitation: Social, cognitive, and developmental aspects of imitative processes in nonhuman primates. In: *Imitation in infancy: Progress and prospects of current research*, ed. J. Nadel & G. Butterworth. Cambridge University Press. [KAB]

Baron-Cohen, S. (1991) Precursors to a theory of mind: Understanding attention in others. In: *Natural theories of mind*, ed. A. Whiten. Basil Blackwell. [VS]

Baron-Cohen, S., ed. (1995) *Mindblindness: An essay on autism and theory of mind*. MIT Press. [rCMH, AP]

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985) Does the autistic child have a "theory of mind"? *Cognition* 21:37–46. [SMG]

Barresi, J. & Moore, C. (1996) Intentional relations and social understanding. *Behavioral and Brain Sciences* 19:107–22. [GC, VS]

Barton, R. A. (1996) Neocortex size and behavioral ecology in primates. *Proceedings of the Royal Society of London Series B--Biological Sciences* 263:173–77. [SFW]

Beck, B. B. (1976) Tool use by captive pigtailed monkeys. *Primates* 17:301–10. [aCMH]

Bennett, J. (1978) Some remarks about concepts. *Behavioral and Brain Sciences* 1(4):57–60. [GC]

Bernstein, I. S. (1988) Metaphor, cognitive belief, and science. *Behavioral and Brain Sciences* 11:247–48. [aCMH]

Bischof-Koehler, D. (1989) *Spiegelbild und Empathie* [Mirror image and empathy]. Hans Huber Verlag. [TS]

Bitterman, M. E. (1988) Creative deception. *Science* 239:1360. [rCMH]

Bjorklund, D. (1995) *Children's thinking*. Brookes/Cole. [HLM]

Blaschke, M. & Ettlinger, G. (1987) Pointing by monkeys. *Animal Behaviour* 35:1520–23. [RWM]

Block, N. (1978) Troubles with functionalism. In: *Perception and cognition: Issues in the foundations of psychology. Minnesota Studies in the Philosophy of Science, vol. 9*, ed. C. W. Savage. University of Minnesota Press. [rCMH]

Boyd, R. (1985) Observations, explanatory power, and simplicity: Toward a non-Humean account. In: *The philosophy of science*, ed. R. Boyd, P. Gasper & J. D. Trout. MIT Press. [aCMH]

Brooks-Gunn, J. & Lewis, M. (1984) The development of early visual self-recognition. *Developmental Review* 4:215–39. [KAB]

Budwig, N. & Bamberg, M. (1996) Language and its role in understanding intentional relations: Research tool or mechanism of development? *Behavioral and Brain Sciences* 19:125–26. [VS]

Byrne, R. W. (1994) The evolution of intelligence. In: *Behaviour and evolution*, ed. P. J. B. Slater & T. R. Halliday. Cambridge University Press. [aCMH, DC]

Byrne, R. W. & Tomasello, M. (1995) Do rats ape? *Animal Behaviour* 50:1417–20. [DC, aCMH]

Byrne, R. W. & Whiten, A. (1985) Tactical deception of familiar individuals in baboons (*Papio ursinus*). *Animal Behaviour* 33:669–73. [RWB]

(1987) The thinking primate's guide to deception. *New Scientist* 116(1589):54–57. [RWB]

(1988) Towards the next generation in data quality: A new survey of primate tactical deception. *Behavioral and Brain Sciences* 11:267–73. [RWB]

(1990) Tactical deception in primates: The 1990 data-base. *Primate Report* 27:1–101. [RWB]

(1991) Computation and mindreading in primate tactical deception. In: *Natural theories of mind*, ed. A. Whiten. Basil Blackwell. [RWB]

(1992) Cognitive evolution in primates: Evidence from tactical deception. *Man* 27:609–27. [RWB]

Byrne, R. W. & Whiten, A., eds. (1988) *Machiavellian intelligence*. Oxford University Press. [aCMH]

Calhoun, S. & Thompson, R. L. (1988) Long-term retention of self-recognition by chimpanzees. *American Journal of Primatology* 15:361–65. [aCMH]

Campbell, D. T. (1953) Operational delineation of "what is learned" via the transposition experiment. *Psychological Review* 61:167–74. [aCMH]

(1969) Ethnocentrism of disciplines and the fishscale model of omniscience. In: *Interdisciplinary relationships in the social sciences*, ed. M. Sherif & C. W. Sherif. Aldine. [aCMH]

(1986) Science's social system of validity-enhancing collective belief change and the problems of the social sciences. In: *Metatheory in social science: Pluralisms and subjectivities*, ed. D. W. Fiske & R. A. Shweder. University of Chicago Press. [rCMH]

(1997) From evolutionary epistemology via selection theory to a sociology of scientific validity. *Evolution and Cognition* 3:5–38. [rCMH]

