Performance-based social comparisons in humans and long-tailed macaques

Stefanie Keupp[1,2], Farhan Abedin[2], Lena Jeanson[2], Carolin Kade[2], Josefine Kalbitz[2], Rowan Titchener[2], Thomas Mussweiler[3], Thomas Bugnyar[4], Julia Fischer[2,5,6]

Accepted for publication in *Animal Behavior and Cognition* on 27th September 2020.
This is the un-corrected accepted version of the paper.

[1]University College London, Department of Experimental Psychology, London, United Kingdom, [2]German Primate Center, Cognitive Ethology Laboratory, Goettingen, Germany, [3]London Business School, London, United Kingdom, [4]University of Vienna, Department of Cognitive Biology, Vienna, Austria, [5]Department of Primate Cognition, Georg-August-University Goettingen, Goettingen, Germany, [6]Leibniz ScienceCampus Primate Cognition, Goettingen

Corresponding author details:
Email: s.keupp@ucl.ac.uk
ORCID: 0000-0002-5451-4256

Abstract

Social comparisons are a fundamental feature of human thinking and affect self-evaluations and task performance. Little is known about the evolutionary origins of social comparison processes, however. Previous studies that investigated performance-based social comparisons in nonhuman primates yielded mixed results. We report three experiments that aimed (a) to explore how the task type may contribute to performance in monkeys, and (b) how a competitive set-up affects monkeys compared to humans. In a co-action touchscreen task, monkeys were neither influenced by nor interested in the performance of the partner. This may indicate that the experimental set-up was not sufficiently relevant to trigger social comparisons. In a novel co-action foraging task, monkeys increased their feeding speed in competitive and co-active conditions, but not in relation to the degree of competition. In an analogue of the foraging task, human participants were affected by partner performance and experimental context, indicating that the task is suitable to elicit social comparisons in humans. Our studies indicate that specifics of task and experimental setting are relevant to draw the monkeys' attention to a co-actor and that, in line with previous research, a competitive element was crucial. We highlight the need to explore what constitutes "relevant" social comparison situations for monkeys as well as nonhuman animals in general, and point out factors that we think are crucial in this respect (e.g. task type, physical closeness, and the species' ecology). We discuss that early forms of social comparisons evolved in purely competitive environments with increasing social tolerance and cooperative motivations allowing for more fine-grained processing of social information. Competition driven effects on task performance might constitute the foundation for the more elaborate social comparison processes found in humans, which may involve context-dependent information processing and metacognitive monitoring.

Keywords: co-action, competition, nonhuman primates, social comparison, task relevance

Introduction

Humans frequently compare themselves to others and such social comparisons affect how we feel, perform or attend to a task (Festinger, 1954; Mussweiler, 2003; Tesser, 1988; Zajonc, 1965). Who and what we compare to, i.e. the comparison standards, may be chosen deliberately or unconsciously, and a number of factors appear to shape these comparison processes (Mussweiler, 2003; Tesser, 1988). For instance, when presented with pictures of highly attractive or highly athletic comparison standards, subjects subsequently rated themselves as less attractive or athletic than when they had seen unattractive or non-athletic comparison standards (Brown et al., 1992; Mussweiler et al., 2004). Social comparisons can also influence task performance. For example, Seta (1982) presented almost identical tasks to pairs of participants who sat across from one another and who could infer how their co-acting partner was performing from the number and frequency of success tones. Crucially, the researchers manipulated after how many button presses a success tone would appear which resulted in differing perceptions of how well the other participant was performing. Following Festinger's (1954) argumentation that social comparisons are oriented upwards and most likely to occur for slightly better comparison standards, Seta predicted that subjects should improve their performance when paired with a slightly better participant, but not when the participant was extremely better, worse, or performing equally; these predictions were met.

In humans, social comparisons involve sophisticated cognitive processes that are tightly linked to our self-concept and self-other distinctions (Mussweiler, 2003). But social comparisons are also important for other animals. To evaluate how one fares in relation to others is important for intra-species competition, for instance when assessing the resource holding potential of other males in the competition for females (Clutton-Brock & Albon, 1979; Fischer et al., 2004;

Kitchen et al., 2003) as well as in intergroup competition, when the number of opponents needs to be compared to one's own group size (McComb et al., 1994; Wilson et al., 2001).

Little is known about the cognitive processes underlying self-other comparisons in nonhuman species, though this information is crucial to understand the evolution of this important mechanism. Before we continue, it is useful to distinguish between different types of comparison processes because the term can be used in a variety of readings. Self-other comparisons may refer either to comparing the *outcome* of a given action, e.g. when subjects receive different rewards for the same task (e.g., Brosnan & De Waal, 2003), or the comparison of the *effort* needed to achieve a certain reward (Wascher & Bugnyar, 2013). The comparison may also concern the actual *task performance*, whether someone else performs better or worse in comparison to the self, and how this affects subject performance. In the following, we will focus on performance-based comparisons, because we think it captures best what social comparison processes in humans are about.

Schmitt and colleagues (Schmitt et al., 2016) investigated performance-based comparison processes in nonhuman primates. In their study, long-tailed macaques (*Macaca fascicularis*) performed a touch-screen based picture discrimination task in the presence or absence of a conspecific social partner (i.e., the comparison standard) in the adjacent cage. Partners were either close affiliates with strong social bonds to the subject or non-affiliates. Subjects received acoustic information about the alleged performance of the co-actor, but had no visual access to the partner's performance. The study aimed to test predictions derived from research in humans (Mussweiler et al., 2004), namely that subjects should (i) assimilate to moderate standards and contrast away from extreme standards, and (ii) assimilate to socially close others and contrast away from socially distant others. This should result in an interaction of direction and extremity as well as direction and similarity (realized via bond strength category). These specific predicted interactions were not found in the monkeys. There was an effect of relationship quality on accuracy performance in the social control condition: subjects performed better in the presence of an affiliative partner who was not working at the task than when a non-affiliative partner was present. For reaction time, Schmitt et al. found an interaction of relationship quality and standard direction that affected the location of the upper quantiles: Slow responses occurred more frequently when subjects were paired with a non-affiliate who was performing worse than themselves (Schmitt et al., 2016). Based on these findings, Schmitt and colleagues discussed social comparison effects might involve different processes in monkeys than in humans or might even be restricted to humans. Dumas and colleagues (Dumas et al., 2017) challenged this idea. They assessed the role of task complexity for the occurrence and direction of social comparison effects in Guinea baboons (*Papio papio*) and found an interaction of similarity and comparison direction for the simple version of their task – a contextual cuing task where subjects had to find a target among several distractor stimuli on a touchscreen. There were some important differences in experimental design between the two studies pertaining to how comparison standard information was provided and how similarity was defined. Dumas and colleagues categorized individual pairings as "self better" or "other better" pairs, based on the difference of number of rewards that the two individuals in a dyad had obtained independently in the month prior to collection of the test data. Instead of bond strength, as in Tesser et al. (1988) and Schmitt et al. (2016), they used sex composition to gauge similarity. Choice of comparison standards, the way subjects learn about them, experimental task, and study species are among the differences that make it difficult to compare the outcomes of the two studies. For example, the task of Schmitt et al. might have been too demanding and bound all of the subjects' attention preventing them to process comparison standard information, or Guinea baboons might be more prone to engage in social comparisons due to their relatively relaxed social system as compared to long-tailed macaques (see also General discussion in this paper), or visual access to a co-actor during task performance might be crucial to elicit performance-based social comparisons (this was not the case in Schmitt et al.).

Against the background of a growing interest among researchers of comparative cognition to study animal's behaviour and performance in social interactive settings, it is necessary to carefully explore similarities and differences of social perception of others and the social test situation in general to allow appropriate interpretation of results. Our current experiments aimed at delineating some preconditions for the study of social comparisons. A major motivation to publish this particular package of experiments is that they constitute important steps on our endeavour to study social comparison processes in nonhuman primates. The experiments build on each other historically rather than adhering to the standards of a perfectly designed and balanced experimental plan. Yet, we believe they provide valuable insights for other researchers who are interested in this or related topics, as it is equally relevant to learn what did not work as it is to learn about successful paradigms. We report our attempts to study social comparisons in long-tailed macaques with different paradigms, the problems we encountered, and the current picture that emerged from it. We highlight the need to explore what constitutes "relevant" social comparison situations for monkeys as well as nonhuman animals in general, and point out factors that we think are crucial in this respect (e.g. task type, physical closeness, and the species' ecology). Importantly, we cannot close our paper with a clear result pattern but our take home message is rather that it would be premature to draw strong conclusions regarding general presence or absence of human-like social comparison processes in other animals.

In this paper, we addressed the question of task relevance in a series of experiments that explored the effects of task type (touch screen task vs. manual foraging task) and co-action type (competitive vs. co-active) in long-tailed macaques. Experiment 1 was similar to the two previously discussed papers regarding task type (touch-screen based picture discrimination task) but for the first time allowed subjects to directly observe the co-actors' task performance, including which task they actually had to solve. We aimed to add to the findings of Schmitt et al. (2016) by testing whether the comparison standard's direction influences task performance in macaques using a more visible presentation of comparison standard. We also wanted to test whether they pay attention to the co-actor and thus perceive the standard manipulation at all. To this end, we monitored looking and other behaviours of the monkeys during testing to assess how interested they were in the co-acting partner.

The findings of Experiment 1 led to the question if the monkeys' lack of attention to the partner was a consequence of this particular task or if competition might be a crucial factor to elicit interest in a co-acting partner. In Experiments 2a and 2b, we used a different paradigm to increase the potential relevance of social partners and their performance for the monkeys. Feeding represents a highly relevant activity that naturally draws the monkeys' attention. We therefore designed a co-feeding situation in which subjects either competed for the same food resource or co-fed next to a human partner from a different resource. We also manipulated how well the human partner performed the task by presenting slow and fast foragers. In Experiment 3, we presented an equivalent foraging task to a group of adult human participants.

If social comparison processes in long-tailed macaques mirror those of humans (Festinger, 1954; Mussweiler, 2003; Seta, 1982; Tesser, 1988), the monkeys should adapt their behaviour in response to a co-actors' performance if the setting is sufficiently transparent and relevant to elicit social comparisons. In Experiment 1, this would result in increased accuracy and faster response latencies in the picture discrimination task when the co-actor is performing better compared to when she is performing worse. Experiments 2a and 2b did not test social comparison effects proper but aimed to delineate what constitutes sufficiently relevant contexts for the monkeys. A relevant context would result in the monkeys adjusting their feeding behaviour to a competitor's or co-actor's action speed. Alternatively, the monkeys' behaviour might be purely driven by self-concern, i.e., they strive to maximize their reward outcome while ignoring the details of partner performance. As we know from a previous study (Seta, 1982) that humans adjusted their performance towards slightly better co-actors we predicted the same for our Experiment 3. We further expected that they would increase performance when directly *competing* with a *fast* partner, but had no clear predictions regarding the *slow* competitor condition.

Experiment 1

Extending previous findings on social comparisons in non-human subjects, we tested long-tailed macaques (*Macaca fascicularis*) with the same picture categorization task used by Schmitt et al. (2016) and a slightly modified procedure. We aimed at reducing cognitive demands of the co-action situation by making partner performance directly visible to the subjects, thus reducing the inferential demands of this task and potentially increasing the relevance of the comparison standard. Against the background that Schmitt et al. (2016) did not find an effect of standard extremity and to keep the analysis simple, we only manipulated direction but not extremity of the comparison standards. Monkeys were paired with a conspecific partner who appeared to perform the same picture categorization task at another close-by touch screen. We also included two control conditions without comparison standard information to test for the effect of partner presence, since previous studies showed that the mere presence of conspecifics can influence the performance of non-human primates (e.g., Huguet et al., 2014). If social comparison processes in long-tailed macaques mirror those of humans, monkeys should assimilate their behaviour in response to a co-actor's performance if the setting is sufficiently transparent and relevant to elicit social comparisons. This would result in increased accuracy performance and faster response latencies in the picture discrimination task when the co-actor is performing slightly better compared to when she is performing slightly worse.

Methods

Subjects

General information across Experiments 1 and 2

The subjects came from two study populations which were both housed at the German Primate Center (see section ‘Compliance with ethical standards’ for more details). All monkeys participated voluntarily in the experiments. They were not food or water deprived for testing and were fed their normal diet of monkey chow, fruits and vegetables twice a day. Water was available ad libitum. The monkeys of group 1 were housed in a group of ca. 35 individuals and had access to indoor and outdoor enclosures (49 m2 and 141 m2 respectively), which were equipped with various enrichment objects, wooden platforms, fire hoses, and a water basin during the warm months. Testing took place in a designated testing area (2.60 m × 2.25 m ×1.25 m; height × width × depth), which could be subdivided into six experimental compartments; the compartment was adjacent to the monkeys’ indoor enclosure. The group was used to behavioural testing taking place on a regular basis, however some of the monkeys had never shown any interest in participating. The actual pool of potential subjects ranged from around 12 to 18 individuals. The subjects of group 2 came from a study population of 14 individuals which is divided in three smaller groups. The animals were housed in three adjacent identically built and sized indoor enclosures (7.5 m2) with access to outdoor enclosures (6.4 m2) for each group. Each enclosure was equipped with various enrichment objects, wooden platforms, fire hoses and plastic boxes. Testing took place in two designated testing areas (each:

height x width x depth = 190 cm x 170 cm x 85 cm) adjacent to the inner enclosures which could be divided into 4 smaller rooms each (each: height x width x depth = 95 cm x 85 cm x 85 cm). During testing, they were separated from the group but visual as well as acoustical contact remained.

