DOI: 10.1177/0956797616644070

Children's experiment

Methods

Participants

We tested a total of 96 participants, N= 24 pairs per age group, which is a conventional sample size in developmental studies. All parents provided written informed consent for their children's participation in the study. All participants were native German speakers and most of them came from middle- to upper-class families. The study was approved by the Ethics Committee at the MPI for evolutionary Anthropology.

Materials, apparatus and set-up

The two children in a dyad stand across each other with the apparatus in between. The apparatus consisted of a wooden box (85cm x 52cm x 50cm) with a transparent cover. The box contained two interconnected parallel trays (22 cm apart and 7cm wide each) that could be moved to the left and right from each child's perspective. There was a single rope threaded through each tray and extending across the width of the apparatus. Each end of the rope was sticking out of the box through holes at the front of the apparatus, and each child had to hold one end of the rope and pull simultaneously with the other to move the left tray to the left and the right tray to the right. If only one child pulled her end of the rope, the rope came out of the box, becoming disconnected from the tray. Each tray had two tilted plastic opaque pipes (one to each side of the apparatus) where a maximum of six sticker-containing balls could be placed in line. The first ball in the pipeline was visible and resting in a hollow cavity. Children had to agree sliding the left tray to the left (or the right tray to the right, from any child's perspective) in order to align the balls with the access holes in the corner of the box, falling down and becoming accessible for the children. The rewards of the second tray

DOI: 10.1177/0956797616644070

dropped down through the holes in the center of the apparatus and were not accessible for the children anymore. A weight brought the trays back into the start position at the same time that the next balls in the pipeline took the previous balls' position. In this way, the apparatus re-filled itself automatically after each completed trial.

Experimental procedure

Dominance test

Children were approached in their group by the two experimenters and asked if they wanted to participate in a game together with another child. After talking to them for some time they were asked to come with the experimenters (E1 and E2) into the testing room.

The dominance test consisted in giving the children two "surprises", one per trial, and observing them while they explored and manipulated them. They were left alone in the room during the test but observed through an external monitor placed outside the room but connected to three video-cameras placed in the testing room. Several measures related to who monopolized the surprises and/or tried to control the situation were coded (see analyses' section below). The first surprise was a tablecloth with a picture on it. In order to see the picture, children had to unfold the tablecloth spreading it out. Once children were finished unfolding it and had seen the surprise picture, the experimenter entered again and gave them a "box with a surprise inside". Children were told to wait until the experimenter had left the room to open the box. The experimenter left the room and returned after one minute had elapsed. After finishing, children were brought outside the room, while E2 prepared the room for the introduction phase.

DOI: 10.1177/0956797616644070

Introduction to apparatus

Before entering the testing room E1 showed the rewards to the children. E1 told the children that E2 had prepared a game, in which they could win many stickers and invited them to enter the room to join E2 and play the game.

Demonstration trials: E1 and E2 approached the apparatus together with the children. One tray of the apparatus contained golden balls, whereas the other tray contained transparent balls. E1 commented that in order to get the balls one had to pull the ropes. In a 1st demonstration trial, E1 pulled alone unthreading the rope and being unsuccessful. E1 lamented that the apparatus was now broken and E2 answered that she would repair it while they waited outside. When E1 returned with the children, E1 commented to E2 that they could try to pull together. E1 and E2 started pulling simultaneously the tray with the golden balls slowly until the transparent (empty) balls fell into the center hole (reward-trap hole). In that moment E2 commented that although they lost the transparent balls it didn't matter since they were empty. The experimenters continued pulling until the golden balls became accessible. At this stage children were introduced to a colour-marked but transparent tube, in which they could keep their balls. The colour of the tubes matched the colour of the apparatus' side children were at. Each experimenter placed her golden ball in one of the tubes and E1 explained the children that they would keep the balls in these tubes until the break, time at which they could open the balls and get the stickers. In the next trial E1 and E2 pulled the tray containing transparent balls in order to show the children that it was important to pay attention to the trays, since they could contain golden or transparent balls. In the next 4 trials, E1 & E2 always pulled the tray containing the golden balls and asked the children to help them pulling in order to familiarize them with the apparatus more directly. Which tray (left or right) contained golden or transparent balls was counterbalanced across trials (by positioning the balls in the pipeline in a certain order), but the two individuals in the dyad had never

conflicting preferences in this phase, since both sides of the tray were filled equally (see Fig. S1). After emptying the apparatus (6 trials), E1 and E2 opened the balls with the children. The stickers were put in the sticker book and E1 left the room with the children while E2 reloaded the apparatus.

