Anim. Behav., 1995, **50**, 1533–1542

Self-recognition in primates: further reflections create a hall of mirrors

C. M. HEYES
Department of Psychology, University College London

(Received 2 December 1994; initial acceptance 7 February 1995; final acceptance 18 May 1995; MS. number: A7175)

Abstract. Gallup et al.'s (1995, *Anim. Behav.*, **50**, 1525–1532) defence of the view that Gallup's mark test has provided evidence of self-recognition in apes is countered point by point. The plausibility of the defence is criticized because of poorly designed experiments, unreliable experimental effects, illegitimate cross-experimental comparisons, false inferences and anecdotal observations. A recent attempt to fortify the case using developmental data (Povinelli et al. 1993, *J. comp. Psychol.*, **107**, 347–372) was unsuccessful because it failed to find a reliable relationship between age and mark test performance. Consequently, there is still no convincing evidence of self-recognition or mirror-guided body inspection in animals. An alternative research strategy is recommended in which the self-recognition hypothesis is tested directly through complete, carefully designed experiments.

Heyes (1994a) argued (1) that studies involving mirrors, and in particular Gallup's mark test procedure (e.g. Gallup 1970), had failed to provide convincing evidence that any non-human primates can use mirrors to derive information about their own bodies, and (2) that such a capacity, whether described as 'self-recognition' or 'mirror-guided body inspection', even if it were clearly demonstrated, would not imply self-awareness or the possession of a self-concept. In a petitionary response, Gallup et al. (1995) mistook the first of these claims, and declined to address the second. Although in my earlier paper I stated clearly the expectation that 'evidence of mirror-guided body inspection in primates will eventually be obtained' (Heyes 1994a, page 914), Gallup et al. (1995) took me to have confused absence of evidence for evidence of absence, and to have asserted that chimpanzees and other primates, are not capable of mirror-guided body inspection. Discussion of the second issue, the significance or implications of mirror-guided body inspection, was deferred by Gallup et al. (1995) 'for purposes of clarity' to be resolved by 'future research'.

In spite of these concerns, I welcome much of what Gallup et al.'s paper contained. Their re-analyses of existing data, and additional studies, represent some development of the 'more critical attitude towards mirror experiments' that my arguments were intended to promote (Heyes 1994a, page 909). Neither this contribution from Gallup et al., however, nor other studies cited by them and published since the paper by Heyes (1994a) was written (e.g. Povinelli et al. 1993), have provided convincing evidence of mirror-guided body inspection in non-human primates. To explain why the data remain unconvincing, this paper answers each of the many points and objections raised by Gallup et al. To facilitate cross-reference, the issues are considered in the same order, and under the same headings, as in the paper by Gallup et al. Both the number and the nature of their objections indicate that Gallup et al. favour a research strategy involving cross-experimental comparisons, partial post-hoc re-analyses of data sets, and circumstantial or anecdotal evidence. These merely create a hall of mirrors in which it seems plausible that certain apes are capable of self-recognition. In contrast, I should like to see the self-recognition hypothesis tested directly in whole, new, carefully designed experiments.

Is Self-recognition an Artefact of Anaesthetization?

Gallup's mark test procedure consists of (1) exposing an animal to a mirror for some period of time, (2) marking, under anaesthetic, parts of the animal's body that it cannot see directly with a

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

0003–3472/95/121533+10 $12.00/0

substance that can be detected only through vision and (3) comparing the frequency with which the animal touches the marked areas in the absence and then in the presence of a mirror (e.g. Gallup 1970). A higher frequency of mark touching in the mirror-present than in the mirror-absent condition has been recorded for chimpanzees, *Pan troglodytes*, and orang-utans, *Pongo pygmaeus*, but not for other species, which has been interpreted as evidence that chimpanzees and orang-utans are capable of self-recognition or mirror-guided body inspection (e.g. Gallup 1970, 1977; Gallup & Suarez 1991). Heyes (1994a) questioned this interpretation, pointing out that in every case where Gallup's mark test had yielded a positive result, the animals had been observed in the mirror-absent condition before the mirror-present condition, and were likely to have had a relatively high baseline or ambient frequency of face-touching behaviour. The animals were tested only a few hours after the anaesthetic was administered, and it is therefore possible that they touched their marks more in the mirror-present condition, not because they used the mirror to detect the marks, but because they had recovered a little more from the anaesthetic, and their level of spontaneous face touching had returned more nearly to normal.

