Title: Chimpanzees consider humans´ psychological states when drawing statistical inferences

Johanna Eckert*[1,2,3], Hannes Rakoczy[2,3] Josep Call[1,4], Esther Herrmann[1], Daniel Hanus[1]

[1] Department of Developmental and Comparative Psychology, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

[2] Department of Developmental Psychology, University of Göttingen, Waldweg 26, 37073 Göttingen, Germany

[3] Leibniz ScienceCampus ”Primate Cognition”, German Primate Center/Leibniz Institute for Primate Research, Kellnerweg 4, 37077 Göttingen, Germany

[4] School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK

*Lead Contact: eckert.johanna@gmail.com

Summary

Great apes have been shown to be intuitive statisticians: they can use proportional information within a population to make intuitive probability judgments about randomly drawn samples [1,2]. Humans, from early infancy onwards, functionally integrate intuitive statistics with other cognitive domains to judge the randomness of an event [3-7]. To date, nothing is known about such cross-domain integration in any nonhuman animal, leaving uncertainty about the origins of human statistical abilities. We investigated whether chimpanzees take into account information about psychological states of experimenters (their biases and visual access) when drawing statistical inferences. We tested 21 sanctuary-living chimpanzees in a previously established paradigm that required subjects to infer which of two mixed populations of preferred and non-preferred food items was more likely to lead to a desired outcome for the subject. In a series of three experiments we found that chimpanzees chose based on proportional information alone when they had no information about experimenters´ preferences and (to a lesser extent) when experimenters had biases for certain food types but drew blindly. By contrast, when biased experimenters had visual access, subjects ignored statistical information and instead chose based on experimenters´ biases. Lastly, chimpanzees intuitively used a violation of statistical likelihoods as indication for biased sampling. Our results suggest that chimpanzees have a random sampling assumption that can be overridden under the appropriate circumstances and they are able to use mental state information to judge whether this is necessary. This provides further evidence for a shared statistical inference mechanism in apes and humans.

Keywords: Intuitive statistics, probabilistic reasoning, mental states, random sampling, nonhuman primates, great apes

Results

We used an established paradigm [1] in which chimpanzees faced two mixed populations of preferred and non-preferred food items and could choose from which of the two populations they wanted to receive a sample. In contrast to previous studies where drawing was always random, we here varied whether sampling was random or not (method adapted from [4]). To examine whether chimpanzees could integrate knowledge about others´ choice biases and visual access into their statistical inferences, we first demonstrated to them that two experimenters E1 and E2 had specific and opposing biases regarding two types of food in Experiment 1: E1 preferred the type of food liked less by the apes themselves (carrot), whereas E2 showed the same preferences as the apes (peanut). These choice biases were established as follows: E1 repeatedly drew only carrots from a population with mostly peanuts (200 peanuts and 20 carrot pieces) and E2 showed the reverse patterns, repeatedly drawing only peanuts from a population with mostly carrots (20 peanuts and 200 carrot pieces). During the subsequent two test conditions, subjects witnessed the two experimenters sampling from their respective populations and were allowed to pick one of the samples. As the sample itself was hidden inside E1's/E2's fist, they had to infer from which population/experimenter they would most likely receive a favorable food item as a sample. The crucial manipulation between conditions was whether the experimenters looked into the bucket during sampling (visual access condition, see Figure 1A and B) or drew blindly (no visual access condition, see Figure 1C and D). The order of these two test conditions was counterbalanced across subjects.

To examine chimpanzees´ baseline performance in this task without any prior information about experimenters´choice biases, we tested them in Experiment 2 with new food types in the same proportions as before (200:20 vs. 20:200). Similar to the no visual access test, both experimenters drew blindly from the populations.