Practise trials: the apparatus was prepared as in the demonstration trials so that children did not have conflicting preferences regarding which tray to pull (Fig.S1). Children participated in 6 trials, in which they had to pull by themselves but were occasionally encouraged or reminded by the experimenters about the apparatus and things they had to pay attention to.

TT-test

They participated in four blocks of six trials each. During these six trials the experimenter left the room, since it was important that the children wouldn't feel observed and coerced in behaving in a particular way. Therefore, experimenters had no influence over the duration of each trial (only 1% of all the children's trials lasted more than 60s). After each block E1 stayed outside the room with the children and opened the balls placing the stickers in the book, while E2 prepared the apparatus for the next block of six trials. Which side of the trays was filled with golden balls or left empty alternated across blocks. Since children had their side of the apparatus marked with a specific colour, we made sure that they always returned to the same side after the breaks between blocks.

Coding and analyses

To establish the dominance relationship between the two children in a dyad, we coded several measures allocating "dominance points" to each child. In trial 1 (tablecloth), we coded (a) who monopolized or spent longer unfolding the cloth (i.e. child that took the lead unfolding, using both hands and manipulating a larger portion of the cloth or child that

manipulated the cloth the longest) (+1). In trial 2 (surprise box), we coded (b) who took first the toy inside (if partner took away the toy from this child within 1s, then the partner was given the point) (+1) and (c) who spent longer playing with the toy (+1). When the difference between the children in duration unfolding the cloth (a) or playing with the toy (c) was less than 5s, they were considered equal. The child that got the highest number of points was categorized as the more dominant child in the dyad. In the TT-test, we coded within each trial the outcome (whether or not children cooperated and which child obtained the reward). We also coded the latency to make a decision which tray to pull. The start of the trial was defined as the moment, in which both children were in their position oriented and paying attention to the apparatus, the previous rewards had already been placed in their tubes, and the apparatus had returned to the start position refilling itself. We considered a decision to be made when children started pulling simultaneously one of the trays. In addition, we coded instances of communication between the children before agreeing to pull in one direction. Specifically we coded verbal demands to pull to their own advantage (e.g. “here”, while pointing/indicating their side, “my side”), and suggestions to take turns which refer at least to one future trial (e.g. “one for me and one for you, ok?”, “we alternate”, “but next one is for me, ok?”). All coding was done from videotape by P. G. Inter-observer reliability comparisons were performed on a randomly chosen 25% of the trials of the dyads who succeeded at least once, with the second coder blind to the hypotheses (Dominance: measure (a) Cohen’s kappa= 0.87, N=12, measure (b) Cohen’s kappa= 1, N = 12, measure (c) Cohen’s kappa= 1, N = 12; Outcome: Cohen’s kappa = 0.983, N= 233; Latency: $r_s = 0.846$, N =233, $P < 0.01$; Verbal demands: Cohen’s kappa=0.794, N=234; Turn-taking verbal suggestions: Cohen’s kappa=0.77, N=39).

The data were analysed using Generalized Linear Mixed Models (Baayen, 2008). Models were fitted in R (version 3.0.2, R Core Team, 2013) using the function glmer of the

package lme4 (Bates, Maechler, Bolker & Walker 2013). To analyze whether children took turns across trials we conducted a GLMM with binomial error structure and logit link function that modelled the effect of age, trial and age on the likelihood of a "turn" (alternation of who gets the reward in two consecutive trials). We also included the interaction of trial and age. We included dyad identity as a random effect to control for repeated measures and as random slope trial within the random effect of dyad identity. We did not include the correlations between random slopes and the random intercept because, according to Barr, Levy, Scheepers & Tily (2013), neglecting such random slopes does not affect model validity (i.e., type *I* error rates). Nevertheless, in order to check this, we also fitted models including the correlation between random slopes and intercepts and the results were virtually identical. The fixed effects predictor trial was standardized to a mean of 0 and a standard deviation of 1 (Schielzeth 2010). We excluded dyads that did not cooperate (N=8).