In an initial response to these concerns, Gallup et al. (1995) referred to the results of a control procedure in which animals were subjected to the second and third components of the mark test without prior mirror exposure, and stressed both that no one would 'knowingly' test for self-recognition in animals that were still sedated, and that the investigators always waited until the animals 'appeared' to have fully recovered from the anaesthetic. The inadequacy of this control procedure was discussed by Heyes (1994a): I accused no one of deliberate misconduct and I did not doubt that, upon casual visual inspection, the animals seemed to have recovered.

Temporal Parameters of the Mark Test

Gallup et al. (1995) described as 'dramatic' the differences in mark touching between the mirror-absent and mirror-present conditions that are typically reported, and argued that they were unlikely to have been linked to anaesthetic recovery because (1) the recovery gradient would have to have been very steep, (2) most mark-directed responses occurred shortly after the mirror was reintroduced, and (3) injection-test intervals of between 2 and 7 h have been used with positive results.

Mark test effects may sometimes appear dramatic, but they are seldom statistically reliable. Much of the apparent drama derives from the fact that Gallup and his associates (e.g. Gallup 1970; Gallup et al. 1971; Suarez & Gallup 1981) have presented their results in the form of 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. This practice made it impossible for the reader to assess the reliability of the effect and its magnitude for individual animals. The modesty of the effect is, however, apparent in the data reported by other authors: Calhoun & Thompson (1988) found that, after failing to touch their marks at all during the mirror-absent period, each of the two chimpanzees that they tested made just two responses in the mirror-present condition. Thirty chimpanzees tested by Povinelli et al. (1993, their experiment 4) touched their marks, on average ($\pm$ SD), 2.5 ± 3.7 times in the absence of the mirror and 3.9 ± 8.0 times in its presence. Swartz & Evans (1991) reported that only one of 11 chimpanzees touched its marks more in the mirror-present condition and that, on average, 3.3 ± 3.7 touches occurred while the mirror was absent, whereas 2.9 ± 7.19 were made when it was present. In all three experiments, the mirror-present and mirror-absent periods were each of 30 min duration. Thus, it would not be necessary for an anaesthetic recovery gradient to be improbably steep, or especially uniform across animals, to account for the mark-touching effects typically observed.

Regarding the temporal distribution of responses within observation periods, 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. If such data were available, they would be equally consistent with Gallup's hypothesis and with the hypothesis that mirror introduction elevates arousal and thereby produces an increase in the frequency of a range of behaviour patterns.

Gallup et al. cited two studies, by Calhoun & Thompson (1988) and Povinelli et al. (1993), to support the claim that the mark test effect is robust across a 2–7-h range of injection-test intervals. Calhoun & Thompson allowed 'about 3 hours' and Povinelli et al. 'about 5 hours' for recovery, and, even if the range had been as broad on average as Gallup et al. suggested, the contrast would not have been instructive because Povinelli et al. used a higher dose of anaesthetic (10 mg/kg ketamine) than did Calhoun & Thompson (5–7 mg/kg ketamine).

Responses to Control Body Marks

'Heyes also failed to consider data from studies that have applied control marks to body parts that the subjects can see directly (e.g. Suarez & Gallup 1981)' (Gallup et al. 1995, page 1527). These studies, in which marks were applied to chimpanzees wrists as well as their heads, were discussed in Heyes (1994a, pages 912–913); namely, that (1) wrist touching is likely to require less energy and coordination than head touching, (2) the frequency of wrist touching prior to mirror introduction was low and (3) any decline in frequency of wrist touching between mirror-present and mirror-absent conditions could have resulted from simple habituation to the wrist mark, rather than diversion by the head mark.

Studies Without Anaesthesia

A review of all published results of the mark test revealed that, as the anaesthetic hypothesis would predict, all of the animals that had passed the test were members of species with a relatively high baseline frequency of face touching, and had been marked under anaesthetic, not, for example, in the course of playful interaction or natural sleep (Heyes 1994a). Gallup et al. (1995) claimed that (1) the negative test outcomes I cited (Anderson 1983; Robert 1986; Povinelli 1989) had 'no bearing on [the anaesthetic] hypothesis' (page 1527), and could be accommodated by Gallup's own hypothesis, and (2) there is 'evidence of mirror-mediated mark-directed behaviour without the use of anaesthesia (e.g. Lethmate & Dücker 1973; Patterson 1984; Calhoun & Thompson 1988; Miles 1994)' (page 1527).