We found that subjects´ choice in Experiment 1 was significantly influenced by conditions (GLMM; X^2=44.26, df=1, P<0.001). More specifically, in the visual access condition, when experimenters looked into the buckets, chimpanzees preferentially picked the sample drawn from the less favorable population ($Mean_{favorable\ population}$=33.8.%), significantly different from what would be expected by chance (t=-3.58, df=19, P=0.002). Thus, subjects based their choice on the experimenters´ choice biases rather than on the proportional composition of the population. In contrast, when the same experimenters sampled blindly in the no visual access condition, subjects´ choice was different: Here they tended to choose the sample from the more favorable population more often, albeit not above what would be expected by chance ($Mean_{favorable\ population}$=57.1% of trials; t=1.37, df=19, P=0.187). Yet, a comparison of the two conditions revealed that subjects chose the proportionally favorable population significantly less often in the visual access condition than in the visual access condition (Estimate$\pm$SE=-1.083$\pm$0.204, df=2, P<0.001, CI(-1.714,-0.496), see Figure 2). This pattern was not due to any order effects, since it held equally for both orders of presentation of the test conditions (X^2=0.007, df=1, P=0.931). Moreover, the effect was not driven by single individuals: Apart from one young female showing the opposite pattern, and two subjects showing no difference between conditions, all remaining 17 individuals chose the sample from the more attractive population numerically more often in the no visual access condition.

In Experiment 2, when subjects did not have any prior information about potential choice biases and drawing was random, chimpanzees chose the sample from the more favorable population at the highest levels ($\mathrm{Mean}_{\mathrm{favorable\ population}}$=88.9% of trials), significantly above chance level (t=11.78, df=17, P<0.001) and significantly more often than in the visual access condition (Estimate$\pm$SE=3.261$\pm$0.355, df=2, P<0.001, CI(2.416,4.548)) and in the no visual access condition of Experiment 1 (Estimate$\pm$SE=2.177$\pm$0.352, df=2, P<0.001, CI(1.234,3.317; see Figure 2).

We did not find any effect of trial number within the conditions for the two experiments, indicating that chimpanzees did not learn within a session which of the two populations/experimenters was the rewarded one (X^2=2.693, df=2, P=0.260). First trial performance confirmed the choice pattern: 45% of subjects chose the sample coming from the more attractive population in the first trial of the visual access condition compared to 60% in the no visual access condition and 78% in the random condition. The identity of the experimenter did not influence the chimpanzees' choice (X^2=1.130, df=1, P=0.264; see also Table S2 and S3 for detailed results of Experiment 1 and 2).

To control for potential associative learning explanations, we lastly tested chimpanzees in Experiment 3, again using populations of new food types (100:10 vs. 10:100). Before the test, subjects experienced that both experimenters would always draw preferred food items out of their population. However, while E1 sampled blindly from the more favorable population, E2 sampled from the less favorable one while looking into the bucket. In the subsequent test, both experimenters drew in the same manner as before, but this time from identical populations containing equal proportions of preferred to non-preferred food items (55:55 vs. 55:55). We found that chimpanzees preferred the sample drawn by the experimenter who had before sampled the statistically unlikely (preferred) food type significantly above chance level ($Mean_{favorable\ experimenter}$=64.8% of trials; t=4.438, df=17, P<0.001; CI(0.577, 0.718); see Figure 3). Again, we did not find an effect of trial number (X^2=0.007, df=1, P=0.933), indicating that subjects did not learn within the test session which experimenter was favorable (see also trial 1 performance: 66.7%). Moreover, we did not find an effect of experimenter´s ID, neither when considering only Experiment 3 (X^2=0.803, df=1, P=0.370), nor when considering whether it was the same or the opposite one compared to Experiment 1 (X^2=1.142, df=1, P=0.286), indicating that subjects did not perform better when the positive experimenter was the same as in the first experiment (also see Table S4).