To analyze whether the more dominant children in each dyad were obtaining more rewards than their partners we conducted another GLMM with binomial error structure and logit link function with "dominant got the reward" (yes/no) on a given trial as binary response term. We included the same fixed effect predictors as in the previous model (age, trial and gender), as random effect the dyad identity, and as random slope trial within the random effect "dyad identity". The information whether dominants obtained more rewards than their less dominant partners could be drawn from the intercept of the model. An estimate for the intercept being > 0 indicates that the dominant received more rewards (est. < 0 that the subordinate received more rewards). The respective significance test is somewhat problematic because the estimate is frequently biased away from zero, which makes it anti-conservative (R. Mundry, pers. Comm). Hence, if the test of the intercept revealed significance, the predictors were adjusted such that the intercept was not biased anymore (by adding a constant to them). This adjustment does not affect the estimates of the fixed effects

but only that of the intercept and shrinks it towards zero. Also this model we re-analyzed with the correlations between the random slopes and the random intercept included, and, again, the results were virtually identical.

The models' stability was assessed by excluding dyads one at a time from the data and comparing the coefficients derived with those revealed from all data. The intercept optimization and model stability were done with a function written by Roger Mundry (available upon request). We excluded dyads that did not cooperate (N=8) and dyads in which we found no dominance asymmetry between children (N=4).

We also examined the latencies to obtain the rewards by fitting the data with a general linear mixed model with negative binomial error structure and log link function for latency as the response. We included the same random effect as in the previous models (dyad identity) and as fixed effects age, trial and gender, as well as the interaction age and trial. The model was fitted using the function glmmadmb of the R package glmmADMB (Fournier et al. 2012; Skaug et al.2014). Fitting a model including the correlations between the random sopes and the random intercept was computationally not feasible since the model did not converge.

Finally, we conducted two analyses to investigate whether verbal demands were more likely after a trial, in which the partner had just obtained the reward, and whether verbal demands were more successful after a trial, in which the partner had just obtained the reward. Therefore, we conducted two more generalized linear mixed models. The response measures for these two models were whether each child made a demand to pull her side on a given trial (yes/no) and whether the demand was successful (yes/no). We used a binomial error distribution with a logit link function. In both models we included as fixed effect predictors the trial number, age group (3/5), gender (m/f) and whether the child had obtained the reward on the previous trial (yes/no). The fixed effect predictor trial was standardized to a mean of 0

and a standard deviation of 1 (Schielzeth, 2010). We also included in the models the random effect of subject, and as random slopes trial within subject and outcome of previous trial (reward or not) within subject. The models' stability was assessed by excluding levels of the random effect one at a time from the data and comparing the coefficients derived with those revealed from all data. As for the other models, including the correlation between the random slopes and the random intercept did not influence the results.

Results

There were 8 dyads of 3.5 year-olds that were unable to find any kind of solution to the dilemma during the first block of trials. We categorized these children as "unsuccessful". One pair started fighting, so that experimenters had to intervene and stop the test. Children in the other 7 pairs did one or several of the following things: they remained passive, without doing anything for at least 3 minutes on more than one occasion, they pulled without coordinating with the partner, unthreading the rope more than twice, or they refused to pull and left the testing room complaining to the experimenters more than twice during these first trials. When any or more than one of these things happened, the pair was considered unsuccessful, the test was interrupted, and experimenters joined them to play the game with them.

We investigated the effect of dominance on the likelihood of obtaining the rewards. For this analysis we fitted the data in a GLMM with binomial error structure and logit link function with "dominant got the reward" (yes/no) on a given trial as binary response term. We included fixed (trial, age and gender) and random (dyad identity) effects but were mainly interested in the test intercept. The results showed that overall dominant children did not monopolize more rewards than the less dominant ones (Estimate ± SE =-0.054 ±0.438, Z=-0.124, p=0.902).

DOI: 10.1177/0956797616644070

Chimpanzees' experiment

Methods

Subjects

Eight of the twelve chimpanzees lived in a social group of 17 individuals (Group 1). The other four individuals belong to a second social group of 6 individuals (Group 2). We tested all subjects from the WKRC in the Leipzig Zoo that could be tested individually (younger individuals could not be separated from their mothers) and were generally motivated and interested in participating in problem-solving tasks. Each group has its own sleeping rooms, observation rooms, indoor and outdoor enclosures. The indoor and outdoor facilities contain natural vegetation and structures to climb. The breeding program at the zoo is framed within the global strategy of the European Endangered Species Program (EEP) and all subjects are treated in accordance with the EEP guidelines.

Table S1. Subjects included in the study and their two partners.

