Subjects

Long–tailed macaques (*Macaca fascicularis*) live in sizeable and socially complex multi–male, multi–female groups. These groups are highly hierarchical and feature matrilines – groupings of individuals with a common maternal lineage which behave preferentially towards one another compared to non-matriline group members, for instance aiding one another in conflict (1). In the wild females stay with the natal group for the duration of their lives whilst males, following maturation, relocate to other groups (2).

Test subjects of the current study were housed at the German Primate Center. As testing commenced the subjects were living in a social group of 36 individuals comprised of 9 males (4+ years in age), 17 females (3+ years in age) and 10 juveniles (under 3 years in age). The monkeys were housed in an enclosure which has both an indoor (49 m^2) and an outdoor space (141 m^2). From the indoor enclosure monkeys were able to enter a testing area (2.6 m*2.25 m*1.25 m; h * w * d) which can be subdivided into six compartments. Monkeys could choose to enter the testing area. They were neither food nor water deprived.

Fifteen monkeys, ranging in age from 1 year to 15 years of age, initially took part in the study. The final sample size comprised 10 subjects (age range = 1 – 7 years) with one subject refusing to participate in the social competition condition. Two additional monkeys served as partners and were not test subjects (see table ESM 1 for more details). One individual died before the end of data collection and two individuals did not participate regularly enough to complete the initial training.

Table ESM 1
Information about subjects and partners

	Subjects	
	Sex	Date of birth
Ilja	m	29.12.2012
Linus	m	16.01.2013
Lukas	m	09.03.2015
Mara	f	30.12.2015
Max	m	01.02.2013
Meiwi	f	18.11.2015
Mila	f	07.04.2012
Milka	f	29.12.2014
Moritz	m	16.07.2016
Sambia	f	18.02.2015
	Partners	
Mars	m	17.01.2014
Snickers	m	12.01.2014

Setup

Two different double-tiered drawer-sets were constructed and these were interchangeably mounted between two testing cages. One drawer-set created a competition setting (CO) – a monkey pulling a drawer from one side removed this drawer from the reach of a monkey on the opposing side – the two monkeys could forage simultaneously from the same reward pool. The other drawer-set created a co-action setting (CoA) – a monkey interacting with one side of the drawers did not interfere with the foraging efforts of a monkey on the opposite side – the two monkeys could forage simultaneously from separate reward pools. Figure ESM 1 shows details of the test setup and dimensions of the drawer-sets. Each tier of drawers was covered by a sheet of plexiglass which had holes to allow for baiting of drawers. During data collection drawers were baited with cut raisins. Transparent plexiglass sliders were used to grant or deny monkeys access to the drawers as per trial requirements (details of procedure and conditions see next section). The effort which the partner monkeys had to expend to open drawers was manipulatable; either the partner pulled drawers of equal weight to the subject (~ 30 g) or the partner pulled drawers which weighed more (2 kg). Subjects always pulled 30 g drawers.