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Carruthers, P. & Smith, P., eds. (1996) *Theories of theories of mind*. Cambridge University Press. [RMG, AW]
Cheney, D. L. & Seyfarth, R. M. (1980) Vocal recognition in free-ranging vervet monkeys. *Animal Behaviour* 288:362–67. [aCMH]
(1991) Truth and deception in animal communication. In: *Cognitive ethology: The minds of other animals*, ed. C. A. Ristau. Erlbaum. [aCMH]
(1992) Précis of *How monkeys see the world*. *Behavioral and Brain Sciences* 15:135–82. [aCMH, VS]
Cheney, D. L. & Seyfarth, R. M., eds. (1990a) *How monkeys see the world*. Chicago University Press. [GGG, aCMH]
(1990b) Attending to behaviour versus attending to knowledge: Examining monkeys' attribution of mental states. *Animal Behaviour* 40:742–53. [aCMH]
Cheney, D. L., Seyfarth, R. M. & Smuts, B. (1986) Social relationships and social cognition in nonhuman primates. *Science* 234:1361–66. [aCMH]
Chudler, E. H., Sugiyama, K. & Dong, W. K. (1995) Multisensory convergence and integration in the neostriatum and globus-pallidus of the rat. *Brain Research* 674:33–45. [SFW]
Custance, D. (1994) Social learning and imitation in human and non-human primates. Ph. D. thesis. University of St. Andrews. [DC]
Custance, D. & Bard, K. A. (1994) The comparative and developmental study of self-recognition and imitation: The importance of social factors. In: *Self-awareness in animals and humans*, ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [KAB]
Custance, D. M., Whiten, A. & Bard, K. A. (1995) Can young chimpanzees imitate arbitrary actions? Hayes and Hayes (1952) revisited. *Behaviour* 132:837–59. [KAB, DC, aCMH, HLM, TRZ]
Dasser, V. (1988) A social concept in Java monkeys. *Animal Behaviour* 36:225–30. [aCMH]
Davies, M. & Stone, T., eds. (1995a) *Folk psychology: The theory of mind debate*. Blackwell. [RMG]
(1995b) *Mental simulation: Evaluations and applications*. Blackwell. [RMG]
Davis, L.H. (1978) Intentions, awareness, and awareness thereof. *Behavioral and Brain Sciences* 1(4):566–67. [GC]
Dawson, G. & McKissock, F. (1984) Self-recognition in autistic children. *Journal of Autism and Developmental Disorders* 14:383–94. [rCMH]
Dennett, D. C. (1978a) Beliefs about beliefs. *Behavioral and Brain Sciences* 1(4):568–69. [GC, VS]
(1978b) *Brainstorms*. Bradford Books. [DK]
(1980) The milk of human intentionality (commentary on Searle). *Behavioral and Brain Sciences* 3:428–30. [aCMH]
(1983) Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *Behavioral and Brain Sciences* 6:343–90. [aCMH]
(1987) *The intentional stance*. MIT Press. [WMB]
(1989) Cognitive ethology: Hunting for bargains or a wild goose chase. In: *Goals, no-goals and own goals: A debate on goal-directed and intentional behaviour*, ed. A. C. Montefiore & D. Noble. Unwin Hyman. [aCMH]
(1996) *Kinds of minds*. Basic Books. [AW]
Dickinson, A. (1980) *Contemporary animal learning theory.* Cambridge University Press. [aCMH]
Dimond, S. & Harries, R. (1984) Face touching in monkeys, apes and man: Evolutionary origins and cerebral asymmetry. *Neuropsychologia* 22:227–33. [aCMH, TRZ]
Eddy, T. J., Gallup, G. G., Jr. & Povinelli, D. J. (1996) Age differences in the ability of chimpanzees to distinguish between mirror-images of the self from video images of others. *Journal of Comparative Psychology* 110:38–44. [GGG]
Eglash, A. R. & Snowdon, C. T. (1983) Mirror-image responses in pygmy marmosets. *American Journal of Primatology* 5:211–19. [aCMH]
Epstein, R., Lanza, R. P. & Skinner, B. F. (1981) "Self-awareness" in the pigeon. *Science* 212:695–96. [aCMH, RWM]
Fadiga, L., Fogassi, L., Pavesi, G. & Rizzolatti, G. (1995) Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology* 73:2608–11. [SFW]
Ferrari, M. & Mathews, W. (1983) Self-recognition deficits in autism: Syndrome-specific or general developmental delay? *Journal of Autism and Developmental Disorders* 13:317–24. [rCMH]
Flanagan, O. (1992) *Consciousness reconsidered*. MIT Press. [HLM]
Flavell, J. (1988) The development of children's knowledge about the mind: From cognitive connections to mental representations. In: *Developing theories of mind*, ed. J. Astington, P. Harris & D. Olson. Cambridge University Press. [VS]
Flavell, J., Everett, B., Croft, K. & Flavell, E. (1981) Young children's knowledge about visual perception: Further evidence for the Level 1-Level 2 distinction. *Developmental Psychology* 17:99–103. [VS]
Fletcher, P. C., Hae, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J. & Frith, C. D. (1995) Other minds in the brain: A functional imaging study of theory of mind in story comprehension. *Cognition* 57:109–28. [SFW]
Fodor, J. A. (1992) A theory of the child's theory of mind. *Cognition* 44:283–96. [rCMH, VS]
Fragaszy, D. M. & Visalberghi, E. (1989) Social influences on the acquisition and use of tools in tufted capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology* 103:159–70. [aCMH]
Fragaszy, D. M. & Visalberghi, R. (1990) Social processes affecting the appearance of innovative behaviors in capuchin monkeys. *Folia Primatologica* 3–4:54. [aCMH]
Frye, D. & Moore, C. (1991) The acquisition and utility of theories of mind. In: *Children's theories of mind*, ed. D. Frye & C. Moore. Lawrence Erlbaum. [HLM]
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. T. R. Zentall & B. G. Galef, Jr. Erlbaum. [aCMH]
(1992) The question of animal culture. *Human Nature* 3:157–78. [aCMH]
Galef, B. G., Manzig, L. A. & Field, R. M. (1986) Imitation learning in budgerigars: Dawson and Foss (1965) revisited. *Behavioral Processes* 13:191–202. [DC, aCMH, TRZ]
Gallese, V., Fadiga, L., Fogassi, L. & Rizzolatti, G. (1996) Action recognition in the premotor cortex. *Brain* 119:593–609. [SFW]
Gallup, G. G., Jr. (1970) Chimpanzees: Self-recognition. *Science* 167:86–87. [SMG, aCMH, TRZ]
(1977) Self-recognition in primates. *American Psychologist* 32:329–38. [aCMH]
(1982) Self-awareness and the emergence of mind in primates. *American Journal of Primatology* 2:237–48. [GGG, aCMH, RWM]
(1983) Toward a comparative psychology of mind. In: *Animal cognition and behavior*, ed. R. E. Mellgren. Elsevier. [aCMH]
(1985) Do minds exist in species other than our own? *Neuroscience and Biobehavioral Reviews* 9:631–41. [aCMH]
(1988) Toward a taxonomy of mind in primates. Commentary on Whiten & Byrne. *Behavioral and Brain Sciences* 11:255–56. [aCMH]
(1994) Self-recognition: Research strategies and experimental design. In: *Self-awareness in animals and humans*, ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [GGG]
(1996) Rhesus monkeys are radical behaviorists. *Behavioral and Brain Sciences* 19:129. [VS]
Gallup, G. G., McClure, M. K., Hill, S. D. & Bundy, R. A. (1971) Capacity for self-recognition in differentially reared chimpanzees. *Psychological Record* 21:69–74. [aCMH]
Gallup, G. G., Povinelli, D. J., Suarez, S. D., Anderson, J. R., Lethmate, J. & Menzel, E. W. (1995) Further reflections on self-recognition in primates. *Animal Behaviour* 50:1525–32. [GGG, aCMH, RWM, IMP]
Gallup, G. G. & Suarez, S. D. (1991) Social responding to mirrors in rhesus monkeys: Effects of temporary mirror removal. *Journal of Comparative Psychology* 105:376–79. [aCMH]
Goldman, A. (1993) The psychology of folk psychology. *Behavioral and Brain Sciences* 16:15–28. [GC, aCMH]
Gómez, J. C. (1996) Nonhuman primate theories of (non-human primate) minds. In: *Theories of theories of mind*, ed. P. Carruthers & P. K. Smith. Cambridge University Press. [JCG]
Gómez, J. C. & Teixidor, P. (in preparation) Theory of mind in an orangutan: A "key" experiment for the non-verbal study of mindreading. [JCG]
Goodall, J., ed. (1986) *The chimpanzees of Gombe: Patterns of behavior*. Belknap Press. [aCMH, HLM]
Gopnik, A. (1993) How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16:1–14. [aCMH, VS]
Gopnik, A. & Astington, J. W. (1988) Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development* 59:26–37. [TS]
Gopnik, A. & Wellman, H. M. (1992) Why the child's theory of mind really is a theory. *Mind and Language* 7:145–71. Reprinted in Davies & Stone 1995a. [RMG]
(1994) The theory theory. In: *Mapping the mind: Domain specificity in cognition and culture*, eds. L. A. Hirschfeld & S. A. Gelman. Cambridge University Press. [aCMH]
Grafton, S. T., Arbib, M. A., Fadiga, L. & Rizzolatti, G. (1996) Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Experimental Brain Research* 112:103–11. [SFW]
Graziano, M. S. A., Yap, G. S. & Gross, C. G. (1994) Coding of visual space by premotor neurons. *Science* 266:1054–57. [SFW]
Guillaume, P. (1926/1971) *Imitation in children*, 2nd. ed. University of Chicago Press. [RWM]
Harman, G. (1978) Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences* 1(4):576–77. [GC]