Negative results without anaesthesia

The anaesthetic artefact hypothesis could be disconfirmed by reliable evidence that animals can pass the mark test (i.e. touch their marks more in the mirror-present than in the mirror-absent condition) when they have a low baseline frequency of mark touching, or have been properly marked without anaesthetic (i.e. such that they could not have detected either mark application, or the marks themselves, using tactile, olfactory or direct, rather than mirror-mediated, visual cues). Although it is not clear that Povinelli's (1989) elephants, *Elaphus maximus*, were properly marked, each of the stumptailed macaques, *Macaca artoides*, tested by Anderson (1983) and the chimpanzee and orang-utan tested by Robert (1986) apparently met at least one of these conditions. Therefore, the results of these studies could have, but did not, disconfirm the anaesthetic artefact hypothesis, and even Popper (e.g. 1959), the principal proponent of falsificationist methodology, acknowledged that failure to falsify carries epistemological weight.

Povinelli's (1989) and Anderson's (1983) data are consistent, in a straightforward way, with both the anaesthetic artefact hypothesis and Gallup's claim that his mark test detects mirror-guided body inspection or self-recognition. Because Robert's (1986) chimpanzee and orang-utan were not anaesthetized, the anaesthetic hypothesis also directly predicts their failure on the mark test. In contrast, Gallup et al. had to engage in some special pleading to explain these failures; they said that Robert's subjects were too young to pass the mark test. This constitutes special pleading because the most thorough developmental study to date (Povinelli et al. 1993) found no reliable indication of a relationship between age and mark test performance in chimpanzees.

In the course of two experiments, Povinelli et al. (1993, their experiments 3 and 4) mark-tested 38 chimpanzees, ranging in age from 3 years 4 months to 31 years 8 months, and recorded for each the number of times the marks were touched in mirror-absent (control) and mirror-present (test) conditions. In the latter condition, responding was also categorized as 'mirror-monitored' (the animal was facing the mirror and looking ahead) or 'non-mirror monitored' (the animal was not facing the mirror, or 'facing the mirror but not looking into it'). The data from these tests (derived from Tables 4 and 5 in Povinelli et al. 1993) are represented in Fig. 1, where animals were assigned to the age categories used by

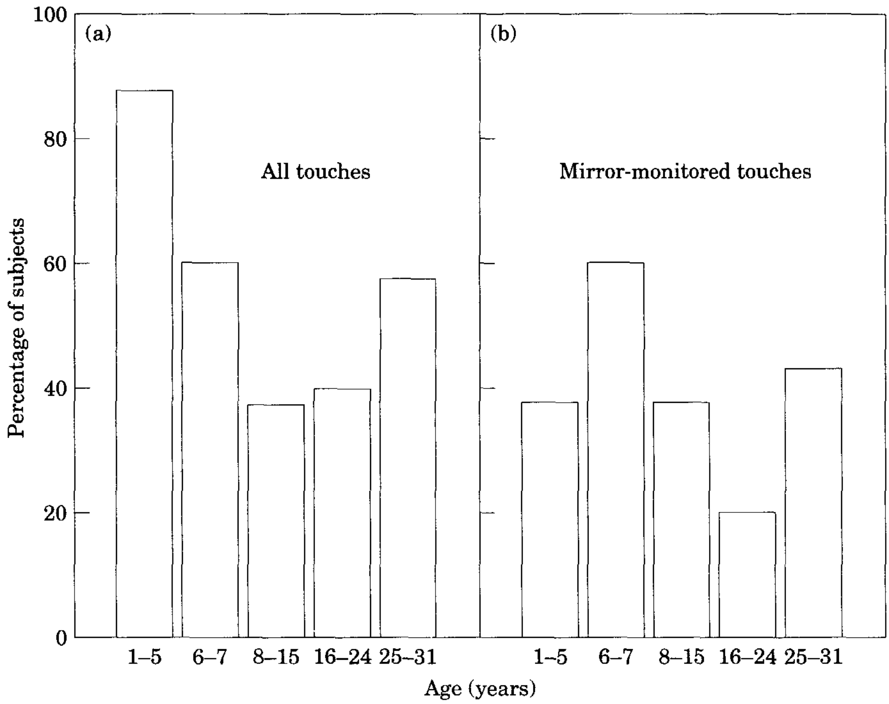

Figure 1. Percentage of chimpanzees in five age categories that passed the mark test, as measured (a) by more mark touches in the mirror-present than in the mirror-absent condition, and (b) by more mirror-mediated mark touches in the mirror-present condition than touches in the mirror-absent condition. Data from Povinelli et al. (1993, their Tables 4 and 5) and D. Povinelli (personal communication).

Povinelli et al. (1993) in their analysis of other self-recognition data.

