

Table S1. Details of individuals included in the study: sex, age, zoo/group.

Individual	Sex	Age (in years)	Average age in years per group (Mean ± SD)	Zoo/Group
Gusta	female	40	36.8 ± 4.0	Berlin
Pedro	male	39		
Lilly	female	38		
Karel	male	37		
Soko	female	30		
Bangi	female	45	44.5 ± 0.7	Halle
Sebastian	male	44		
Heidi	female	48	46.7 ± 1.5	Heidelberg
Susi	female	47		
Lulu	female	45		
Robert	male	43	23.4 ± 14.8	Leipzig A
Fraukje	female	43		
Corrie	female	42		
Riet	female	41		
Natascha	female	39		
Dorien	female	38		
Maja	female	33		
Sandra	female	26		
Frodo	male	25		
Swela	female	23		
Bambari	female	18		
Lome	male	17		
Tai	female	16		
Kisha	female	15		
Lobo	male	15		
Azibo	male	4.5		
Ohini	male	3.5		
Makeni	male	1.5		
Youma	female	1.5		
Jeudi	female	53		
Frederike	female	45		
Daza	female	33		
Hope	female	28		
Zira	female	22		
Alex	male	18		
Nana	female	33	19.7 ± 7.2	Magdeburg
Mumin	female	29		
Kananga	male	23		
Ubangi	male	21		
Sambala	female	20		
Mufasa	male	19		
Minga	female	14		
Sokoto	male	14		
Kofi	male	14		
Bangolo	male	10		
Epulu	male	51	43.5 ± 10.6	Wuppertal
Kitoto	female	36		

Table S2. Ethogram and operational definitions implemented in the present coding scheme, adapted from Povinelli et al. (1993). (See SV1 and SV2 for video examples)

Behavior	Operational Definition	Evidence for MSR?
Hold or carry mirror	Individual holds the mirror in the hand or carries it without engaging in any other mirror-related behaviors	No
Object manipulation	Individual manipulates the mirror as an object (not related to the reflective surface), including object play and use as a tool	
Mirror exploration	Individual visually or tactily explores the mirror, including brief looks at the mirror whether at the reflection or not (<5s)	
Look or grab behind	Individual attempts to look over or behind the mirror or reaches and/or grabs behind a mirror while looking at the specular reflection	
Social response	Individual responses towards the own specular reflection as to a conspecific, including aggressive (e.g. display, charge, hit, kick) and affiliative behavior (e.g. present genitals, play invitation, begging gesture, kiss)	
Watch reflection	Individual intensely looks at the own specular reflection without contingent movements or self-exploration	Not considered compelling evidence in this study
Contingent movements	Individual engages in contingent body movements or facial contortions while observing the specular reflection, including repetition of one movement type or sequences, open mouth, tongue wiggling etc., but without mirror-guided self-directed behavior/exploration	
Self-exploration	Individual manipulates/touches body parts otherwise not or barely visible, e.g. facial areas, oral cavity, anal-genital region or back with fingers while looking into the mirror; or individual holds a small mirror in a way that allows visual inspection of otherwise not or barely visible areas, e.g. oral cavity, back, ano-genital region, while looking into the mirror	Yes

Uncorrected

#### Methods: Additional information on data analyses

To test whether the proportion of time spent with mirror-related behaviors in the total duration of a session was influenced by *Condition* (large mirror vs. small mirrors), we ran a Generalized Linear Mixed Model with Beta distribution, logit link function and zero-inflation (R function `glmmTMB`, Brooks et al., 2017) with *Condition* as a fixed effect. We decided on a zero-inflation model because of many zero values (54 out of 188). The zero-inflation component of the model tests if the likelihood for showing no MRB at all (=zero) is influenced by variables included in the model. We further included *Session* (i.e. session number per condition: 1, 2) to control for a potential decrease of motivation to interact with a mirror over time, and also *Age* and *Sex* as further fixed effects to control for their potential influences. Additionally, we included *Individual* and *Group* as random effects to account for the possibility that motivation to interact with a mirror varied across chimpanzees and for group specific characteristics potentially influencing the mirror-related behaviors, respectively. Due to convergence issues, it was not possible to include random slopes into the model.

Age was z-transformed to a mean of zero and a standard deviation of one. The model was fitted in R (R Core Team, 2020) using the function `glmmTMB` of the package `glmmTMB` (Brooks et al., 2017). We checked for model stability by excluding individuals and groups, respectively, one at a time from the data and comparing the model estimates derived from these data with those derived from the full data set. This indicated no individuals or groups as influential.

As an overall test, we compared the full model with a null model that included all elements but the test predictor using a likelihood ratio test (Dobson & Barnett, 2018). We used likelihood ratio tests (R function `drop1`, 'test'="Chisq") to test for the influence of the particular variables (Barr et al., 2013). Statistical significance was assessed at the  $\alpha$ -level of 0.05. Confidence intervals were calculated using the R function `confint`.

References:

- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278.  
<https://doi.org/10.1016/j.jml.2012.11.001>
- 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). GlmmTMB 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>
- Dobson, A. J., & Barnett, A. G. (2018). *An introduction to generalized linear models*. CRC press.
- Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (3rd ed.). Sage.  
<https://socialsciences.mcmaster.ca/jfox/Books/Companion/>
- 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.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>