Animal Behavior and Cognition
ABC 2025, 12(3):330-359
DOI: https://doi.org/10.26451/abc.12.03.02.2025

Choose Your Partner: Social Evaluation of Skillfulness at Cooperative Co-Action Tasks in Tonkean Macaques (*Macaca tonkeana*)

Marie Hirel[1,2], Hélène Meunier[4,5], Roger Mundry[1,2,3], Hannes Rakoczy[3,6], Julia Fischer[1,2,3], and Stefanie Keupp[1,2,3]

[1] Georg-August-Universität Göttingen, Johann-Friedrich Blumenbach Institute, Department for Primate Cognition, Göttingen, Germany
[2] German Primate Center – Leibniz Institute for Primate Research, Cognitive Ethology Laboratory, Göttingen, Germany
[3] Leibniz ScienceCampus, German Primate Center – Leibniz Institute for Primate Research, Göttingen, Germany
[4] Centre de Primatologie de l'Université de Strasbourg, Niederhausbergen, France
[5] Laboratoire de Neurosciences Cognitives et Adaptatives, Strasbourg, France
[6] Georg-August-Universität Göttingen, Georg-Elias-Müller Institute for Psychology, Department for Cognitive Developmental Psychology, Göttingen, Germany

*Corresponding author (Email: mhirel@dpz.eu)

Citation – Hirel, M., Meunier, H., Mundry, R., Rakoczy, H., Fischer, J., & Keupp, S. (2025). Choose your partner: Social evaluation of skillfulness at cooperative co-action tasks in Tonkean macaques (*Macaca tonkeana*). *Animal Behavior and Cognition, 12*(3), 330-359. https://doi.org/10.26451/abc.12.03.02.2025

Abstract – Social evaluation – inferring individual characteristics of others from their past behaviors – is an adaptive strategy that helps to inform social decisions. However, how nonhuman primates form and use impressions about others to select their social partners strategically is still unclear. In this study, we investigated whether Tonkean macaques, *Macaca tonkeana*, can spontaneously use information, acquired by observation, to choose the optimal human partners for cooperation in two co-action tasks from the same domain. The subjects (N=5) did not prefer to cooperate with the skillful human partner compared to the unskilled human partner, irrespective of the task or how much attention they paid to the human partners' actions solving a solo version of one task prior to the test. The probability of optimal choices did not increase through trials either, indicating no learning by experience with the human partners across the 16 test trials. Our results contradict findings of previous studies that tested monkeys in different domains of competence and contexts, and thus encourage further investigation on monkeys' social evaluation abilities. Our experimental design presents a promising way to investigate different contexts, the inter-individual differences, the different types of social information involved, and the cognitive mechanisms underlying social evaluation and partner selectivity in nonhuman primates.

Keywords – Social evaluation, Cooperation, Partner choice, Nonhuman primates, Social cognition, Co-action tasks

Many nonhuman primates live gregarious lifestyles in which they are confronted with social decisions on a daily basis. For example, they choose conspecific partners in agonistic contexts such as recruitments for coalitions (Perry et al., 2004; Silk, 1999; Silk et al., 2004; Slocombe & Zuberbühler, 2007; Young et al., 2014), in affiliative contexts such as grooming (Mielke et al., 2018) as well as for cooperation (Molesti & Majolo, 2016; Suchak et al., 2014), or to learn foraging techniques from (Horner et al., 2010; Kendal et al., 2015). Such decisions may be based on existing social bonds and the individual's knowledge of its group members' relationships (Borgeaud et al., 2013; Whitehouse & Meunier, 2020; Wittig et al., 2007, 2014). In addition, social decisions may be based on the assessment of other individuals' characteristics. For example, knowing about physical strength or skills of a conspecific is advantageous for deciding whether to enter a competition over resources. In some situations, such as recruiting a partner for cooperation, picking a model to learn from, or choosing a partner to form coalitions with, individual characteristics may be as important as relationship status for optimal outcomes. Information about others can be obtained from interactions, observations, or, specifically in humans, gossip from third-party individuals (Sommerfeld et al., 2007). Social evaluations are an important part of human cognitive capacities and inform decisions on whom to trust, whom to work with, whom to learn from, or whom to compare to in current as well as future encounters (Festinger, 1954; Hermes et al., 2016, 2018). Yet, whether and to what extent social information is gathered by nonhuman primates and how it affects their partner selectivity is not yet fully understood.