Subjects in Experiment 1

Sixteen monkeys (all from group 1) participated in Experiment 1. One individual only participated in the partner role. Of the remaining fifteen monkeys, one died and six did not participate regularly enough to reach the training criterion in time. Thus, the final sample contained data of eight subjects (see Appendix Table S1). Only two subjects did not have experience with touch-screens but learned to use it quickly through the course of the training sessions. This experiment was only the second experiment in which the monkeys worked on a touchscreen and where they worked in pairs (the first being the experiment of Schmitt et al., 2016).

Setup and procedure

Setup

The testing area was divided into separate compartments for the subject and the co-actor. In front of each compartment, a folding table was attached, on which a laptop with touch-screen could be fixed (Lenovo IdeaPad Flex 2-15 – 15.6" 1920x1080 Notebook with Full-HD 16:9 Multitouch LED IPS Display, see Figures 1 and S1). The table could be adjusted according to required angle of the screen and the monkeys could reach through the cage bars to touch the screen. The monkeys could see each other and the other’s screen from their respective compartment, i.e. they had full visual access to each other’s behaviour in the front part of the compartment, task performance and rewards being given. Importantly, however, it was not possible to determine from the distance which of the two pictures on the partner's screen showed a man and which showed a woman. Stimuli were presented with the software E-Prime (E-Studio, Version 2.0 Professional). An experimenter stood in front of the table and provided a raisin when the monkey made a correct response.

----Insert Figure 1 about here----

Procedure

The general procedure was adapted from the study by Schmitt et al. (2016) and had a training phase and a test phase. The training phase served to familiarize the monkeys with the new setup and the task. In the first training phase, the monkeys learned to touch a circle or triangle (depending on reward category) in a two-choice discrimination task. In the second training phase, they had to discriminate pictures of men and women (see Appendix for more details of the training procedure). For both training and test procedure, each session consisted of 20 trials, where trial refers to the presentation of a picture pair. Once a monkey performed correct in >14 of 20 trials in two consecutive sessions, the test phase began. During the training phase, the monkeys not only learned the task but also about visual feedback contingencies of correct and incorrect choices, which they could later use to assess partner performance by aid of visual cues (screen colour, absence/presence of rewards).

Test procedure.

A test session consisted of 20 two-choice discrimination trials, including ten familiar male/female picture pairs (from the training stimuli pool) and ten novel pairs, which appeared in random order. While the monkeys were always working alone during the training sessions, in the critical test conditions, they were paired with a designated co-actor whose alleged performance was experimentally manipulated. Each subject was tested in two experimental conditions ("high" and "low" standard conditions) and two control conditions ("social" and "non-social" control). In the experimental conditions, they were working alongside the co-actor, who was engaged in the same task on a second laptop in the adjacent test compartment. We manipulated the co-actor's performance (for details see section "standard induction" below) resulting in 18 of 20 correct decisions in high standard sessions and 10 of 20 correct decisions in low standard sessions. In the social control condition, the co-actor was present but not working. In the non-social control condition, the co-actor was not present. All subjects received two sessions of each condition, i.e. eight sessions in total. The first and last two sessions consisted of the control conditions and in sessions 3-6 experimental conditions were presented, with half of the subjects starting with high standard condition and the other half with the low standard condition.

Standard induction.

We manipulated the alleged performance of the co-actor by assigning the experimenter's keyboard instead of the co-actor's touch screen to be the valid input device. From her position in front of the cage the experimenter had a good view on the co-actor and pressed the key when she saw the monkey touching a stimulus. The experimenter produced correct and incorrect responses according to a randomized pre-designed schedule and rewarded the co-actor for correct trials.

For the two experimental conditions, a test session consisted of two consecutive phases – standard induction and co-action phase. First, the subject could watch the co-actor responding to 20 picture discrimination trials, thereby getting the chance to gather information about her performance. Performance level could be inferred by visual information of screen colour (red after incorrect choice, white after correct choice) and whether a food reward was provided. A keen subject could thus learn from observing the partner during this stage that the partner received a reward almost all the time and saw a red screen only twice after touching a picture (high standard condition) or that the partner received a reward in only half of the cases and saw a red screen in the other half (low standard condition). Subsequently, subject and co-actor worked side by side simultaneously. This co-action phase followed immediately after the standard induction and two experimenters were involved, each of who was attending to one monkey only. For the co-actor, the performance was again manipulated according to the same schedule. We recorded number of correct responses and latencies of touches from the subject. For the control conditions, no standard induction phase was needed because only the subject was working. No laptop was present on the partner side in the control conditions.

Coding and analysis

Experiments were filmed from a central frontal perspective, allowing us to see subject and co-actor simultaneously but not their task performance, which was logged automatically by the E-Prime program. We coded the following behaviours from video, for each of the test phases:

i. Attention to co-actor's performance (subject looks to co-actor, or interacts by lip-smacking, threats, etc. with co-actor); this was only coded for high and low standard conditions
ii. Attention away from co-actor/experimental setup (subject leaves the front area of the cage where they can see the co-actor, or visibly engages in other activities like exploration of their cage or the attached table or interacts with group members in the nearby indoor enclosure); for the co-action phase this applied also to being distracted from their own task

Coding was done using Mangold Interact by a research assistant who was not involved in data collection. A second coder who was blind to conditions and study rationale coded 25% of the videos. Observer agreement was good (Pearson correlation coefficient r = 0.83) for interest in the co-actor and moderate (r = 0.7) for attention away from co-actor.

Performance was assessed as the number of correct responses and touch latencies. E-Prime registered and automatically logged the position and timing of touches on the screen. Latency to touch was calculated from the time the stimulus appeared on the screen until it was touched and disappeared.

All Analyses were performed with R statistical computing environment. To investigate what influenced the probability of responding correctly, we used a Generalized Linear Mixed Model (GLMM, Baayen, 2008) with binomial error structure and logit link function. The model was fitted using the function glmer of the package lme4 (Bates et al., 2016). We included condition as a fixed predictor of interest and stimulus novelty and trial as fixed control predictors. We included random slopes of condition, stimulus novelty and trial number within subject, but not the correlation parameters among random intercept and random slopes terms (Barr, 2013; Schielzeth & Forstmeier, 2008). Trial number was z-transformed. We compared this full model with a null model comprising only the control predictors (using likelihood ratio tests with the anova function). The effect of the predictors on response latency was assessed in a Linear Mixed Model of the same model structure as specified for the accuracy response. The model was fitted using the function lmer of the lme4 package. We additionally analyzed response latencies for correct and incorrect trials separately. Log-transformed latencies were used as response variable. The model structure was the same as for the above analyses.

We additionally analyzed subjects' behaviour with a special focus on their interest in the partner and her performance during standard induction phase. We assessed the subjects' attention towards and attention away from the partner as a function of the fixed predictors condition and trial, random effect subject and random slopes of condition and trial within subject (using lmer function of the lme4 package).

For all analyses in this paper, we assessed the assumption of normally distributed and homogenous residuals by inspecting a qq plot and the residuals plotted against fitted values, 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 (VIF, Field, 2005) for a linear model excluding the random effects. Unless reported otherwise in the respective study, there were no obvious deviations from assumptions, no indications for model instability, and no problematic issues with variance inflation. We provide 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 (Barton, 2017)).

Results

Task Performance

In the test sessions, we were interested whether the monkeys' performance would change as a function of condition. Table 1 gives an overview of success rates and reaction times per condition. 1278 observations of 8 individuals were included in this dataset. All comparisons of full and respective null models revealed no significant differences, indicating that neither accuracy ($\chi^2 = 4.06$, df = 3, $p = .256$) nor reaction times ($\chi^2 = 1.98$, df = 6, $p = .921$) changed as a function of condition. An additional explorative comparison of the full model with a reduced model comprising only trial number as control predictor revealed no significant differences for both response measures, indicating that stimulus novelty had no systematic influence on the monkeys' performance (accuracy: χ^2 = 6.26, df = 4, p = .18, latency: χ^2 = 4.42, df = 7, p = .73). We found the same pattern for correct and incorrect trials.

---- insert Table 1 about here ---

Behavioural observations

Regarding the subjects' behaviour, we found that they only paid attention to the partner's performance on average a quarter of the duration of the standard induction phase (proportion of time spent with attending to the partner on average in high standard condition: *M* = .25, range = .07 - .42; and low standard condition: *M* = .24, range = .07 - .58). In contrast, they spent on average over two-thirds of the time with other activities – classified as "attention away from the partner or setup" (proportion of time not attending to the partner on average in high standard condition: *M* = .66, range = .45 - .88; and low standard condition: *M* = .69, range = .36 - .86). The amount of time spent with each of the types of behaviours did not differ between the experimental conditions, as evident from the full and null models being not significantly different (attention to partner: χ^2 = 0.11, df = 1, *p* = .738 and attention away from partner: χ^2 = 0.98, df = 1, *p* = .320).

Discussion

In Experiment 1, we found that neither the number of correct responses nor the reaction time differed as a function of condition in long-tailed macaques. The monkeys performed at equal levels when working next to a better performing or a worse performing conspecific, when working in the presence of a non-working conspecific in the adjacent cage, or when no partner was in the adjacent cage. We additionally coded their attention to the co-actor and found that they only occasionally attended to the partner's performance. In contrast, they spent over two-thirds of the time on average with other activities—classified as "attention away from the partner or setup". The amount of time spent with each of the types of behaviours did not differ between experimental conditions. It thus seems that our subjects were not particularly interested in what the partner was doing and how well she performed. The monkeys' looking patterns resembled occasional looks rather than periods of intense long observation followed by a loss of interest (most attention events were below 2 seconds duration). We are not saying that there was a complete lack of interest in the other monkey; subjects have surely observed a few responses (including their conditional rewarding), but their observations were not consistent enough to be able to distinguish between chance or above chance performance of the partner. This lack of interest in the partner's actions might indicate that the experimental setup and task may have been too abstract and irrelevant to catch the monkeys' attention and evoke interest in a partner's performance. Importantly, we aimed to make it very clear that no competition was to be expected from the partner monkey (separate food sources, closed cage separation, and even two different experimenters provided the food for the two monkeys). We chose this paradigm to mirror the non-competitive nature of default social comparison paradigms in studies with humans. In the case of non-human primates, however, it might result in social comparison processes not being activated. Given that long-tailed macaques live in a quite competitive environment (e.g., food and mate competition and strict social dominance hierarchies) it is well possible that they only engage in social comparisons when the consequences of a conspecific's actions are directly relevant for their own outcome. Consequently, our experimental paradigm might not have captured the relevant aspects to trigger social comparison processes in the monkeys. We suspect that some level of competition might be needed to draw monkeys' attention to a partner's actions in a co-action situation.

In Experiment 2a, we changed both task type and competitive nature of the context, as this seemed the combination most likely to reveal if the monkeys care at all about a partner working with them in parallel. If this is the case, more manipulations regarding competitiveness and partner performance levels can be devised with this paradigm to have a closer look at those effects.

Experiment 2a

In Experiment 2a, we presented the monkeys with a competitive foraging task from a limited food resource. Two human experimenters played the roles of a fast and a slow competitor, who would take food items from the shared resource. While we are aware that conspecifics might make for more salient comparison standards, we opted for human partners to allow manipulation of partner performance in this straightforward task. Humans have been used as interaction partners in experiments on social cognition before with results showing test subjects to be sensitive to the human's behaviour (for example, see findings of third-party social evaluations in i) chimpanzees: Herrmann et al. (2013); ii) capuchin monkeys: Anderson et al. (2013); or findings of unwilling-unable discrimination in i) chimpanzees; Call et al. (2004); ii) capuchin monkeys: Phillips et al. (2009)). We predicted two possible scenarios: first, the monkeys might increase their feeding speed irrespective of a competitor's actual foraging performance. This would be a first indication that the task is sufficiently relevant for the monkeys to pay attention to a performing partner. Second, the monkeys might adapt their feeding speed according to the speed at which the competitor depletes the resource. This means that an increased feeding speed is not only the result of the competitive situation but that they attend to the actual foraging performance of the competitor in more detail.

Methods

Subjects

Eight monkeys from group 1 (six males, two females, see Table S3) completed the study and only their data is included in the final dataset. One additional female lost interest to participate after giving birth and we stopped testing with five additional monkeys due to time constraints in the testing schedule after 5, 9, 10, 12 and 14 trials respectively because they did not participate regularly enough.

