

Small mirrors do the trick: A simple, but effective method to study mirror self-recognition in chimpanzees

Kathrin S. Kopp^{1*^}, Sonja J. Ebel^{1,2*}, Roman M. Wittig¹, Daniel B. M. Haun¹, Catherine Crockford¹

Accepted for publication in *Animal Behavior and Cognition* (17th October 2020)

This is the uncorrected accepted version of the paper.

¹Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

²Lifespan Psychology, Institute of Psychology I, University of Lübeck, Lübeck, Germany

*These authors contributed equally to the work.

[^]Corresponding author; e-mail: kathrin_kopp@eva.mpg.de

Abstract

Mirror self-recognition (*MSR*) is considered an indicator of self-awareness. Standardized mirror tests reveal compelling evidence for *MSR* in a few non-human species, including all great apes. However, substantial inter-individual variation of *MSR* within species resulted in an ongoing methodological controversy, questioning the appropriateness of standard *MSR* tests for cross-species comparisons. Especially lack of motivation is discussed as one possible cause for false negative results. Here, we compare the spontaneous behavioral response of 47 zoo-housed chimpanzees (*Pan troglodytes*) to (i) standard body-sized, stationary mirrors and (ii) small, portable hand mirrors. We predicted that the monopolizability and maneuverability of small mirrors increase the chances of identifying *MSR* across a larger proportion of individuals. Chimpanzees both revealed a substantially higher frequency of general mirror-related behaviors and engaged in significantly more and longer behaviors specifically indicating *MSR* when provided with small mirrors compared to a large mirror. Handheld mirrors provide a more sensitive measure for *MSR* within and likely between primate species than the traditional large mirrors, and thereby are a potentially valuable tool for studying self-awareness across species.

Keywords: *MSR*, comparative cognitive research, primates, self-awareness, mirror test, cognition

Introduction

Humans across cultures recognize their specular reflection as their own (*mirror self-recognition, MSR*) (Broesch et al., 2011; Kärtner et al., 2012; Keller et al., 2004; Priel & de Schonen, 1986), which is widely considered an indicator of self-awareness (Asendorpf et al., 1996; Bard et al., 2006; Rochat et al., 2012; Suddendorf & Butler, 2013). The ability for *MSR* emerges around the second year of life, indicated by the shift from mirror-directed behavior to self-directed actions (*mirror-guided self-exploration, SE*) (Amsterdam, 1972). Ontogenetic trajectories of behaviors indicating *MSR* vary depending on socio-cultural influences, e.g., parenting styles (Kärtner et al., 2012; Keller et al., 2004), though not on mirror familiarity (Priel & de Schonen, 1986).

Other animals, including fish, birds, and mammals also spontaneously respond to their image in mirrors and other reflective surfaces, predominantly with social behaviors commensurate with those shown to conspecifics or with other mirror-directed behaviors like, e.g., exploring the mirror (Gallup Jr., 1968). However, in most species, *SE* does not occur (Gallup Jr., 1968; Gallup Jr. & Anderson, 2019) – even after long periods of mirror exposure (Anderson, 1984). These observations have supported claims that *MSR* is a uniquely human ability. However, mirror test results with non-human great apes (from now on referred to as ‘great apes’) may contradict these claims (Reiss & Morrison, 2017). In contrast to lesser apes (Hylobatidae) and monkeys (Anderson, 1984; Suddendorf & Collier-Baker, 2009), chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*) and orangutans (*Pongo spp.*) show spontaneous *SE* of otherwise hard to see body parts and/or explore a mark surreptitiously placed on such a body part, the so-called ‘mark test’ (Bard et al., 2006; Gallup Jr., 1970; Krachun et al., 2019; Lethmate & Dücker, 1973; Lin et al., 1992; Povinelli et al., 1993; Robert, 1986; Suarez & Gallup, 1981; Walraven et al., 1995; Westergaard & Hyatt, 1994). Amongst all tested non-human animals, great apes demonstrate the most consistent and well-replicated evidence for *MSR* (except for gorillas (*Gorilla gorilla*); see e.g., Anderson & Gallup Jr., 2015; Swartz & Evans, 1994).

While *MSR* is a universal human trait, intriguingly, in most studies, less than half of great apes demonstrated compelling evidence for *MSR*. This raised an ongoing controversy concerning methodological issues of *MSR* tests possibly generating false negatives (Anderson & Gallup Jr., 2015; de Veer & van den Bos, 1999; de Waal, 2019; Gallup Jr. et al., 1995; Gallup Jr. & Anderson, 2019; Heyes, 1994, 1995; Kohda et al., 2019; Schilhab, 2004). Here, we investigate a modified method suitable for comparative studies with primates, which aims to reduce false negatives.

Methodological criticisms regarding *MSR* tests frequently refer to: (i) criteria for *MSR* being ambiguous, (ii) the possibility of systematic false negatives due to test designs ignoring species-specific characteristics, and (iii) varying motivation to interact with mirrors having a significant impact on detecting *MSR* abilities (de Veer & van den Bos, 1999).

Several studies in different species considered direct physical access to the mirror as an important factor stimulating *MRB* because it allowed tactile examination and looking behind the mirror (Anderson, 1984; Gallup Jr., 1994; Reiss & Morrison, 2017). In addition, the mirror size was usually adapted to the body size of the study species (e.g., Anderson & Roeder, 1989; Baragli et al., 2017; Gallup Jr., 1968; Plotnik et al., 2006; Prior & Schwarz, 2008; Reiss & Marino, 2001; Vanhooland et al., 2020). In contrast, other characteristics of the mirrors, especially their mobility and controllability by the individual, which Gallup (1994, p. 40) referred to as an “obvious extension”, are rarely considered. One of the few informative studies with primates in this respect showed that small, stationary mirrors elicited less social responses than large, vertical mirrors in capuchins (*Cebus apella*) (Anderson & Roeder, 1989). Moreover, small, portable mirrors stimulated more frequent and diverse mirror-related behaviors, *MRB*. This is also described in an early study with patas monkeys (*Erythrocebus patas*; Hall, 1962). Furthermore, a century ago, Köhler reported a high motivation in chimpanzees to manipulate and interact with small mirrors (Köhler, 1926, pp. 329–331). However, Köhler did not describe frequencies of *MRB* or compare the chimpanzees’ interactions with small mirrors to those with large mirrors.