Name	Age (years)	Sex	Rearing history	Partners
Robert[1]	35	M	Nursery	Patrick/Sandra
Sandra[1]	17	F	Mother	Frodo/Robert
Frodo[1]	17	M	Mother	Sandra/Patrick
Patrick[1]	13	M	Mother	Robert/Frodo
Ulla[1]	33	F	Nursery	Lobo/Lome
Lobo[1]	6	M	Mother	Ulla/Pia
Pia[1]	11	F	Mother	Lome/Lobo
Lome[1]	9	M	Mother	Pia/Ulla

DOI: 10.1177/0956797616644070

Jahaga[2]	17	F	Mother	Fifi/Trudy
Fifi[2]	17	F	Mother	Jahaga/Alex
Gertruida[2] (Trudy)	17	F	Mother	Alex/Jahaga
Alex[2]	9	M	Nursery	Fifi/Trudy

[1]Group 1, [2] Group 2.

Apparatus and set-up

Subjects were tested in pairs in familiar testing rooms with a familiar experimenter. The two individuals in a dyad were separated from each other in two adjacent rooms with a rectangular testing booth in between them (see Figure 1 of the main manuscript), but we chose pairs based on the zoo keepers' knowledge of their social relationships, choosing tolerant pairs and avoiding pairs which were known to engage in aggressive interactions outside the testing context. The booth had a frontal window and two Plexiglas panels (73 x 55 cm) on either side. Subjects sat across each other (98cm apart) with the apparatus placed in the booth between them. The apparatus consisted of two interconnected parallel trays (96cm long x 8cm wide each) that rested on platforms attached to each subject's window (the two trays were 6.5 cm apart). The two trays extended from window to window and could be moved to the left and right from the subjects' perspective. Each tray had two blocking pieces (one on each end) sticking into a ratchet bar underneath of and perpendicular to the tray. Each blocking piece had a rope that extended through the plexiglas window into the subjects' room. The ratchet prevented the movement of the trays, so that the trays were only movable by pulling both ropes of one tray simultaneously and lifting up the blocks out of the ratchet. The trays did not move when one subject alone pulled from one rope, or if subjects pulled

different trays. Each tray had a transparent food pipe on each end accessible to the subjects once the tray was aligned (through simultaneous pulling from both subjects) with the holes in the Plexiglas panel. Once subjects agreed to move one of the trays to one side (left or right), subjects could reach through the Plexiglas panel and obtain the reward placed on the tray. The food in the opposite tray dropped down through a reward-trap hole in the center of the platform and was neither accessible nor visible to the subjects anymore.

Experimental Procedure

Pretests

In the first two pretests subjects were individually introduced to the apparatus. In pretests III and IV they were introduced to the social aspect of the task, i.e. the need to coordinate with the partner.

1. Pretest I: Introduction to apparatus I. Only one of the trays was baited with one food reward on one side. Only the ropes of the baited tray were reachable for the subject. The ropes of the 2nd tray were visible but not reachable. The door between the two testing rooms was opened, so that subjects could move back and forth between the two sides of the booth (See Figure S2). When the back tray was baited, the subject could grab both ropes and pull by herself. When the front tray was baited, the experimenter pulled with the subject. We counterbalanced the tray and side of the tray that was baited across trials. Subjects received sessions of 8 trials each until they reached the criterion of successfully pulling the baited tray in six consecutive trials. Each trial lasted 60sec. Subjects participated in a minimum of one and a maximum of three sessions to reach criterion (Number of trials: Mean=10, Range=8-24 trials).

DOI: 10.1177/0956797616644070

2. Pretest II: Introduction to apparatus II. Only one of the trays was baited, but this time both sides of the tray were baited. Now, the ropes of both trays were now reachable. The door between the two testing rooms was opened, so that subjects could move between the two sides of the booth as they wished (See Figure S2). When the back tray was baited, the subject could grab both ropes and pull by herself. When the front tray was baited, the experimenter pulled with the subject. We counterbalanced which tray (back or front tray) was baited across trials. Subjects received sessions of 8 trials each until they reached the criterion of successfully pulling the baited tray in six consecutive trials without touching the rope of the empty tray. Each trial lasted 60sec. Subjects participated in a minimum of one and a maximum of twelve sessions to reach criterion (Number of trials: Mean=35.3, Range=8-96 trials).
3. Pretest III: Coordination I. The door between the two testing rooms was closed, and there was one subject in each side of the booth. Both trays were baited, but one tray was baited with high quality food (banana or pellets) and the other one with low quality food (carrot). We counterbalanced which tray (back or front tray) was baited with the high (or low) quality food across trials. When subjects coordinated pulling in one direction, they saw how the food rewards of the other tray dropped down in a non-accessible container through the centre hole on the tray (See Figure S2). Dyads received sessions of 8 trials each until they reached the criterion of successfully pulling the tray with high quality food in three consecutive trials. Each trial lasted 60sec. Dyads participated in a minimum of one and a maximum of five sessions to reach criterion (Number of trials: Mean=26, Range=8-40 trials).
4. Pretest IV: Coordination II. This pretest was as pretest III with the exception that now both trays were baited with high quality food (See Figure S2). Subjects had to coordinate their pulling in eight consecutive trials to pass to the TT-test. Each trial