Several nonhuman primates seem to possess social evaluation skills at least to some degree. For example, they differentiated between prosocial/cooperative and antisocial/uncooperative individuals. Chimpanzees (*Pan troglodytes*) preferentially chose “nice” over “mean” experimenters (Herrmann et al., 2013; Russell et al., 2008; Subiaul et al., 2008), brown capuchins (*Sapajus apella*) preferentially avoided non-helpful experimenters (Anderson et al., 2013), and common marmosets (*Callithrix jacchus*) preferred to approach individuals who engaged in positive cooperative vocal interactions (Brügger et al., 2021). Common marmosets, brown capuchins, and squirrel monkeys (*Saimiri sciuerus*) also discriminated against non-reciprocators by accepting less food from a human who refused to exchange food with a third party before (Anderson et al., 2016; Anderson et al., 2013; Kawai et al., 2014). This preference for prosocial and cooperative social agents supports the hypothesis that partner choice based on social evaluation is a key adaptive strategy promoting and maintaining cooperation (Manrique et al., 2021; Wu et al., 2016). For successful cooperation, however, skillfulness in the task at hand is also important; consequently, it’s advantageous to evaluate and consider other individuals’ skill sets when recruiting collaborators.

Nonhuman primates indeed seem to form impressions about the expertise of others and target their attention towards these ‘experts’ in non-cooperative contexts. For example, long-tailed macaques (*Macaca fascicularis*), vervet monkeys (*Chlorocebus pygerythrus*), and Guinea baboons (*Papio papio*) changed their behavior toward the only individual in the group that was able to provide food in experimental situations (Fruteau et al., 2009; O’Hearn et al., 2025; Stammbach, 1988). Long-tailed macaques also differentiated between an experimenter who repeatedly proved capable of opening a box with food rewards and an unsuccessful experimenter who failed at it (Placì et al., 2019). Brown capuchins preferentially observed the more skillful nutcrackers within their group, rather than the lesser skilled ones, independently of social affiliations or social proximity (Ottoni et al., 2005; Resende et al., 2004), and chimpanzees preferentially copied the most expert members of their group at a foraging technique (Horner et al., 2010; Kendal et al., 2015). Yet, whether nonhuman primates can evaluate the skillfulness of partners to select them strategically for cooperation is still an open question. Only chimpanzees have been tested in this context and they strategically selected a conspecific cooperator who had successfully helped them in the same situation in the past (Melis et al., 2006, 2008). A recent study extended these findings by showing that directly experiencing outcomes of successful interactions was not necessary for chimpanzees to evaluate conspecifics’ skills in cooperative tasks (Keupp & Herrmann, 2024).

Despite these findings, the cognitive mechanisms underlying nonhuman primates’ social evaluation skills require further investigation. Particularly, it is still an open question whether primates can infer future behavior only for similar matching behaviors or whether they can make wider generalizations of a model’s characteristics and resulting behaviors across contexts. From the human literature, we know about at least three inference types that play a role in social decision-making: behavior matching, i.e., inferences based on the similarity of the past and current situations; global impression formation, i.e., making wider generalization of a model’s characteristics across contexts; and rational trait reasoning, i.e., rational trait-like inferences in context-specific ways (Fusaro et al., 2011; Hermes et al., 2015, 2016; Titchener et al., 2023). For nonhuman primates, it is difficult to disentangle whether they express social reasoning skills corresponding to either of these inferences or rather to associative learning strategies from their past interactions (Heyes, 2012). For instance, the response pattern of chimpanzees in Melis et al. (2006) spoke in favor of decisions based on success in the previous trial rather than in favor of a general evaluation of skillfulness. Chimpanzees indeed appeared to rely on both direct interactions (Melis et al., 2006) and indirect third-party observations (Herrmann et al., 2013; Russell et al., 2008; Subiaul et al., 2008). By prompting chimpanzees to recruit conspecific partners to perform with at different co-action tasks in cooperation as well as competition, Keupp & Herrmann (2024) recently provided evidence that chimpanzees can draw domain-specific inferences about conspecifics' skills.

In the current study, we investigated whether Tonkean macaques (*Macaca tonkeana*) can choose optimal partners in cooperative co-action tasks and whether they spontaneously (i.e., from the first trials) use information acquired by observation to do so. Subjects had several opportunities to sample information about two human partners' skillfulness (demonstration phase). One human partner was a skillful performer whereas the other failed most of the time. Subjects then had to choose one of the human partners to cooperate with on a familiar co-action task (subjects observed the human partners operating this task) and on a novel one (subjects had never seen the human partners operating this task before). We were interested in their partner choices – the optimal strategy to maximize reward outcomes being to pick the human partner who was formerly skillful. Optimal choices from the first trial on would indicate a spontaneous preference for the skillful human partner based solely on observation, suggesting an evaluation of the human partners' skills, while optimal choices increasing with trials would indicate a learning effect based on direct experience with the human partners. We also investigated whether the subjects' attention during the demonstration phase was predictive of the probability to make optimal partner choices. Contrasting subjects' partner choices in a novel task condition and a familiar task condition allowed us to investigate whether they generalized their social evaluation across different tasks from the same domain.