Discussion

The current study shows that chimpanzees were able to flexibly adapt their choice as a function of statistical and psychological information in a paradigm that required them to reason probabilistically from population to sample. In the visual access condition of Experiment 1, when biased experimenters drew samples while looking into the bucket, chimpanzees preferred the sample drawn by the experimenter with the preference for the favorable food type, mostly disregarding the proportional composition of the populations. This suggests that subjects expected the drawing to be based on the experimenters´ choice biases and therefore non-random in this condition. When the same biased experimenters drew samples from the same populations in the no visual access condition blindly, subjects switched and now showed a slight preference for the proportion-wise more favorable population, despite the experimenters´ biases. Hence, depending on whether or not the experimenters had visual access to the buckets while drawing, subjects based their choice either on the experimenters´ choice biases or rather on the mere proportional composition of the population. In Experiment 2, when chimpanzees did not have any information about potential biases of the experimenters and they drew blindly, subjects chose the sample drawn from the favorable population at higher levels than in both conditions of Experiment 1. Results of these two experiments suggest that chimpanzees, without any prior information, assumed random sampling and expected the sample to reflect the population´s distribution. If they, however, had reason to assume that the experimenters were biased, subjects´ choice reflected these biases; the severity of this influence was dependent on whether the experimenters had visual access or not.

However, despite the fact that we did not find any indication for learning within test sessions, we cannot exclude that subjects might have learned during the demonstration of Experiment 1 to simply associate one of the populations/experimenters positively or negatively, and pick/avoid this one in the visual access condition, where the setup was identical to the demonstration. The difference between conditions could congruently be explained by a change in setup in case of the no visual access condition (presence of a barrier) or the elapsed time in case of Experiment 2. We believe this scenario is unlikely considering that chimpanzees and other nonhuman primates are known to have severe difficulties learning rules that clash with their natural predisposition to choose the larger of two (preferred) food amounts [8-10]. Furthermore, the shortness of the demonstration exposure makes any rule-learning explanation additionally implausible. Nevertheless, we sought to address this alternative explanation in Experiment 3, in which chimpanzees were required to infer an experimenter´s choice bias from statistical information (and according behavioral cues), without being differentially rewarded in the demonstration. In the test subjects intuitively preferred the sample drawn by the experimenter who had previously drawn the statistically unlikely (and preferred) food type in the demonstration over the experimenter who drew blindly (and therefore randomly). This suggests that chimpanzees were able to use statistical information, in particular a violation of statistical likelihoods, to infer an experimenter´s choice biases and draw conclusions about the sampling process. At the same time, it corroborates our hypothesis that subjects do not rely on associatively learned rules in this kind of task. It should be noted that, even though there is evidence that great apes have some understanding about human preferences or desires [11], we do not intend to make any strong claims about how chimpanzees interpreted the experimenter´s choice bias in our study. It is possible for example that the subjects inferred that Experimenter 1 seems to like (drawing/giving) peanuts. It is similarly possible that they simply noticed that Experimenter 1, for whatever reason, draws peanuts when she has the possibility to do so. We cannot disentangle these two possibilities and for the interpretation of our data it is sufficient to assume the latter option.

While chimpanzees showed a remarkable flexibility and sophistication in this study, one may wonder why they did not perform better in the no visual access condition of Experiment 1. Subjects in this condition chose the sample of the proportionally attractive population on average in 57% of trials as compared to 89% in Experiment 2, although we used the exact same proportions in both experiments. The most likely explanation for this difference is that chimpanzees in Experiment 2 did not have any information about potential biases of the experimenters, which left the randomness of the draw the only aspect to consider (results of this experiment also demonstrated that subjects had not remembered any “good/bad”-labels for the experimenter from the previous experiment). By contrast, in the no visual access condition of Experiment 1, chimpanzees had to overcome what they had just experienced, namely, that E1 always extracted carrots from the peanut-population, and E2 always extracted peanuts from the carrot-population. This information was even repeated (in reminder trials) right before the no visual access condition. Hence, this condition required two extra steps compared to Experiment 2: Chimpanzees had to recognize and understand the indicators of blind drawing and they had to weigh the indicators of “biased sampling” and “blind sampling” against each other and choose accordingly. Therefore, our task design required a fair amount of cognitive flexibility which might have been too demanding for some of the subjects.