Setup and procedure

Setup

The setup consisted of a vertical feeding board (32.5 *40.5*3.5 cm dimensions) with 36 compartments (6 x 6 arrangement), which was attached to the outside of the testing cage (Fig. 2). The four upper rows of the board were baited with small pieces of raisins, resulting in 24 food items that could be obtained in a trial. The monkeys could reach through the mesh and take the food items with their hands. Depending on condition, an experimenter stood next to the cage and either took food items from the board (competitive conditions) or was merely present but did not take food.

---- insert Figure 2 about here ----

Procedure

During the initial familiarization, every subject could explore the feeding board on which some food items were accessible. They also experienced that they could not reach food items when an opaque plastic panel was inserted between the feeding board and the mesh. Once this occluder was lifted by the experimenter (E1) the monkeys could access and feed from the baited compartments. After the familiarization, we proceeded to establish the baseline feeding rate. We used the first 30 test trials to assess how quickly they ate the food items. Based on the average feeding speed of 1.3 seconds, the rate at which a slow and a fast human competitor would take food items from the board in the competition conditions was set at 2 items/s for the fast and 0.25 items/s for the slow condition. The monkeys were then randomly assigned to one of two orders of conditions, in which we tested the *baseline*, *social control*, *slow competition* and *fast competition* conditions in alternating turns (see Table S4). Each individual received five *baseline/alone* trials, three *social control* trials, six *fast competition* trials and six *slow competition* trials. We presented two trials in a row during a test slot which resulted in a maximum of 4 trials per day (up to two slots per day were available). For all conditions, the main experimenter (E1) was present and baited the board, moved the monkeys and lifted the occluder to give the monkeys access to the food items. During the baseline trials, no other person was present. During social control conditions, one of the two human partners (E2 or E3) was additionally present and stood next to the testing cage. Finally, during the competition conditions, either E2 or E3 was present and started to feed from the board once E1 had lifted the occluder. Throughout the experiment, E2 played the role of the slow competitor, i.e. she took a raisin every 4 seconds and E3 was the fast competitor, i.e. she took a raisin every 0.5 seconds.

Coding and analysis

We coded when and by whom each food item was taken. SK coded all of the videos and a second coder who was blind to the hypothesis of the study coded 25 % of the videos. Observer agreement was very good (Pearson correlation coefficient $r = 0.98$).

We calculated the latencies between taking consecutive raisins within each trial. In each trial, the board was baited with 24 raisins, which resulted in a maximum of 24 retrieval events per trial and thus in a maximum of 23 latencies between taking consecutive raisins per trial. The number of retrieval latencies differed between trials in the competition condition, because the number of items taken by monkey and human was different for every trial. We used log-transformed average trial latencies as outcome variable in a Linear Mixed Model. The model was fitted using the function lmer of the package lme4. We included *condition* as a fixed predictor of interest and *trial* as fixed control predictor. We included *random slopes of condition and trial within subject*, but not the correlation parameters among random intercept and random slopes terms. Trial number was z-transformed. We compared this full model with a null model comprising only the control predictors (using likelihood ratio tests with the anova function) to determine if the data is better explained by the latter. We provide 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 (Barton, 2017)). We ran planned pairwise comparisons for different levels of the factor condition (using the glht function of the package multcomp (Hothorn et al., 2008)) when the model comparison revealed a significant difference between full and null model.

Results

Figure 3 gives an overview of mean latencies to take the next item per condition. The monkeys obtained on average 10.2 (range: 6 – 15 raisins) raisins in the fast competition condition and 19.5 (range: 15 – 22 raisins) raisins in the slow competition condition. The model comparison revealed the full model to be significantly different from the null model ($\chi^2 = 18.86$, $df = 3$, $p < .001$, *conditional* $R^2 = 0.502$). We found a significant effect of condition ($\chi^2 = 18.87$, $df = 3$, $p < .001$). The negative coefficient of the trial estimate indicates that response latencies decreased with increasing trial number (see Table 2 for summary of the full model). Pairwise comparisons revealed no difference between social control condition and baseline condition and no difference between high standard condition and low standard condition. But high and low standard conditions were both different from social control and baseline condition, indicating that the monkeys increased their feeding speed in response to a competing partner, (see confidence intervals for pairwise comparisons in table S5). Following Reviewer suggestions, we further explored the effect of decreasing response latencies in Exp. 2a and 2b (see Appendix for details on these analyses). Comparing baseline condition and social control condition across Exp. 2a corroborated the effect of decreased retrieval latencies. A comparison of latencies between last trials of each condition found no difference between conditions.

---- insert Table 2 about here ---

--- insert Figure 3 about here ---

Discussion

The monkeys fed faster in both competition conditions compared to when the partner was absent, whereas feeding speed was similar in social control and baseline conditions. Thus, the co-feeding setup was clearly a relevant context, in which the monkeys paid attention to a performing partner. Given that we did not find a difference between fast and slow condition, we could only conclude that the monkeys' performance was driven by a self-concern to maximize their own outcome. To this end, increasing one's feeding speed as much as possible whenever in a competitive situation (however, not when a partner is merely passively present) is the most successful strategy. Retrieval latencies decreased across conditions through the course of the experiment and seemed to align towards the end. We take this as a sign that the monkeys experienced increasing uncertainty how E2 will behave next and preventively increased their feeding speed irrespective of condition. To address the possibility that the competition factor was too dominant and interfered with a potentially more differentiated sensibility for a partner's actions, we introduced a non-competitive co-action condition in addition to a competition condition, in Experiment 2b.

Experiment 2b

In Exp. 2b, our goal was to explore if the monkeys would react to a co-feeding partner in similar ways as when a partner was in direct competition with them. The co-action condition was similar to Exp. 1 with respect to the partner's task being independent from the subject's task. It was similar to Exp. 2a, however, regarding the relevant nature of the task. Following the reasoning of Exp. 2a, if the monkeys adapt their feeding speed not only in the competition but also in the co-action condition, it would be a first indication that their performance is driven by more than self-concern and that they might be sensitive to the actual foraging performance of the partner.

Methods

Subjects

We expanded data collection in Exp. 2b to a new group of long-tailed macaques who had not participated in Exp. 1 or 2a to increase our sample size and thus statistical power, and to include naïve monkeys who had not participated in this foraging task before. Eleven monkeys (6 males, 5 females) of group 1 and ten monkeys of group 2 (all female) participated in Exp. 2b. Seven of the group 1 subjects had also participated in Exp. 2a with a 3-months break between the studies. Three subjects of group 2 refused to participate regularly in the competition condition and their data were excluded from statistical analysis.

Setup and procedure

The setup consisted of a variation of the feeding board of Exp. 2a. A second identical board was added perpendicular to the original one but out of reach for the monkeys (see Figures 2B and S3). Depending on condition, the human partner stood next to the cage and either took food items from the frontal board (competitive condition), from the added left board (co-action condition) or was merely present but did not take food (social control condition). The number of raisins per trial was reduced to 20 per board, due to an additional panel that served as a barrier for the monkeys to reach the partner's raisins and which blocked some of the compartments that were formerly baited.

Independent of their participation in Exp. 2a, all subjects received some familiarization experience with the setup prior to the beginning of the study during which they experienced that they could not reach food items on the left board (i.e. the "experimenter's" board) or when an opaque plastic panel was inserted between the feeding board and the mesh. After the familiarization, we proceeded to establish the baseline feeding rate, which resulted in an average feeding rate of 1.05 raisins per second. Based on this foraging speed, we chose a feeding rate of 2 raisins per second for the human partner, identical to the high standard in Exp. 2a.

Each monkey was tested with all conditions (baseline, social control, co-action, and competition). The conditions were presented block-wise this time and we counterbalanced the order of conditions across subjects (for more details see Table S6). The procedure was identical to Exp. 2a regarding the roles of the experimenters. In the new condition (co-action), E2 started to feed from the left board (instead of the frontal board as during competition) once E1 had lifted the occluder. In all conditions, both boards were baited to hold the total number of food items constant across conditions. E2 left the area in front of the cage once all food items were gone on the frontal board (competition and social control conditions) or left board (co-action condition).

A different counterbalance design was used for group 1 and group 2 individuals. The reason is that we started this experiment with group 1 and had the impression that experiencing direct competition with the human partner might have influenced the monkeys' subsequent behaviour. We opted for an ABA design for group 2 to increase the number of trials during which individuals were naïve to a direct competition scenario. Responses are pooled for the main analysis but we also looked at naïve trials separately.

Coding and analysis

The same coding scheme was used as in Exp. 2a. The board was baited with 20 raisins in Exp. 2b which resulted in a maximum of 19 retrieval latencies per trial. Importantly, for the co-action condition, we had to account for the fact that the human partner fed faster than the monkeys and left before the monkey had finished eating. Since we were interested in co-action effects, the presence of a feeding partner is crucial and thus we only included the latencies of the first 10 raisins in our analysis (note that the human feeding rate was chosen to be roughly twice the baseline feeding rate of the monkeys, hence this makes a good estimate of raisins consumed during partner presence). Each individual received three *baseline* trials, six *social control* trials, and six *competition* trials*;* Group 1 individuals received six *co-action* trials, and group 2 individuals received 12 co-action trials (6 trials before and 6 trials after the competition trials, see differences in experimental design). We presented two trials in a row during a test slot which resulted in a maximum of 4 trials per day (when morning and afternoon slots were available). As in Exp. 2a, number of retrievals differed between trials in the competition condition, because the number of items taken by monkey and human was different for every trial. The monkeys obtained on average 7.5 raisins (range: 2 – 11 raisins) in the competition condition. RT coded all of the videos from group 1 and LJ from group 2. A second coder who was blind to the hypothesis of the study coded 25 % of the videos of group 1. Reliability was assessed using Pearson correlation coefficient, which was 1.0 for the timing, i.e. when a monkey took a food item. Data analysis approach was equivalent to Exp. 2a.

Results

Figure 4 gives an overview of mean latencies to take the next item per condition. The model comparison revealed the full model to be significantly different from the null model (χ^2 = 32.423, *df* = 3, *p* < .001, *conditional* R^2 = 0.544). We found a significant effect of condition (χ^2 = 32.193, *df* = 3, *p* < .001, see table S7 for detailed summary of the full model). Pairwise comparisons revealed a significant difference between competition and baseline as well as between co-action and baseline condition: In both conditions, the monkeys increased their feeding speed compared to baseline (see confidence intervals for pairwise comparisons in table S8). To address the possibility that the increased feeding speed in the co-action condition was merely a consequence of experienced direct competition, we separately assessed the responses of only those events where a monkey had not yet experienced a competing human partner. We found that response latencies of naïve individuals were faster in the co-action condition (*m* = 0.82, *SEM* = 0.05) compared to baseline (*m* = 1.08, *SEM* = 0.05), indicating that the increased feeding rate is not simply a consequence of a carry-over effect from experiencing a competing human partner. Furthermore, experiencing a human competitor affected social control conditions similarly to what we saw in Exp. 2a: Comparison of the first and second block of social control trials in group 2 showed that the monkeys tended to feed faster in the second compared to the first block (see Appendix for more details). Comparison of the first respective block of social control and co-action condition (i.e., before the monkeys experienced food loss by E2) showed that the monkeys fed faster in co-action compared to social control trials (see Appendix for more details).

--- insert Figure 4 about here ---

Discussion

In Exp. 2b, we aimed at exploring if the monkeys reacted differently when in direct competition compared to a situation where a partner was merely feeding in proximity but not from the same food source. We reasoned that if the monkeys adapted their feeding speed not only in the competition but also in the co-action condition, this would be a first indication that their performance was driven by more than "self-concern" (i.e., by more than a mere focus on their own food intake) and that they were sensitive to the foraging performance of the partner. We found that the monkeys increased their feeding speed compared to baseline when they were in direct competition with a human as well as when the human performed the same feeding behaviour on a different food board but not when the human partner was merely present. There was no difference between co-action and competition condition. Also naïve individuals, who had not yet experienced E2 as a food competitor, fed faster in the co-action condition than in the baseline and social control condition. This might be explained by social facilitation whereby a dominant response (here: retrieving the food items) is facilitated by the co-feeding situation (Zajonc, 1965). It would be interesting to compare changes in feeding speed between a slow co-actor condition and a fast co-actor condition. If subjects increase their feeding speed similarly in both conditions, this would indicate social facilitation rather than performance-dependent social comparison effects.

It is also possible that the monkeys perceived E2 as a potential competitor because E2 had the physical possibility to access the monkey's raisins. We have reason to believe, however, that at least some of the monkeys perceived the competitive condition differently from the co-action condition. Three monkeys outright refused to compete directly with the experimenter, while they were fine to approach the setup when E2 was feeding at the same distance but oriented towards the other feeding board.

