Anim. Behav., 1994, **48**, 242–244

Cues, convergence and a curmudgeon: a reply to Povinelli

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

(Received 7 January 1994; accepted 14 January 1994; MS. number: SC-976)

In reply to my comments (Heyes 1993) on a study of mental state attribution in chimpanzees, *Pan troglodytes*, by Povinelli et al. (1990), Povinelli (1994) has (1) provided additional data from the experiment, (2) queried my suggestions for improvements in the experimental design, (3) proposed an experiment which is supposed to overcome the problems encountered by the earlier study, and (4) reaffirmed the need for convergent evidence of mental state attribution in chimpanzees. Since my original article singled out the experiment by Povinelli et al. as embodying a relatively promising approach to the investigation of mental state attribution in animals, it is not surprising that I agree with much of what he has said in his reply. In particular, I welcome the additional data, admire the design of the proposed experiment, and agree whole-heartedly about the need for convergent evidence. However, as I respond to each of Povinelli's points (1–4 above), it will become apparent that we take transfer procedures to have different sorts of limitations in the study of mental state attribution, and that we may use different criteria in assessing convergent evidence.

Povinelli et al. (1990) reasoned that if, instead of attributing mental states, their chimpanzees learned in the first two phases of discrimination training to select the cup indicated by the person who had done the baiting, or who was in the room during baiting, then the animals' performance would drop to chance levels when the test procedure was changed in phase three. The additional data reported by Povinelli (1994) show that this is exactly what happened and therefore, even when viewed within their own framework, the results of Povinelli et al.'s (1990) experiment are more consistent with a 'traditional learning theory account of the chimpanzees' performance' (page 207) than with an account in terms of mental state attribution. Thus, the additional data show that the results of the experiment were not, after all, 'consistent with the hypothesis that chimpanzees are capable of modeling the experiences of others', or of supporting a 'working hypothesis' to that effect. In this they are enlightening, and I am glad if my curmudgeonly comments (Heyes 1993) played some part in their publication.

Assuming, on the basis of the original data, that the chimpanzees' performance was above chance at the beginning of the third phase, I suggested that interpretation of a result like this would be less problematic if the Guesser's view of baiting was obscured by a screen, rather than a paper bag, or if both the Guesser and Knower wore bags, the former during baiting, and the latter after baiting but before the chimpanzee made its choice. With respect to the former suggestion, Povinelli (1994) took me to be proposing the use of a cardboard screen covering the Guesser's face, and wondered how the effect of such a screen could possibly differ from that of a paper bag. This change would, of course, be trivial, but what I was in fact suggesting was use of the same screening procedure that had been employed to prevent the chimpanzee from seeing the containers during baiting. Thus, a cardboard screen would stand at some distance from the Guesser, but would none the less obstruct his view of the baiting procedure. If good transfer were observed under these conditions, or when the Knower had put a bag over his head after baiting, it could not be attributed to fear of a previously 'bagged' trainer leading the chimpanzees to avoid looking at him and, by default, choosing the container indicated by the Knower (see Heyes 1993, page 183).

While the minor methodological changes that I recommended may help to eliminate this specific, fear-related 'killjoy' (Dennett 1983) explanation for successful transfer performance, I agree with Povinelli (1994) that neither they, nor any other transfer conditions that I can imagine, could show definitively that successful performance is due to mental state attribution rather than the use of an observable cue. However, this is not, as Povinelli suggests, because the animals could always be assumed to have learned to use a specific observable cue during the transfer phase. Suppose, for

0003–3472/94/070242+03 $08.00/0

example, that we trained chimpanzees under conditions in which the Guesser leaves the room during baiting, and then tested them when the Guesser's view of baiting is obstructed by a screen. If the animals were rewarded on test for selecting the container indicated by the Knower, and if their initial performance in the second condition was no better than chance, then, like the chimpanzees tested by Povinelli et al. (1990), we could readily attribute later success to learning to 'pick the person without the screen in front of them' (Povinelli 1994). However, if the subjects provided genuine evidence of transfer, that is, if they showed above chance performance at the very beginning of a test in which they were rewarded for selecting the Knower, or if they developed a preference for the Knower in spite of being tested without differential reinforcement, this interpretation could be ruled out.