Conclusions

Taken together the results of our three experiments suggest that chimpanzees did consider information about the experimenters´ choice biases and visual access when drawing statistical inferences. Subjects were not only able to recognize that sampling would be non-random when biased experimenters had visual access while drawing, they also knew to some extent that when visual access was blocked, the choice bias information was rendered irrelevant and could therefore be dismissed. Moreover, chimpanzees were able to draw inferences about the experimenter and the sampling process from the given statistical information even without being differentially rewarded: when samples were unambiguously non-representative of a populations´ distribution and the experimenter looked into the population while sampling, subjects seemed to infer that the sampling person must have a bias for drawing one of the food types and act accordingly in the test condition. While previous studies have shown that chimpanzees can reason probabilistically from population to sample [1, 2] and are sensitive to what others can and cannot see (both conspecifics [e.g. 12] and human experimenters [e.g. 13]), our study is the first to suggest that chimpanzees are able to flexibly combine these two sources of information to make rational decisions under uncertainty. Our results resemble findings on human infants: Just as the chimpanzees in our study, 11-month old infants were shown to be sensitive to whether a sample was drawn randomly from a population or not on the basis of information about the drawing agent's psychological states (her preference and visual access) [4]. Similar to our apes, infants were also able use statistical information (in particular a violation of likelihoods), to draw conclusions about the sampling agent and the sampling process [14, 15].

Our study therefore gives further reason to assume that human statistical reasoning might be grounded in a cognitive mechanism that is utilized from early infancy onwards and shared with our closest living relatives.

Acknowledgements

This work was supported by a research grant of the German Science Foundation DFG (grant # RA 2155/3-1) to Hannes Rakoczy and Josep Call. Further, we acknowledge support by the Leibniz Association through funding for the Leibniz ScienceCampus Primate Cognition. We are thankful to Chimpanzee Sanctuary and Wildlife Conservation Trust and all staff involved in our stay, especially keepers, for providing us with the opportunity to test at Ngamba Island. We appreciate permission from the Ugandan National Council for Science and Technology and the Uganda Wildlife Authority. Our special thanks goes to Shona Duguid for acting as second experimenter and for valuable discussion. Further we would like to thank Marika Skowronski and Ines G. Stark for reliability coding, Laura Därr for preparing the figures and Roger Mundry for statistical support.

Author contributions

Conceptualization: D.H., J.E.; Methodology: D.H., J.E.; Investigation: J.E., Formal Analysis: J.E; Writing-original Draft: J.E.; Writing-Review &Editing: J.E., D.H., H.R., J.C., E.H.; Supervision: H.R., J.C.; Funding Acquisition: H.R., J.C., J.E.; Resources: E.H.

Declaration of interests

The authors declare no competing interests.

References

1. Rakoczy, H., Cluver, A., Saucke, L. , Stoffregen, N., Grabener, A., Migura, J., and Call, J. (2014). "Apes are intuitive statisticians." Cognition 131(1), 60-68.

2. Eckert, J., Call, J., Herrmann, E., and Rakoczy, H. (invited revision and resubmission). “Signature limits of intuitive statistics in chimpanzees.”

3. Teglas, E., Girotto, V., Gonzalez, M., and Bonatti, L. L. (2007). "Intuitions of probabilities shape expectations about the future at 12 months and beyond." Proc Natl Acad Sci U S A 104(48), 19156-19159.

4. Xu, F., and S. Denison (2009). "Statistical inference and sensitivity to sampling in 11-month-old infants." Cognition 112(1), 97-104.

5. Denison, S. and Xu, F. (2010). "Integrating physical constraints in statistical inference by 11-month-old infants." Cogn Sci 34(5), 885-908.

6. Teglas, E., Vul, E. , Girotto, V. , Gonzalez, M. , Tenenbaum, J. B., and Bonatti, L. L. (2011). "Pure reasoning in 12-month-old infants as probabilistic inference." Science 332(6033), 1054-1059.

7. Denison, S., Trikutam, P., and Xu, F. (2014). "Probability versus representativeness in infancy: can infants use naive physics to adjust population base rates in probabilistic inference?" Dev Psychol 50(8), 2009-2019.