Another factor that is likely to have some impact refers to the actual context, i.e., whether the test is conducted with individuals being separated from their social group or within their social group. The former provides more controlled conditions in terms of reduced opportunities for distraction and guaranteed access to the mirror. However, the latter offers the advantage that participants can observe the *MRB* of other individuals as well as potentially others' reflections, which might facilitate learning about mirror usage (Gallup Jr., 1994; Gallup Jr. et al., 1980). Other, and in our view even more important, advantages of a social setting are that no stress is induced due to separations, and that the animals can explore a mirror like any other new enrichment object. Furthermore, when provided with several mirrors at a time, data of more than one individual can be collected simultaneously. Methodological challenges of group tests, such as non-independency of individuals within a group or potential, hierarchy-based bias in access to a mirror, can be met, for example, by appropriate statistical methods (e.g., mixed model approaches). Additionally, providing several mirrors can reduce competition (for a discussion of tests in social settings, see Cronin et al., 2017). Especially, but not exclusively, those studies applying the mark test (Gallup Jr., 1970) followed the single-individual/separation approach (e.g., Gallup Jr., 1970; Povinelli, 1989; Prior & Schwarz, 2008; Robert, 1986), while others chose the social setting (e.g., de Veer et al., 2003; Lin et al., 1992; Povinelli et al., 1993; Reiss & Marino, 2001; Westergaard & Hyatt, 1994). Some studies also used a combination of both settings, e.g., the social setting for an initial mirror exposure and the single-individual setting for the mark test (Vanhooland et al., 2020), or changed from a single-individual to a social setting when separations caused stress in an individual (Mahovetz et al., 2016).

We suggest that using small, portable mirrors provides a simple, but promising way to address methodological problems discussed for mirror tests with primates: (i) Individuals can actively reposition handheld mirrors while engaging in *SE* to maximize visibility of the target body part. This coordinated hand-eye action provides a clear criterion for differentiating *SE* from other self-directed behaviors. (ii) Presenting small instead of body-sized mirrors might be less likely to trigger aggressive mirror-directed

responses, which might prevent individuals from further interactions with the mirror because small mirrors do not reflect the entire body. (iii) Being portable and controllable objects, small mirrors might increase the interest in exploring and manipulating them. Furthermore, if a group of primates is provided with enough hand mirrors, each individual will be able to monopolize one of them. This reduces competition over possession and enables investigation at one's own pace and at the preferred location. Taken together, mirrors that are directly accessible, monopolizable and maneuverable provided in a social setting might facilitate spontaneous *MRB* (Gallup Jr., 1994). These advantages may increase self-motivated and self-reinforced learning of the contingencies of mirror use (Reiss & Morrison, 2017), which might increase the likelihood of recognizing the reflection as their own, expressed as *SE*.

In the present study, we compared the behavioral responses of socially housed chimpanzees ($N = 47$) towards a large, stationary mirror with those towards small hand mirrors. Based on the considerations outlined above, we hypothesized that small, portable and maneuverable mirrors are more likely to elicit *MRB* in general and *SE* in particular than large, stationary mirrors. Accordingly, we made the following predictions: (P1) Chimpanzees demonstrate a greater proportion of time engaging with hand mirrors as compared to large mirrors. (P2) More chimpanzees engage in *SE* and for a greater proportion of time with small mirrors as compared to large mirrors. We suggest that changing the methodology by using portable and controllable small mirrors instead of a stationary large mirror will provide more informative results in future studies on *MSR* abilities in primates and other species that are able to maneuver hand mirrors.

Method

Subjects

We conducted the mirror tests in seven groups of chimpanzees housed in six German zoos in Berlin, Halle, Heidelberg, Leipzig, Magdeburg and Wuppertal from July to December 2019, with group sizes ranging from two to 19 individuals. In total, forty-seven chimpanzees (29 females, 18 males; age range: 1.5-53

years, *Mean* = 29 years) participated in this study. All but four chimpanzees were older than 9 years when tested and thereby above the age in which *MSR* is expected to emerge (Povinelli et al., 1993). Of the four immature individuals, two were 1.5 years old and the others were 3.5 and 4.5 years old. Although we did not expect the youngest infants to demonstrate *SE*, we did not exclude them from the analyses because we were interested in all *MRB* that large and small mirrors elicited. (See Table S1 for details about the individuals.)

With regard to prior experience, we cannot claim that the individuals were truly naïve with respect to their specular image or reflective surfaces in general, especially given that 2/3 of the individuals had already been living in human care for more than 20 years when tested. Most of them probably have had access to a mirror or other highly reflective surfaces at some time in their lives, which is likely true for all zoo populations. In addition, all enclosures had at least one glass windowpane, which was slightly reflective. However, both test setups used in the present study were new to the chimpanzees.

Materials

For the *Large mirror* condition, we attached two plain, rectangular, rimless mirrors directly next to each other to the outside of an enclosure's glass pane using duct tape to create one body-sized mirror surface (minimum 100 x 80 cm, maximum 120 x 100 cm). For the *Small mirrors* condition, we distributed shatterproof hand mirrors throughout the enclosure. The hand mirrors consisted of a rectangular transparent pane of polycarbonate (12 x 8 x 0.6 cm; edges rounded), with mirror foil stuck to one side. In order to limit competition over possession, the number of mirrors equaled the number of individuals per group.

We recorded all test sessions on video using observer operated digital video cameras on tripods (Panasonic HC-V777, HDC-X929, SD9CJ and JVC Everio), and additionally a GoPro Hero 3 attached to a

window pane of the enclosure for some groups. The number of cameras used in a session depended on the group and enclosure size. Video files were saved in MP4/720p and MOD format, respectively.