Figure 1a gives the percentage of animals in each age category that touched their marks more in the mirror-present than in the mirror-absent condition. Using this criterion of success on the mark test, younger chimpanzees were, if anything, more likely than older chimpanzees to pass the mark test. Kendall's S revealed a reliable trend when applied to the first three age groups ($z=1.76$, $P=0.05$), but not when applied to all five groups ($z=1.34$, $P=0.09$). Figure 1b gives the percentage of chimpanzees in each age group that made more mirror-monitored touches in the experimental condition than touches in the control condition. Application of this criterion failed to indicate an age-related trend in test performance; Kendall's S did not reveal a trend when applied to all five age categories ($z=0.38$, $P=0.70$) or to the first three ($z=0.32$, $P=0.74$), and chi-squared analysis failed to show a difference between the 1–5- and the 6–7-year-olds ($\chi^2=0.63$). Similar analyses in which age was treated as a continuous variable, and attempts to correlate age with the magnitude of any difference between responding in experimental and control conditions, also failed to find a relationship between age and mark test performance.

Thus, Povinelli et al. (1993) did not provide reliable evidence that younger chimpanzees are less likely to pass the mark test than older chimpanzees, and, even if evidence of such a relationship were obtained, it would in no way favour Gallup's hypothesis over the anaesthetic hypothesis. The latter identifies rate of recovery from anaesthesia and baseline frequency of face touching as the variables determining mark test outcome, and it is plausible that younger chimpanzees take longer to recover from an anaesthetic, and/or have a lower baseline rate of face touching, than older chimpanzees. Far from being discredited by the study of face touching reported by Gallup et al (1995; see Ambient Face Touching below), the latter possibility is implicated by evidence that younger chimpanzees engage in less self-exploratory behaviour than older members of the same species (Povinelli et al. 1993).

If there were reliable evidence that young chimpanzees typically fail the mark test, then Gallup's early success with the technique would be surprising. In the studies that founded this research field, successful mark test performance was reported for four 'preadolescent' (Gallup 1970) and four 3–6-year-old (Gallup et al. 1971) chimpanzees.

Positive results without anaesthesia

Gallup et al. cited four studies, involving chimpanzees, gorillas and orang-utans, as evidence against the anaesthetic hypothesis. They claimed that these studies 'obtained evidence of mirror-mediated mark-directed behaviour without the use of anaesthesia' (page 1527). In three of the studies (Lethmate & Dücker 1973; Patterson 1984; Miles 1994), the animals were improperly marked (i.e. such that they could have detected either mark application, or the marks themselves, using tactile, olfactory or direct, rather than mirror-mediated, visual cues), and the frequency with which they touched their marks in the presence of a mirror was compared with the frequency of touching the marked area before the marks were applied. Consequently, it is entirely possible that the animals in these studies touched their marks because their attention had been drawn to the marked area by the marking operation, by tactile cues received when the marked areas were initially touched by chance, and/or by visual inspection of traces of pigment adhering to their fingers after chance contact. The latter two possibilities are implicated by the fact that, under comparable test conditions, Lethmate & Dücker (1973) failed to find evidence of persistent mark touching in monkeys, which have a relatively low baseline frequency of face touching (Dimond & Harries 1984; see Species Differences in Face Touching below), and are therefore less likely than chimpanzees to have touched their marks by chance.

The fourth study put forward by Gallup et al. as evidence of mark test success without anaesthetic (Calhoun & Thompson 1988) had a promising design. Under physical restraint, Calhoun & Thompson marked a chimpanzee, Barash, on one ear with a solvent containing a dye and on the other ear with the solvent alone. Presumably their intention was to compare the frequency with which the animal touched its ears in mirror-present and mirror-absent conditions but, if they did this, their paper did not report the outcome. No quantitative data were given. The reader was told merely that 'positive results (mark-directed responses) were recorded', which amounts to no more than an assurance that it was the authors' impression that Barash (a 2-year-old chimpanzee) recognized himself in the mirror.

Specificity of Mark-directed Responses

Gallup et al. (1995) claimed that Heyes (1994a) failed to provide 'constructive suggestions for future research' (page 1531), but also presented, in their study of Megan's face-touching behaviour, a (partial) implementation of one of my proposals (Heyes 1994a, page 911). A study is undoubtedly needed that compares responding to marked and comparable unmarked areas. Gallup et al.'s analysis of Megan's behaviour was not instructive, however, because it failed to demonstrate that Megan touched her marked areas more because she had detected the marks using the mirror, or that she was in any way representative of other chimpanzees given the mark test.