8. Boysen, S. T., and Berntson, G. G. (1995). "Responses to Quantity - Perceptual Versus Cognitive Mechanisms in Chimpanzees (Pan-Troglodytes)." J Exp Psychol Anim Behav Process 21(1), 82-86.

9. Schmitt, V., and J. Fischer (2011). "Representational format determines numerical competence in monkeys." Nat Commun 2, 257.

10. Vlamings, P. H., Uher, J., and Call, J. (2006). "How the great apes (Pan troglodytes, Pongo pygmaeus, Pan paniscus, and Gorilla gorilla) perform on the reversed contingency task: The effects of food quantity and food visibility." J Exp Psychol Anim Behav Process 32(1), 60-70.

11. Buttelmann, D., Call, J., and Tomasello, M. (2009). “Do great apes use emotional expressions to infer desires?” Dev Sci 12(5), 688–698.

12. Hare, B., Call, J., and Tomasello, M. (2001). "Do chimpanzees know what conspecifics know?" Anim Behav 61(1), 139-151.

13. Melis, A. P., Call, J., and Tomasello, M. (2006). "Chimpanzees (Pan troglodytes) conceal visual and auditory information from others." J Comp Psychol 120(2), 154-162.

14. Wellman, H. M., Kushnir, T., Xu, F., and Brink, K. A. (2016). "Infants Use Statistical Sampling to Understand the Psychological World." Infancy 21(5), 668-676.

15. Gweon, H., Tenenbaum, J. B., and Schulz, L. E. (2010). „Infants consider both the sample and the sampling process in inductive generalization.” Proc Natl Acad Sci USA, 107(20), 9066-9071. doi:10.1073/pnas.1003095107

16. Baayen, R. H. (2008). “Analyzing Linguistic Data: A Practical Introduction to Statistics.” Cambridge: Cambridge University Press.

17. Schielzeth, H., and Forstmeier, W. (2009). "Conclusions beyond support: overconfident estimates in mixed models." Behav Ecol 20(2), 416-420.

18. Barr, D. J. (2013). "Random effects structure for testing interactions in linear mixed-effects models." Front Psychol 4, 328.

19. Field, A. (2005). "Discovering statistics using SPSS" Sage Publications, London

20. Fox, J., and Weisberg, S. (2011). "An R Companion to Applied Regression" Second Edition. Thousand Oaks CA: Sage. URL: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

21. Dobson, A.J. (2002). "An Introduction to Generalized Linear Models." (ed. C. Hall/CRC). Boca Raton

22. Forstmeier, W., and Schielzeth, H. (2011). "Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse." Behav Ecol Sociobiol, 65(1), 47-55. doi:10.1007/s00265-010-1038-5

23. R Core Team, 2016. "A language and environment for statistical computing". R Foundation for Statistical Computing, Vienna, Austria.

24. Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). "Fitting Linear Mixed-Effects Models Using lme4." J Stat Softw, 67(1), 1-48. doi:10.18637/jss.v067.i01>

Main text figure legends

Figure 1

Illustration of the procedures of the two test conditions in Experiment 1. In the visual access condition, experimenters looked into the buckets while sampling (A) before offering the subject a choice between the two samples hidden in their fists (B). In the no visual access condition (C and D), a screen was placed between experimenters and buckets, blocking the experimenters´ view into the populations. Moreover, in this condition the experimenters´ faces and bodies were oriented away from the table, further emphasizing a lack of visual access to the buckets during sampling.

Figure 2

Proportion of trials in which subjects chose the sample drawn from the more favorable population in Experiment 1 and 2. Size of the dots represents number of subjects performing at the same level. Bold horizontal lines depict the mean probability predicted by the model and grey dotted vertical lines depict bootstrapped 95% confidence intervals. Also see Table S2 and S3.

Figure 3

Proportion of trials in which subjects chose the sample drawn by the experimenter with a bias for the preferred food type (instead of the sample drawn by the blind experimenter) in Experiment 3. Size of the dots represents number of subjects performing at the same level. Bold horizontal line depicts the mean value for all subjects and grey dotted vertical lines depict 95% confidence intervals. Dashed horizontal line indicates chance level (i.e. indifference between both experimenters). See also Table S4.