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Harris, P. (1992) From simulation to folk psychology: The case for development. *Mind and Language* 7:120–44. Reprinted in Davies & Stone 1995a. [RMG]

Hauser, M. D. (1988) Invention and social transmission: New data from wild vervet monkeys. In: *Machiavellian intelligence*, ed. R. Byrne & A. Whiten. Clarendon Press. [aCMH]

Hauser, M. D., Kralik, J., Botto-Mahan, C., Garrett, M. & Oser, J. (1995) Self-recognition in primates: Phylogeny and the salience of species-typical features. *Proceedings of the National Academy of Sciences (USA)* 92:10811–14. [rCMH]

Hayes, K. J. & Hayes, C. (1952) Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology* 45:450–59. [aCMH, HLM, TRZ]

Hess, J., Novak, M. A. & Povinelli, D. J. (1993) 'Natural pointing' in a rhesus monkey, but no evidence of empathy. *Animal Behaviour* 46:1023–25. [aCMH]

Heyes, C. M. (1987) Contrasting approaches to the legitimation of intentional language within comparative psychology. *Behaviorism* 15:41–50. [aCMH]

(1993) Anecdotes, training, trapping, and triangulating: Do animals attribute mental states? *Animal Behaviour*, 46:177–88. [RWB, aCMH]

(1994a) Imitation, culture and cognition. *Animal Behaviour* 46:999–1010. [aCMH]

(1994b) Social cognition in primates. In: *Animal learning and cognition*, ed. N. J. Mackintosh. Academic Press. [aCMH]

(1994c) Reflections on self-recognition in primates. *Animal Behaviour* 47:909–19. [KAB, aCMH]

(1994d) Cues, convergence and a curmudgeon: A reply to Povinelli. *Animal Behaviour* 48:242–44. [aCMH]

(1995a) Imitation and flattery: A reply to Byrne and Tomasello. *Animal Behaviour* 50:1421–24. [aCMH]

(1995b) Self-recognition in primates: Further reflections create a hall of mirrors. *Animal Behaviour* 50:1533–42. [aCMH]

(1995c) Self-recognition in primates: Irreverence, irrelevance and irony. *Animal Behaviour* 51:470–73. [aCMH]