Instead of having detected the marks using the mirror, Megan may have touched her marked eye and ear more than her unmarked eye and ear by chance. The statistical analysis presented by Gallup et al. showed that the observers are unlikely to have distributed their ratings by chance, but not that Megan is unlikely to have distributed her touches by chance. And even if Gallup et al. had shown that Megan had a systematic bias in favour of touching the marked areas, it may have had nothing to do with the presence of a mirror. Perhaps those areas were itching, or it may simply have been her habit to touch one ear and eye more than the other. To support the hypothesis that an animal can detect a mark using a mirror, it would be necessary to show, not only that marked areas are touched more than comparable unmarked areas when a mirror is present, but also that this effect diminishes or disappears when a mirror is absent. The hypothesis clearly predicts an interaction between these two variables.

A final possibility, which also has implications for the conduct of future experiments, is that

Megan touched her marked ear and eye because she was subtly and inadvertently rewarded for doing so by someone who was with her when she was tested. Gallup et al. showed a laudable concern about the possibility of response bias among their video raters, but did not mention any precautions against an observer of the live performance emitting the kind of cues that led to the celebration of Clever Hans (Pfungst 1965).

In contrast to Gallup et al.'s study of Megan, the following experiment would effectively test the self-recognition hypothesis. In the usual way, each chimpanzee would be given several days' exposure to a large mirror, and then marked, under anaesthetic, with an odourless, non-irritant dye on one eyebrow ridge and the opposite ear. Several hours later, the animals' behaviour would be recorded on video for 30 min when no mirror was present, and for a further 30 min when the mirror had been reintroduced. To guard against the possibility of cueing, no observers would be present during the test periods. Viewing the videotapes later, observers would record the number of times that each chimpanzee touched (1) the marked ear and eye in the mirror-absent condition, (2) the unmarked ear and eye in the mirror-absent condition, (3) the marked ear and eye in the mirror-present condition and (4) the unmarked ear and eye in the mirror-present condition.

Ideally, the observers should be unaware of which areas are marked when they produce their ratings. It may be possible to achieve this by using a dye that cannot be detected on a video recording made with a filter lens. Alternatively, one might seek expectancy effects by comparing the ratings of observers who tend to believe that chimpanzees are capable of mirror self-recognition, to those who are more sceptical. If the ratings of these observers conformed to their expectations, or if such a test were not performed, it would be prudent to use a conservative measure in which a chimpanzee is judged to have touched an area on any given occasion only if all raters agreed that it had done so.

If statistical analysis of the data from this experiment indicated that the chimpanzees touched 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.

Ambient Face Touching

Citing Povinelli et al. (1993), Gallup et al. (1995) claimed that there is a 'rather striking developmental transition among chimpanzees, with animals younger than 6 years of age rarely showing self-recognition. Most adolescents show compelling signs of self-recognition; however, among adults aged 16 years or older there is a decline in the proportion of chimpanzees that appear to recognize themselves in mirrors' (page 1529). They then assumed that the anaesthetic artefact hypothesis would predict a parallel developmental curve in spontaneous, baseline or ambient face touching, and presented as evidence disconfirming this prediction a study that failed to find a relationship between age and frequency of spontaneous face touching in chimpanzees.

There are two major problems with this argument. First, the anaesthetic artefact hypothesis currently does not make predictions regarding the development of spontaneous face touching because, as its name suggests, it is a possible explanation for the results of anaesthetic mark tests, and there is no evidence that performance in these tests is a function of age (see Studies Without Anaesthesia above).

Second, like most null results, those of the face-touching study reported by Gallup et al. (1995, their Table I) are ambiguous, and certainly do not support the conclusion that face touching does not vary with age in chimpanzees. The study may have failed to detect an effect of age due to rater or sampling error. Only one person scored the behaviour, no information was given about the conditions in which observations were made or how these conditions may have varied across subjects, and each animal was watched for only 10 min. The latter feature is surprising because Suarez & Gallup (1986) criticized Dimond & Harries (1984) for sampling face touching in each of their animals for no more than 20 min.

Thus, Gallup et al. presented a null result of a methodologically weak study as disconfirmation of a non-existent prediction.

Other Mark Test Considerations

In discussing other mark test considerations, Gallup et al. attempted to show that 'The anaesthetization hypothesis is also contrary to the fact that subjects that pass the mark test typically show compelling patterns of mirror-mediated self-directed responding prior to anaesthetization' (page 1530). In fact, the anaesthetic hypothesis is perfectly consistent with the available evidence of a relationship between mark test performance and self-directed behaviour.