We studied this question in Tonkean macaques, a species living in multi-male multi-female societies which is characterized as highly tolerant and can display cooperative behaviors (Petit et al., 1992; Riley, 2010; Thierry et al., 1994). In a tolerant social system, individuals can frequently interact with group members, also outside their own matriline, without risking severe aggression (Duboscq et al., 2013; Thierry et al., 1994, 2000). Consequently, they can interact with a broad range of individuals in diverse situations, where knowledge about characteristics of these individuals might come in handy. Tonkean macaques have also demonstrated good knowledge of individuals' social relationships in their group (Whitehouse & Meunier, 2020) as well as expressed other social cognitive skills (Canteloup et al., 2016, 2017; Canteloup & Meunier, 2017). In light of the social and cognitive characteristics of this species and previous findings revealing social evaluation abilities in other nonhuman primates, we predicted that our subjects would preferably choose the skillful human partner from the very first trial on, in both tasks. We also expected that the more an individual observed the human partners during the demonstration phase, the more relevant information this individual can obtain and, therefore, the greater the probability that this individual makes optimal choices afterward.

Methods

Ethics Statement

This study respects the European ethical standards and regulations (Directive 2010/63/UE) and has been approved by the internal ethical committee of the Centre de Primatologie – Silabe de l'Université de Strasbourg (SBEA 2022-03, registration n° B6732636). We only used positive reinforcement and all individuals participated on a voluntary basis, i.e., they were never forced to enter, be isolated, or to perform the tasks in the experimental rooms. Their daily feeding regime was not affected by cognitive testing and water was available *ad libitum* at all times.

Subjects and Study Site

This study was conducted at the Centre de Primatologie – Silabe de l'Université de Strasbourg, in France, between March 2022 and July 2023. All the subjects lived in one social group of 31 individuals (17 females), aged from three months to 27 years, living in a wooded outdoor enclosure of 3788 m^2 with constant access to an indoor room of 20 m^2. Individuals were fed once per day with dry pellets and once per week with fruits and vegetables, and had access to water *ad libitum*. They participated in cognitive experiments on a voluntary basis in experimental rooms situated next to their outdoor enclosure (see Figure S1 in the Supplementary Materials). All the individuals in the group could participate in the experiment without specific criteria, with the exception of being comfortable being separated from their group in the experimental rooms. 15 individuals (7 females) started the training phase (information on the training procedure in the Supplementary Materials). These 15 individuals were aged from 2 to 23 years at the start of the training, all captive-born and mother-raised in their social group in the same enclosure at the Centre de Primatologie – Silabe de l'Université de Strasbourg (Table S1). Only five of them (one female; age range at the start of the experiment: 10-24 years; Table 1) were coming regularly enough in the experimental rooms and managed to pass all training steps to be tested as subjects in the experiment. The tested individuals had participated in several ethological studies before (e.g., Ballesta & Meunier, 2023; Canteloup & Meunier, 2017; Hirel et al., 2020; Miss et al., 2022; Whitehouse & Meunier, 2020), but they had never been tested with social evaluation experiments before.

Table 1

Overview of the Subjects and their Choices

Name	Sex	Age (Years)	Initial Preference (# Skillful)	Skillful Partner	First Task	Test Session 1 (# Skillful)	Test Session 2 (# Skillful)	Averages ± SDs of # Skillful
Abricot	M	10	2/8	Actor A	Novel	4/8	6/8	5 ± 1.4
Alaryc	M	10	1/8	Actor A	Familiar	2/8	3/8	2.5 ± 0.7
Barnabé	M	9	4/8	Actor B	Novel	2/8	6/8	4 ± 2.8
César	M	8	3/8	Actor B	Familiar	3/8	4/8	3.5 ± 0.7
Néréis	F	24	3/8	Actor B	Novel	6/8	6/8	6 ± 0

Note. Presented are subject characteristics, the identity of the human partners, test condition (familiar or novel), and how often subjects chose the skillful partner during initial preference assessment and test trials.