(1996) Genuine imitation. In: *Social learning in animals: The roots of culture*, ed. C. M. Heyes & B. G. Galef. Academic Press. [aCMH]

Heyes, C. M. & Dawson, G. R. (1990) A demonstration of observational learning using a bidirectional control. *Quarterly Journal of Experimental Psychology* 42B:59–71. [aCMH, TRZ]

Heyes, C. M., Dawson, G. R. & Nokes, T. (1992) Imitation in rats: Initial responding and transfer evidence *Quarterly Journal of Experimental Psychology* 45B:81–92. [DC, aCMH]

Hietanen, J. K. & Perrett, D. I. (1996) A comparison of visual responses to object-motion and egomotion in the macaque superior temporal polysensory area. *Experimental Brain Research* 108:341–45. [SFW]

Hull, D. (1988a) *Science as a process: An evolutionary account of the social and conceptual development of science.* University of Chicago Press. [rCMH]

(1988b) A mechanism and its metaphysics: An evolutionary account of the social and conceptual development of science. *Biology and Philosophy* 3:123–55. [rCMH]

Hume, D. (1948/1748) *An inquiry concerning human understanding*. Bobbs Merrill. Originally published 1748. [aCMH]

Humphrey, N. K. (1976) The social function of intellect. In: *Growing points in ethology*, ed. P. P. G. Bateson & R. A. Hinde. Cambridge University Press. [SFW]

Hyatt, C. W. & Hopkins, W. D. (1994) Self-awareness in bonobos and chimpanzees: A comparative perspective. In: *Self-awareness in animals and humans,* ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [KAB]

Itakura, S. (1994) Symbolic representation of possession in a chimpanzee. In: *Self-awareness in animals and humans: Developmental perspectives*, ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [HLM]

Jeannerod, M., Arbib, M. A., Rizzolatti, G. & Sakata, H. (1995) Grasping objects: The cortical mechanisms of visuomotor transformations. *Trends in Neurosciences* 18:314–20. [SFW]

Johnson, P. B., Ferraina, S., Bianchi, L. & Caminiti, R. (1996) Cortical networks for visual reaching: Physiological and anatomical organization of frontal and parietal lobe arm regions. *Cerebral Cortex* 6:102–19. [SFW]

Jolly, A. (1985) *The evolution of primate behaviour*. Macmillan. [RWB, aCMH]

(1991) Conscious chimpanzees? A review of recent literature. In: *Cognitive ethology: The minds of other animals*, ed. C. R. Ristau. Erlbaum. [GGG, aCMH]

Karmiloff-Smith, A., Klima, E., Bellugi, U., Grant, J. & Baron-Cohen, S. (1995) Is there a social module: Language, face processing, and theory of mind – in individuals with Williams syndrome. *Journal of Cognitive Neuroscience* 7:196–208. [SFW]

Kellog, W. N. (1968) Communication and language in the home-raised chimpanzee. *Science* 162:423–27. [JEP]

Kertzman, C., Schwarz, U., Zeffiro, T. A. & Hallett, M. (1997) The role of posterior parietal cortex in visually guided reaching movements in humans. *Experimental Brain Research* 114:170–83. [SFW]

Kitchen, A., Denton, D. & Brent, L. (1996) Self-recognition and abstraction abilities in the common chimpanzee studied with distorting mirrors. *Proceedings of the National Academy of Sciences* 93:7405–08. [rCMH]

Kîhler, W. (1927) *The mentality of apes* (trans. E. Winter). Routledge & Kegan Paul. (Original work published 1917). [TS]

Krebs, J. R. & Dawkins, R. (1984) Animal signals: Mind reading and manipulation. In: *Behavioural ecology*, ed. J. R. Krebs & N. B. Davies. Blackwell Scientific Publications. [aCMH]

Kugiumutzakis, G. (in press) Development of early infant imitation to facial and vocal models. In: *Imitation in infancy*, ed. J. Nadel & G. Butterworth. Cambridge University Press. [KAB]

Kummer, H., Dasser, V. & Hoyningen-Huene, P. (1990) Exploring primate social cognition: Some critical remarks. *Behaviour* 112:84–98. [aCMH]

Leavens, D. A. & Hopkins, W. D. (in press) Intentional communication by chimpanzees: A cross-sectional study of the use of referential gestures. *Developmental Psychology.* [DAL]

Leavens, D. A., Hopkins, W. D. & Bard, K. A. (1996) Indexical and referential pointing in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology* 110:346–53. [rCMH, DAL]

Ledbetter, D. H. & Basen, J. A. (1982) Failure to demonstrate self-recognition in gorillas. *American Journal of Primatology* 2:307–10. [aCMH]

Leslie, A. (1991) The theory of mind impairment in autism: Evidence for a modular mechanism of development. In: *Natural theories of mind*, ed. A. Whiten & R. Byrne. Blackwell. [aCMH]

Lin, A. C., Bard, K. A. & Anderson, J. R. (1992) Development of self-recognition in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology* 106:120–27. [KAB, aCMH]

Loveland, K. A., Tunalikotoski, B., Pearson, D. A., Brelsford, K. A., Ortegon, J. & Chen, R. (1994) Imitation and expression of facial affect in autism. *Development and Psychopathology* 6:433–44. [SFW]

Mason, W. A. & Hollis, J. H. (1962) Communication between young rhesus monkeys. *Animal Behaviour* 10:211–21. [aCMH]

Mealey, L. (1992) Are monkeys nomothetic or idiographic? *Behavioral and Brain Sciences* 15:162. [VS]