Procedure

Data collection

We tested each group on four separate days for one hour per day. Tests took mainly place during the morning. On test days one and two, we applied the *Large mirror* condition, on test days three and four, the *Small mirrors* condition. Since the present study was part of a larger project that required a standardized procedure with the same test order for all individuals, we did not counterbalance the order of the conditions across groups. Moreover, our seven study groups varied considerably with respect to group size (ranging from 2 to 19 individuals) and age structure ($Mean_{age} \pm SD$ ranging from 19.7 ± 7.2 years to 46.7 ± 1.5 years), and all immatures were in one group (Table S1). Given that these variables were unbalanced on the group level, we did not consider varying the order of conditions across groups as a methodological improvement for this study. Hence, we accounted for two alternative potential biases that could arise in the results: First, increased familiarity with a mirror surface over time might increase the likelihood for *SE* to occur. Second, providing novel objects like portable small mirrors was expected to trigger mirror-related exploration and manipulation, especially in such a curious species as chimpanzees (Anderson & Roeder, 1989; Gallup Jr., 1994; Köhler, 1926; Paquette & Prescott, 1988; Westergaard & Fragaszy, 1985). This in turn might increase the likelihood to detect contingencies between the own body/face movements and the mirror image and thereby potentially facilitate *SE* (Gallup Jr., 1994; Reiss & Morrison, 2017). Considering that (i) all chimpanzees sampled had most probably some experience with their reflections, (ii) interest in a mirror was likely to decrease rather quickly (Anderson & Roeder, 1989; Povinelli et al., 1993), and (iii) most of the chimpanzees who revealed compelling *MSR* in the Povinelli et al. (1993) study performed *SE* within the first 20 min of mirror exposure, we expected a potential novelty effect of small mirrors having a greater impact on *MRB* than an increased familiarity with the reflection

over time. Therefore, in order to prevent a bias towards the predicted outcome, we presented the large mirror first. In our statistical analysis, we compared the chimpanzees' interest in interacting with a mirror on the one hand between conditions and on the other hand between sessions within conditions to account for both potential order effects.

For the *Large mirror* condition, we attached the mirror outside of a glass pane facing the enclosure – well visible from the chimpanzees' perspective and with enough space in front of it to allow several individuals at a time to access the mirror. For the *Small mirrors* condition, we distributed as many mirrors as individuals were in the respective group throughout the enclosure. Thus, all chimpanzees had potentially the opportunity to pick up a mirror.

A session started when the chimpanzees entered the enclosures and ended after one hour. In some cases when the chimpanzees did not exchange the mirrors for a reward after one hour, the mirrors remained inside the enclosure until the group went to a different enclosure (e.g., the outdoor enclosure) shortly afterwards. Each session was continuously video-recorded by one to four observers, depending on group size.

Data coding

We video-coded spontaneous *MRB* using the *all occurrences sampling* method (Altmann, 1974) and a coding scheme adapted from Povinelli et al. (1993) and implemented in BORIS vers. 7.9 (Friard & Gamba, 2016). Behaviors included, e.g., holding, manipulating or exploring the mirror, social responses to the reflection, watching the reflection, contingent movements and *SE* (see Table S2 for details). We coded the onset and offset for each *MRB*, so we could calculate both the frequencies of *MRB* and their respective duration. Furthermore, we coded the time at which an individual entered and left the range of a mirror where *MRB* was possible. An individual was considered as being *within range* when located within a

predefined three-dimensional area (2 x 5 x 2 m) in front of a large mirror (adapted from Povinelli et al., 1993) or within arm's reach of a non-occupied small mirror.

We classified only *SE* as a convincing criterion for *MSR* (Povinelli et al., 1993). Given the considerably shorter time of mirror exposure in our study (4 x 1 h) compared to previous studies (e.g., five days à 6-8 h, Povinelli et al., 1993), we did not differentiate between compelling and ambiguous *MSR* evidence depending on a minimum number and total duration of *SE* instances. Instead, we suggest that those individuals that demonstrated *SE* only once in this short time were likely to show compelling evidence of *MSR* with additional testing.

Following an intensive training phase, 30 min of video footage was coded by all four coders in order to assess inter-rater reliability. We used a function implemented in BORIS, which calculated the inter-rater agreement for *SE*, other *MRB* and the *location relative to the range of a mirror* based on a time interval comparison (time interval = 1 s). There was very good agreement among the raters (Cohen's kappa coefficient κ (Cohen, 1960) with unclear cases being considered: $Mean_{\kappa} \pm SD = 0.92 \pm 0.05$; with unclear cases being excluded: $Mean_{\kappa} \pm SD = 0.94 \pm 0.02$).

Data analysis

For data analysis, we used statistics software R vers.3.6.3 (R Core Team, 2020) with the additional packages *car*, *glmmTMB*, *exactRankTests*, *sfsmisc* and *ggplot2* (Brooks et al., 2017; Fox & Weisberg, 2019; Hothorn & Hornik, 2019; Maechler, 2020; Wickham, 2016). Due to skewed distributed data, we report medians (*Mdn*) and interquartile ranges (*IQR*) unless indicated otherwise.

We used a GLMM with Beta distribution, logit link function and zero-inflation (R function *glmmTMB*, Brooks et al., 2017) to investigate possible influences on the proportion of time spent on *MRB* in the total time of a session (conditional component) and on whether *MRB* is demonstrated at all (zero-inflation component). Into the model, we included *Condition (Large mirror/Small mirrors)* as the test

predictor and controlled for *Session* (first/second), *Age* and *Sex* (fixed effects) as well as *Individual* and *Group* (random effects). *Session* was included to control for possible novelty effects within a condition as interest in a presented mirror has been reported to decrease over time (Povinelli et al., 1993). (See Supplementary material for details)

For each group, we calculated the total time when no individual was located within range of the large mirror and the mirror was potentially accessible by all individuals. Thus, we controlled for the possibility that – especially in larger groups – the amount of time an individual spent interacting with the large mirror was determined by the opportunity to access the mirror rather than by interest in *MRB*.

Due to the small number of observations, a respective model with the proportion of *SE* as response did not converge. Alternatively, we conducted exact Wilcoxon signed-rank tests to test for the differences between the conditions for both the number of *SE* instances and the proportional duration of *SE* behavior and calculated Pearson's correlation coefficient *r* for effect sizes (Cohen, 1992).

Statistical significance was assessed at the α -level of 0.05.

Ethical considerations

This study was approved by the joint ethical committee of the Max Planck Institute for Evolutionary Anthropology and Leipzig Zoo, and by the curators and head keepers at the participating zoos in Berlin, Halle, Heidelberg, Magdeburg and Wuppertal. Animal husbandry and research comply with the EAZA Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria and the WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums. Interactions with the mirrors were voluntary. The chimpanzees were never deprived of water, nor separated from their group due to our study. In general, there were no alterations to the daily feeding schedule.