The findings of Exp. 2b indicate that the monkeys' attention in this manual feeding task was drawn to a co-actor's performance more than in the touch-screen task, they were not merely focused on their own task performance and reward. We cannot conclude from Exp. 2b whether this is due to the situation that the human partner had potential access to the monkeys' food and was perceived as a competitor as soon as she showed interest in obtaining raisins or due to the task itself. For example, task difficulty has been shown to play a role in social comparisons in humans and baboons: Tesser (Tesser, 1988; Tesser et al., 1988) found a three-way interaction of social bond category, comparison direction and task difficulty in humans and Dumas et al. (2017) found this interaction in baboons (with the interaction of social bond and comparison direction being significant for the simple but not the complex task). Applied to the current context, one could argue that collecting raisins from a board is a simple task and discriminating artificial categories on a touchscreen is a more complex task and was perhaps not suitable to elicit social comparisons. Unfortunately, we didn't get the chance to further disentangle effects of competition and task difficulty by running the touchscreen task of Exp. 1 in a slightly more competitive setup, or by presenting slow versus fast co-actors in the co-feeding task. It would be interesting to conduct these experiments with individuals who have no prior experience with the task and social comparison setups.

In Exp. 3 we gave the same foraging task of Exp. 2 to adult human participants to test if the paradigm is feasible at all to test for classic social comparison effects.

Experiment 3

Performance-based social comparisons affected task performance of humans in various experimental settings (Allport, 1920; Seta, 1982; Tesser, 1988; Triplett, 1898; Whittemore, 1924; Zajonc, 1965). The goal of Exp. 3 in this paper was to provide a proof of concept for the foraging paradigm in Exp. 2a+b, i.e. to test if it is suitable to elicit social comparison effects in human participants. We tested adult participants' performance in competitive, co-active, and alone situations and we manipulated the performance level of the partner.

Methods

Subjects

Participants were recruited via leaflets in cafeterias and bulletin boards around campus at the University of Göttingen and via a local online forum. They were invited to a quiet room at the German Primate Center and participated in one experimental session of 40 minutes. Each participant received 10 EUR as compensation for their time. The current experiment was one of two experiments conducted in the same session as part of a M.Sc. thesis. The other task was conducted on computers and was about how participants perceived interacting with another human or a computer program. We measured response time and touch patterns of how participants touched stimuli on a touchscreen. Prior to the experiments, all participants received a description of the two tasks. They were informed that they could quit the experiment anytime without providing reasons. All gave their written consent to participate, gave permission to videotape the procedure for purposes of data analysis, and consent to their anonymized data being used for scientific purposes.

Our final sample comprised 87 participants (51 females, 36 males, mean age = 26.3 years, age range = 19 to 51 years). Sixteen additional participants were tested in a pilot phase to determine feasibility of different comparison standards and procedural details. Due to camera failure, we have no video footage of some trials of 11 participants, and one participant received an additional trial. We have data of at least two trials per condition for all but one participant.

Setup and procedure

Task

The same type of plastic grid board was used as in the monkey studies. Participants' task was to collect small wooden blocks (2x2x2 cm) from the board compartments instead of food items. Depending on condition, either one board (alone and competition conditions) or two boards (co-action condition) were placed on a table. Participants sat opposite of their partner (a confederate of the experimenter) and had good view of both their own and the partner's board.

--- insert Figure 5 about here ---

Half of the participants were assigned to the fast comparison standard group (20 male, 27 female) and the other half to the slow comparison standard group (16 male, 24 female). They were paired with the same (slow or a fast) partner during all experimental conditions.

Pilot phase

Prior to data collection, we ran a pilot phase to determine feasible comparison standards, i.e. the speed at which the partner collected their blocks, and to finetune experimental procedures.

In a first step, four prospective confederates provided data to determine the maximum speed at which a trained person can collect blocks from the board. Their performance stabilized at a rate of one block per 0.7 seconds and this performance level was subsequently used as the fast comparison standard during the main experiment.

The next step was to find a comparison standard that is perceived as different from the fast standard, yet sufficiently realistic to not raise suspicion in future participants, who we wanted to perceive the confederate as a real other participant. To this end, we asked seven pilot participants to provide feedback regarding how they experienced the fast as well as two different slow retrieval rates (one block every 1 and every 2 seconds) in competitive and co-active conditions. All participants indicated that they perceived the 2 seconds retrieval speed as unrealistic. Retrieval speed of both 0.7 and 1 seconds between consecutive block retrievals were perceived as realistic and different from each other. Based on this preliminary assessment, we used a retrieval speed of 1 block per second as the slow comparison standard performance in the main experiment.

Finally, nine additional participants were tested in all three experimental conditions (alone, competition, co-action) and confirmed these impressions. Their data is not included in the final analysis because we made substantial changes in the experimental procedure (pertaining to rebaiting of grid boards and number of trials per condition) after receiving their feedback. Our final sample consisted of 87 participants (40 in the slow condition and 47 in the fast condition), a number that resulted from practical reasons of what was possible in the course of a semester project rather than considerations and of power and effect size.

General procedure

Upon arrival, participants were greeted by the main experimenter at the entrance of the building and were led to the experiment room where the confederate was already waiting. Confederates (henceforth sometimes also referred to as the partner) were of the same gender as participants and were introduced by the experimenter as another participant who had arrived earlier and had already started with introduction and parts of the experiment. The latter information served as explanation later on during the experiment as to why only the real participant was engaged in the *alone* condition when confederate and experimenter left the room. The experimenter then explained task and general procedure and obtained informed consent from the participant before the start of the experiment.

We presented participants with three conditions: *alone*, *competition*, and *co-action*. Each condition comprised a block of three trials, where a trial is defined as presentation of a loaded grid board. A board was loaded with 30 wooden blocks and we used the latencies between taking consecutive items (we did not include the latency between ringing the start bell and taking the first item). As such, one trial resulted in up to 29 reaction time data points depending on how many blocks a participant obtained in this particular trial. The order of conditions was counterbalanced across participants.

Participants and confederates were told that their task was to retrieve wooden blocks from the grid board and that some rules applied regarding how the blocks must be collected. They were only allowed to use one hand (their preferred hand) and had to place the blocks on the table in front of them. They were instructed to ring a bell on the table to indicate start and end of their item collection in each trial. They were also told that sometimes they would work alone and sometimes with a partner. The experimenter emptied the table and provided a new loaded board for each next trial. Depending on condition participant and partner collected blocks from different boards or from the same board.

Alone

During this condition experimenter and confederate both left the room. Participants were instructed to begin their trial only after the experimenter had left the room.

Competition

Participant and partner were told that they would be working on the same board. They were seated facing each other at a table with the grid board between them, such that the blocks could be retrieved from either side. The experimenter retreated to the back of the room during this condition and gave the start sign upon which the participants could start the trial by ringing their bells simultaneously. Only the person who obtained the last block was asked to ring their bell. Both counted their blocks and the experimenter “rewarded” the one who had the most blocks with a token. Ultimately, there was no extra reward for these tokens, they functioned as markers that a round was won with the goal to enhance motivation in this competitive scenario.

Co-action

Participant and partner were told that they would be working alongside each other on two separate boards. They were seated facing each other but slightly shifted to the side at a table and each had their own board in front of them. The experimenter retreated to the back of the room during this condition and gave the start sign upon which the participants could start the trial by ringing their bells simultaneously.

Manipulation Check

After participants had finished the tasks, they answered a number of questions about the experiment. To check if the standard manipulation worked, we included questions about how they perceived the performance of the confederate in comparison to their own performance. All participants who were paired with a slow comparison standard reported they thought they were faster compared to the partner. 41 of 52 participants who were paired with a fast comparison standard answered they thought they were slower compared to partner, 9 estimated they were equally fast, 2 estimated they were faster (this includes the 9 beta phase participants). This indicates that our manipulation has worked and standards were perceived as intended.

Coding and analysis

Coding was similar to the monkey studies. The measure of interest was participants' speed of item retrieval from the grid board and we assessed latencies between taking consecutive blocks within each trial. FA coded all of the videos and a second coder who was naïve to conditions coded 21% of the videos. Reliability was assessed using Pearson correlation coefficient, which resulted in very good coder agreement of .99. To investigate the influence of comparison standard on the participants' reaction time, we built a Linear Mixed Model comprising comparison standard, action context (co-action or competition), and their interaction as predictors of interest and trial as a fixed control predictor. We included random slopes of trial and standard condition within subjects. We compared this model with a null model comprising the predictors comparison standard and trial. By keeping comparison standard as a predictor in the null model, we can conclude two things in case the model comparison reveals a difference: first, that the effect of condition is significant; second, that a significant interaction indicates that latencies in the different conditions are affected differently for the two comparison standard groups.

Results

Figure 6 shows the average response latencies. The model comparison revealed the full model to be significantly different from the null model ($\chi^2 = 90.779$, $df = 4$, $p < .001$, *conditional* $R^2 = 0.933$) thus showing that comparison standard has an effect on participants' performance. The interaction of standard condition and action context was significant ($\chi^2 = 73.002$, $df = 2$, $p < .001$), indicating that latencies in the different conditions were affected differently. Also the effect of trial was significant ($\chi^2 = 48.341$, $df = 1$, $p < .001$) with estimates decreasing with increasing trial number (see Table 3 for results of the full model).

--- insert Table 3 about here ---

--- insert Figure 6 about here ---

We found that participants' responses in the alone condition differed between fast and slow group. Therefore, we additionally checked if this was a general difference between the two groups or a consequence of prior experience with the different comparison standards. There was no difference between the groups when participants saw the alone condition first (m_{slow} = 0.890 s, m_{fast} = 0.850 s). In contrast, when participants had performed in the respective co-action or competition context before the alone condition, the groups differed significantly (Welch two-sample t-test: $t = 12.159$, $df = 118.32$, $p < .001$) with slower reaction times in the slow standard compared to the fast standard group (m_{slow} = 0.934 s, m_{fast} = 0.657 s).

Discussion

In Exp. 3 we gave human participants an item-retrieval task and assessed the effect of competition, co-actor presence and co-actor performance level on participants' task performance. Participants performed slower when paired with a slow partner than when paired with a fast partner. This effect also carried over to the non-social control condition, where participants in the slow condition performed slower than participants in the fast condition. These results are in accordance with previous findings showing an increase in task performance when participants were paired with a slightly better performing co-actor (Seta, 1982). It is less clear whether our results also replicate Seta's finding that participants' performance did not decrease when they were paired with a worse performing co-actor. Our alone condition was originally meant to represent a neutral control condition against which the social comparison conditions could have been compared. Since participants' performance in this condition was affected by their comparison standard assignment, we cannot conclude whether the general response pattern shows an assimilation towards a slow or fast comparison standard or both. On the one hand, a look at condition means indicates that assimilation towards the comparison standard was stronger in the fast condition than in the slow condition. On the other hand, several participants reported that they clearly noticed the slow performance of their partner, that they were slightly puzzled by it and deliberately slowed down their own actions. Consequently, we cannot unequivocally conclude whether our findings differ from these previous findings in regards to the role of a slow comparison standard. But we noticed some differences in methodologies that we think are important and warrant attention in future studies. Participants in Seta's experiment did not see the actual responses of their experimental partners because the effect buttons were hidden under an opaque screen. His participants only received acoustic feedback about partner performance. Two of our participants reported being slightly confused about the slow performance of the confederate and suspected there might be a hidden goal they had not yet found out about. As a consequence, they slowed down their own responses. Another participant reported feeling sympathy for the confederate and slowed down because they did not want to make the other person feel bad for being so slow. This hints at the possibility that additional processes are activated and underlie overt social comparison effects in this study and probably many social comparison scenarios. For example, a social norm to avoid humiliating others might stand in conflict with a drive for personal improvement and upward comparisons and might alter the resulting behaviour patterns. These effects possibly emerge stronger in transparent scenarios in close proximity of both co-actors (such as the current paradigm). But even if additional processes were at work causing the behaviour patterns in our participants these processes would rest on an initial comparison of the standard's performance with participants' own behaviour.

Another interesting aspect warranting more systematic attention is how comparison standard information is presented. A previous study, which presented both upward and downward comparison standards and assessed task performance in human participants, found that participants performed better in a simple task when engaging in upward comparisons with a friend and they performed worse when engaging in downward comparisons with a friend (Tesser et al., 1988). In that study, participants received verbal information about their performance relative to a co-actor in an unrelated task (answering questions about social sensitivity and creativity) before performing the test task of typing a numerical sequence. This feedback, despite being about information in an unrelated task, was unequivocal (self-better vs. other-better) and thus participants had a clear idea of the direction of the comparison. This touches two different aspects: (i) How easy or difficult it is for the participant to assign a value to co-actor performance in relation to own performance might matter for social comparison effects. (ii) Recent research in human decision-making showed that people behaved differently when they were engaged in experience-based based decision making compared to knowledge-based decision making ('description-experience gap', see e.g. Hertwig & Erev, 2009). Similar influences might be relevant during social comparisons and lead to different result patterns depending on how information about a comparison standard is presented.