Quantitative data were reported in only one of the studies cited by Gallup et al. as evidence of a relationship between mark test performance and self-directed behaviour; the remaining observations were anecdotal and, in some cases, also unpublished. The quantitative data showed that, compared to chimpanzees who failed the mark test, those who passed had previously shown a higher frequency of self-directed behaviour in the presence of a mirror (Povinelli et al. 1993). This result suggests that the mark test and Povinelli et al.'s test of self-directed behaviour in the presence of a mirror were, at least in part, measuring the same thing. Supporters of the self-recognition hypotheses have assumed that the common target or quantity was the animals' capacity for self-recognition. Povinelli et al. (1993) treated face touching as a component of self-directed behaviour, however; therefore the anaesthetic hypothesis would plausibly suggest that the common target was the animals' propensity to engage in spontaneous or ambient face touching. Thus, it is possible that in both tests administered by Povinelli et al., the chimpanzees' self-directed behaviour had nothing whatever to do with the mirror.

Following Gallup et al., the present discussion focuses on the quality of existing evidence that primates are capable of mirror self-recognition, not on the other major issue raised by Heyes (1994a), the putative relationship between self-recognition and self-conception. However, before leaving the subject of self-directed behaviour, it may be worth noting that, if the occurrence of this behaviour in the presence of a mirror is assumed to be an index of self-recognition, then the developmental trends in self-directed behaviour reported by Povinelli et al. (1993) and quoted by Gallup et al. (1995) further undermine the view that self-recognition or mirror-guided body inspection correlates with the possession of a self-concept. Although there are no age-related trends in mark test performance (see Studies Without Anaesthesia above), Povinelli et al. showed that self-directed behaviour in the presence of a mirror increases between the ages of 1 and 15 years, and then declines sharply in mature adult chimpanzees. Is it plausible that chimpanzees acquire a self-concept or self-awareness as they grow up, and then promptly lose it upon reaching adulthood?

Is Self-directed Behaviour an Artefact of Normal Self-grooming?

On the basis of an unpublished study by T. J. Eddy, D. J. Povinelli & G. G. Gallup, Gallup et al. (1995) claimed that 'contrary to Heyes' thesis, mirror-mediated patterns of self-directed behaviour in chimpanzees old enough to show self-recognition can be reliably distinguished from instances of normal self-grooming, general scratching or body touching' (page 1530). In fact, given the design of the experiment by T. J. Eddy et al., the most that it could have shown is that, under the conditions used in the study, human observers can reliably distinguish, in terms of quantity (not quality) of self-directed behaviour, 7–10-year-old chimpanzees that are in the presence of mirrors from 7–10-year-old chimpanzees that are in the presence of a video of conspecifics. This result will have been shown if there was no non-behavioural basis for the discrimination (i.e. the mirror-versus-video variable was not confounded, and the observers were truly blind with respect to group assignment), if inter-rater reliability was good, and if the difference between conditions in the number of reported self-directed behaviour patterns was statistically reliable. If these standard conditions are met, there will be reason to believe that the 7–10-year-old chimpanzees showed more self-directed behaviour in the mirror condition. This quantitative difference between mirror and video conditions would not imply, however, that the self-directed behaviour exhibited in the mirror condition was of a different quality; that is, it would not imply that the behaviour which occurred in the presence of the mirror was 'mirror-mediated', or, therefore, that mirror-mediated self-directed behaviour had been distinguished from non-mirror-mediated self-directed behaviour. A simpler explanation for the effect is that the chimpanzees engaged in self-directed behaviour with their usual, ambient, frequency in the mirror condition, and that such behaviour dropped below baseline levels in the video condition because it was displaced by social responding to the conspecifics in the film.

If social responses to the video displaced self-directed behaviour in the 7–10-year-olds, then why not in the 3-year-olds tested by T. J. Eddy et al.? Without seeing the data, it is impossible to tell whether there really was an effect of age (i.e. a statistically reliable interaction between age and condition, with age unconfounded) and, if so, whether in both conditions the frequency of self-directed behaviour in 3-year-olds was high or low relative to that of older chimpanzees. The developmental data reported by Povinelli et al. (1993, their Figure 3) suggest that the frequency is likely to have been relatively low, and therefore a floor effect may have interfered with the detection of a difference between mirror and video conditions in the 3-year-olds.

Thus, the study by T. J. Eddy et al. leaves a fundamental question unanswered: is it possible to tell, simply by looking at an animal with access to a mirror, whether its behaviour, self-directed or otherwise, is or is not mirror-mediated? It may be possible to tell whether the animal is facing the mirror, whether there is an unobstructed line between its eyes and the mirror, or even whether its behaviour is 'mirror-monitored', if that means no more than that the behaviour is such as to give a human observer the impression of mirror use. However, to say that a behaviour is mirror-mediated implies that information from the mirror image was motivating or guiding body movements, and both 'folk' and cognitive psychology assume that the uptake of information by a cognitive system is something that may be inferred from, but not directly observed in, behaviour (Heyes, in press).