Meltzhoff, A. (1996) The human infant as imitative generalist. In: *Social learning in animals: The roots of culture*, ed. B. G. Galef & C. M. Heyes. Academic Press. [KAB]

Meltzoff, A. & Moore, M. (1983) Newborn infants imitate adult facial gestures. *Child Development* 54:702–09. [aCMH]

Menzel, E. W., Jr. (1974) A group of young chimpanzees in a one-acre field. In: *Behavior of nonhuman primates, vol. 5*, ed. A. Schrier & F. Stollnitz. Academic Press. [MDM]

Mignault, C. (1985) Transition between sensorimotor and symbolic activities in nursery-reared chimpanzees (*Pan troglodytes*). *Journal of Human Evolution* 14:747–58. [aCMH]

Miles, H. L. (1994) Me Chantek: The development of self-awareness in a signing orangutan. In: *Self-awareness in animals and humans*, ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [KAB, rCMH. HLM]

(1997) Anthropomorphism, apes and language. In: *Anthropomorphism, anecdotes, and animals*, ed. R. W. Mitchell, N. S. Thompson & H. L. Miles. State University of New York Press. [HLM]

Miles, H. L., Mitchell, R. W. & Harper, S. (1996) Simon says: The development of imitation in an enculturated orangutan. In: *Reaching into thought: The minds of great apes,* ed. A. E. Russon, K. A. Bard & S. T. Parker. Cambridge University Press. [rCMH, HLM, RWM]

Miller, N. E. & Dollard, J. (1941) *Social learning and imitation*. Yale University Press. [aCMH]

Mitchell, R. W. (1986) A framework for discussing deception. In: *Deception: Perspectives on human and nonhuman deceit*, ed. R. W. Mitchell & N. S. Thompson. State University of New York Press. [RWB, RWM]

(1993) Mental models of mirror-self-recognition: Two theories. *New Ideas in Psychology* 11:295–325. [KAB, RWM]

(1996) Self-recognition, methodology and explanation: A reply to Heyes. *Animal Behaviour* 51:467–69. [rCMH, RWM]

(1997) A comparison of the self-awareness and kinesthetic-visual matching theories of self-recognition: Autistic children and others. *New York Academy of Sciences* 818:39–62. [RWM]

Mitchell, R. W. & Anderson, J. R. (1993) Discrimination learning of scratching, but failure to obtain imitation and self-recognition in a long-tailed macaque. *Primates* 34:301–09. [RWM, TRZ]

(1997) Communicative and deceptive pointing in cebus monkeys (*Cebus apella*). *Journal of Comparative Psychology* 111(4):351–61. [RWM]