DOI: 10.1177/0956797616644070

lasted 60sec. Dyads participated in a minimum of two and a maximum of three sessions (Number of trials: Mean=18.7, Range=16-24 trials).

Test phase

Only one side of each tray was baited (1-0 vs. 0-1). The back tray was always baited on the right side of the booth, whereas the front tray on the left side of the booth, but subjects exchanged positions every session, so that they were all the same number of trials on the left side and right side of the booth. The apparatus was refilled after every trial by the experimenter. Each dyad was tested in a total of 48 trials administered in four sessions of 12 trials each. Each subject was tested with two different partners, and the identity of the partner alternated every 2 sessions. In total each subjects participated in 96 trials (48 trials with each partner). Each trial lasted 60s. If subjects had not pulled any tray within this time limit, the experimenter removed the rewards and started a new trial.

Coding and Analyses

All trials were filmed and recorded with three video cameras: one in front of each subject and one in the centre behind the experimenter. We coded in each trial whether or not subjects cooperated pulling one of the trays and which tray was pulled (i.e. who obtained the reward). We also coded the latencies to cooperate pulling one of the trays. A trial started in the moment in which E placed the two pieces of food on the trays and ended when one subject started reaching for the accessible piece of food. Additionally, we coded whether or not subjects used any attention-getters during each trial (yes/no). The following behaviours were coded as attention-getters: (1) Stomping/banging: the subject stomps once or several times in a row with her feet on the ground or against the plexiglass, or hits several times in a row the plexiglass or metal mesh with her hands, (3) Clapping: subject claps with his/her hands producing a loud clapping sound, (4) Vocalizing: subjects hoots, grunts or whines, (5)

Rope shaking: shaking the rope in a powerful way up and down more than once. Finally, we coded subjects' behaviour during the unsuccessful trials. We coded whether subjects were during the last 10 seconds of the trial at the apparatus (and so could potentially pull) or had left the apparatus. All trials were coded live by J. K, who also verified the live coding by watching the videofiles. The subjects' behaviour during the unsuccessful trials and the attention-getters were coded by watching the video files by J. K. and a research assistant respectively. Inter-observer reliability comparisons were performed on a randomly chosen 20-25% of the trials, with the second coder blind to the hypotheses (cooperation (yes/no): Cohen's kappa = 0.97, N=144, P < 0.001; tray pulled: Cohen's kappa = 0.94, N=144, P < 0.001, latencies: Pearson's r=0.864, N=117, P < 0.001; attention-getters: Cohen's kappa = 0.722, N=288, p<0.001; behaviour unsuccessful trials (subject present or absent): Cohen's kappa = 0.89, N= 106, p<0.001).

The data was analysed using Generalized Linear Mixed Models (Baayen, 2008). Models were run in R (version 3.0.2, R Core Team. 2013) using the function glmer of the package lme4 (Bates, Maechler, Bolker, & Walker, 2013). In order to examine the levels of cooperation and the distribution of rewards between individuals in a dyad, we conducted 4 different models.

First, we examined the effect of session (1-4) and trial (within session) on the likelihood of cooperation (cooperation: yes/no) using a GLMM with binomial error structure and logit link function. The fixed effect predictors (session and trial) were standardized to a mean of 0 and a standard deviation of 1 (Schielzeth, 2010). We included in the model three random effects (random intercepts), dyad, dominant and subordinate (each individual in a dyad was categorized as dominant or subordinate), and as random slopes those of all fixed effects within all three random effects (Barr, Levy, Scheepers & Tily, 2013; Schielzeth & Forstmeier, 2009). Including the correlations between random slopes and intercepts into this

model was computationally not feasible (i.e., the model did not converge). However, since Barr, Levy, Scheepers & Tily (2013) found that neglecting such correlation components does not affect model validity (also in terms of type *I* error rate) we are confident, that the model can be trusted.