General Discussion

In a series of experiments, we asked whether and how long-tailed macaques adapted their task performance as a function of the presence and performance of a social partner. Specifically, we investigated whether subjects' performance changed as a function of the performance of a co-actor (Exp. 1 and 2b) or competitor (Exp. 2a and 2b). In Exp. 3 we gave an equivalent task to human adult participants – who are known to engage in social comparisons in other established paradigms – to compare performance-based social comparison outcomes with the behaviour patterns of the monkeys.

In Exp. 1, we found that neither the presence nor the performance of a conspecific partner affected the monkeys' performance (accuracy and response time) in a touchscreen task. Additional assessment of their behaviour during the test sessions indicated, however, that the monkeys were not particularly interested in the co-actor's task performance in the first place. We take this as a hint that, in contrast to humans, only tasks in which the behaviour of the co-actor has potentially relevant consequences for the monkeys themselves will attract their attention and might potentially trigger social comparison processes. In Exp. 2a and 2b, we aimed at presenting the monkeys with a more relevant and salient setting than the touchscreen setup. When confronted with a new foraging task, the monkeys increased their feeding speed in response to a competing as well as co-acting human partner but they did not adjust the speed to different competitors' feeding rates (Exp. 2a) or when the human partner was merely present but remained passive (Exp. 2a and 2b). Although the underlying cause of the monkeys' faster food retrieval in competition and co-action conditions is unclear (social facilitation or competition), we found the setup in Exp. 2 to be a promising route to study social comparisons in these monkeys. Further fine-tuning of situational parameters and experimental design is necessary to find an optimal procedure that is both relevant enough to elicit potential social comparisons while at the same time avoiding confounds with effects of direct food competition. There are several possibilities to address such fine-tuning.

- A first step could be to test if speed of a co-actor's performance affects retrieval latencies of subjects in a similar setup as the current Exp. 2 – importantly without the partner ever directly competing for food with subjects.

- A systematic manipulation of the strength of competitive threat via variation of physical distance between the food sources and co-actors, or varying the quality of the food rewards might further help to disentangle co-action and competition effects.
- Furthermore, longer trials in which a continuously performing co-actor is present might be needed to allow for co-action effects to manifest in general. This might be especially true for detecting more subtle variations in partner performance, such as a moderately or extremely better performing partner for co-action effects to manifest subsequently.
- It is also possible that we have not found the optimal way, yet, to introduce comparison standards to the monkeys. Different approaches to introduce comparison standards are possible: first, one can provide online feedback about partner performance (as was the case, for example, in the study with humans participants by Seta (1982) as well as in the long-tailed macaque study of Schmitt et al. (2016) and Exp. 1 of this paper); second, short-term exposition to partner performance (for example, similar to the standard exposition phase in Exp. 1 of this paper); finally, long-term exposition to partner performance (for example, similar to the baboon study by Dumas et al. (2017)). The latter approach might be especially relevant for non-human primates, who might form long-term general impressions of their group members and need longer exposition to comparison standards pertaining to a particular domain of competence than humans. This presents an interesting topic for future studies, for both humans and nonhuman animals: What are the effects of long-term and short-term exposition to comparison standards for performing a task – preferably a task that was introduced for the purpose of the experiment, thus no prior information about partner competence is present at start.

An obvious avenue for future research in nonhuman primates is to use conspecifics in a similar task and address more systematically how bond strength or similarity (e.g. same-sex vs different-sex pairings) affect subjects' performance. Assuming that conspecifics are both more relevant and similar to the subjects, co-action effects might look different: Conspecifics are part of the subjects' social network and matter beyond the experimental situation. They are also more "equal" interaction partners in that they are subject to the same experimental restrictions as the subjects themselves (unlike humans, who are usually the creators of those restrictions). While Dumas et al. (2017) and Schmitt et al. (2016) implemented *conspecific* performance as comparison standards, both studies have shortcomings that make it difficult to draw strong conclusions regarding the presence or absence of social comparison processes. For example, we have no information about whether the subjects paid attention to the relevant information and whether they were aware that the other monkey was performing the same task as them and is thus a suitable candidate to compare one's own performance to. Performance-based social comparisons are not very meaningful if evaluations of self and other are based on information from different domains. For example, a monkey who is engaged in a touchscreen game might notice another monkey close-by who is performing particularly impressive acrobatics or who is engaged successfully in an enrichment food retrieval activity, but the first monkey can't engage in a comparison of touchscreen game performance based on the currently available information. After data of the current set of studies was collected, we ran a study with monkeys from group 1 to learn more about the role of a conspecific comparison standard in a simple task and transparent setup: In Keupp et al. (2019), we tested the monkeys with essentially the same simple task as in Exp. 2 using a setup that allowed to test conspecific partners in full view of each other. We presented the subjects with very similar competitive and non-competitive food retrieval situations and with a slow and fast competitor. The monkeys were only affected when a partner's presence and/or actions had potential consequences for food availability (then they retrieved items faster), but not when the partner had no access to the apparatus or when the partner fed from the opposite side of the apparatus, which was out of reach for the subjects (and vice versa the subject's food was out of reach for the partner). The study could not answer whether the monkeys' performance was influenced by different partner performance levels and our sample size did not allow to test for effects of rank and bond strength – hence these remain open questions in need of further exploration.

Another relevant aspect might be a species' social ecology. Primates differ in how tolerant they are with having other group members in close proximity and how individuals interact depending on rank and social bond strength (e.g., Fischer et al., 2017; Thierry, 2007). Such interaction patterns arguably make it more or less useful to collect information about others depending on whether one can actually put that knowledge to use. If tolerance is very limited and social hierarchies inflexible then it might not pay to compare yourself to others on any other dimension than dominance because dominance will determine the outcome of most interactions. For species with more lenient interaction patterns, it might be useful to attend to a larger variety of others' characteristics and behaviours. This acquired knowledge can form the basis for social comparisons. To this end, comparisons between response patterns of more and less tolerant species will be informative. Species ecology might be of interest also in regards to a slightly different yet related topic: it has long been suggested that equity and fairness concerns are based on social comparisons and play a role in the evolution of cooperation (Brosnan, 2011; Fehr & Schmidt, 1999; Silk & House, 2011). Such concerns are very prevalent in humans and this has been studied intensely in fields such as economics and psychology (Fehr & Fischbacher, 2004; Gintis & Fehr, 2012; Güth & Tietz, 1990). Other animals have been found to react to situations where they are worse off than others as well, for example in token exchange paradigms where one individual gets fewer or less preferred rewards than another individual (Rhesus macaques: Hopper et al., 2013; Chimpanzees: Hopper et al., 2014; Long-tailed macaques: Massen et al., 2012; Corvids: Wascher & Bugnyar, 2013). While the underlying cognitive mechanisms of these findings are disputed (Bräuer et al., 2009;

Engelmann et al., 2017), it seems clear that in some test conditions subjects have at least registered the difference in outcomes between what they get and what others get and thus have engaged in some form of comparison. An interesting question is then whether more tolerant or more cooperative species engage in such comparison to higher degrees and consequently react stronger to inequity than less tolerant or less cooperative species.

In Exp. 3 we found that human participants performed slower when paired with a slow partner than when paired with a fast partner. From an evolutionary perspective, performance decrease is not expected, because deliberately opting to forgo one's attainable outcome would be hard to explain in this context. Such behaviour only makes sense considering additional processes, for example taking pity on the partner or complying with cultural norms of not humiliating others, conformity effects, or to signal affiliative motivations. It is thus especially interesting that this is what we saw in our human participants in Exp. 3, where participants differed in their responses to upward and downward comparisons and in fact some explicitly reported engaging in such additional considerations. After performing a self-other comparison, humans might deliberately adjust their behaviour to meet social or normative demands of a particular situation. In addition, such effects might be especially strong in close spatial proximity with the partner and when the incentive structure of the task has no intrinsic value to participants, as was the case in our setup.

On a broader scale, we are facing the question what makes humans so interested in others that even subtle exposition to comparison standards can have an effect on our behaviour and cognitive processing (Mussweiler et al., 2004)? One crucial characteristic of humans is that we have evolved unique cooperative social motivations, something that, according to one theory, was driven by the evolutionary pressure to cooperate (e.g., Tomasello, 2016). Most extant animals, on the other hand, operate in predominantly competitive environments. Early human's need for cooperation had at least two consequences: they needed to look for good cooperation partners and thus for attributes in others detached from acutely competitive interactions – hence, a larger variety of information became relevant. Second, early humans' dependence on each other lead to an understanding of self-other equivalence (Tomasello, 2016, Chapter 2). This broadening of perspective might have fostered an increasing ability to represent others and their performance, to ascribe a certain value to it, and to represent the difference, that is to evaluate this relative to some standard. In addition, human adults appear to automatically process and co-represent the perspective of others in addition to their own (Samson et al., 2010) and represent their own and others' actions in functionally equivalent ways (Sebanz et al., 2003), an ability that emerges at around 4 years of age in children (Milward et al., 2014). Given the reliance of much of human psychology on self-other relations, it appears valid to suggest that the extent to which other animals engage in self-other comparisons might thus be limited by their ability to relate self and other in general.

For the current argument this raises the question: Are we dealing with a multi-layered architecture of social comparison processes in humans, where more sophisticated forms of social comparisons built on a shared competitive component, or are social comparisons in humans and nonhuman animals fundamentally different processes? We propose the following account: Social comparison processes are rooted in a competitive component that ranges from concrete physical competition or direct competition over the same resources to expected or potential competition. This component is shared among humans and other animals, and likely also activated in many of the classic experiments in social comparison research in humans. While nonhuman primates like macaques consider how they fare in relation to others only in immediately competitive conditions, humans (and perhaps some other animals) are – in addition – evaluating how their own performance compares to those of others (or their own expectations about their own performance), and this evaluation process may be mediated by e.g. the relationship with the partner or task relevance for a person’s self-image. For example, Seta (1982) suggested that participants might feel the need to achieve an implicitly estimated acceptable performance level, such as the performance standard set by the co-actor, to please a third party (e.g. the experimenter). Further, Cottrell et al. (1968) have demonstrated that the apprehension of being evaluated by an audience affected participants’ performance in a pseudo-word recognition task. There is even some indication that giving participants the possibility to compare with other participants can increase competitive behaviour (McClintock & McNeel, 1966; McClintock & Nuttin, 1969). Thus, competition-driven social comparisons remain relevant for humans but they can take different forms, for example, when one feels one’s reputation is at stake if one does not perform well. Taken together, this multi-layered conception of social comparison processes provides a framework for exploring how these develop in human children, how adult humans fare under different cognitively demanding conditions, and how the immediacy of competition affects both human and nonhuman subjects.

Compliance with ethical standards

The subjects came from two study populations housed at the German Primate Center and participated voluntarily in the experiments. They were not food or water deprived for testing. During testing, they were separated from the group but visual as well as acoustical contact remained. The experiments were approved by the ethics committee of the Animal Welfare Body of the German Primate Center (permit numbers E7-16 and E4-17) and were classified as non-invasive and exempt from requiring an animal test license by the Lower Saxony State Office for Consumer Protection and Food Safety (LAVES Documents 33.19-42502-04 and 33.19-42502-04-16/2278).

Authors' contributions

SK designed and coordinated the studies, collected data, analyzed the data and drafted the manuscript. JK, CK, LJ and RT substantially contributed to data collection. TM and TB contributed to the development of the research question as well as to the manuscript. JF supported designing the study and contributed to the manuscript. All authors gave final approval for publication.

Data availability

Accompanying data and R-code are available here:

https://osf.io/yk3es/?view_only=2df46ec7d76b42d49115777660282725

Competing interests

The authors have no competing interests.

Acknowledgments

We thank Lukas Schad and Vanessa Wilson for help with the data collection, Ludwig Ehrenreich for technical support, Holger Sennhenn-Reulen for statistical support and Natalia Barbarroja Capdevila and Laura Ahlves for help with coding. We also thank the editor and three anonymous reviewers for their helpful comments on previous versions of the manuscript. This research was funded by the DFG (FI707/18-1 as part of the DFG Forschergruppe FOR2150)

References

Allport, F. H. (1920). The influence of the group upon association and thought. *Journal of Experimental Psychology*, *3*(3), 159-182. https://psycnet.apa.org/doi/10.1037/h0067891

Anderson, J. R., Kuroshima, H., Takimoto, A., & Fujita, K. (2013). Third-party social evaluation of humans by monkeys. *Nature Communications*, *4*, 1561. https://doi.org/10.1038/ncomms2495

Baayen, R. (2008). Analyzing linguistic data. Cambridge, UK.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328. https://doi.org/10.3389/fpsyg.2013.00328

Barton, K. (2017). Package 'MuMIn' Multi-Model Inference. R package version 1.40.0. *https://CRAN.R-project.org/package=MuMIn*.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2016). Package 'lme4'. In *R Package Version 1.1–10*.