Results

Ad (P1): More frequent and longer interactions with small rather than large mirrors

Confirming our prediction, chimpanzees spent more time interacting with small mirrors than with the large mirror, with interaction times varying considerably across individuals (Table 1, Figure 1A).

Table 1 Summary of descriptive statistics on MRB and SE (per individual) for both test conditions.

	Large Mirror Condition				Small Mirror Condition			
	Absolute values (min)		Proportions ⁽¹⁾		Absolute values (min)		Proportions ⁽¹⁾	
	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>	<i>Mdn</i>	<i>IQR</i>
Time spent on MRB	0.66	2.46	0.005	0.021	8.06	30.80	0.067	0.255
Time spent on SE ⁽²⁾	0.00	0.00	0.000	0.000	0.00	0.42	0.000	0.004
Number of SE instances ⁽²⁾	0.00	0.00	---	---	0.00	1.50	---	---
	Average of absolute values		Range of absolute values		Average of absolute values		Range of absolute values	
	<i>Mean</i>	<i>SD</i>	Min.	Max.	<i>Mean</i>	<i>SD</i>	Min.	Max.
Time spent on SE (min)	0.36	1.03	0.00	5.56	1.43	3.88	0.00	21.70
Number of SE instances	1.13	2.74	0	12	3.28	8.71	0	53

⁽¹⁾ Proportion in test time per condition

⁽²⁾ In total, 20 individuals out of 47 engaged in SE, which explains a median being 0 for both conditions. To provide more information, the respective means and ranges are additionally reported.

Overall, the full model was clearly significant as compared to the null model (likelihood ratio test, LRT: $\chi^2 = 24.77$, $df = 2$, $P < 0.001$). Specifically, *Condition* had a significant effect with the proportion of time spent interacting with small mirrors being greater than with large mirrors (LRT: $\chi^2 = 17.99$, $df = 1$, $P < .001$). Also performing MRB was more likely with small rather than with large mirrors (zero-inflation component: $z = -2.56$, $P = .011$). Furthermore, the effect of *Session* was significant, with a greater proportion of MRB in the respective first sessions compared to the second sessions (LTR: $\chi^2 = 7.33$, $df = 1$, $P = .007$, Table 2, Figure 1A)

Table 2 The impact of large vs. small mirrors on chimpanzee response. Full model results (zero-inflated Beta GLMM) for fixed effects⁽¹⁾.

Conditional Model						
Response: Proportion of time spent on MRB						
Effect	Estimate	Standard Error	CI _{lower}	CI _{upper}	χ^2	P
Intercept	-1.49	0.38	-2.23	-1.38	(2)	(2)
Condition ⁽³⁾	0.93	0.24	0.45	1.41	17.99	<0.001
Session ⁽⁴⁾	-0.52	0.19	-0.90	-0.15	7.33	0.007
Sex ⁽³⁾	-0.12	0.21	-0.54	0.29	0.33	0.568
Age ⁽⁵⁾	-0.05	0.13	-0.32	0.21	0.14	0.705
Zero-inflation Model						
Test for impact on likelihood for response being zero						
Effect	Estimate	Standard Error	CI _{lower}	CI _{upper}	z value	P
Intercept	-0.60	0.56	-1.69	0.50	(2)	(2)
Condition ⁽³⁾	-0.86	0.34	-1.52	-0.20	-2.56	0.011
Session ⁽⁴⁾	0.11	0.33	-0.54	0.76	0.33	0.742
Sex ⁽³⁾	-0.25	0.37	-0.97	0.48	-0.67	0.505
Age ⁽⁵⁾	-0.22	0.18	-0.57	0.13	-1.24	0.216

⁽¹⁾ Estimated coefficients, standard error, confidence intervals (lower CI at 2.5%, upper CI at 95.5%), results of LRT (χ^2 , P) and test statistics (z value and related P)
⁽²⁾ not shown because of having a very limited interpretation
⁽³⁾ *Condition* and *Sex* were dummy coded with *Large mirror* and *Female*, respectively, as the reference category
⁽⁴⁾ *Session*: continuous variable with first Session=1, second Session=2
⁽⁵⁾ *Age* was z-transformed. *Mean* \pm *SD* of the original values: 347.28 \pm 171.54 months of age (i.e. 29 \pm 14 years of age)

There was no overt competition over access to either the large or the small mirrors. In particular, the area in front of the large mirror was unoccupied on average 73% of the total test time across groups (*Mean* \pm *SD* = 0.73 \pm 0.14). While seven males and two females (39% of males, 7% of females) performed dominance displays and/or brief instances of aggression (charge, hit, kick) towards the large mirror, no individual performed such behaviors towards the small mirrors.