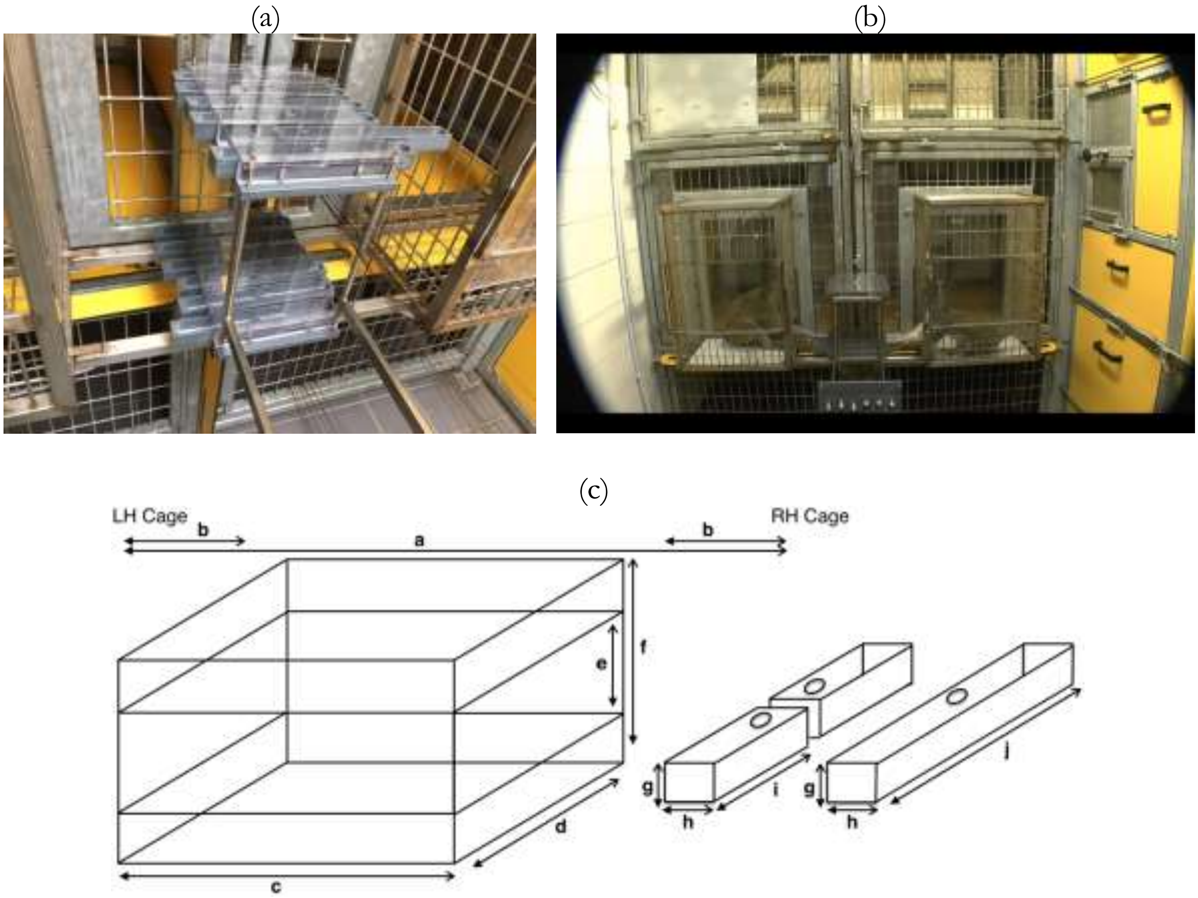

Figure ESM 1 Illustration and details of the drawer setup. (a) competition setup, (b) overview of test setup with two monkeys pulling drawers, (c) apparatus dimensions (in cm): a = 57, b (distance between edge of drawers and cage bars) = 15, c = 20, d = 30, e = 25, f = 32, h = 2.5, g = 3, I = 13 and j = 26. (Figure is not drawn to scale).

Procedure

Familiarization

All monkeys received extensive familiarization training with both drawer-sets prior to experiencing the test conditions to ensure that they understood the workings of both setups and differences between them.

First, they learned to operate the drawers when alone in the testing area. This phase included (i) exploration of the setup from both left and right cage (cage separation panel was removed) and (ii) making a detour to reach the other cage via two adjacent cages above or below the testing cages. Human interaction was not permitted during detour trials. Criterion for participation in the study was that a monkey should take the detour spontaneously at least two times per drawer–set type. There was no set limit as to the number of training trials which an animal could undergo. Partner monkeys underwent the same training phases as the prospective test subjects. In the case of the partner monkeys sometimes 2 kg weights were attached, sometimes 30 g. This was done to ensure that they would not be surprised by the increased effort condition during the actual test conditions.

Second, monkeys received 4 *baseline* trials with each drawer-set. We use "trial" to refer to a fully baited drawer-set, i.e. a monkey could receive up to 12 raisins per trial. The left–hand cage was kept empty and the subject was placed in the right–hand cage with unlimited time to retrieve raisins. Partners received 16 baseline trials: 4 trials for each possible weight and drawer-set combination. Presentation order was balanced; half the subjects received CO baseline trials first, the other half CoA baseline trials first. Partners were tested in the left-hand cage. Baseline trials were used to compare the partners' drawer-pulling latencies between heavy and normal conditions. The heavy weights resulted in significantly slower pulling latencies (m_{normal} = 5.53 s; m_{heavy} = 9.95 s; Wilcoxon test: $p > .001$).