Materials

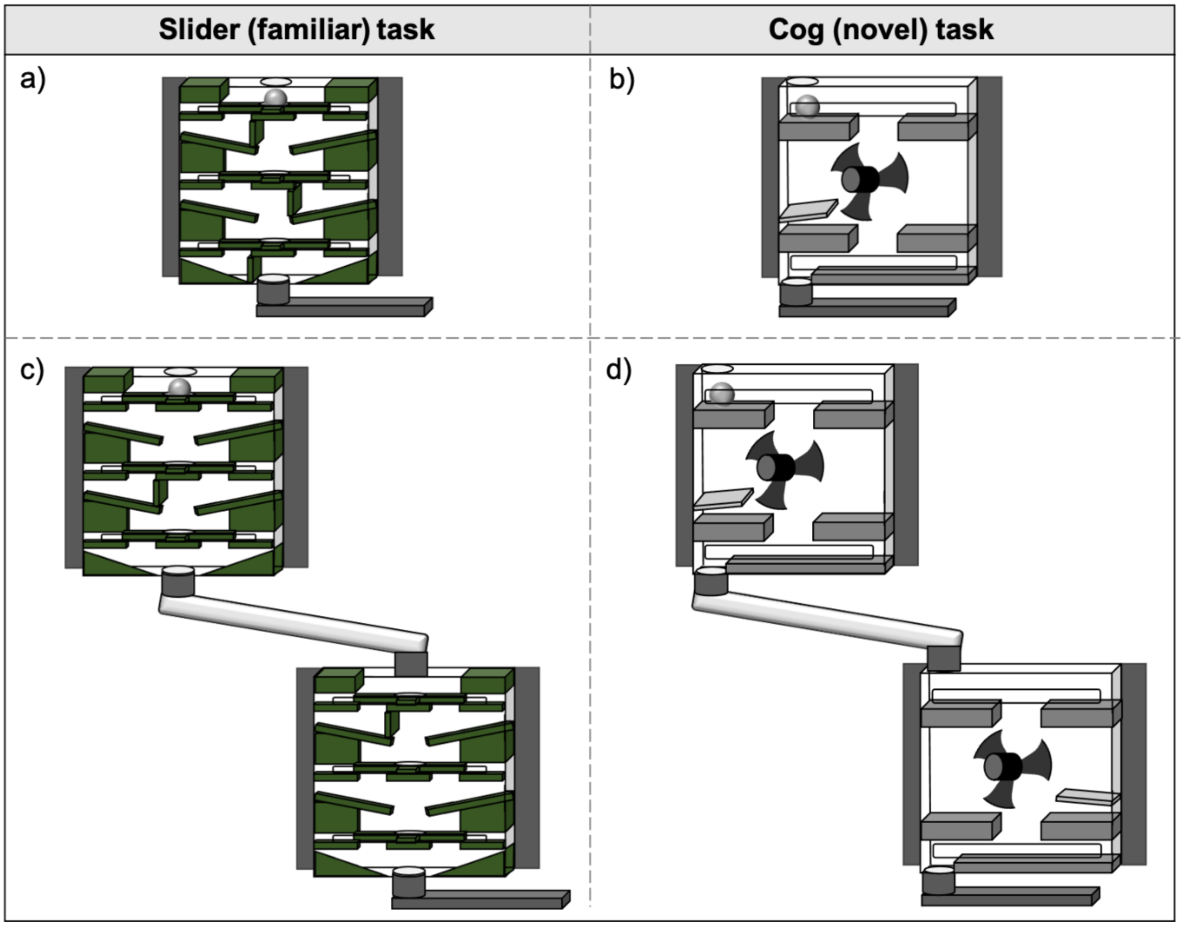
Puzzle Box Tasks

Subjects were tested with two different types of puzzle boxes: the Slider task and the Cog task (Figures 1a and 1b). These tasks were the same as those used in the study on chimpanzees by Keupp & Herrmann (2024). These tasks required subjects to navigate a ball inside the boxes from top to bottom and to decide where to move the ball to overcome obstacles and traps. The subjects were rewarded with a piece of fruit once the ball fell into an aluminum tray at the bottom of the box. The apparatuses had a size of 0.25 m x 0.25 m and were fixed to the mesh of the experimental rooms from the outside. The monkeys could use their fingers to operate the ball through the mesh (size of the apertures in the mesh: 5 x 5 cm). A ball of 28 mm diameter could be inserted on top of the boxes.

The puzzle boxes required different actions. The Slider task was a transparent puzzle box with three levels. On each level, the ball was caught on a slider. By moving the slider to the left or to the right, the ball could be transported towards a gap that let it fall down to the next level. Under each level, traps could be inserted on either side to block the ball from falling down to the next one. If the ball was navigated to the blocked side, it was trapped and the attempt was considered a failure. The Cog task required moving the ball from the top level into the middle level, where a cog wheel had to be turned left or right to move the ball further down. Traps could be inserted on either side of the cog wheel: once placed, they blocked the ball from falling down on the respective side and the ball could not be retrieved.