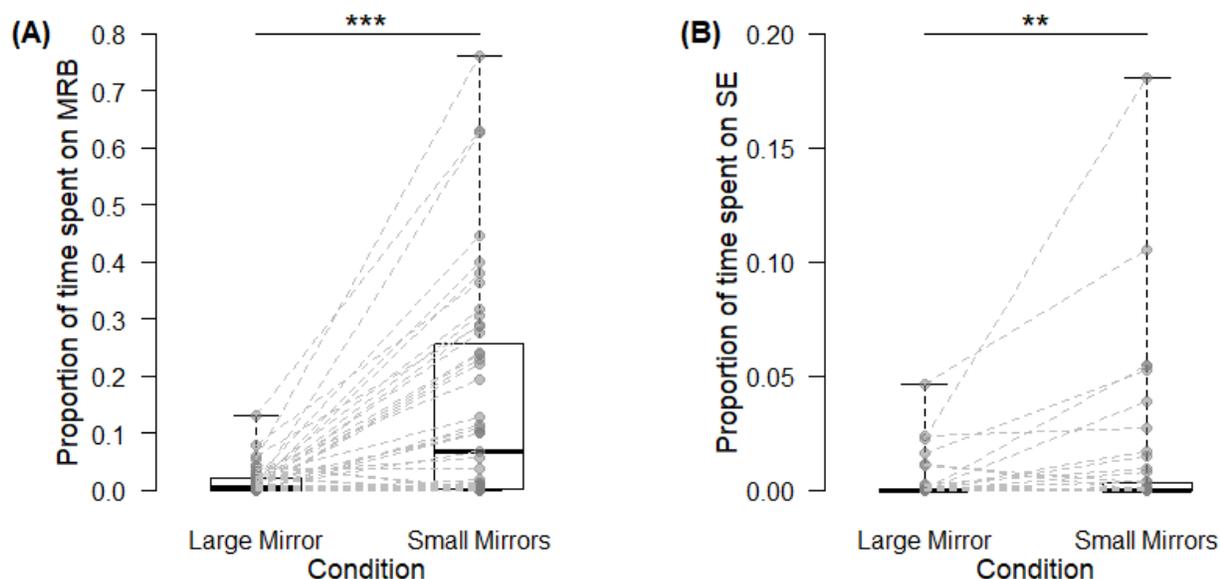


Figure 1 Comparison of the proportion of time individuals spent on MRB in general (A) and SE (B) per condition. Boxplots: horizontal lines=medians, boxes=IQR, whiskers=minima and maxima; Points: individuals, dashed lines connect data points for the same individual. Asterisks indicate statistical significance (***) $P < .001$; ** $P < .01$)

Despite the short time (4 x 1h), we observed the whole spectrum of *MRB* defined in the ethogram (See Table S2, also SV1 and SV2 for video examples). Small mirrors were often held in changing angles and/or moved close to or around the face/body while intensely looking at the reflection, with or without accompanying contingent facial/body movements or *SE* of, e.g., teeth and tongue (Figure 2, SV1 and SV2).



Figure 2 Examples of interactions with a large mirror (A) and with a small mirror (B). (For video examples of the different types of mirror-related behavior, see SV1 and SV2.)

Ad (P2): More individuals engaged in *SE* with small rather than large mirrors

As predicted, chimpanzees engaged in *SE* with small mirrors both significantly (a) more often and (b) for a longer time compared to the large mirror (Wilcoxon exact: (a) $T^* = 30$, $N = 19$ (28 ties), $P = .007$; effect size: $r = .60$; (b) $T^* = 33$, $N = 20$ (27 ties), $P = .005$; effect size: $r = .60$; Table 1, Figure 1B).

In total, we observed 207 instances of *SE* by 20 out of 47 individuals (43%) across all sessions with varying number and duration of *SE* (range of per capita number: 1-65 instances of *SE*, $Mdn = 5$, $IQR = 11.25$; range of *SE* duration: 1-444 seconds, $Mdn = 10$ seconds, $IQR = 22$ seconds). *SE* mainly consisted of exploration of face and mouth (90% of all *SE* instances).

The number of individuals demonstrating *SE* differed between the two conditions ($N_{Large} = 11$, $N_{Small} = 17$). Eight individuals showed *SE* behavior in both conditions, three showed *SE* only in the *Large mirror* condition and nine only in the *Small mirrors* condition. Using differences in behavior expressed between conditions as a measure of false negatives, the *Small mirrors* condition thus produced a lower minimum number of false negatives ($N_{MinFalseNeg} = 3$, 6.4%) than the *Large mirror* condition ($N_{MinFalseNeg} = 9$, 19.1%).

Discussion

In the current study, we compared the traditional *MSR* test setup using a stationary, body-sized mirror with an alternative setup using several portable, small hand mirrors. Corroborating our predictions, chimpanzees interacted more with the small mirrors than with the large mirror, reflected by a higher proportion of time spent with mirror-related behaviors and self-exploration. We found 12.8% more individuals engaging in *SE* with the small mirrors than with the large mirrors in the short period of only two hours for each condition. Taken together, our results support the hypothesis that this method provides a more sensitive measure to detect *MSR* ability in chimpanzees and potentially other species capable of maneuvering a mirror than the traditional approach. Including small, portable mirrors in standardized comparative *MSR* tests will likely help to verify and potentially re-assess variation in *MSR* previously found across and within species (de Veer & van den Bos, 1999; Gallup Jr. & Anderson, 2019).

As predicted in (P1) and corroborating early scarce descriptions of the behavior of capuchins and chimpanzees (Anderson & Roeder, 1989; Köhler, 1926), our study revealed a substantially higher tendency of chimpanzees to interact with maneuverable hand mirrors compared to a stationary body-sized mirror. Being provided with unfamiliar, portable objects increases exploration and manipulation behavior in captive primates (Paquette & Prescott, 1988; Westergaard & Fragaszy, 1985). Carrying a mirror away from other chimpanzees to a place of their choice might facilitate intense mirror exploration with limited social distraction. This opportunity might be reduced with a large, stationary mirror either because dominant individuals can potentially monopolize it or because it is harder to position one's body to observe difficult-to-see body parts in a large mirror than to maneuver the handheld mirror to observe stationary body parts. Related to the former, one might argue that the difference found was a result of having less opportunity to get access to the large stationary mirror, especially for bigger social groups, than to one of several small mirrors. However, our observations do not support this explanation. While small mirrors were frequently picked up, carried around and explored, the area in front of the large mirror remained

completely unoccupied on average 73% of the test time across groups. This suggests that the observed difference is much more likely to be due to a greater interest in portable and maneuverable reflective objects compared to a large stationary reflective surface, rather than to limited access to the large mirror because of monopolization by particular individuals. Additionally, stationary mirrors force individuals to spend time in a predefined area to engage with them, something not all individuals are keen to do. Mobile mirrors allow for taking them to preferred spots and engaging with them at one's own pace.

Furthermore, the different size of the two mirror types in relation to the body size of a chimpanzee might have had an influence on their *MRB*. Unlike large mirrors, small mirrors do not reflect the whole body, thus, the social stimulus properties of mirror-images (Anderson & Roeder, 1989) should be less strong and therefore, less likely to trigger aggressive or fearful responses, which might constrain mirror-related interactions including *SE* (de Veer & van den Bos, 1999). Our finding that aggressive responses were directed only towards large mirrors, predominantly by adult males, supports this interpretation. This is important to consider when examining *MSR* abilities in species with aggressive tendencies towards rivals or potential mates.