Third, *social control* trials were used to gauge the foraging behaviour of subjects when under proximate conspecific surveillance. Monkeys received 4 social control trials with each drawer-set. As with the baseline trials, presentation order was counterbalanced. During social control trials a partner was placed in the left-hand testing cage but access to the apparatus was prevented by leaving the slider in place on the partner's side. The subject was placed in the right–hand cage with unlimited foraging time. The drawers on the left-hand side were not baited during CoA trials. The reason for this was that it was thought that presentation of extra food would create a point of difference between CoA and CO trials; in CoA, if the left–hand side was baited 24 raisin pieces would have been visible, 12 more than in the CO condition. The monkeys received another block of social control trials after the test conditions.

Finally, two types of *staggered* trials were used to demonstrate clearly to subjects that the CO and the CoA drawer-sets differed from one another with regard to reward certainty. In both of the staggered conditions subject access to drawers was delayed – the subject slider remained in place and the left–hand drawers were pulled open. In the *staggered–ghost* condition the left–hand drawers were pulled open via strings by an experimenter (obscured from view) - a "ghost" drawer–puller. In the *staggered–social* condition the left–hand drawers were pulled by one of the partner monkeys. After the last left–hand drawer had been tugged open the slider on the right-hand cage was removed and the subject was granted drawer access. In the case of the CO drawer-set the subject had one minute to interact with the apparatus and establish that no food items could be salvaged. In the case of the CoA drawer-set monkeys had unlimited time to forage. The left-hand drawers were baited during staggered–ghost trials in order to minimize the differences between staggered–ghost and staggered–social trials (i.e. outlay of food was kept constant). Monkeys received 2 staggered–ghost conditions (1 per drawer-set) and two staggered–social conditions (again 1 per drawer-set). Presentation order was balanced; half the participants received staggered–ghost trials first, the other half received staggered–social trials first.

Test

Subjects were presented with 3 test conditions per drawer-set: *normal* (partner pulled 30 g drawers), *slow* (partner pulled 2 kg drawers) and *ghost*. Subjects always pulled 30 g drawers. For ghost trials strings were attached to the left–hand side of the drawer–set and these drawers were pulled open at 5 second intervals by an experimenter obscured from view. Despite no animal being present in the left-hand cage during ghost trials the plastic slider was still lifted and replaced to signify trial start and end. Ghost trials were included to enable comparison of subjects' foraging behaviour when conspecifics were present versus absent whilst holding factors such as slider–movement, bait and drawer–movement constant. Conditions were presented in blocks of 6 trials each and order of conditions was counterbalanced across subjects. Subjects always received two trials in a row and were tested up to twice per day – resulting in a maximum of 4 trials per subject per day.

Coding

All test sessions were videotaped and were coded using the Mangold Interact software (version 9.0.7). We coded pull-latency as the time in seconds between the full extension of drawers pulled. Drawer-pulls were coded for both subject and partner in each trial. RT coded all of the videos and a second coder who was blind to the hypothesis of the study coded 20 % of the videos. Observer agreement was very good (Pearson correlation coefficient $r = 1$).

Analysis

We originally wanted to test our predictions by analyzing the effects of drawer-set (CoA vs CO), comparison standard (slow vs normal) and their interaction on pull-latency. However, despite an initial manipulation check after the baseline trials standard manipulation was not sustained over the course of the experiment. As a consequence, slow and normal condition did not result in overall slower performance in the slow condition. As a result of this we used actual partner performance (rather than condition assignment) as a predictor. It is perhaps important to mention that partner performance still had sufficient range to theoretically allow adaptation of one's behavior in individual trials to a comparison standard: CoA: $m = 3.03$ sec (1.39 sec – 8.42 sec), CO: $m = 2.71$ sec (1.45 sec – 6.3 sec). Furthermore, in cases in which both monkeys reached for the same drawer, reaction times were automatically delayed as it was necessary for one of the monkeys to switch to another drawer. We also noticed that monkeys sometimes accidentally pushed a raisin further away when trying to retrieve it from a drawer in the CO setup, which resulted in longer retrieval latencies and consequently longer latency until the next drawer was pulled. This could not happen with the CoA setup because the distant end of the drawer functioned as a natural barrier and the raisin could not be pushed further away. With this in mind we decided that the CO and CoA contexts should not be compared directly.