Species Differences in Face Touching

Following Suarez & Gallup (1986), Gallup et al. attempted to discredit Dimond & Harries' (1984) evidence of species differences in frequency of spontaneous face touching by identifying flaws in the way in which Dimond & Harries presented their results, and pointing out that Suarez & Gallup had found higher levels of face touching in monkeys than those reported for monkeys, and comparable to those found in chimpanzees by Dimond & Harries.

It is not clear why Suarez & Gallup attributed the differences between their own results and those of Dimond & Harries to methodological weaknesses in the latter study. The two studies examined different species of monkey in different groupings (small, single-species versus large, mixed-species groups), in different settings (indoors versus outdoors), and using different sampling procedures (six 10-min periods versus one 20-min period per animal). Any one or combination of these variables could have been responsible for the different estimates of face-touching frequency, and in no case would it imply that one study had been conducted properly while the other had not.

Table I. Mean number of face touches per 5 min of observation per animal

	Dimond & Harries (1984)	Suarez & Gallup (1986)	Gallup et al. (1995)
Monkeys	0.3	4.2	—
Chimpanzees	6.1	—	11.6

Data are calculated from one 20-min period of observation (Dimond & Harries), six 10-min periods (Suarez & Gallup) and two 5-min periods (Gallup et al.).

If Suarez & Gallup thought that Dimond & Harries' methodology was weak and therefore doubted the conclusion that chimpanzees touch their faces more than monkeys, then the appropriate course of action would have been to compare chimpanzees and monkeys in a single experiment, using a single methodology. Instead, Suarez & Gallup compared their own estimate of face touching in monkeys with Dimond & Harries' estimate of face touching in chimpanzees and concluded that, because there was little numerical difference between the two, monkeys must touch their faces almost as much as chimpanzees. This approach was illegitimate because, according to Suarez & Gallup, they had already shown by comparing their monkey data with those of Dimond & Harries that the latter used unreliable measurement techniques.

Gallup et al. (1995) appear to have provided the only data on spontaneous face touching in chimpanzees since those published by Dimond & Harries (1984). Table I represents the data on face touching in chimpanzees and monkeys from Gallup et al. (1995), Suarez & Gallup (1986) and Dimond & Harries (1984). For those willing to trust cross-experimental comparisons, and numerical rather than statistical differences, Table I may suggest that the studies by Suarez &

Gallup and Gallup et al. not only corroborated Dimond & Harries' claim that chimpanzees touch their faces more than monkeys, but also implied that the degree of discrepancy may have been underestimated.

The Mark Test Revisited

Gallup et al. (1995) seemed to suggest, upon 'revisiting' the mark test, that one should not rely too heavily on it as an index of self-recognition, and that mark test evidence is no more compelling than anecdotal reports of mirror self-recognition. I cannot disagree.

Lack of Convergent Evidence

Contrary to what is claimed by Gallup et al. (1995), Menzel et al. (1985) did not include a 'control condition' in which their chimpanzee subjects were exposed to a pre-recorded video showing the door and their target, but not their hands. It is true that, at some unspecified interval after collection of the data discussed by Heyes (1994a), Menzel et al. watched the same chimpanzees' reactions to a pre-recorded video of this kind, and reported the observers' impressions that the animals found the target more readily when attending to live recordings. This did not constitute a control condition, however, because performance in relation to the live and pre-recorded videos was not compared in any disciplined or quantitative way, and the two kinds of video were not viewed under the same conditions; in the follow-up study the chimpanzees had a choice of monitors to watch, whereas in the main study only one was available. Whether a procedure is a control depends on how it fits, or is related to, the remainder of the experiment.

Conclusions

In their closing paragraph, Gallup et al. asserted that 'it is important to note that theorizing about self-recognition, its implications for self-conception and a corresponding capacity for mental state attribution has been a rich source of testable hypotheses about both the evolution and ontogeny of social intelligence' (page 1531). Theorizing about self-recognition has certainly been influential, and that is why it is worthwhile to examine closely its empirical and conceptual underpinnings. If these are as weak as I suggest (Heyes 1993, 1994b), then theorizing about self-recognition has already led to a considerable waste of time and resources. These costs cannot be recovered, but it may be possible to mitigate any further damage.

A clear answer to the question of whether apes and other non-human animals are capable of mirror-guided body inspection is more likely to come from complete, new, carefully designed experiments that test the hypothesis, than from partial follow-up studies and piecemeal, post-hoc reanalyses of existing data. Therefore, I urge those with the resources for primate research and an interest in self-recognition, to conduct such experiments, to examine the subject with a steady eye, and not to linger in the hall of mirrors.