In accordance with our prediction (P2), we observed more individuals engaging in *SE* and for a greater proportion of time with the hand mirrors compared to the large mirrors. The chimpanzees were exploring and interacting more frequently and extensively with the small mirrors compared to the large ones. The small mirrors were directly accessible and could actively be manipulated and moved relative to one's own body. In doing so, chimpanzees had potentially more and varying opportunities to investigate the reflective properties of the small mirrors and to detect contingencies between their own or other's movements and the reflection compared to large mirrors (Gallup Jr., 1994; Reiss & Morrison, 2017). This in turn might have resulted in more spontaneous *SE* in the *Small mirrors* than in the *Large mirror* condition, an indicator of *MSR* in great apes (Anderson & Gallup Jr., 2015; Gallup Jr., 1970, p. 1970; Lethmate & Dürcker, 1973; Povinelli et al., 1993). The objective of our study, however, was not to assess *MSR* abilities

in chimpanzees, which has been done elsewhere (e.g., de Veer et al., 2003; Gallup Jr., 1970; Lin et al., 1992; Mahovetz et al., 2016; Povinelli et al., 1993). Hence, we did not include control conditions to determine *MSR* abilities, such as presenting mirror sized object without a reflective surface. Our aim was to compare the effects of the *Large mirror* condition and the *Small mirror* condition on mirror-related behaviors and thereby their suitability for *MSR* studies particularly in primates.

In that respect, we suggest that maneuverable hand mirrors offer a better tool to identify spontaneous *SE*. When chimpanzees used the small mirrors for *SE*, they did not only touch otherwise not or barely visible body parts (the classic criterion for *SE*), but held the mirror, moved and adjusted its position, presumably to see the target body part. These directed mirror movements in combination with manual exploration or actively opening of the mouth while looking at the reflection and following the mirror with the gaze, are less ambiguous criteria to distinguish *SE* from not-mirror-guided self-directed behavior (Heyes, 1994). Furthermore, the maneuverable small mirrors themselves had a highly reflective and a non-reflective side. This offered a possibility to control whether self-directed behavior was mirror-guided or not by analyzing instances of self-directed behavior while looking at one side or the other.

Given the short period of total mirror exposure in our study compared to other studies (e.g., Povinelli et al., 1993), it is notable that we observed spontaneous *SE* in 42.6% of the chimpanzees (with four immatures too young to show *SE* included in sample). However, we would not claim that we demonstrated compelling evidence for *MSR* in all these individuals, especially because some of them engaged in *SE* only once or twice and we did not include explicit control conditions. The point we want to make here is that those individuals that showed *SE* only once and especially with a small mirror are likely to demonstrate compelling evidence for *MSR* when studied with hand mirrors for a longer period of time.

Our results revealed an increased interest in the mirrors in the second compared to the first condition, i.e., in the *Small mirrors* condition compared to the *Large mirror* condition, but a drop of

interest in interacting with a mirror in the second session compared to the first session within conditions. These findings are in line with our expectations and previous reports of novelty effects, resulting in higher interest in interacting with novel objects and decreasing interest in the mirror over time in chimpanzees and other primates (Anderson & Roeder, 1989; Gallup Jr., 1994; Povinelli et al., 1993). In contrast, they do not suggest a substantial impact of a possible familiarity effect. We therefore argue that the greater number of individuals demonstrating *SE* and the higher proportion of time spent with *SE* in the *Small mirrors* condition compared to the *Large mirror* condition is more likely due to the substantially higher amount of interactions with the mirror – and thereby increased opportunity to learn about the mirror – than due to familiarity with the reflection accumulated during the presentation of the large mirror. However, future studies on *MSR* abilities need to consider these potential influences and include appropriate control measures.

Finally, there are practical advantages of using small hand mirrors. The method is easy to apply in both simple mirror tests and the mark test (Gallup Jr., 1970). It is applicable in a social setting in the usual enclosure and offers enrichment opportunities (Cronin et al., 2017). It is suitable for tests in various primate species and potentially in other species capable of maneuvering a small mirror, as the size and weight of the mirrors can be adapted to species-specific characteristics.

However, we are aware that the method proposed here is not generally applicable across different taxa, because being able to hold and maneuver the mirror – either by hand, trunk or other means – is a predisposition, which is not or not easily fulfilled in many species of interest, e.g., in marine mammals, canines, ungulates without a trunk or birds (Vonk, 2020). While acknowledging this limitation, we suggest that the proposed method has the potential to inform comparative research and to draw more representative data of the capacity to recognize oneself in the mirror than large-mirror setups, not only in chimpanzees, but across the primate order and potentially other species able to maneuver a hand

mirror. Adequate cross-species methods are needed to aid our understanding of the evolutionary origins of human self-awareness.

Acknowledgements

We thank all the keepers, curators and directors at the participating zoos for their extraordinary support for this study, especially André Schüle, Christian Aust and Ruben Gralki at Zoo Berlin, Karoline Albig, Sven Zeidler and Kerstin Gribbe at Zoo Halle, Sandra Reichler, Anke Fuchs and Dominik Winkel at Zoo Heidelberg, Daniel Geissler, Stefan Leideritz, Nico Schenk, Evelyn Kanzler and René Berger at Zoo Leipzig, Kai Perret, Susann Paelecke and Julia Schrage at Zoo Magdeburg, Severin Dressen, Matthias Schmitz and Julian Kusak at Zoo Wuppertal as well as Daniel Hanus and Hanna Petschauer at the Wolfgang Köhler Primate Research Center. Many thanks go to Isabella O’Neal, Pieter Nyssen and Peter Niehoff for data collection and coding, to Raik Pieszek for manufacturing the small mirrors, to Cédric Girard-Buttoz for statistical advice and to all colleagues within the *Evolution of Brain Connectivity* project for valuable input and discussion.

Funding

This work was funded by the Max Planck Society (Presidential funds to the Research Unit ‘Evolution of Brain Connectivity’).

Conflict of Interests

We declare we have no competing interests.

Authors’ contributions

All authors contributed to the concept and the design of the study. KK and SE coordinated and participated in the data collection. KK carried out the statistical analyses. All authors contributed to the interpretation

of the data. KK and SE drafted the manuscript. All authors critically revised the manuscript and gave final approval for publication.