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

Mitchell, R. W., Thompson, N. S. & Miles, H. L., ed. (1997) *Anthropomorphism, anecdotes, and animals*. State University of New York Press. [HLM]
Menzel, E. W., Jr. (1974) A group of young chimpanzees in a one-acre field. In: *Behavior of non-human primates, vol. 5*, ed. A. M. Schrier & F. Stolnitz. Academic Press. [HLM]
Moore, C. & Dunham, P., eds. (1995) *Joint attention: Its origins and role in development*. Erlbaum. [VS]
Morgan, C. L. (1894) *An introduction to comparative psychology*. Walter Scott. [RWM]
Morton, A. (1980) *Frames of mind*. Oxford University Press. [rCMH]
Murata, A., Gallese, V., Kaseda, M. & Sakata, H. (1996) Parietal neurons related to memory-guided hand manipulation. *Journal of Neurophysiology* 75:2180–86. [SFW]
Nelson, K. (1993) The psychological and social origins of autobiographical memory. *Psychological Science* 4:7–14. [AP]
(1996) Four-year-old humans are different: Why? *Behavioral and Brain Sciences* 19:134–35. [DAL, VS]
Nishida, T. (1986) Local traditions and cultural transmission. In: *Primate societies*, ed. B. B. Smuts, D. L. Cheney, R. M. Seyfarth, R. W. Wrangham & T. T. Struhsaker. University of Chicago Press. [aCMH]
Novey, M. S. (1975) The development of knowledge of others' ability to see. Unpublished doctoral dissertation, Harvard University. [aCMH]
Oram, M. W. & Perrett, D. I. (1994) Responses of anterior superior temporal polysensory (STPa) neurons to biological motion stimuli. *Journal of Cognitive Neuroscience* 6:99–116. [SFW]
Parker, A. & Gaffan, D. (1997a) Mamillary body lesions in monkeys impair object-in-place memory: Functional unity of the formix-mamillary system. *Journal of Cognitive Neuroscience* 9:512–21. [AP]
(1997b) The effect of anterior thalamic and cingulate cortex lesions on object-in-place memory. *Neuropsychologia* 35:1093–1102. [AP]
(1997c) Memory systems in primates: Episodic, semantic, and perceptual learning. In: *Comparative neuropsychology*, ed. A. D. Milner. Oxford University Press. [AP]
(1997d) Frontal/temporal disconnection in monkeys: Memory for strategies and memory for visual objects. *Society for Neuroscience Abstracts* 23:11. [AP]
Parker, S. T., Mitchell, R. W. & Boccia, M. L., eds. (1994) *Self-awareness in animals and humans*. Cambridge University Press. [RWM]
Passingham, R. (1997, in press) The specializations of the human neocortex. In: *Comparative neuropsychology*, ed. A. D. Milner. Oxford University Press. [AP]
Patterson, F. G. P. & Cohn, R. H. (1994) Self-recognition and self-awareness in lowland gorillas. In: *Self-awareness in animals and humans*, ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [KAB, HLM]
Pepperberg, I. M. (1990) Some cognitive capacities of an African Grey parrot (*Psittacus erithacus*). In: *Advances in the study of behavior, vol. 19*, ed. P. J. B. Slater, J. S. Rosenblatt & C. Beer. Academic Press. [IMP]
(1994) Evidence for numerical competence in an African Grey parrot (*Psittacus erithacus*). *Journal of Comparative Psychology* 108:36–44. [IMP]
(1996) Categorical class formation by an African Grey parrot (*Psittacus erithacus*). In: *Stimulus class formation in humans and animals*, ed. T. R. Zentall and P. R. Smeets. Elsevier. [IMP]
Perner, J. (1991) *Understanding the representational mind*. Bradford Books, MIT Press. [aCMH, TS, VS]
(1996) Simulation as explicitation of prediction-implicit knowledge about the mind: Arguments for a simulation-theory mix. In: *Theories of theories of mind*, ed. P. Carruthers & P. Smith. Cambridge University Press. [RMG]
Perner, J., Leekam, S. R. & Wimmer, H. (1987) Three-year-olds' difficulty with false belief. *British Journal of Developmental Psychology* 5:125–37. [DAL]
Perner, J. & Ruffman, T. (1995) Episodic memory and autogenetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology* 59:516–48. [TS]
Peskin, J. (1996) Guise and guile. *Child Development* 4:1735–51. [DK]
Peterson, C. C. & Siegal, M. (1995) Deafness, conversation and theory of mind. *Journal of Child Psychology and Psychiatry* 36:459–74. [CG, VS]
Petrides, M. & Pandya, D. N. (1994) Comparative architectonic analysis of the human and the macaque frontal cortex. In: *Handbook of neuropsychology, vol. 9*, ed. F. Boller & J. Grafman. Elsevier. [AP]
Piaget, J. (1962) *Play, dreams and imitation in childhood*. Norton. [aCMH, HLM, TRZ]
Platt, M. M. & Thompson, R. L. (1985) Mirror responses in Japanese macaque troop. *Primates* 26:300–14. [aCMH]
Povinelli, D. J. (1987) Monkeys, apes, mirrors and minds: The evolution of self-awareness in primates. *Human Evolution* 2:493–509. [aCMH, RWM]