To analyze whether dominants obtained more food than subordinates and whether subjects in a dyad alternated pulling for each other, we conducted two more generalized linear mixed models. Since the response measure for these two models was whether the dominant got food (yes/no; trials where none of the two got the food were excluded) we used a binomial error distribution with a logit link function. The one model included as fixed effect predictors the side of the apparatus at which the dominant was placed, the session number (1-4) and the trial number within session. The other model included the same fixed effects and, in addition, whether the dominant got the food in the previous trials. All fixed effects predictors were standardized to a mean of 0 and a standard deviation of 1 (Schielzeth, 2010). We also included into the models the three random effects dyad, dominant and subordinate (each individual in a dyad was categorized as dominant or subordinate) and as random slopes all fixed effects within all three random effects (but no correlations between random slopes and intercepts, though; for the reasons see above). As with the children, the information whether the food obtained by the dominant was significantly above the food obtained by the subordinate was drawn from the intercept of the model. An estimate for the intercept being > 0 indicates that the dominant received more food (est. < 0 that the subordinate received more food). If the test of the intercept revealed significance, the predictors were adjusted such that the intercept was not biased anymore (as explained in the children section above). The models' stability was assessed by excluding levels of the random effects one at a time from the data and comparing the coefficients derived with those revealed from all data. The

DOI: 10.1177/0956797616644070

intercept optimization and model stability were conducted with a function written by Roger Mundry (available upon request).

We also examined the latencies to obtain the food by fitting the data with a general linear mixed model with Gaussian error structure and identity link for latency as the response. We included the same random effects as in the previous models and as fixed effects session (1-4) and trial number (within-session). The latencies were square root transformed. The normality and homogeneity of the residuals was assessed by visual inspection of a qq-plot and residuals plotted against fitted values (Quinn & Keough, 2002; Zuur, Ieno, Elphick, & 2010) which didn't reveal any obvious problems.

Finally, we examined whether attention-getters increased subjects' likelihood of obtaining the reward. We focused on those trials, in which only one subject used an attention getter and at least one of them obtained the reward. The response measure for this model was whether the subject who used the attention-getter (we excluded those trials in which both subjects used attention getters or none of them did) got the reward, and we used a binomial error distribution with a logit link function. If using an attention getter is effective one would expect that in the majority of the trials the subject actually got the reward; hence, in this model the intercept was the key predictor and we wished to know whether it was significantly larger than zero. As control predictors with fixed effects we included whether the subject (who used the attention getter) was dominant over its partner (manually dummy coded and then centered to a mean of zero) and the successive trial number (i.e., across all trials per dyad; z-transformed to a mean of zero and a standard deviation of one). As control predictors with random effects (random intercepts) we included the identity of the subject and its partner and also the particular dyad. We also included random slopes of trial number within all three random effects (more random slopes we could not include because their showed little

variation (if at all) within the levels of the random effects. We also did not include correlations between random intercepts and random slopes (see above).

We assessed model stability by excluding the levels of the random effects one at a time and comparing the estimates derived with those obtained from the full data set. Since we were specifically interested in the significance of the intercept we had to adjust the two fixed effects such that the intercept, after being turned into probability space, reveals the empirical probability of the subject using an attention getter receiving the reward. This is required because in more complex models the intercept frequently does not translate directly into this probability, even after centering all fixed effects. We achieved that by adding a value to both fixed effects such that the intercept (i.e., the link space probability of receiving the reward at both fixed effects being zero) equalled the empirical probability of receiving the reward (Mundry in prep.).

Follow-up

To facilitate the emergence of a turn-taking strategy, we slightly modified the apparatus, so that subjects could take turns within trials (and after one of them had already obtained her reward). The reward-trap in the centre of the apparatus was blocked so that after pulling one of the trays, the reward of the not-chosen-tray wouldn't fall down the trap becoming inaccessible, and subjects could still pull the other tray in the other direction, obtaining the 2nd reward in that same trial.