References

- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, 49(3/4), 227–267. <https://doi.org/10.1163/156853974X00534>
- Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology*, 5(4), 297–305. <https://doi.org/10.1002/dev.420050403>
- Anderson, J. R. (1984). Monkeys with mirrors: Some questions for primate psychology. *International Journal of Primatology*, 5(1), 81–98. <https://doi.org/10.1007/BF02735149>
- Anderson, J. R., & Gallup Jr., G. G. (2015). Mirror self-recognition: A review and critique of attempts to promote and engineer self-recognition in primates. *Primates*, 56(4), 317–326. <https://doi.org/10.1007/s10329-015-0488-9>
- Anderson, J. R., & Roeder, J.-J. (1989). Responses of capuchin monkeys (*Cebus apella*) to different conditions of mirror-image stimulation. *Primates*, 30(4), 581–587. <https://doi.org/10.1007/BF02380884>
- Asendorpf, J. B., Warkentin, V., & Baudonnière, P.-M. (1996). Self-awareness and other-awareness. II: Mirror self-recognition, social contingency awareness, and synchronic imitation. *Developmental Psychology*, 32(2), 313. <https://doi.org/10.1037/0012-1649.32.2.313>
- Baragli, P., Demuru, E., Scopa, C., & Palagi, E. (2017). Are horses capable of mirror self-recognition? A pilot study. *PLOS ONE*, 12(5), e0176717. <https://doi.org/10.1371/journal.pone.0176717>
- Bard, K. A., Todd, B. K., Bernier, C., Love, J., & Leavens, D. A. (2006). Self-Awareness in human and chimpanzee infants: What is measured and what is meant by the mark and mirror test? *Infancy*, 9(2), 191–219. https://doi.org/10.1207/s15327078in0902_6

- Broesch, T., Callaghan, T., Henrich, J., Murphy, C., & Rochat, P. (2011). Cultural variations in children's mirror self-recognition. *Journal of Cross-Cultural Psychology*, 42(6), 1018–1029. <https://doi.org/10.1177/0022022110381114>
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). GlimmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378–400. <https://doi.org/10.3929/ethz-b-000240890>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J. (1992). Statistical Power Analysis. *Current Directions in Psychological Science*, 1(3), 98–101. <https://doi.org/10.1111/1467-8721.ep10768783>
- Cronin, K. A., Jacobson, S. L., Bonnie, K. E., & Hopper, L. M. (2017). Studying primate cognition in a social setting to improve validity and welfare: A literature review highlighting successful approaches. *PeerJ*, 5, e3649. <https://doi.org/10.7717/peerj.3649>
- de Veer, M. W., Gallup Jr., G. G., Theall, L. A., van den Bos, R., & Povinelli, D. J. (2003). An 8-year longitudinal study of mirror self-recognition in chimpanzees (*Pan troglodytes*). *Neuropsychologia*, 41(2), 229–234. [https://doi.org/10.1016/S0028-3932\(02\)00153-7](https://doi.org/10.1016/S0028-3932(02)00153-7)
- de Veer, M. W., & van den Bos, R. (1999). A critical review of methodology and interpretation of mirror self-recognition research in nonhuman primates. *Animal Behaviour*, 58(3), 459–468. <https://doi.org/10.1006/anbe.1999.1166>
- de Waal, F. B. M. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *PLOS Biology*, 17(2), e3000112. <https://doi.org/10.1371/journal.pbio.3000112>
- Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (3rd ed.). Sage. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>

- Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, 7(11), 1325–1330.
<https://doi.org/10.1111/2041-210X.12584>
- Gallup Jr., G. G. (1968). Mirror-image stimulation. *Psychological Bulletin*, 70(6, Pt.1), 782–793.
<https://doi.org/10.1037/h0026777>
- Gallup Jr., G. G. (1970). Chimpanzees: Self-Recognition. *Science*, 167(3914), 86–87.
<https://doi.org/10.1126/science.167.3914.86>
- Gallup Jr., G. G. (1994). Self-recognition: Research strategies and experimental design. In M. L. Boccia, R. W. Mitchell, & S. T. Parker (Eds.), *Self-Awareness in Animals and Humans: Developmental Perspectives* (pp. 35–50). Cambridge University Press; Cambridge Core.
<https://doi.org/10.1017/CBO9780511565526.005>
- Gallup Jr., G. G., & Anderson, J. R. (2019). Self-recognition in animals: Where do we stand 50 years later? Lessons from cleaner wrasse and other species. *Psychology of Consciousness: Theory, Research, and Practice*. <https://doi.org/10.1037/cns0000206>
- Gallup Jr., 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(6), 1525–1532.
[https://doi.org/10.1016/0003-3472\(95\)80008-5](https://doi.org/10.1016/0003-3472(95)80008-5)
- Gallup Jr., G. G., Wallnau, L. B., & Suarez, S. D. (1980). Failure to find self-recognition in mother-infant and infant-infant rhesus monkey pairs. *Folia Primatologica*, 33(3), 210–219.
<https://doi.org/10.1159/000155935>
- Hall, K. R. L. (1962). Behaviour of monkeys towards mirror-images. *Nature*, 196(4861), 1258–1261.
<https://doi.org/10.1038/1961258a0>
- Heyes, C. M. (1994). Reflections on self-recognition in primates. *Animal Behaviour*, 47(4), 909–919.
<https://doi.org/10.1006/anbe.1994.1123>