STAR Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Johanna Eckert (eckert.johanna@gmail.com).

Experimental Model and Subject Details

In total, we tested 21 chimpanzees (9 females) with estimated ages between 7 and 33 years (Mean=18.7). Twenty of those participated in Experiment 1, 18 participated in Experiment 2 and 3 (see Table S1 for more detailed information about the subjects). All individuals were wild born orphans that lived in a social group of 49 individuals at Ngamba Island Chimpanzee Sanctuary, Uganda. In accordance with the recommendations of the Weatherall report 'The use of nonhuman primates in research' chimpanzees were allowed to roam freely on the 40 ha island covered with tropical rainforest during the day and voluntarily spend the night in seven interconnected sleeping rooms (approx. 140 m2) with regular feedings and water ad libitum. Subjects participated in the study on voluntary basis and were never food or water deprived. Research strictly adhered to the legal requirements of Uganda and was reviewed and approved by the Ugandan Wildlife Authorities and the Ugandan National Council for Science and Technology. The study was ethically approved by committees of the Max Planck Institute for Evolutionary Anthropology and the Chimpanzee Sanctuary & Wildlife Conservation Trust. Animal husbandry and research comply with the 'PASA Primate Veterinary Healthcare Manual' and the 'Guidelines for the Treatment of Animals in Behavioral Research and Teaching' of the Association for the Study of Animal Behavior.

Method Details

The study consisted of three experiments. Experiment 1 was conducted in May 2017, Experiment 2 and 3 were conducted consecutively in January 2018. In all Experiments, subjects were tested individually in their sleeping rooms and two experimenters E1 and E2 were seated at a table (L/W/H: 73cm/40cm/48cm) in front of the subject close to the mesh. As stimuli we used mixed populations of preferred and non-preferred food items that were presented in two transparent buckets (∅ 21.5 cm; height 19 cm). Food items of both types were of roughly equal size.

Experiment 1

Experiment 1 consisted of three phases: the demonstration phase, the visual access condition and the no visual access condition. In all phases, E1 presented a transparent bucket filled with 200 peanuts and 20 carrot pieces (P1), E2 presented a bucket filled with 20 peanuts and 200 carrot pieces (P2). We knew from previous studies (e.g. Eckert et al., under revision) that all tested individuals clearly preferred peanuts over carrots; hence, P1 was considered the more attractive population. The identity of E1 and E2 was counterbalanced across subjects. While all individuals started with the demonstration phase, the order of presentation of the two test phases was counterbalanced to avoid potential order effects. All three phases were tested on consecutive days. Twenty subjects participated in this experiment.

Demonstration phase

In the demonstration phase subjects experienced that both experimenters had a bias to sample items of the minority type in their bucket, i.e. E1 for carrot pieces (from population P1) and E2 for peanuts (from population P2). In other words, E2 had the same preference as the subject, while E1 had the opposing preference. A demonstration trial started with the right experimenter presenting her bucket by shaking it, tilting it and turning it around to give the subject a good overview about the content. Subsequently, she looked into her bucket, searched for three seconds using one hand and then visibly drew one item (of the minority type) and handed it to the subject. In the next trial, the left experimenter did the same with her bucket. In one session, subjects received ten demonstration trials per experimenter, with both experimenters sampling in alternating order. The side on which the experimenters were seated was counterbalanced. Chimpanzees received a total of two demonstration sessions on two consecutive days.

Test phase

Each test condition (visual access condition and no visual access condition) was administered in a single session consisting of 12 trials. The order of presentation of the two conditions was counterbalanced. Before a test session, each subject received three reminder trials per experimenter using the same procedure as in the demonstration trials.