Bräuer, J., Call, J., & Tomasello, M. (2009). Are apes inequity averse? New data on the token-exchange paradigm. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *71*(2), 175-181. https://doi.org/10.1002/ajp.20639

Brosnan, S. F. (2011). A hypothesis of the co-evolution of cooperation and responses to inequity. *Frontiers in Neuroscience*, *5*, 43. https://doi.org/10.3389/fnins.2011.00043

Brosnan, S. F., & De Waal, F. B. (2003). Monkeys reject unequal pay. *Nature*, *425*(6955), 297-299. https://doi.org/10.1038/nature01963

Brown, J. D., Novick, N. J., Lord, K. A., & Richards, J. M. (1992). When Gulliver travels: Social context, psychological closeness, and self-appraisals. *Journal of Personality and Social Psychology*, *62*(5), 717-727. https://psycnet.apa.org/doi/10.1037/0022-3514.62.5.717

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). 'Unwilling'versus 'unable': chimpanzees' understanding of human intentional action. *Developmental Science*, *7*(4), 488-498. https://doi.org/10.1111/j.1467-7687.2004.00368.x

Clutton-Brock, T. H., & Albon, S. D. (1979). The roaring of red deer and the evolution of honest advertisement. *Behaviour*, *69*(3), 145-170. https://doi.org/10.1163/156853979X00449

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, *9*(3), 245-250. https://psycnet.apa.org/doi/10.1037/h0025902

Dumas, F., Fagot, J., Davranche, K., & Claidiere, N. (2017, May 31). Other better versus self better in baboons: an evolutionary approach of social comparison. *Proceedings of the Royal Society B: Biological Sciences*, *284*:20170248. https://doi.org/10.1098/rspb.2017.0248

Engelmann, J. M., Clift, J. B., Herrmann, E., & Tomasello, M. (2017). Social disappointment explains chimpanzees' behaviour in the inequity aversion task. *Proceedings of the Royal Society B: Biological Sciences*, *284*:20171502. https://doi.org/10.1098/rspb.2017.1502

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences, 8*(4), 185-190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817-868. https://doi.org/10.1162/003355399556151

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117-140.

Field, A. (2005). *Discovering statistics using SPSS*. Thousand Oaks, CA, US.

Fischer, J., Kitchen, D. M., Seyfarth, R. M., & Cheney, D. L. (2004). Baboon loud calls advertise male quality: acoustic features and their relation to rank, age, and exhaustion. *Behavioral Ecology and Sociobiology*, *56*(2), 140-148. https://doi.org/10.1007/s00265-003-0739-4

Fischer, J., Kopp, G. H., Dal Pesco, F., Goffe, A., Hammerschmidt, K., Kalbitzer, U., Klapproth, M., Maciej, P., Ndao, I., & Patzelt, A. (2017). Charting the neglected West: The social system of Guinea baboons. *American Journal of Physical Anthropology*, *162*, 15-31. https://doi.org/10.1002/ajpa.23144

Gintis, H., & Fehr, E. (2012). The social structure of cooperation and punishment. *Behavioral and Brain Sciences*, *35*(1), 28-29. https://doi.org/10.1017/S0140525X11000914

Güth, W., & Tietz, R. (1990). Ultimatum bargaining behavior: A survey and comparison of experimental results. *Journal of Economic Psychology*, *11*(3), 417-449. https://doi.org/10.1016/0167-4870(90)90021-Z

Herrmann, E., Keupp, S., Hare, B., Vaish, A., & Tomasello, M. (2013). Direct and indirect reputation formation in nonhuman great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus) and human children (Homo sapiens). *Journal of*

Comparative Psychology, *127*(1), 63-75. https://psycnet.apa.org/doi/10.1037/a0028929

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517-523. https://doi.org/10.1016/j.tics.2009.09.004

Hopper, L.M., Lambeth, S.P., Schapiro, S.J., Bernacky, B.J. & Brosnan, S.F. (2013). The ontogeny of social comparisons by rhesus macaques (*Macaca mulatta*). *Journal of Primatology*, *2*(109). doi: 10.4172/2167-6801.1000109

Hopper, L. M., Lambeth, S. P., Schapiro, S. J., & Brosnan, S. F. (2014). Social comparison mediates chimpanzees' responses to loss, not frustration. *Animal Cognition*, *17*(6), 1303-1311. DOI 10.1007/s10071-014-0765-9

Hothorn, T., Bretz, F., & Westfall, P. (2008, Jun). Simultaneous inference in general parametric models. *Biom J*, *50*(3), 346-363. https://doi.org/10.1002/bimj.200810425

Huguet, P., Barbet, I., Belletier, C., Monteil, J.-M., & Fagot, J. (2014). Cognitive control under social influence in baboons. *Journal of Experimental Psychology: General*, *143*(6), 2067-2073. https://psycnet.apa.org/doi/10.1037/xge0000026

Keupp, S., Titchener, R., Bugnyar, T., Mussweiler, T., & Fischer, J. (2019). Competition is crucial for social comparison processes in long-tailed macaques. *Biology Letters*, *15*(3). https://doi.org/10.1098/rsbl.2018.0784

Kitchen, D. M., Seyfarth, R. M., Fischer, J., & Cheney, D. L. (2003). Loud calls as indicators of dominance in male baboons (Papio cynocephalus ursinus). *Behavioral Ecology and Sociobiology*, *53*(6), 374-384. https://doi.org/10.1007/s00265-003-0588-1

Massen, J. J., Van Den Berg, L. M., Spruijt, B. M., & Sterck, E. H. (2012). Inequity aversion in relation to effort and relationship quality in long-tailed macaques (Macaca fascicularis). *American Journal of Primatology*, *74*(2), 145-156. https://doi.org/10.1002/ajp.21014

McClintock, C. G., & McNeel, S. P. (1966). Reward and score feedback as determinants of cooperative and competitive game behavior. *Journal of Personality and Social Psychology*, *4*(6), 606-613. https://psycnet.apa.org/doi/10.1037/h0023986

McClintock, C. G., & Nuttin, J. M. J. (1969). Development of competitive game behavior in children across two cultures. *Journal of Experimental Social Psychology*, *5*(2), 203-218. https://doi.org/10.1016/0022-1031(69)90047-X

McComb, K., Packer, C., & Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, Panthera leo. *Animal Behaviour*, *47*(2), 379-387. https://doi.org/10.1006/anbe.1994.1052

Milward, S. J., Kita, S., & Apperly, I. A. (2014). The development of co-representation effects in a joint task: Do children represent a co-actor? *Cognition*, *132*(3), 269-279. https://doi.org/10.1016/j.cognition.2014.04.008

Mussweiler, T. (2003, Jul). Comparison processes in social judgment: mechanisms and consequences. *Psychological Review*, *110*(3), 472-489. https://doi.org/10.1037/0033-295x.110.3.472

Mussweiler, T., Ruter, K., & Epstude, K. (2004, Dec). The ups and downs of social comparison: mechanisms of assimilation and contrast. *Journal of Personality and Social Psychology*, *87*(6), 832-844. https://doi.org/10.1037/0022-3514.87.6.832

Phillips, W., Barnes, J. L., Mahajan, N., Yamaguchi, M., & Santos, L. R. (2009). ‘Unwilling’versus ‘unable’: capuchin monkeys’(Cebus apella) understanding of human intentional action. *Developmental Science*, *12*(6), 938-945. https://doi.org/10.1111/j.1467-7687.2009.00840.x

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255-1266. https://psycnet.apa.org/doi/10.1037/a0018729

Schielzeth, H., & Forstmeier, W. (2008). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416-420. https://doi.org/10.1093/beheco/arn145

Schmitt, V., Federspiel, I., Eckert, J., Keupp, S., Tschernek, L., Faraut, L., Schuster, R., Michels, C., Sennhenn-Reulen, H., Bugnyar, T., Mussweiler, T., & Fischer, J. (2016). Do monkeys compare themselves to others? *Animal Cognition*, *19*(2), 417-428. DOI 10.1007/s10071-015-0943-4

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: just like one's own? *Cognition*, *88*(3), B11-B21. https://doi.org/10.1016/S0010-0277(03)00043-X

Seta, J. J. (1982). The impact of comparison processes on coactors' task performance. *Journal of Personality and Social Psychology*, *42*(2), 281-291. https://psycnet.apa.org/doi/10.1037/0022-3514.42.2.281

Silk, J. B., & House, B. R. (2011). Evolutionary foundations of human prosocial sentiments. *Proceedings of the National Academy of Sciences*, *108*(Supplement 2), 10910-10917. https://doi.org/10.1073/pnas.1100305108

Tesser, A. (1988). Toward a self-evaluation maintenance model of social behavior. *Advances in Experimental Social Psychology*, *21*, 181-227. https://doi.org/10.1016/S0065-2601(08)60227-0

Tesser, A., Millar, M., & Moore, J. (1988). Some affective consequences of social comparison and reflection processes: the pain and pleasure of being close. *Journal of Personality and Social Psychology*, *54*(1), 49-61. https://psycnet.apa.org/doi/10.1037/0022-3514.54.1.49

Thierry, B. (2007). Unity in diversity: lessons from macaque societies. *Evolutionary Anthropology: Issues, News, and Reviews*, *16*(6), 224-238. https://doi.org/10.1002/evan.20147

Tomasello, M. (2016). *A natural history of human morality*. Harvard University Press.

Triplett, N. (1898). The dynamogenic factors in pacemaking and competition. *The American Journal of Psychology*, *9*(4), 507-533. doi:10.2307/1412188

Wascher, C. A., & Bugnyar, T. (2013). Behavioral responses to inequity in reward distribution and working effort in crows and ravens. *PLoS One*, *8*(2), e56885. https://doi.org/10.1371/journal.pone.0056885

Whittemore, I. C. (1924). The influence of competition on performance: An experimental study. *The Journal of Abnormal Psychology and Social Psychology*, *19*(3), 236-253. https://psycnet.apa.org/doi/10.1037/h0063946

Wilson, M. L., Hauser, M. D., & Wrangham, R. W. (2001). Does participation in intergroup conflict depend on numerical assessment, range location, or rank for wild chimpanzees? *Animal Behaviour*, *61*(6), 1203-1216. https://doi.org/10.1006/anbe.2000.1706

Zajonc, R. B. (1965, Jul 16). Social Facilitation. *Science*, *149*(3681), 269-274. https://doi.org/10.1126/science.149.3681.269

Figures and Tables

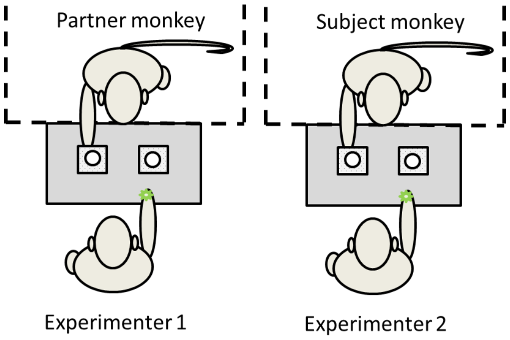

Figure 1 Schematic view of the test setup of Study 1. Each monkey performed the picture discrimination task on a touchscreen mounted in roughly 45° horizontal angle in front of each cage and were rewarded by an experimenter for correct responses, see also Figure S1 for pictures of the setup and the visibility of the neighbor's touchscreen.

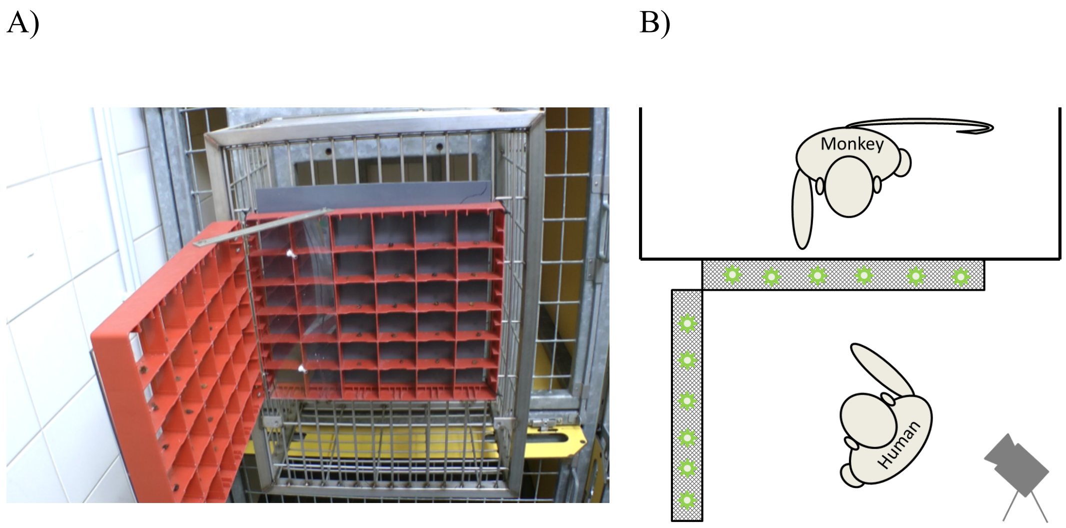

Figure 2 Experimental Setup of Exp. 2b. A) shows a picture of the baited feeding boards with closed occluder. B) is a bird-eye view schematic depiction of the setup with positions of monkey and human partner. In Exp. 2a, only the frontal feeding board was in place.