Figure 1

Single and Cooperative Co-Action Setups of the Puzzle Box Tasks



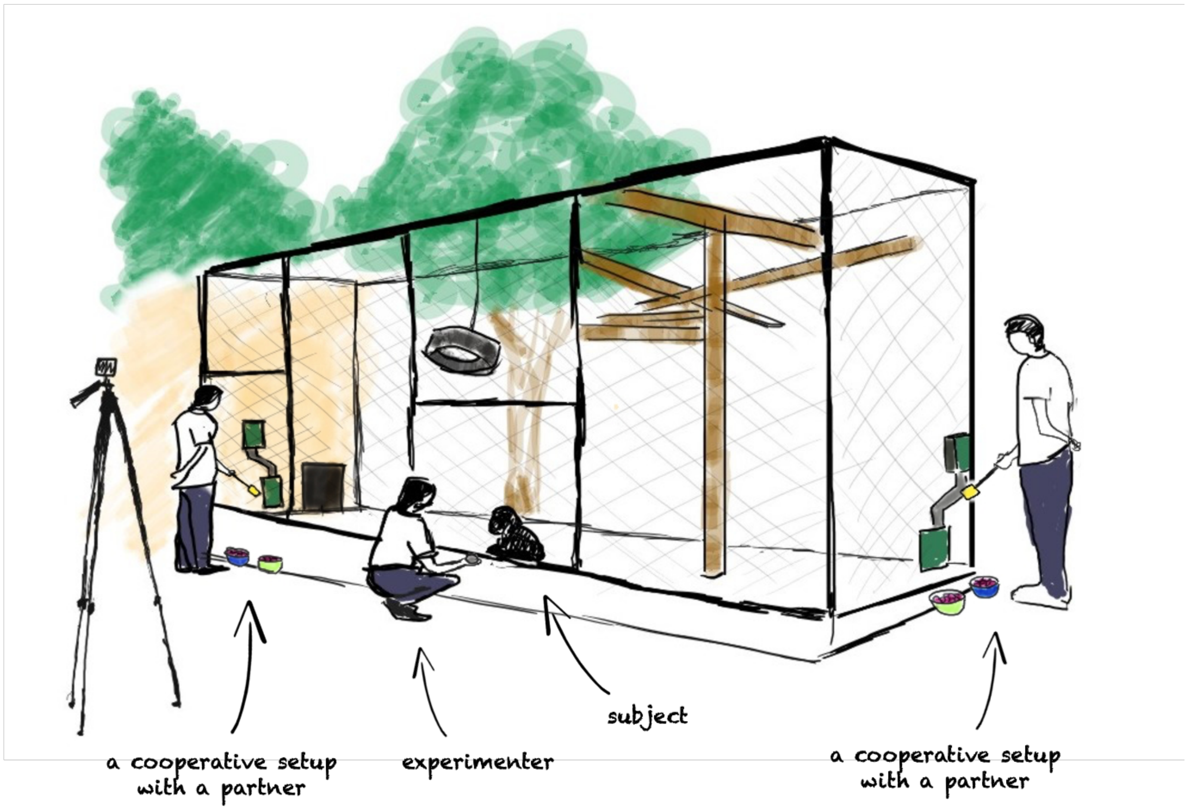
Note. The single setup (a) of the Slider (familiar) task and (b) of the Cog (novel) task, and the cooperative co-action setup (c) of the Slider (familiar) task, and (d) of the Cog (novel) task. For the Slider task, traps could be inserted in the box under each level, on either side, to block the ball from falling down to the next level. For the Cog task, a trap could be inserted in the box under either side of the cog wheel to block the ball from falling down.

Cooperative Testing Setups

A cooperative setup consisted of two puzzle boxes connected by a transparent hose (Figure 1c). One box was mounted slightly higher such that a ball could roll from the upper box into the lower box. If the ball was navigated successfully to the bottom of the upper box, it rolled into the lower box. If the ball was navigated into a trap in the upper box, it never reached the lower box. The two boxes of a cooperative setup were accessible from different positions/rooms, such that each player only had access to one of them and the overall success depended on both players' successful task performance. During the test phase, two cooperative setups of the same task were mounted on each side of the experimental room (approximately 4 m apart; Figure 2). The subjects had to decide with which of two human partners they wanted to perform the task by approaching the respective location and touching the target presented by the human partner.

Figure 2

Schematic Depiction of the Enclosure at the Start of a Test Trial



Note. Two cooperative setups were fixed on the left and right side, the experimenter and the subject were situated in the middle equidistant to the two setups, the two human partners moved in front of the setups and presented their target (yellow pane). The subject then had to move toward one of the human partners and touch the target, indicating its choice to perform the task in cooperation with this human partner. Illustration by J. Desriac.