Visual access condition

Each trial started with the right experimenter presenting her population by shaking, tilting and turning the bucket. Then the left experimenter did the same with her bucket. Subsequently, E1 and E2 simultaneously looked into their bucket, searched for three seconds using one hand and drew one item each without the subject seeing which item they had extracted (see Figure 1A). Just as in the demonstration, both experimenters always sampled an item of the minority type, i.e. E1 drew a less favorable item (carrot) out of the more favorable population (mostly peanuts), E2 drew a more favorable item (peanut) out of the less favorable population (mostly carrots). Both experimenters kept the sample hidden in their fist and presented the closed fist to the subject (see Figure 1B). The subject then indicated a choice between the two samples by pointing to one of the fists and immediately received the chosen sample as reward. Again, the side on which E1 sat was counterbalanced.

<u>No visual access condition</u>

The procedure was the same as for the visual access condition with the following modification: After having presented their buckets with the populations, the experimenters placed an opaque screen (L/W/H: 60cm/15cm/37cm) in between themselves and the buckets thereby blocking their view into the buckets. To further emphasize their lack of visual access, experimenters´ body orientation and gaze was directed away from the buckets during the sampling process (see Figure 1C). Next, both experimenters drew quickly and randomly the first item they could grasp in the bucket. Subsequently, the subject was offered a choice between the two hidden samples (Figure 1D).

Experiment 2

In Experiment 2, subjects were tested in a single test condition and did not experience any demonstration beforehand (and accordingly no reminder-trials). In order to avoid carry-over effects from Experiment 1, we used new types of food, their preference hierarchy was established in preference tests before and after the experiment (see Table S1 for more detailed information). The proportions within the populations remained the same as before (200:20 vs. 20:200). The procedure was similar to the no visual access condition of Experiment 2 with the following two modifications: 1. Experimenters were not assigned to one of the buckets across trials; instead, we counterbalanced the number of trials in which each experimenter presented and sampled from each of the populations. Thereby we hoped to minimize chances that subjects would form good/bad associations with the experimenters (while we were at the same time able to detect such potential effects statistically post hoc). 2. We did not use the barrier to indicate blind drawing; instead, the experimenters just turned away from the buckets and directed their gaze towards the ceiling. In doing so we wanted to examine whether these cues are sufficient for the apes to assume random drawing, which was important for the subsequent Experiment 3. Two subjects changed preferences over the course of the experiment (showing the opposite preference in the food preference test after the experiment compared to before). Accordingly, their data was excluded from the analysis. One further subject that had participated in Experiment 1 did not show any preference for one of the food types and was therefore not tested. In total, we included 18 subjects in the analysis.

Experiment 3

Experiment 3 consisted of two phases, the demonstration phase and the test phase, which were administered on consecutive days (only one subject did not enter the sleeping room the day after the demonstration phase and therefore had a one-day-break between demonstration and test). We again used new types of food in order to avoid carry over effects from the previous experiments (see Table S1). One individual had to be excluded because of lack of motivation; two further subjects could not be tested because they either did not enter the sleeping rooms within our data collection period or because they did not show a clear preference for one of the food types. In total, we included 18 subjects in our data analysis.

Demonstration phase

In the demonstration phase subjects experienced that E1 would blindly draw preferred items from the more favorable population (P1: 100 preferred and 10 non-preferred items), while E2 would intentionally draw preferred food items from the less favorable population (P2: 10 preferred and 100 non-preferred items). In each trial, consecutively, E1 turned away, directed her gaze towards the ceiling and drew one item quickly from her population and handed it over to the subject; E2 looked into her bucket and searched for three seconds before she handed over a preferred food item to the subject. Both experimenters always drew the preferred food type (except for one trial each for two subjects in which E2 accidentally drew a non-preferred item. Note that we ran a second analysis without these two subjects which did not change the significance of the results). Per session, subjects saw ten demonstration trials per experimenter, with both experimenters sampling in alternating order. For half of the subjects the identity of E1 and E2 remained the same as in Experiment 1, for the other half identities were swopped, which allowed us to test for carry over effects in our analysis. The side on which experimenters were seated was counterbalanced and the experimenter on the right always started sampling. Chimpanzees received a total of two demonstration sessions on two consecutive days.