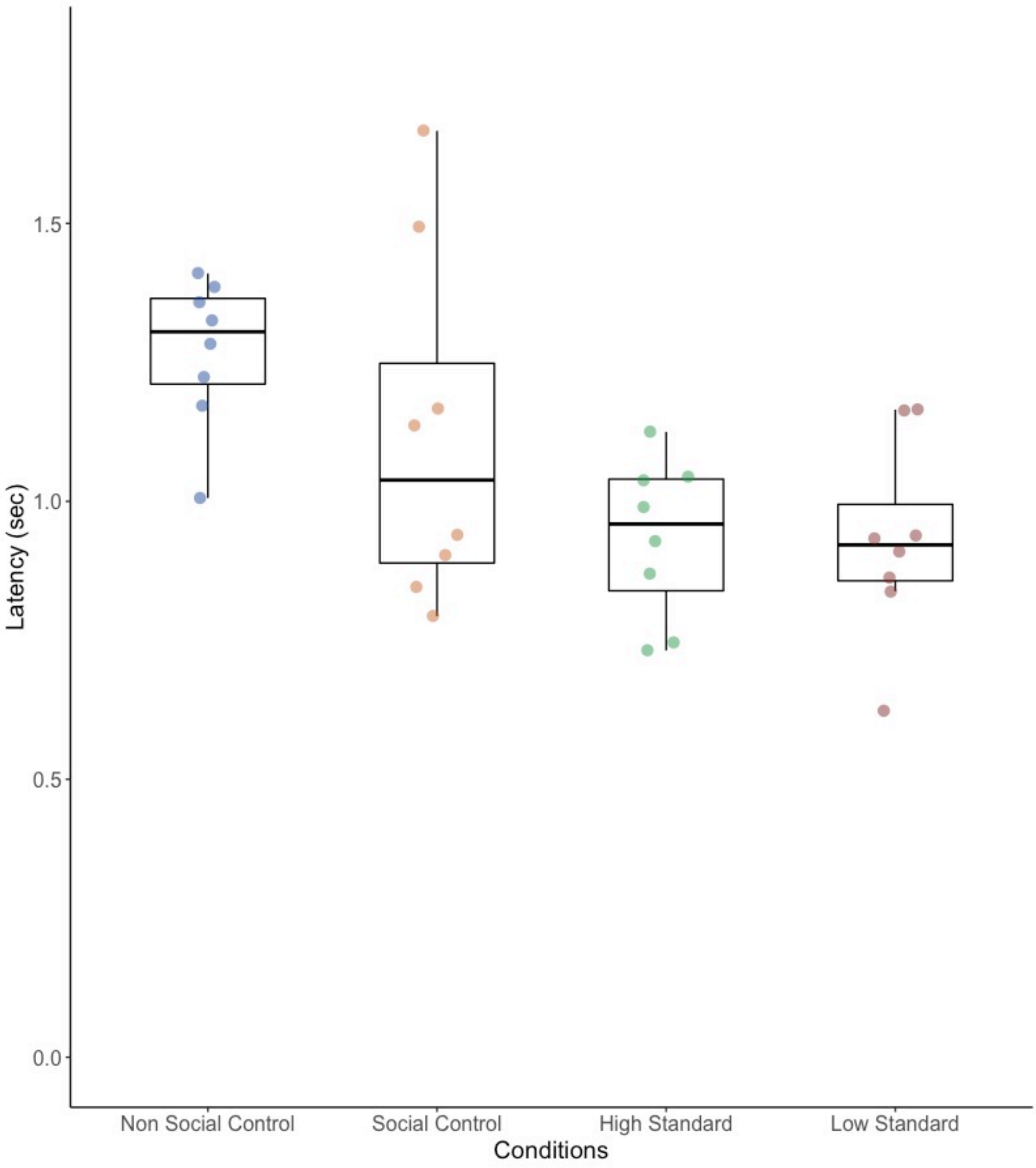

Figure 3. Mean latencies to take the next food item of each individual in each condition in Experiment 2a. In this and all other boxplots in this manuscript, horizontal lines represent median (thick line) and 25th & 75th percentiles; Whiskers extend to smallest and largest value within 1.5 * interquartile range; colored points represent the average latency per participant per condition.

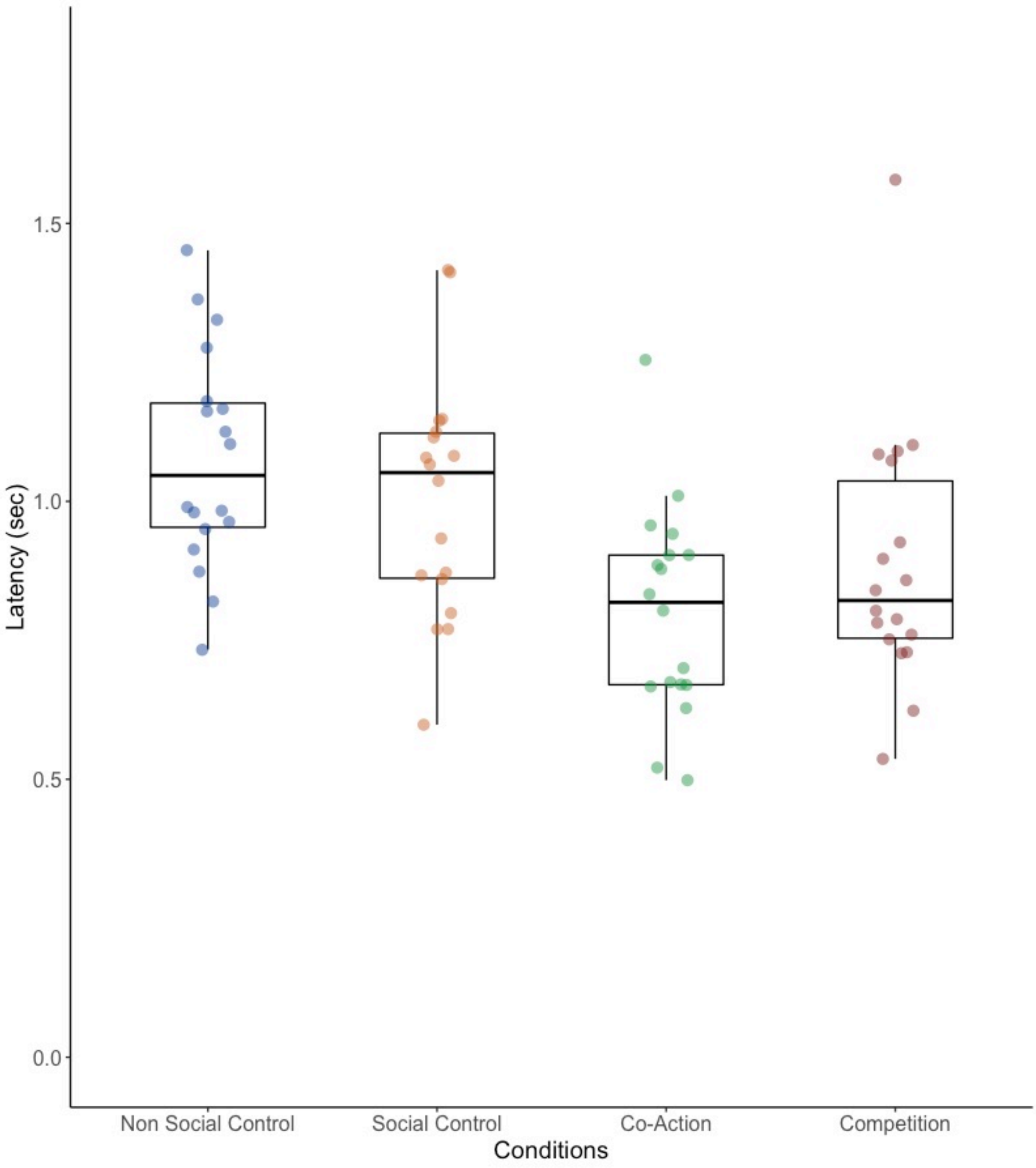

Figure 4. Mean latencies to take the next food item for each individual in each condition in Exp. 2b.

Figure 5. Setup for co-action condition in Experiment 3 (SA1 indicates location of participant, SA 2 indicates location of confederate).

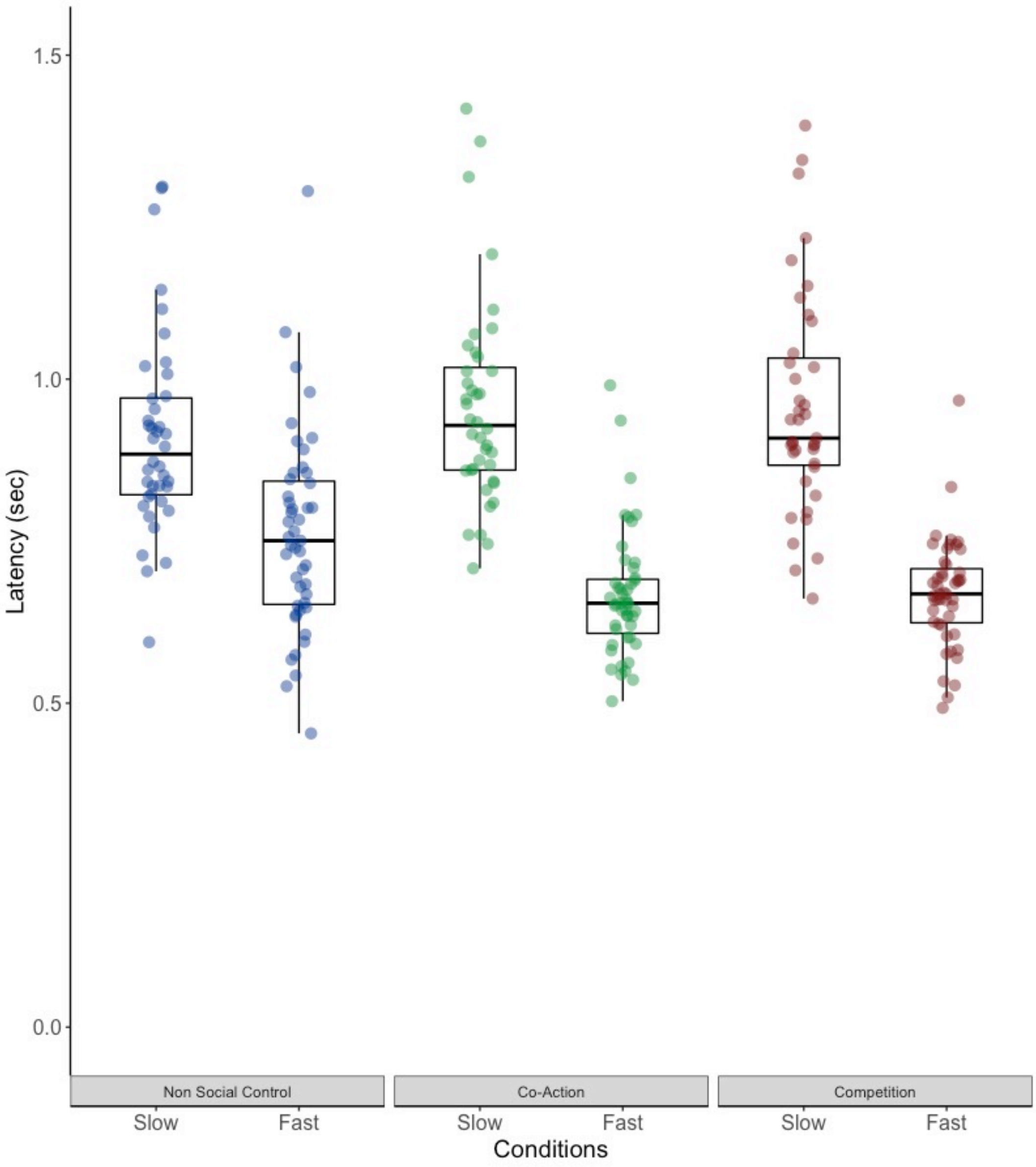

Figure 6. Response latencies as a function of standard condition and action context for Exp. 3.

Table 1

Descriptive statistics for accuracy and reaction time by condition in Experiment 1.