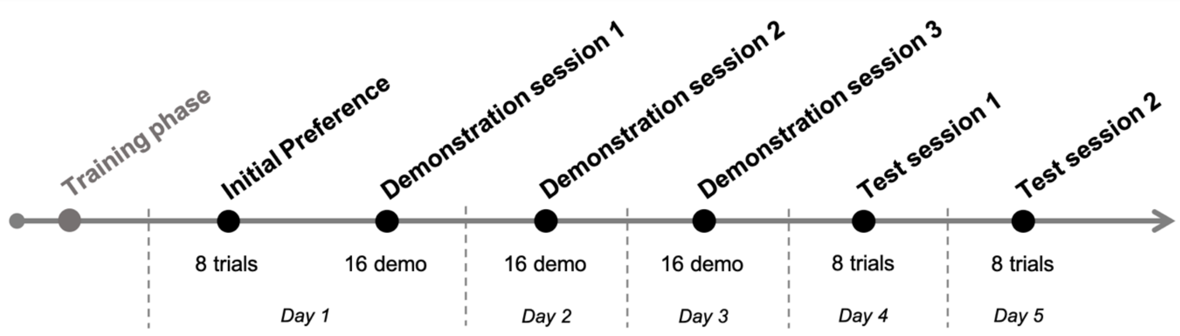
Procedure and Design

Subjects underwent an extensive training phase with comprehension checks to ensure they understood the tasks (for details, see the Supplementary Materials). Specifically, they learned to operate the single setups (i.e., individual box), to navigate the cooperative setup by operating the boxes sequentially, and to indicate their choice between two cooperative setups associated with one human partner each by touching a target at the desired location. They also had to demonstrate their principal understanding of the consequences of choosing a cooperative setup with or without a co-action human partner in place by choosing the correct setup, i.e., they had to reliably choose the location with a human partner who could access the upper box of the cooperative setup rather than the location with no partner in place. The humans who acted as partners during the training phase were different from those used during the experiment.

The experiment was split into three different steps conducted on five consecutive days: the initial preference assessment, the demonstration phase, and the test phase (Figure 3). Humans rather than conspecifics were used as partners for logistic reasons and to allow us to control for the potential effects of pre-existing relationships between subjects and partners on their choices during the experiment. Two humans played the role of the partners, who were the same throughout the experiment and for all the subjects. They had never interacted with the subjects before and were equally unfamiliar to them. Before starting the experiment, we checked that each subject was comfortable enough to approach and take food from the hands of these humans. The attribution of the role of the skillful or the unskilled partner to these humans depended on each subject on the initial preference assessment (see below). The experimenter was the same person for all the subjects throughout the training phase and the experiment and was equally familiar to all the subjects.

Figure 3

Chronological Order of the Experimental Steps



Initial Preference Assessment

To assess whether subjects had an initial preference for one of the human partners, we conducted eight initial preference trials prior to experimental manipulation. The experimenter lured the subject in the middle of the test room before each trial (to be at an equal distance from the partners). Then, both human partners approached the mesh and presented a raisin on the palm of their hand. The subject could approach one of the human partners to take their raisin; this approach was considered as a choice for this human partner. The partners' position was randomized across trials, with an equal number of times on each side. If the subjects demonstrated no preference for one human partner, the role of the human partners was randomly assigned to the subjects and counterbalanced between subjects. If the subject chose one human partner five times or more out of the eight trials, we assigned to this 'preferred' human partner the role of the unskilled partner for this subject. In this way, the subjects had to change their preference toward the other human partner to succeed in the cooperative test. As no group's initial preference for the same person was found, the roles of the human partners could be counterbalanced between subjects (Table 1).

Demonstration Phase

Subjects were able to observe the human partners individually performing at the single setups of the Slider task to gather information on the human partners' skills. Therefore, the Slider task is later referred to as the "familiar" task, whereas the Cog task is referred to as the "novel" task – subjects had no information about human partner performance at the Cog task. Our decision on the amount of exposure to human partners' skills was based on other studies such as Melis et al. (2008) and Subiaul et al. (2008), in which chimpanzees required at least 12 trials of interactions with each partner. Since we were specifically interested in whether monkeys spontaneously use information acquired merely by observation to pick a co-action human partner, we went beyond this minimum amount and used 24 information sampling events per human partner. These demonstration events were divided into three sessions of eight demonstration trials per human partner, with one session conducted per day on three consecutive days (Figure 3).

During each demonstration session, the human partners were playing alone on the single setups of the Slider task – in contrast to the co-action setups used for the cooperative testing condition. Four single setups of the Slider task were mounted – a bottom box and a top box on each side of the experimental room. Subjects could only access the boxes at the bottom, while the human partners only operated the boxes at the top. A session started with one human partner moving in front of a single setup on one side and with the experimenter attracting the subject toward this same side and inserting the ball into the subject's box (i.e., bottom box). The subject and the human partner took alternating turns with each performing the task twice in a row (i.e., subject (2x), partner (2x), subject (2x), partner (2x)). Subject and human partner performed at the task alternatingly to keep subjects motivated and spatially close to the human partners, and thus to maximize their attention. This procedure was repeated twice with both human partners, such that each subject could observe each human partner performing four times at each location, eight times in total, in a demonstration session. The presentation order and location of the human partners were alternated across blocks and pseudo-randomized across sessions. The first and the last human partner observed were also counterbalanced across subjects.

