Supplementary Material for "Chimpanzees trust conspecifics to engage in low-cost reciprocity"

STUDY 1

Method

Subjects and period of data collection.

Fifteen chimpanzees (8 females and 7 males), ranging in age from 7 to 25 years ($M = 15.3$ years), and living at Sweetwaters Chimpanzee Sanctuary, Kenya, participated in this study. Subjects were tested in dyads with nine chimpanzees solely acting as subjects ($M = 16.3$ years) and six chimpanzees exclusively as partners ($M = 13.8$ years).

Chimpanzees had access to a large outdoor enclosure (29 ha) during the day with regular feedings, daily enrichment, and water ad lib. Subjects voluntarily participated in the study and were never food or water deprived.

Data collection for Studies 1, 2, and 3 took place between August and November 2013. Study 2 started two days after the end of Study 1; and Study 3 started two days after the end of Study 2.

Procedure and Design.

Familiarization. All subjects were individually introduced to the experimental setup. Throughout the familiarization and experiments, individuals participated in one session per day. Familiarization consisted of four subsequent steps. During the first step, subjects were exposed to the trust rope and no-trust rope separately. Subjects first had to successfully pull the no-trust rope within 60 seconds eight out of ten times in one session. Having succeeded, subjects were then exposed only to the trust rope

and again had to successfully pull that rope eight out of ten times in one session. This involved pulling the trust rope (within 60 seconds), moving to room 3 via the overhead raceway (see Figure 1), eating half of the high-quality food, sending the vehicle back, moving back to room 2, and eating the other half of the high-quality food. During the second step of familiarization, both ropes were present but only one of the options was baited. Subjects had to pull the correct tray (containing the food) in each of two subsequent sessions of ten trials at least eight times. Once subjects started pulling one rope, an experimenter removed the other rope. Chimpanzees thus learnt that they had to make a choice and that only one rope could be pulled during each trial. The third step of familiarization was included to expose subjects to a setup closely matching that of the actual testing phase and to give subjects a full experience of the apparatus. Specifically, both ropes were present and both trays were baited. However, the overhead raceway between rooms 2 and 3 (Fig.1) was open and so subjects could move freely between those two rooms. Subjects participated in two sessions of ten trials. There was no criterion. During step 4, each subject received ten trials in one session with the overhead raceway closed and access to both ropes. This step was included to familiarize subjects with the actual setup of the test (up to this point the overhead raceway had been open during each step of the familiarization).

Importantly, during familiarization chimpanzees were not in any way trained to pull the trust rope. During steps 1 and 2, they pulled the no-trust and the trust rope equal amount of times. During step 3, they were rewarded when pulling either the trust or the no-trust rope. And during step 4 they were only rewarded when pulling the no-trust rope.

Coding.

All trials were videotaped with 2 cameras. The first author coded all trials live as well as later from videotape. Pulling the trust rope was coded as trusting behavior and pulling the no-trust rope was coded as non-trusting behavior. A research assistant, who was unaware of the study design and hypothesis, independently coded 25% of all trials. Interrater agreement was excellent (Cohen′s κ = 1).

STUDY 2

Coding.

Coding was done in the same way as in Study 1. A research assistant, who was unaware of the study design and hypothesis, independently coded 25% of all trials. Interrater agreement was excellent (Cohen′s κ = 1).

Statistical Analysis.

To test whether the level of trust changed over the course of trials depending on the behavior of the partner we used a Generalized Linear Mixed Model (GLMM, Baayen 2008). Since the reponse was binary (trust or not) this model was fitted with binomial error structure and logit link function (McCullagh & Nelder 2008). To test our key prediction of differential change of trust depending on the partner's behavior we included trial number, the behavior of the partner and their interaction as fixed effects into the model. To further control for potential carry over effects, we also included the test day as a fixed effect into the model. As random effects we first included random intercepts for the tested subject, the partner and the dyad. Furthermore, to keep type I error rates at the nominal level of 0.05 we included

various random slopes (Schielzeth & Forstmeier 2009; Barr et al. 2013), namely those of trial number, partner behavior and their interaction as well as day, within both the subject and the partner, and also that of trial number within dyad. We did not include correlations between random intercepts and random slopes or correlations among the random slopes to avoid an excessively complex model; additionally, Barr et al. (2013) have shown that these do not affect type I error rate.

The model was fitted in R (R Core Team 2014) using the function glmer of the package lme4 (Bates et al. 2014). Prior to running the model we z-transformed trial number and test day to a mean of zero and a standard deviation of one to easier interpretable estimates (Aiken & West 1991; Schielzeth 2010). To test the effect of the interaction we compared the deviance full model with that of the reduced model but comprising all other terms present in the full model using a likelihood ratio test (Dobson 2002). The sample size for this model was a total of 540 trials conducted with nine subjects, six partners, combined in 54 dyads. To check for model stability we excluded each level of each of the three random effects, one at a time, and compared the estimates derived for the full models fitted to these reduced data sets with those obtained from the model on all data. This revealed no influential subjects, partners or dyad.

STUDY 3

Coding.

The same coding scheme was implemented as in Studies 1 and 2. In addition, a trusting relationship was defined as choosing the trust option significantly more often than expected by chance over the course of five days and fifty trials. A research

4

assistant, who was unaware of the study design and hypothesis, independently coded 25% of all trials. Interrater agreement was excellent (Cohen's κ = 1).

Statistical Analysis.

To test whether the overall level of trust as shown in study 3 was above chance level, we fitted a further GLMM with binomial error structure and logit link function. This included trial number (z-transformed) as the only fixed effect and the identity of the dyad (with four levels; no individual occurred in more than a single dyad) as the only random effect. In addition to the random intercept of dyad we also included the random slope of trial number within dyad. The crucial term in this model was the intercept since this informs about whether the level of trust is above chance. However, with additional fixed effects in the model the intercept is usually slightly biased in the sense that its inverse logit transform does not reveal the average response (i.e., probability of trusting) but a value slightly further away from zero. To account for this we adjusted the values of trial by adding a constant, such that the intercept corresponded to average trust probability (Mundry in prep). Apart from that, the model was implemented as in Study 2. However, to establish model stability we tested each of the four dyads separately using a binomial test.

To test whether trusting in a given trial depended on the partner's behavior in the preceding trial we ran a GLMM with partner behavior in the preceding trial (trustworthy yes/no) and trial number as fixed effects and dyad as the single random effect (the factor dyad had four levels, and no individual occurred in more than a single dyad). We also included the random slope of partner behavior in the previous trial within dyad. The model was fitted with binomial error structure and logit link

function and implemented as described for the previous two models. The model stability check did not reveal serious issues.

References

Aiken, L.S. & West, S.G. (1991). Multiple Regression: Testing and Interpreting Interactions, (Sage. Newbury Park).

Baayen, R.H. (2008). Analyzing Linguistic Data, (Cambridge: Cambridge University Press).

Barr, D.J., Levy, R., Scheepers, C., & Tily, H.J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Memory Lang*, 68, 255–278.

Dobson, A.J. (2002). An Introduction to Generalized Linear Models, (Boca Raton: Chapman & Hall/CRC).

McCullagh, P. & Nelder, J.A. (2008). Generalized linear models, (London: Chapman and Hall).

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Meth. Ecol. Evol.*, 1, 103–113.

Schielzeth, H & Forstmeier, W. 2009. Conclusions beyond support: overconfident estimates in mixed models. *Behav Ecol*, 20, 416-420.