Condition	Proportion of correct responses (SD)	Mean latencies (SD) in msec	
		Correct trials	Incorrect trials
High standard	.62 (.49)	1502 (943)	1338 (843)
Low standard	.63 (.48)	1516 (856)	1409 (711)
Nonsocial control	.69 (.46)	1543 (851)	1495 (905)
Social control	.65 (.47)	1473 (874)	1681 (1507)

Table 2

Experiment 2a: Results for individual predictors for latency (reference category: baseline)

Term	Estimate	SE	$CI_{2.5}$	$CI_{97.5}$	χ^2	Df	p-value
Intercept	0.071	0.064	-0.049	0.193			
High standard(1)	-0.180	0.058	-0.303	-0.076	18.86	3	< .001
Low standard(1)	-0.182	0.058	-0.298	-0.064			
Social control(1)	0.074	0.078	-0.072	0.219			
Trial	-0.186	0.029	-0.242	-0.132			

(1) Condition was dummy coded with baseline as the reference category. The indicated test refers to the overall effect of condition

Table 3

Experiment 3: Results of full model for individual predictors for latency (reference category: slow group alone condition)

Term	Estimate	SE	CI$_{2.5}$	CI$_{97.5}$	χ^2	Df	p-value
Intercept	-0.103	0.023	-0.150	-0.060	[1]		[1]
Context Co-action	0.045	0.013	0.018	0.007	[1]		[1]
Context Competition	0.039	0.018	0.002	0.077	[1]		[1]
Standard fast	-0.191	0.032	-0.247	-0.128	[1]		[1]
Trial	-0.049	0.006	-0.060	-0.037	48.341	1	<.001
Co-action × fast[2]	-0.166	0.018	-0.206	-0.128	73.002	2	<.001
Competition × fast[2]	-0.154	0.025	-0.211	-0.104			

[1] not shown because of limited interpretation

[2] the indicated test was obtained from a likelihood ratio test comparing the full model with a model not comprising the interaction term

Experiment 1

Table S1

List of subjects in Study 1

Name	Sex	Birthday	Reward category
Ilja	m	29.12.2012	Male
Snickers	m	12.1.2014	Male
Isaak	m	10.04.2011	Female
Lenny	m	10.04.2009	Male
Mars	m	17.01.2014	Female
Max	m	01.02.2013	Female
Paul	m	05.06.2007	Female
Sophie	f	03.04.2009	Female
Mila (co-actor)	f	07.04.2012	Male

Figure S1 Test setup with two adjacent compartments and folding tables. A) Left compartment: co-actor, right compartment: subject. B) Monkeys could reach outside the cage to touch the screen (stimuli always appeared at the same locations and could be touched through holes in a transparent security panel, ∅ = 1cm, 17 cm apart).

Training procedure

Circle/Triangle discrimination task

For each trial, a circle and a triangle were presented simultaneously on the screen. Each stimulus was surrounded by a 269 x 269 pixel-sized black frame. We used six different colors (RGB values: blue 0-128-255, yellow 255-242-0, green 34-177-76, orange 255-127-39, red 255-0-0, and violet 128-0-255). Half of the monkeys were rewarded for touching circles and the other half for touching triangles. They learned that an incorrect response was followed by a red screen lasting for 5 seconds and a correct response was followed by a white screen and a food reward from the experimenter. The next trial was started upon a key press by the experimenter after she had given the food reward, or after the timeout had ended.

Once an individual reached the training criterion (>14 of 20 correct responses) in two consecutive sessions, they moved on to the next training stage. Those individuals who succeeded in the circle/triangle discrimination training needed on average 146 trials to reach criterion (range: 40 - 260 trials).

Male/Female discrimination task

The male/female training was identical to the circle/triangle task but instead of circles and triangles, the monkeys saw pictures of Caucasian men and women. We presented three categories of pictures: full body, half upper body or face only. We re-used some of the pictures of Schmitt et al. (2016) and gathered additional pictures from various Internet sources (see Table S2). Each stimulus was cropped, presented on a white background and surrounded by a 269 x 269 pixel-sized black frame (see figure S2). Originally, for five individuals, the male pictures served as the positive stimuli (S+) and for seven individuals, the female pictures. Of the remaining subjects in our final sample, three were in the S+ male and five in the S+ female reward category. Those individuals who were successful in the male/female training needed on average 311 trials to reach criterion (range: 140 – 680 trials).

Figure S2 Examples of female test stimuli. We presented pictures of full body, half upper body and face.

Table S2

Sources of stimuli for the male/female discrimination task

Faces	Half body	Full body
Park aging mind laboratory http://agingmind.utdallas.edu/stimuli	http://www.witt-weiden.de	http://store-de.hugoboss.com
pics.psych.stir.ac.uk (Aberdeen)	http://www.baur.de	http://www.hm.com
	http://www.sheego.de	http://www.esprit.de
	http://www.hm.com	http://www.c-and-a.com
	http://www.pullandbear.com	http://www.kleider-kunst.de
	http://www.urbanoutfitters.de	http://www.soliver.de
	http://www.albamoda.de	http://www.michaelax.de
	http://store.americanapparel.net	http://www.burdastyle.de
	http://www.bandyshirt.com	http://www.wellner.modehaus.de
	http://www.peterhahn.de	http://dfb.de
	http://www.easy+D66youngfashion.de	www.pepejeans.com
	http://www.sachenonlinekaufen.de	https://www.bugatti-fashion.com/
	http://content.yancor.de	http://www.tchibo.de/
	http://www.arqueonautas.de	
	http://www.trachten24.eu	
	http://www.kademo.de	
	http://c-and-a.com	
	http://witt-weiden.de	
	http://www.adidas.de/	

Experiment 2a

Table S3

List of subjects for Experiment 2a

Name	Sex	Birthday
Ilja	m	29.12.2012
Snickers	m	12.01.2014
Mars	m	17.01.2014
Max	m	01.02.2013
Linus	m	06.01.2013
Lord	m	04.02.2014
Mila	f	07.04.2012
Milka	f	29.12.2014

Table S4

Experiment 2a: Presentation order of conditions (note that Sessions 10 and 18 are named 'Baseline' because no co-actor was present, but were not used for calculation of initial baselines)

Session	Order 1 (n = 3)	Order 2 (n = 5)
1-3	Baseline	Baseline
4	Social control	Social control
5	Fast	Slow
6	Slow	Fast
7	Slow	Fast
8	Fast	Slow
9	Fast	Slow
10	Baseline	Baseline
11	Slow	Fast
12	Fast	Slow
13	Fast	Slow
14	Slow	Fast
15	Slow	Fast
16	Fast	Slow
17	Slow	Fast
18	Baseline	Baseline
19-20	Social control	Social control

Table S5

Experiment 2a: Pairwise comparisons between the different conditions
(95 % family wise confidence intervals for pairwise multiple comparisons)

	Estimate	Lower boundary	Upper boundary
Social control - Baseline	0.074	-0.125	0.273
High standard – Baseline*	-0.179	-0.328	-0.031
Low standard – Baseline*	-0.182	-0.329	-0.034
High standard – Low standard	0.002	-0.134	0.137
High standard – Social control*	-0.254	-0.435	-0.072
Low standard – Social control*	-0.256	-0.438	-0.074

* indicates a significant difference between the compared conditions

Additional exploratory analyses

To address the effect of decreasing retrieval latency across trials in Exp. 2a, we additionally assessed how latencies developed for the baseline/alone condition and for the social control condition. We found that latencies differed between baseline trials (χ^2 = 16.981, *df* = 4, *p* = .002, *conditional* R^2 = 0.362). We ran the three post-hoc comparisons that we thought most meaningful, namely between trial 3 (the last baseline trial before monkeys say any other test condition – we think this more meaningful than using trial 1 because performance could stabilize by then), trial 10 (baseline/alone trial midway through the experiment), and trial 18 (the last baseline/alone trial). The monkeys were faster in the intermediate and last trial compared to the first trial; there was no difference between trial 10 and trial 18.

Table S5a

Study 2a: Comparison of baseline/alone trials

(95 % family wise confidence intervals for pairwise multiple comparisons)

	Estimate	Lower boundary	Upper boundary
First – Intermediate*	0.350	0.055	0.645
First – Last*	0.515	0.220	0.809
Intermediate – Last	0.165	0.130	0.460

* indicates a significant difference between the compared conditions

There was also a difference between social control trial at the beginning (*m* = 1.504) versus end of the experiment (*m* = 0.853) as assed by a paired-samples t-test (*t*(7) = 5.063, *p* = .002).

An additional analysis comparing only the last trial of each condition revealed that there was no difference between conditions towards the end of the experiment ($\chi^2 = 5.299$, *df* = 3, *p* = .151, *conditional* $R^2 = 0.525$). The monkeys were on average still faster in both competitive conditions (m_{low}=.734, m_{high} = .921) than in the nonsocial control condition (m_{NSC} = .998), but these differences did not reach statistical significance. Their performance in the social control condition (m_{SC} = .999) similarly didn't differ from the other conditions. This finding should be evaluated with caution, however. It fits with the notion than increasing uncertainty how E2 will behave led the monkeys to preventively increase their feeding speed irrespective of condition. It could, however, also be a consequence of the reduced amount of data and an accompanying lack of power to detect differences between conditions.

Experiment 2b

Table S6

List of subjects for Experiment 2b and presentation order of conditions

Subject	Sex	Date of birth	Group	Order of presented condition blocks[1)]
Ilja	m	29.12.2012	1	Social control – Co-action – Competition
Snickers	m	12.1.2014	1	Co-action– Social control– Competition
Mars	m	17.01.2014	1	Social control– Co-action– Competition
Max	m	01.02.2013	1	Social control– Competition– Co-action
Lord	m	04.02.2014	1	Competition– Social control– Co-action
Mila	f	07.04.2012	1	Co-action– Competition– Social control
Milka	f	29.12.2014	1	Competition– Co-action– Social control
Sambia	f	18.02.2015	1	Co-action– Competition– Social control
Selina	f	20.05.2008	1	Competition– Co-action– Social control
Paul	m	05.06.2007	1	Competition– Social control– Co-action
Maja	f	17.10.2007	1	Co-action– Social control– Competition
Bomby*	f	22.06.1998	2D	Social control – Co-action – Competition – Co-action – Social control
Penny*	f	07.04.2000	2D	Social control – Co-action – Competition – Co-action – Social control
Giulia	f	19.04.2000	2D	Social control – Co-action – Competition – Co-action – Social control
Lilly	f	13.12.2009	2D	Social control – Co-action – Competition – Co-action – Social control
Fenja	f	12.05.2003	2E	Social control – Co-action – Competition – Co-action – Social control
Heidi	f	20.05.2001	2E	Social control – Co-action – Competition – Co-action – Social control
Flocke*	f	21.03.1999	2E	Social control – Co-action – Competition – Co-action – Social control
Bella	f	02.04.2000	2F	Social control – Co-action – Competition – Co-action – Social control
Leni	f	10.08.1994	2F	Social control – Co-action – Competition – Co-action – Social control
Kümmel	f	04.01.2011	2F	Social control – Co-action – Competition – Co-action – Social control

* individuals who refused to participate in the competition condition

[1)] all subjects received baseline conditions prior to the experimental conditions (3 baseline trials for group 1 and 6 baseline trials for group 2); every block consisted of 6 trials

Table S7

Experiment 2b: Results for individual predictors for latency (reference category: baseline)

Term	Estimate	SE	$CI_{2.5}$	$CI_{97.5}$	χ^2	Df	p-value
Intercept	-0.040	0.055	-0.152	0.064			
Co-action(1)	-0.220	0.039	-0.292	-0.152	32.193	3	<.001
Competition(1)	-0.135	0.048	-0.227	-0.037			
Social control(1)	0.013	0.043	-0.073	0.095			
Trial	-0.047	0.015	-0.077	-0.017			

(1) Condition was dummy coded with baseline as the reference category. The indicated test refers to the overall effect of condition

Table S8

Experiment 2b: Pairwise comparisons of different conditions
(95 % family wise confidence intervals for pairwise multiple comparisons)

	Estimate	Lower boundary	Upper boundary
Co-action – Competition	-0.085	-0.186	0.016
Co-action – Baseline*	-0.219	-0.315	-0.124
Competition – Baseline*	-0.135	-0.252	-0.017
Social control – Baseline	0.013	-0.092	0.117

* indicates a significant difference between the compared conditions

Additional exploratory analyses

Additional assessment of latencies in first and second block of social control trials revealed that the monkeys tended to feed faster in block 2 (m = 1.023) compared to block 1 (m = 1.169) ($t(6)$ = 2.418, p = .052). To address the potential influence of experiencing food loss by E2 on later trials, we additionally analysed the first block of each condition for group 2 (due to differences in counterbalance and previous experience with Exp. 2, group 1 data cannot be included in this analysis). We found a significant effect of condition (χ^2 = 11.762, *df* = 2, p = .003, *conditional* R^2 = 0.559). Post hoc tests revealed that latencies were significantly faster in co-action compared to social control conditions (see Table S8a).

Table S8a

Study 2b: Comparison of first block per condition

(95 % family wise confidence intervals for pairwise multiple comparisons)

	Estimate	Lower boundary	Upper boundary
Social control – Co-action*	0.227	0.109	0.345
Competition – Co-action	0.012	-0.119	0.142
Social control – Competition*	0.215	0.085	0.346

* indicates a significant difference between the compared conditions