(1989) Failure to find self-recognition in Asian elephants (*Elephas maximus*) in contrast to their use of mirror cues to discover hidden food. *Journal of Comparative Psychology* 103:122–31. [SFW]
(1993) Reconstructing the evolution of mind. *American Psychologist* 48:493–509. [aCMH, SFW]
(1994) Comparative studies of animal mental state attribution: A reply to Heyes. *Animal Behaviour* 48: 239–41. [aCMH]
(1995) Panmorphism. In: *Anthropomorphism, anecdotes and animals*, ed. R. Mitchell & N. Thompson. University of Nebraska Press. [aCMH, GGG]
Povinelli, D. J. & deBlois, S. (1992) Young children's understanding of knowledge formation in themselves and others. *Journal of Comparative Psychology* 106:228–38. [aCMH]
Povinelli, D. J. & Eddy, T. J. (1996a) What young chimpanzees know about seeing. *Monographs of the Society for Research on Child Development*, Vol. 61, No. 247. [aCMH, DAL, VS]
(1996b) Factors influencing young chimpanzees recognition of attention. *Journal of Comparative Psychology* 110:336–45. [rCMH]
Povinelli, D. J., Gallup, G. G., Jr., Eddy, T. J., Bierschwale, D. T., Engstrom, M. C., Perilloux, H. K. & Taxopeus, I. B. (1997) Chimpanzees recognize themselves in mirrors. *Animal Behaviour* 53:1083–88. [GGG]
Povinelli, D. J., Nelson, K. E. & Boysen, S. T. (1990) Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology* 104:203–10. [CG, aCMH, MDM, JEP, TRZ]
(1992a) Comprehension of role reversal in chimpanzees: Evidence of empathy? *Animal Behaviour* 43:633–40. [aCMH]
Povinelli, D. J., Parks, K. A. & Novak, M. A. (1991) Do rhesus monkeys attribute knowledge and ignorance to others? *Journal of Comparative Psychology* 105:318–25. [aCMH]
(1992b) Role reversal by rhesus monkeys, but no evidence of empathy. *Animal Behaviour* 43:269–81. [aCMH]
Povinelli, D. J. & Preuss, T. M. (1995) Theory of mind: Evolutionary history of a cognitive specialization. *Trends in Neurosciences* 18:418–24. [SFW]
Povinelli, D. J., Rulf, A. B., Landau, K. R. & Bierschwale, D. T. (1993) Self-recognition in chimpanzees: Distribution, ontogeny, and patterns of emergence. *Journal of Comparative Psychology* 107:347–72. [aCMH, RWM]
Premack, D. (1983) Animal cognition. *Annual Review of Psychology* 34:351–62. [aCMH]
(1988) "Does the chimpanzee have a theory of mind?" revisited. In: *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*, ed. R. W. Byrne & A. Whiten. Oxford University Press. [aCMH]
Premack, D. & Dasser, V. (1991) Perceptual origins and conceptual evidence for theory of mind in apes and children. In: *Natural theories of mind*, ed. A. Whiten. Blackwell. [aCMH]
Premack, D. & Premack, A. J. (1983) *The mind of an ape*. Norton. [arCMH]
Premack, D. & Woodruff, G. (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1(4):515–26. [WMB, GC, CG, aCMH, MDM, AP]
Quine, W. V. O. (1951) Two dogmas of empiricism. *Philosophical Review* 60:20–43. [rCMH]
(1969) Epistemology naturalized. In: *Ontological relativity and other essays*, ed. W. V. O. Quine. Columbia University Press. [rCMH]
Reichenbach, H. (1951) *The rise of scientific philosophy*. University of Chicago Press. [aCMH]
Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1996) Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3:131–41. [SFW]
Robert, S. (1986) Ontogeny of mirror behavior in two species of great apes. *American Journal of Primatology* 10:109–17. [aCMH]
Roitblat, H. L., Harley, H. E. & Helweg, D. A. (1993) Cognitive processing in artificial language research. In: *Language and communication: Comparative perspectives*, ed. H. L. Roitblat, L. M. Herman & P. E. Nachtigall. Erlbaum. [JEP]
Rumbaugh, D. M., Hopkins, W. D., Washburn, D. A. & Savage-Rumbaugh, S. (1991) Comparative perspectives of brain, cognition, and language. In: *Biological and behavioral determinants of language development*, ed. N. A. Krasnegor, D. M. Rumbaugh, R. L. Schiefelbusch & M. Studdert-Kennedy. Erlbaum. [JEP]
Russon, A. E. & Galdikas, B. M. F. (1993) Imitation in ex-captive orangutans. *Journal of Comparative Psychology* 107:147–61. [aCMH]
Savage-Rumbaugh, E. S. & Lewin, R. (1994) *Kanzi: The ape at the brink of the human mind*. John Wiley & Sons. [HLM]
Shore, B. (1996) *Culture in mind: Cognition, culture, and the problem of meaning*. Oxford University Press. [HLM]
Smith, I. M. & Bryson, S. E. (1994) Imitation and action in autism – a critical review. *Psychological Bulletin* 116:259–73. [SFW, TRZ]
Sober, E. (1988) *Reconstructing the past: Parsimony, evolution, and inference*. MIT Press. [aCMH]