Our final analysis approach was the following: We first tested if partner ID and performance had an effect on subject performance in general; we did this separately for each context (CoA and CO). Since neither partner performance nor ID had an effect on subject performance we combined the slow and normal speed condition data for the subsequent analyses. Then we tested for the effect of disappearing food and social presence and compared (a) CoA vs. SocialControl-CoA vs. Ghost-CoA and (b) CO vs. SocialControl-CO vs. Ghost-CO. While the test conditions were counterbalanced across individuals, all monkeys received four baseline trials prior to the test conditions. Baseline trials were used to get an impression of the monkeys' spontaneous drawer pull performance and to check if the speed manipulation worked. During baseline trials, no conspecific was present in the other cage and no drawers other than the ones pulled by the subject were moved. Unfortunately, this design did not permit inclusion of baseline trials in the model comparison because systematic order effects cannot be controlled for. However, we thought it informative to run pairwise comparisons of the test and baseline conditions to address whether there was an effect of mere presence of a conspecific. Specifically, it would be interesting to see whether baseline and social control condition differed to a different extent compared to baseline and the other test conditions. Significant differences from baseline will have to be treated carefully because effects might stem from a real difference between baseline and test condition or might be an effect of trial number.

All analyses were conducted in the R statistical computing environment. For each analysis, we assessed latencies between pulling consecutive drawers within each trial and used log-transformed average trial latencies as the outcome variable in a Linear Mixed Model (3). We first tested if partner ID and performance had an effect on subject performance and then tested for the effect of disappearing food and social presence. We analyzed the co-action and competition data separately and all models were fitted using the function lmer of the package lme4 (4). For the first analysis, we included partner performance and partner ID as fixed predictors of interest and trial number as fixed control predictor. We included random slopes of partner performance, partner ID and trial within subject, but not the correlation parameters among random intercept and random slopes terms (5, 6). Trial number and partner performance were z-transformed. For the second analysis, we included condition (ghost, social control, co-action/competition) as fixed predictor and z-transformed trial number as fixed control predictor. We included random slopes of condition and trial within subjects, but not the correlation parameters among random intercept and random slopes terms.

For all analyses in this paper, we checked that the assumption of normally distributed and homogenous residuals was met by inspecting qq plots and the residuals plotted against fitted values. We checked model stability by comparing the estimates from the model based on all data with those from models with the levels of the random effects excluded one at a time, and checked for collinearity by determining the Variance Inflation Factor (7) for a linear model excluding the random effects. There were no obvious deviations from assumptions, no indications of model instability, and no problematic issues with variance inflation. We compared all full models with their respective null model comprising only the control predictors and full random slopes structure (using likelihood ratio tests with the anova function) to determine if the data was better explained by the latter. We have provided conditional R^2 effect sizes for those full models which were significantly different from their respective null models (using the function r.squaredGLMM of the package MuMin (8)). We ran planned pairwise comparisons for different levels of the factor condition (using the glht function of the package multcomp (9)) in case the model comparison revealed a significant difference between full and null model.

References

1. Thierry B. Management Patterns across Macaque Species. Natural conflict resolution. 2000;106.

2. Sueur C, Petit O, De Marco A, Jacobs A, Watanabe K, Thierry B. A comparative network analysis of social style in macaques. Animal Behaviour. 2011;82(4):845-52.

3. Baayen R. Analyzing Linguistic Data: A Practical Introduction to Statistics using R: Cambridge University Press; 2008.

4. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. Package 'lme4'. R Package Version 11–102016.

5. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of memory and language. 2013;68(3):255-78.

6. Schielzeth H, Forstmeier W. Conclusions beyond support: overconfident estimates in mixed models. Behavioral Ecology. 2008;20(2):416-20.

7. Field A. Discovering statistics using SPSS: Sage publications; 2009.

8. Barton K. Package 'MuMIn' Multi-Model Inference. R package version 1.40.0. https://CRANR-projectorg/package=MuMIn. 2017.

9. Hothorn T, Bretz F, Westfall P. Simultaneous inference in general parametric models. Biometrical journal. 2008;50(3):346-63.