In each demonstration session, the skillful human partner was successful in all eight trials whereas the unskilled human partner was instructed to fail in seven out of eight trials, e.g., by navigating the ball into a trap (Videos S1 and S2; order and location of successful trials were randomly assigned at each session). Hence, by the end of the third session, subjects observed the skillful human partner succeeding in 24 out of 24 (100%) trials and the unskilled human partner succeeding in three out of 24 (12.5%) trials. Once the skillful human partner successfully navigated the ball through the puzzle box, the experimenter uttered her praise and excitement to emphasize the positive and desired outcome and boost the difference between successful and unsuccessful trials, followed by a provision of a reward (one grape) to the skillful human partner. The unskilled human partner had 30 s for trying to navigate the ball down without success, followed by the experimenter removing the ball manually from where it was stuck and verbally emphasizing the negative outcome.

Test Phase

Subjects had to choose one of the two human partners (skillful *vs.* unskilled) to cooperate at the co-action tasks. They were tested for two test sessions of eight trials each: one session with the Slider task (familiar condition) and one session with the Cog task (novel condition). We presented one test session per day on two consecutive days (Figure 3). The presentation order of novel and familiar sessions was counterbalanced across subjects: two subjects were first tested with the familiar task, and three subjects were first tested with the novel task (Table 1). During each test session, two cooperative setups of one of the two tasks were installed, one on each side of the experimental room, and the two human partners were present simultaneously, one in front of each cooperative setup (Figure 2). A trial refers to the choice of the subject to perform the co-action task with one of the human partners for one ball. Each subject could choose to cooperate with a human partner eight times per session and task, i.e., a total of 16 test trials. The location in front of the left or right cooperative setup of the skillful and unskilled human partners on the first test trial of each test session was counterbalanced between subjects and then was pseudo-randomized across trials (i.e., with an equal number of times on each side during a session).

A trial began with the experimenter luring the subject to a central position equidistant to each cooperative setup. Then, both human partners moved simultaneously to their designated cooperative setup and presented a target (a long wooden stick with a yellow square at the end). The subjects indicated their choice to cooperate with a human partner by approaching their preferred location and touching the respective target. Only then, did the experimenter move in front of the chosen cooperative setup and insert the ball into the human partner's box (i.e., the top box). The unchosen human partner moved away from the mesh and returned to their initial position. Therefore, once a choice was made, the subject could not switch to the other human partner within a single trial. The subjects were given a maximum of one minute to make a choice, otherwise the trial was aborted. In this case, the same trial (with the same human partners' side configuration) was repeated.

If the skillful human partner was chosen, subject and human partner successfully navigated the ball through the cooperative setup and received a reward (a grape) from the experimenter (Video S3). The skillful human partner received the same reward as the subject to emphasize the cooperativeness of the task. If the unskilled human partner was chosen, the human partner failed at navigating the ball successfully and the ball never reached the subject's part of the cooperative setup (i.e., bottom box). The experimenter waited for 30 s after the insertion of the ball before ending the trial. In this case, neither the unskilled human partner nor the subject were rewarded (Video S4). The skillful human partner was always successful (100%) during these two test sessions whereas the unskilled human partner was never successful (0%).

Data Coding

All sessions were videotaped with one to four cameras (GoPro Hero8) depending on the experimental step to obtain different views of the subjects, the human partners, setups, and experimental rooms. Two different observers, including one who was unaware of the study design and hypothesis, coded independently frame by frame all the videos of the three experimental phases using Behavioral Observation Research Interactive Software ("Boris"; Friard & Gamba, 2016). Videos were coded for: (a) the partner choices of the subjects at the initial and test sessions (inter-coder reliability: Cohen's kappa, $\kappa = 1$, $N = 118$), and (b) the duration of looking behaviors of the subjects (i.e., head orientation, or eye orientation when visible) directed toward a human partner (face, body, or hands) during each demonstration trial (inter-coder reliability: $ICC = 0.893$, $N = 240$). A choice was defined as approaching and taking the food from the human in the initial preference session or touching the target presented by the human partner in the test phase. During the demonstration phase, only one human partner was present at a time. We only coded subjects' looking time when the human partner was performing the task, not when it was the subject's turn (for more details, see the Supplementary Materials).

Data Analyses

Did the Subjects Prefer the Skillful Human Partner over the Unskilled Human Partner?