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press

(1989) Independent evidence about a common cause. *Philosophy of Science* 56:275–87. [aCMH]
Spence, K. W. (1937) Experimental studies of learning and higher mental processes in infra-human primates. *Psychological Bulletin* 34:806–50. [aCMH]
Spreckelsen, C., Schurgpfeiffer, E. & Ewert, J. P. (1995) Responses of retinal and tectal neurons in non-paralyzed toads *Bufo bufo* and *B. Marinus* to the real size versus angular size of objects moved at variable distance. *Neuroscience Letters* 184:105–08. [SFW]
Stammbach, E. (1988) Group responses to specially skilled individuals in a *Macaca fascicularis* group. *Behaviour* 107:241–66. [aCMH]
Stone, T. & Davies, M., eds. (1995) *Mental stimulation: Evaluation and applications*. Basil Blackwell. [AW]
Strawson, P. (1952) *Introduction to logical theory*. Methuen. [aCMH]
Suarez, S. D. & Gallup, G. G. (1981) Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of Human Evolution* 10:175–88. [aCMH, TRZ]
(1986a) Face touching in primates: A closer look. *Neuropsychologia* 24:597–600. [GGG]
(1986b) Social responding to mirrors in rhesus macaques: Effects of changing mirror location. *American Journal of Primatology* 11:239–44. [aCMH]
Suddendorf, T. (in press) The rise of the metamind - beyond the immediately present. In: *Evolution of the hominid mind*, ed. M. C. Corballis & S. Lea. Oxford University Press. [TS]
Suddendorf, T. & Corballis, M. C. (1997) Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs* 123:133–67. [TS]
Suddendorf, T. & Fletcher-Flinn, C. M. (1997) Theory of mind and the origins of divergent thinking. *Journal of Creative Behavior* 31:59–69. [TS]
Suddendorf, T., Fletcher-Flinn, C. M. & Johnston, L. (1996) Pantomime: A form of metarepresentational pretence? Unpublished manuscript. University of Auckland. [TS]
Sumita, K., Kitahara-Frisch, J. & Norikoshi, K. (1985). The acquisition of stone-tool use in captive chimpanzees. *Primates* 26:168–81. [aCMH]
Swartz, K. B. & Evans, S. (1991) Not all chimpanzees show self-recognition. *Primates* 32: 483–96. [aCMH, RWM]
Terrace, H. S., Petitto, L. A., Sanders, R. J. & Bever, T. G. (1979) Can an ape create a sentence? *Science* 206:891–902. [aCMH]
Thompson, R. K. R. & Contie, C. L. (1994) Further reflections on mirror usage by pigeons: Lessons from Winnie-the-Pooh and Pinocchio too. In: *Self-awareness in animals and humans*, ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [RWM]
Thorndike, E. L. (1898) Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monographs*, 2, No. 8. [aCMH]
Timberlake, W. & Grant, D. L. (1975) Autoshaping in rats to the presentation of another rat predicting food. *Science* 190:690–92. [aCMH]
Tomasello, M. (1996) Do apes ape? In: *Social learning: The roots of culture*, ed. C. M. Heyes & B. G. Galef. Academic Press. [aCMH]
Tomasello, M. & Call, J. (1994) Social cognition of monkeys and apes. *Yearbook of Physical Anthropology* 37. [aCMH]
Tomasello, M., Davis-DaSilva, M., Camak, L. & Bard, K. (1987) Observational learning of tool-use by young chimpanzees. *Human Evolution* 2:175–83. [KAB, aCMH]
Tomasello, M., George, B. L., Kruger, A. C., Farrar, M. J. & Evans, A. (1985) The development of gestural communication in young chimpanzees. *Journal of Human Evolution* 14:175–86. [DAL]
Tomasello, M., Savage-Rumbaugh, S. & Kruger, A. C. (1993) Imitative learning of actions on objects by children, chimpanzees, and enculturated chimpanzees. *Child Development* 64:1688–1705. [aCMH]
Trevarthen, C. (In press) The concept and foundations of infant intersubjectivity. In: *Intersubjective communication and emotion in early ontogeny: A source book*, ed. S. Braten. Cambridge University Press. [KAB]
Trick, L. & Pylyshyn, Z. (1989) Subitizing and FNST spatial index model. University of Ontario, COGMEM No. 44. [IMP]
Turing, A. M. (1950) Computing machinery and intelligence. *Mind* (New Series) 59:433–60. [RWM]
Ungerer, J. A. (1989) The early development of autistic children. In: *Autism: Nature, diagnosis and treatment*, ed. G. Dawson. Guilford Press. [aCMH]
Ungerer, J., Zelazo, P., Kearsley, R. & O'Leary, K. (1981) Developmental changes in the representation of objects in symbolic play from 18 to 34 months of age. *Child Development* 52:186–95. [rCMH]
Visalberghi, E. & Fragaszy, D. M. (1992) Do monkeys ape? In: *"Language" and intelligence in monkeys and apes*, ed. S. T. Parker & K. R. Gibson. Cambridge University Press. [aCMH]
Visalberghi, E. & Trinca, L. (1989) Tool use in capuchin monkeys, or Distinguish between performing and understanding. *Primates* 30:511–21. [aCMH]
Vonhofsten, C. & Siddiqui, A. (1993) Using the mother's actions as a reference for object exploration in 6-month-old and 12-month-old infants. *British Journal of Developmental Psychology* 11:61–74. [SFW]
Waal, F. de (1982) *Chimpanzee politics*. Jonathan Cape. [aCMH]
(1991) Complementary methods and convergent evidence in the study of primate social cognition. *Behaviour* 118:297–320. [aCMH]
Watson, J. S. (1994) Detection of self: The perfect algorithm. In: *Self-awareness in animals and humans*, ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [RWM]
Westergaard, G. C. (1988) Lion-tailed macaques (*Macaca silenus*) manufacture and use tools. *Journal of Comparative Psychology* 102:152–59. [aCMH]
Whiten, A. ed. (1991a) *Natural theories of mind*. Basil Blackwell. [aCMH]
Whiten, A. (1991b) The emergence of mindreading: Steps towards an interdisciplinary enterprise. In: *Natural theories of mind*, ed. A. Whiten. Basil Blackwell. [AW]
(1993) Evolving a theory of mind: The nature of non-verbal mentalism in other primates. In: *Understanding other minds*, ed. S. Baron-Cohen, H. Tager-Flusberg & D. J. Cohen. Oxford University Press. [AW]
(1994) Grades of mindreading. In: *Children's early understanding of mind: Origins and development*, ed. C. Lewis & P. Mitchell. Hove. Erlbaum. [aCMH]
(1996a) Imitation, pretense, and mindreading: Secondary representation in comparative primatology and developmental psychology? In: *Reaching into thought*, ed. A. Russon, K. A. Bard & S. T. Parker. Cambridge University Press. [KAB]
(1996b) When does smart behaviour-reading become mind-reading? In: *Theories of theories of mind*, eds. P. Carruthers & P. K. Smith. Cambridge University Press. [aCMH]
(1997) The Machiavellian mindreader. In: *Machiavellian intelligence II*., ed. A. Whiten & R. W. Byrne. Cambridge University Press. [AW]
(in press a) The evolution and development of the mindreading system. In: *Piaget, evolution and development*, ed. J. Langer & M. Killen. Lawrence Erlbaum. [AW]
(in press b) Chimpanzee cognition and the question of mental representation. In: *Metarepresentation*, ed. D. Sperber. Oxford University Press. [JCG, AW]
Whiten, A. & Byrne, R. W. (1988) Tactical deception in primates. *Behavioral and Brain Sciences* 11:233–73. [aCMH]
(1991) The emergence of metarepresentation in human ontogeny and primate phylogeny. In: *Natural theories of mind*, ed. A. Whiten. Basil Blackwell. [aCMH]
Whiten, A. & Custance, D. M. (1996) Studies of imitation in chimpanzees and children. In: *Social learning in animals: The roots of culture*, ed. C. M. Heyes & B. G. Galef. Academic Press. [aCMH]
Whiten, A., Custance, D. M., Gomez, J. C., Teixidor, P. & Bard, K. A. (1996) Imitative learning of artificial fruit processing in children (Homo sapiens) and chimpanzees. *Journal of Comparative Psychology* 110:3–14. [DC, aCMH]
Whorf, B. (1956) *Language, thought, and reality*. MIT Press. [WMB]
Wimmer, H. & Perner, J. (1983) Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13:103–28. [GC, MDM, TS, AW]
Wimmer, H. & Weichbold, V. (1994) Children's theory of mind: Fodor's heuristics examined. *Cognition* 53:45–57. [DK]
Wimsatt, W. C. (1981) Robustness, reliability and overdetermination. In: *Scientific inquiry and the social sciences,* ed. M. B. Brewer & B. E. Collins. Jossey-Bass. [rCMH]
Witte, E. A., Villareal, M. & Marrocco, R. T. (1996) Visual orienting and alerting in rhesus monkeys: Comparison with humans. *Behavioral Brain Research* 82:103–12. [SFW]
Woodruff, G. & Premack, D. (1979) Intentional communication in the chimpanzee: The development of deception. *Cognition* 7:333–62. [aCMH, RWM]
Zentall, T. R., Sutton, J. E. & Sherburne, L. M. (1996) True imitative learning in pigeons. *Psychological Science* 7:343–46. [TRZ]

https://doi.org/10.1017/S0140525X98450702 Published online by Cambridge University Press