The real problem is that, even with genuine evidence of transfer in hand, we would be left with the possibility that the chimpanzees use, during both training and transfer phases, a very subtle observable cue such as 'eye–object line'. That is, the animals may 'pick the person for whom there is/was an unobstructed, notional, straight line between their eyes and the baiting event'. Even the reverse-cue experiment proposed by Povinelli (1994) could not overcome this problem. If chimpanzees rewarded for choosing an unbagged Knower over a bagged Guesser learned faster than those rewarded for choosing a bagged Knower over an unbagged Guesser, it could be because they had learned, during the experiment and/or in the course of their day to day lives, that when an individual has recently been positioned such that there was an unobstructed, notional, straight line between their eyes and some critical object, that individual's behaviour tends to provide a reliable cue to the object's location.

I find this kind of interpretation relatively implausible (more implausible than alternative explanations for other behaviour that has been claimed to provide evidence of mental state attribution), and that is why I judge Povinelli's transfer procedure to be the most promising of those used to date to investigate the possibility of mental state attribution in animals (Heyes 1993, in press). However, since assessments of both plausibility and parsimony are highly idiosyncratic, convergent evidence would be necessary to make any claim based on a transfer experiment persuasive. That is, even if we had evidence in hand of transfer performance in chimpanzees that could not be explained with reference to any observable cue apart from something like eye–object line, it would still be insufficient to support the view that chimpanzees attribute mental states.

The attempts of Povinelli (1993) and his colleagues to obtain convergent evidence are exemplary. Using a procedure similar to the one they used with chimpanzees (Povinelli et al. 1990), they have found that rhesus monkeys, *Macaca mulatta*, fail to learn the first discrimination, and therefore cannot be tested for transfer (Povinelli et al. 1991), and that human children begin to show reliable transfer performance at about the same age as they begin to solve problems that are understood by developmental psychologists to require mental state attribution or 'theory of mind' (Povinelli & deBlois 1992). Both Povinelli and I doubt that these studies provide convergent evidence of mental state attribution in chimpanzees, but for different reasons. Povinelli (1993) is doubtful because the children learned the original Knower–Guesser discrimination in many fewer trials than the chimpanzees. I do not find this worrying because it may have resulted from minor variations in the task requirements imposed on chimpanzees and children, or reflected no more than motivational or attentional differences between the species. I find the convergent evidence unpersuasive for three other reasons. First, and most important, because there is no focal evidence with which it can converge. As emphasized above, the chimpanzee experiment yielded negative results. Second, even if the results of the chimpanzee study had been positive, the monkey data would have no bearing on whether the Knower–Guesser transfer test is a test of mental state attribution because, as it turned out, it was not possible to administer the test to the monkeys. Finally, the results of the experiment with children failed to support the hypothesis that the Knower–Guesser transfer test is a measure of mental state attribution by showing that, when they are asked questions about how they and the trainers derived knowledge of the bait's location, children who pass the tests are no more likely to answer correctly than those who fail.

In summary, I maintain that, while the Knower–Guesser transfer test procedure may be incapable of providing sufficient evidence of mental state attribution in animals, it has greater potential than any other method that has been

used to date by comparative psychologists. However, give the additional data reported by Povinelli (1994), neither the Knower–Guesser transfer test, nor any other procedure, has so far provided even suggestive evidence of mental state attribution in chimpanzees, or any other animals: that is, results that cannot be plausibly explained without reference to mental state attribution. These are, perhaps, bleak and curmudgeonly conclusions, but it is difficult to believe that progress will be stimulated by consensual trust in ineffective methods and unpersuasive results.

REFERENCES

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

Heyes, C. M. 1993. Anecdotes, training, trapping, and triangulating: do animals attribute mental states? *Anim. Behav.*, **46**, 177–188.

Heyes, C. M. In press. Social cognition in primates. In: *Handbook of Perception and Cognition, Vol. 9* (Ed. by N. J. Mackintosh). New York: Academic Press.

Povinelli, D. J. 1993. Reconstructing the evolution of mind. *Am. Psychol.*, **48**, 493–509.

Povinelli, D. J. 1994. Comparative studies of animal mental state attribution: a reply to Heyes. *Anim. Behav.*, **48**, 239–241.

Povinelli, D. J. & deBlois, S. 1992. Young children's understanding of knowledge formation in themselves and others. *J. comp. Psychol.*, **106**, 228–238.

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

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