- Heyes, C. M. (1995). Self-recognition in primates: Further reflections create a hall of mirrors. *Animal Behaviour*, 50(6), 1533–1542. [https://doi.org/10.1016/0003-3472\(95\)80009-3](https://doi.org/10.1016/0003-3472(95)80009-3)
- Hothorn, T., & Hornik, K. (2019). *exactRankTests: Exact Distributions for Rank and Permutation Tests* (R package version 0.8-31) [Computer software]. <https://CRAN.R-project.org/package=exactRankTests>
- Kärtner, J., Keller, H., Chaudhary, N., & Yovsi, R. D. (2012). The Development of mirror self-recognition in different sociocultural contexts. *Monographs of the Society for Research in Child Development*, 77(4), vii–viii, 1–87. JSTOR.
- Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., & Papaligoura, Z. (2004). Developmental consequences of early parenting experiences: Self-recognition and self regulation in three cultural communities. *Child Development*, 75(6), 1745–1760. <https://doi.org/10.1111/j.1467-8624.2004.00814.x>
- Kohda, M., Hotta, T., Takeyama, T., Awata, S., Tanaka, H., Asai, J., & Jordan, A. L. (2019). If a fish can pass the mark test, what are the implications for consciousness and self-awareness testing in animals? *PLOS Biology*, 17(2), e3000021. <https://doi.org/10.1371/journal.pbio.3000021>
- Köhler, W. (1926). *The Mentality of Apes*. Harcourt, Brace & Company, Inc.
- Krachun, C., Lurz, R., Mahovetz, L. M., & Hopkins, W. D. (2019). Mirror self-recognition and its relationship to social cognition in chimpanzees. *Animal Cognition*, 22(6), 1171–1183. <https://doi.org/10.1007/s10071-019-01309-7>
- Lethmate, J., & Dücker, G. (1973). Untersuchungen zum Selbsterkennen im Spiegel bei Orang-Utans und einigen anderen Affenarten. *Zeitschrift Für Tierpsychologie*, 33(3–4), 248–269. <https://doi.org/10.1111/j.1439-0310.1973.tb02094.x>
- Lin, A. C., Bard, K. A., & Anderson, J. R. (1992). Development of self-recognition in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, 106(2), 120–127.

- Maechler, M. (2020). *sfsmisc: Utilities from 'Seminar fuer Statistik' ETH Zurich* (R package version 1.1-6) [Computer software]. <https://CRAN.R-project.org/package=sfsmisc>
- Mahovetz, L. M., Young, L. J., & Hopkins, W. D. (2016). The influence of AVPR1A genotype on individual differences in behaviors during a mirror self-recognition task in chimpanzees (*Pan troglodytes*). *Genes, Brain and Behavior*, *15*(5), 445–452. <https://doi.org/10.1111/gbb.12291>
- Paquette, D., & Prescott, J. (1988). Use of novel objects to enhance environments of captive chimpanzees. *Zoo Biology*, *7*(1), 15–23. <https://doi.org/10.1002/zoo.1430070103>
- Plotnik, J. M., de Waal, F. B. M., & Reiss, D. (2006). Self-recognition in an Asian elephant. *Proceedings of the National Academy of Sciences*, *103*(45), 17053–17057. <https://doi.org/10.1073/pnas.0608062103>
- 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. *Journal of Comparative Psychology*, *103*(2), 122. <https://doi.org/10.1037/0735-7036.103.2.122>
- 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. *Journal of Comparative Psychology*, *107*(4), 347–372.
- Priel, B., & de Schonen, S. (1986). Self-recognition: A study of a population without mirrors. *Journal of Experimental Child Psychology*, *41*(2), 237–250. [https://doi.org/10.1016/0022-0965\(86\)90038-X](https://doi.org/10.1016/0022-0965(86)90038-X)
- Prior, H., & Schwarz, A. (2008). Mirror-Induced behavior in the magpie (*Pica pica*): Evidence of self-recognition. *PLoS Biology*, *6*(8), 9.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

- Reiss, D., & Marino, L. (2001). Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences*, 98(10), 5937–5942. <https://doi.org/10.1073/pnas.101086398>
- Reiss, D., & Morrison, R. (2017). Reflecting on mirror self-recognition: A comparative view. In *APA handbook of comparative psychology: Perception, learning, and cognition, Vol. 2* (pp. 745–763). American Psychological Association. <https://doi.org/10.1037/0000012-033>
- Robert, S. (1986). Ontogeny of mirror behavior in two species of great apes. *American Journal of Primatology*, 10(2), 109–117. <https://doi.org/10.1002/ajp.1350100202>
- Rochat, P., Broesch, T., & Jayne, K. (2012). Social awareness and early self-recognition. *Consciousness and Cognition*, 21(3), 1491–1497. <https://doi.org/10.1016/j.concog.2012.04.007>
- Schilhab, T. S. S. (2004). What mirror self-recognition in nonhumans can tell us about aspects of self. *Biology & Philosophy*, 19(1), 111–126. <https://doi.org/10.1023/B:BIPH.0000013249.23475.1f>
- Suarez, S. D., & Gallup, G. G. (1981). Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of Human Evolution*, 10(2), 175–188. [https://doi.org/10.1016/S0047-2484\(81\)80016-4](https://doi.org/10.1016/S0047-2484(81)80016-4)
- Suddendorf, T., & Butler, D. L. (2013). The nature of visual self-recognition. *Trends in Cognitive Sciences*, 17(3), 121–127.
- Suddendorf, T., & Collier-Baker, E. (2009). The evolution of primate visual self-recognition: Evidence of absence in lesser apes. *Proceedings of the Royal Society B: Biological Sciences*, 276(1662), 1671–1677. <https://doi.org/10.1098/rspb.2008.1754>
- Swartz, K. B., & Evans, S. (1994). Social and cognitive factors in chimpanzee and gorilla mirror behavior and self-recognition. In *Self-awareness in animals and humans: Developmental perspectives* (pp. 189–206). Cambridge University Press. <https://doi.org/10.1017/CBO9780511565526.013>
- Vanhooland, L.-C., Bugnyar, T., & Massen, J. J. (2020). Crows (*Corvus corone ssp.*) check contingency in a mirror yet fail the mirror-mark test. *Journal of Comparative Psychology*, 134(2), 158–169.

- Vonk, J. (2020). A fish eye view of the mirror test. *Learning & Behavior*, 48, 193–194.
<https://doi.org/10.3758/s13420-019-00385-6>
- Walraven, V., van Elsacker, L., & Verheyen, R. (1995). Reactions of a group of pygmy chimpanzees (*Pan paniscus*) to their mirror-images: Evidence of self-recognition. *Primates*, 36(1), 145–150.
<https://doi.org/10.1007/BF02381922>
- Westergaard, G. C., & Fragaszy, D. M. (1985). Effects of manipulatable objects on the activity of captive capuchin monkeys (*Cebus apella*). *Zoo Biology*, 4(4), 317–327.
<https://doi.org/10.1002/zoo.1430040402>
- Westergaard, G. C., & Hyatt, C. W. (1994). The responses of bonobos (*Pan paniscus*) to their mirror images: Evidence of self recognition. *Human Evolution*, 9(4), 273–279.
<https://doi.org/10.1007/BF02435514>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.