ACKNOWLEDGMENTS

The preparation of the manuscript was supported by a grant from the Biotechnology and Biological Sciences Research Council. I am grateful to Daniel Povinelli for supplying the data on subjects' ages included in Fig. 1, and to Chris Mitchell, Richard Darby and Phil Reed for valuable discussion while the manuscript was in preparation.

REFERENCES

Anderson, J. R. 1983. Responses to mirror-image stimulation and assessment of self-recognition in mirror- and peer-reared stumptail macaques. *Q. Jl exp. Psychol.*, **35B**, 201–212.

Calhoun, S. & Thompson, R. L. 1988. Long-term retention of self-recognition by chimpanzees. *Am. J. Primatol.*, **15**, 361–365.

Dimond, S. & Harries, R. 1984. Face touching in monkeys, apes and man: evolutionary origins and cerebral asymmetry. *Neuropsychologia*, **22**, 227–233.

Gallup, G. G., Jr. 1970. Chimpanzees: self-recognition. *Science*, **167**, 86–87.

Gallup, G. G., Jr. 1977. Self-recognition in primates. *Am. Psychol.*, **32**, 329–338.

Gallup, G. G., Jr, McClure, M. K., Hill, S. D. & Bundy, R. A. 1971. Capacity for self-recognition in differentially reared chimpanzees. *Psychol. Rec.*, **21**, 69–74.

Gallup, G. G., Jr, Povinelli, D. J., Suarez, S. D., Anderson, J. R., Lethmate, J. & Menzel, E. W. 1995. Further reflections on self-recognition in primates. *Anim. Behav.*, **50**, 1525–1532.

Gallup, G. G., Jr & Suarez, S. D. 1991. Social responding to mirrors in rhesus monkeys (*Macaca mulatta*):

effects of temporary mirror removal. *J. comp. Psychol.*, **105**, 376–379.

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

Heyes, C. M. 1994a. Reflections on self-recognition in primates. *Anim. Behav.*, **47**, 909–919.

Heyes, C. M. 1994b. Social cognition in primates. In: *Animal Learning and Cognition, Handbook of Perception and Cognition* (Ed. by N. J. Mackintosh), pp. 281–305. New York: Academic Press.

Heyes, C. M. In press. Self-recognition in primates: irreverence, irrelevance and irony. *Anim. Behav.*

Lethmate, J. & Dücker, G. 1973. Untersuchungen zum selbsterkennen im spiegel bei orang-utans und einigen anderen affenarten. *Z. Tierpsychol.*, **33**, 248–269.

Menzel, E. W., Jr, Savage-Rumbaugh, E. S. & Lawson, J. 1985. Chimpanzee (*Pan troglodytes*) spatial problem solving with the use of mirrors and televised equivalents of mirrors. *J. comp. Psychol.*, **99**, 211–217.

Miles, H. L. W. 1994. Me Chantek: the development of self-awareness in a signing orangutan. In: *Self-awareness in Animals and Humans* (Ed. by S. T. Parker, R. W. Mitchell & M. L. Boccia), pp. 254–272. New York: Cambridge University Press.

Patterson, F. G. P. 1984. Self-recognition by *Gorilla gorilla gorilla*. *Gorilla*, **7**, 2–3.

Pfungst, O. 1965. *Clever Hans: The Horse of Mr. von Osten*. New York: Holt.

Popper, K. 1959. *The Logic of Discovery*. London: Hutchinson.

Povinelli, D. J. 1989. Failure to find self-recognition in Asian elephants (*Elephas maximus*) in contrast to their use of mirror cues to discover hidden food. *J. comp. Psychol.*, **103**, 122–131.

Povinelli, D. J., Rulf, A. B., Landau, K. R. & Bierschwale, D. T. 1993. Self-recognition in chimpanzees (*Pan troglodytes*): distribution, ontogeny, and patterns of emergence. *J. comp. Psychol.*, **107**, 347–372.

Robert, S. 1986. Ontogeny of mirror behaviour in two species of great apes. *Am. J. Primatol.*, **10**, 109–117.

Suarez, S. D. & Gallup, G. G., Jr. 1981. Self-recognition in chimpanzees and orangutans, but not gorillas. *J. hum. Evol.*, **10**, 175–188.

Suarez, S. D. & Gallup, G. G., Jr. 1986. Social responding to mirrors in rhesus macaques: effects of changing mirror location. *Am. J. Primatol.*, **11**, 239–244.

Swartz, K. B. & Evans, S. 1991. Not all chimpanzees show self-recognition. *Primates*, **32**, 483–496.