Test phase

Before a test session, each subject received three reminder trials per experimenter using the same procedure as in the demonstration trials. In the test trials both experimenters had the exact same population with the same proportion of preferred to non-preferred food items (55:55). Hence, both populations depicted a 50% chance of leading to a preferred food item as randomly drawn sample. Each trial started with the right experimenter presenting her population by shaking, tilting and turning the bucket. Then the left experimenter did the same with her bucket. Subsequently, E2 looked into her bucket and searched for three seconds, while E1 turned away, directed her gaze towards the ceiling and moved her arm over the bucket. Then, both experimenters simultaneously drew one item without the subject seeing which item they had extracted. Both experimenters kept the sample hidden in their fist and presented the closed fist to the subject. The subject then indicated a choice between the two samples by pointing to one of the fists. Here, E2 always sampled preferred items, while E1 drew truly randomly. Again, the side on which E1 sat was counterbalanced. Chimpanzees received a total of 12 test trials presented in a single session.

Quantification and Statistical Analysis

The apes´ choice was coded live. A second blind observer coded 20% of the trials from video for each experiment. Both raters were in excellent agreement (Experiment 1: K=.96, N=97; Experiment 2: K=1, N=48; Experiment 3: K=.94, N=48). To analyze Experiment 1 and 2, we ran a Generalized Linear Mixed Model (GLMM) [16] with subject's choice (between populations) as dependent variable. As fixed effects we included condition, order of conditions, experimenter ID and trial number (to check for potential learning effects) as well as the three-way-interaction between condition, order and trial number. To control for a potential (linear or non-linear) effect of subjects' age, we included age and age^2 as further fixed effects. Subject ID was included as random effect. To keep type I error rate at the nominal level of 5% [17,18] we included all possible random slopes components (condition, trial number and experimenter ID within subject ID) and the respective correlations between random slopes and intercepts. Trial number, age and age^2 were z-transformed (to a mean of zero and a standard deviation of one). Variance Inflation Factors (VIF) [19] were derived for a standard linear model excluding the random effects and interactions, using the function vif of the R-package car [20] and did not indicate collinearity to be an issue. We assessed model stability by comparing the estimates derived by a model based on all data with those obtained from models with the levels of the random effects excluded one at a time. This revealed the model was stable. The significance of the full model as compared to the null model (comprising only age, age^2 and the random effect subject ID) was established using a likelihood ratio test (R function Anova with argument test set to "Chisq") [21,22]. P-values for the individual effects were based on likelihood ratio tests comparing the full with respective reduced models (R function drop1). The model was fitted in R [23] using the function lmer of the R-package lme4 [24]. To assess whether the average performance of subjects in the different conditions was different from what would be expected by chance, we used two-tailed one-sample t-tests, which were also administered in R.

To analyze Experiment 3 we ran a second GLMM with subject's choice (between experimenters) as dependent variable. As fixed effects we included trial number (to check for potential learning effects) and experimenter ID in Experiment 1 and 3, as well as the interaction between experimenter ID in both experiments (to check whether, e.g. individuals who had the same person as "positive experimenter" in both experiments performed better than those who had the opposite person). To control for a potential (linear or non-linear) effect of subjects' age, we included age and age^2 as further fixed effects. Subject ID was included as random effect. Again, we included all possible random slopes components (trial number within subject ID) and the respective correlations between random slopes and intercepts. Trial number, age and age^2 were z-transformed (to a mean of zero and a standard deviation of one). Again, Variance Inflation Factors (VIF) did not indicate collinearity to be an issue and model the model was found to be stable. The significance of the full model as compared to the null model (comprising only age, age^2 and the random effect subject ID) was again tested using a likelihood ratio test. P-values for the individual effects were based on likelihood ratio tests comparing the full with respective reduced models (R function drop1). Again, the model was fitted in R using the function lmer of the R-package lme4 and to assess whether the average performance of subjects was different from what would be expected by chance, we used a two-tailed one-sample t-test.

Data and Software Availability

Analysis-specific code and data are available by request to the Lead Contact.