As in the previous test, each individual was tested with two different partners, expect Robert and Frodo, since one of their partners (Patrick) had been transferred to a different zoo when this follow-up study was conducted. Each individual received 48 trials with each partner, so that every subject participated in a total of 96 trials. The pair that most often

DOI: 10.1177/0956797616644070

engaged in sequential cooperation or reciprocation within trials was Pia-Lobo (a different pair from the pair that alternated the most in the basic condition). However, similar to the main condition, they did this on their 1st session and then almost never again.

Results

Dominants did not obtain more food than subordinates as revealed by the test of the intercept, which although it was >0 it did not reach significance levels (Table S2), and dyads did not pull faster or slower as the study progressed (Table S3)

Table S2. Factors included in the model looking at whether dominants obtained more food than subordinates.

	Estimate	SE	z	P	min	max
(Intercept)	0.750	0.946	0.792	0.428	-0.233	1.800
Dominant side	1.710	1.138	1.503	0.133	1.008	2.818
Session	1.304	0.670	1.948	0.051	0.781	1.795
Trial	-0.138	0.185	-0.747	0.455	-0.226	0.073

DOI: 10.1177/0956797616644070

Table S3. Factors included in the model looking at latencies to pull any of the trays (latencies were square root transformed).

	Estimate	SE	t	df	chisq	P	min	max
(Intercept)	4.362	0.166	26.210	NA	NA	NA	4.221	4.497
Session	0.131	0.154	0.849	1	0.684	0.408	0.038	0.238
Trial	-0.119	0.084	-1.418	1	2.000	0.157	-0.148	-0.064

The model looking at whether attention getters influenced the likelihood of obtaining a reward revealed that the intercept was positive but it did not reveal significance (Estimate ±SE=0.44±1.93, z=0.23, P=0.82). However, the estimate was associated with large uncertainty as revealed by the model stability estimation (min: -0.777; maximum: 8.151) meaning that this result is highly uncertain.

References

Baayen RH. 2008. Analyzing Linguistic Data. Cambridge University Press. Cambridge.

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

DOI: 10.1177/0956797616644070

Bates D, Maechler M, Bolker B & Walker S. 2013. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-5.

Fournier DA, Skaug HJ, Ancheta J, Ianelli J, Magnusson A, Maunder M, Nielsen A & Sibert J. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models

Quinn GP & Keough MJ. 2002. Experimental Designs and Data Analysis for Biologists. Cambridge University Press. Cambridge.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

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.

Skaug H, Fournier D, Bolker B, Magnusson A & Nielsen A. 2014-05-23. Generalized Linear Mixed Models using AD Model Builder. R package version 0.8.0.

Zuur AF, Ieno EN, Elphick CS. 2010. A protocol for data exploration to avoid common statistical problems. Meth Ecol Evol, 1, 3-14

DOI: 10.1177/0956797616644070

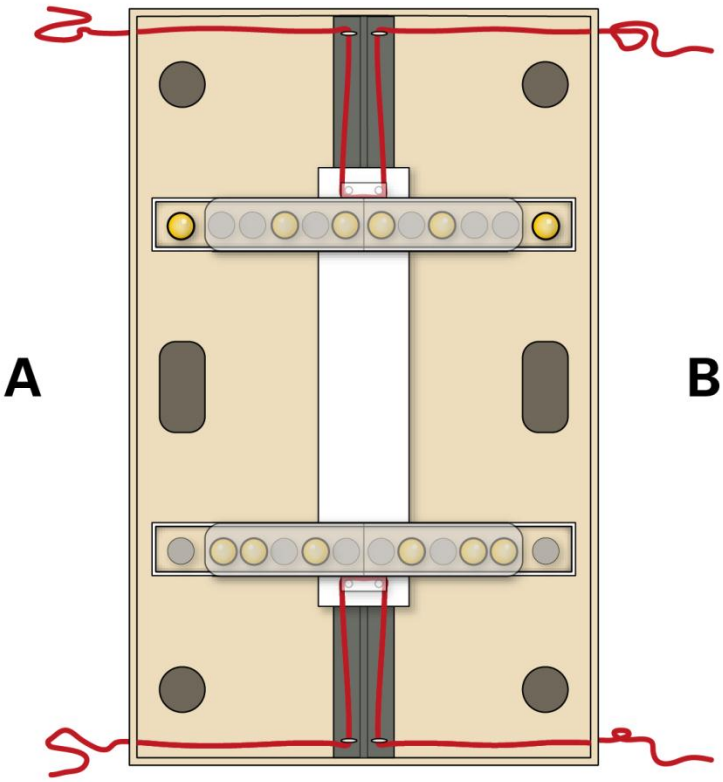

Figure S1. Apparatus in the demonstration and practise trials for the children's experiment. The two trays were filled with golden and transparent balls, so that children had to practise pulling both trays. Since in pretest phase both sides of each tray were always filled in the same manner, children had no conflicting preferences regarding which tray to pull.

DOI: 10.1177/0956797616644070

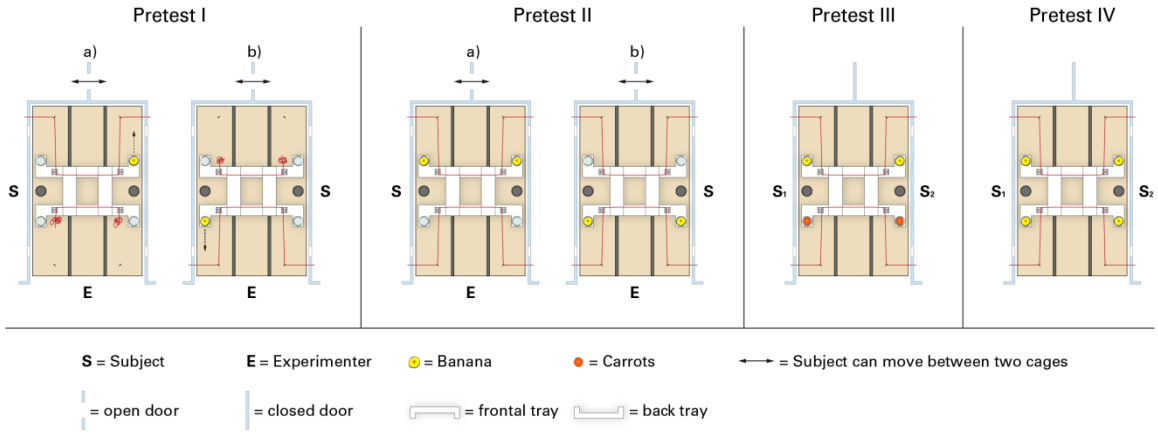

Figure S2. In Pretests I and II the chimpanzees were individually introduced to the apparatus and the door between the two sides of the booth was open. In Pretest I a) only the back tray was baited with a banana piece, whereas in b) the front tray was baited. The ropes of the empty tray were not reachable in this Pretest. In Pretest II, the ropes of both trays were always reachable; in a) the back tray was baited with bananas, one on each side of the booth, and in b) the front tray was baited. In Pretests III and IV the door separating both sides of the booth was closed, and subjects interacted with a partner (S1 and S2 in the figure). In Pretest III one of the trays was baited with bananas and the other one with carrots. In Pretest IV both trays were baited with bananas and subjects had to coordinate pulling from one of them.

DOI: 10.1177/0956797616644070

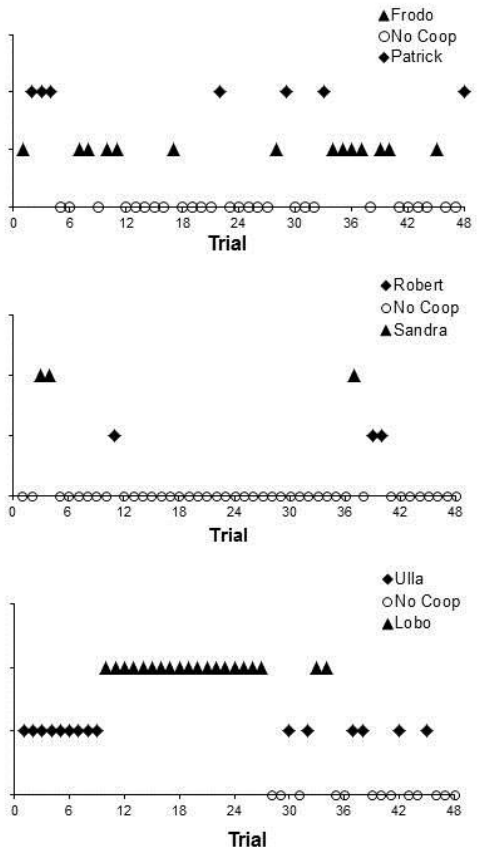

Figure S3**.** Trial by trial outcome of three dyads of chimpanzees. Markers indicate who got the reward and trials in which subjects failed to agree (No coop).