We aimed to investigate whether Tonkean macaques could spontaneously use information acquired by observation to choose optimal partners to cooperate with at a familiar and a novel co-action task. A spontaneous preference for the skillful human partner would manifest in optimal choices at the first trial and no effect of sessions or trials on subjects' choices. A transfer of social evaluation of human partners' skills across the two different tasks would manifest in no effect of task on subjects' choices. However, estimating the effects of task (novel, familiar), session (1, 2), trials, and their potential interaction on the probability of choices of the skillful human partner led to a model too complex for our small final sample size. We hence revised our analysis plan to investigate whether the subjects chose the skillful human partner more often in both test sessions, regardless of the task, and how their choices developed across trials (i.e., with increasing direct experience with the actors). We fitted a model in R (version 4.3.2; R Core Team, 2022) using Generalized Linear Mixed Models (GLMM; Baayen, 2008). We used the function glmer of the package lme4 (version 1.1-35.1; Bates et al., 2015) and a binomial error structure with logit link function (McCullagh & Nelder, 1989). We estimated the effect of session (test session 1, test session 2) on the probability to choose the skillful human partner. To control for their potential effects, we included into this model subjects' initial partner preference and trial number as fixed effects, with trial number in an interaction with session. The reason for including the interaction between session and trial number was that we expected the effect of learning through trials to be more pronounced in the first session than in the second. For instance, while only having information on the human partners' skills from observations when starting the first session, subjects then obtained additional information from direct interactions with the human partners at each trial of the first session that they could use for the next trial. Since the same subjects were tested several times in both sessions and in order to avoid overconfident model estimates and to keep type I error rate at the nominal level of 5%, we included subject ID as a random intercepts effect and all identifiable random slopes (Barr et al., 2013; Schielzeth & Forstmeier, 2009), namely those of trial number and session. As an overall test of the fixed effects and to avoid cryptic multiple testing (Forstmeier & Schielzeth, 2011), we conducted full-null model comparisons using likelihood ratio tests (Dobson, 2018). We compared this full model with a null model lacking the effect of session and its interaction with trial number in the fixed effects part but being otherwise identical. We tested the effect of individual fixed effects by comparing the full model with a reduced model lacking the interaction in the fixed effects' part (Barr et al., 2013). We obtained confidence intervals of model estimates and fitted values by means of a parametric bootstrap (N = 1,000 bootstraps; function bootMer of the package lme4). We checked all the relevant model assumptions and transformed some variables when needed to ease the interpretation of the model estimates and convergence (for more details, see section 5 in the Supplementary Materials).

Was the Attention of the Subjects Predictive of their Partner Choices?

We fitted a similar model except that we estimated the effect of attention during the demonstration on the number of choices of the skillful human partner in the test phase. To potentially obtain the relevant information on the human partners' skills, subjects needed to observe both human partners during the demonstration phase. As a measure of the subjects' attention to both human partners, we hence chose the minimum looking time per subject to the skillful and unskilled human partner, i.e., the time during which each subject observed each human partner, and used this for the following analysis. Into the model, we included the attention during the demonstration as a fixed effect, test session (test session 1, test session 2), and their interaction. The reason for including the interaction between attention and session was that we expected the effect of learning through sessions to be less pronounced for subjects with a higher level of attention during the demonstration sessions compared to less attentive subjects, as attentive subjects could already know at the first test session which human partner is the optimal choice and then do not have to learn from direct experience. We again included subject ID as a random intercepts effect and test session as a random slope within subject. We compared this full model with a null model lacking the effect of proportion of total attention and its interaction with session in the fixed effects part but being otherwise identical and with a reduced model lacking the interaction (for more details, see section 6 in the Supplementary Materials).

Results

Did the Subjects Prefer the Skillful Human Partner over the Unskilled Human Partner?

Subjects chose the skillful human partner 17 out of 40 times (42.5%) in Test session 1 and 25 out of 40 times (62.5%) in Test session 2 (see Table 1 for results for each subject). Overall, their probability to choose the skillful human partner was not significantly influenced by session and its interaction with trial number (full-null model comparison: χ^2 = 4.55, *df* = 2, *p* = .103; Tables 2 and 3), indicating that our subjects' probability to choose the skillful human partner did not change over the course of trials and sessions (i.e., with the acquisition of direct experience with the human partners). Although all subjects chose the skillful human partner more often in the second test session, this increase was marginally non-significant (*p* = .069; Table 3, Figure 4 and Figure S2 in the Supplementary Materials). In addition, the probability to choose the skillful human partner at the very first trial was not different from chance (Figure 4). Post-hoc additional analysis also revealed no significant effect of the choices on the first trial on the probability of the subjects of choosing the skillful human partner during the rest of the test session (full-null model comparison: χ^2 = 0.802, *df* = 1, *p* = .371; Supplementary Materials section 6, Table S6). However, one subject stood out in choosing the skillful human partner in six out of the eight trials in both test sessions (i.e., familiar and novel tasks), and beginning from the first trial. This subject was presented with the novel task at her first test session. She also successfully reversed her actors' preference from the initial preference session in which she had chosen the human who would play the skillful actor three out of eight times (Table 1). On a group level, the strength of the subjects' initial preference for one actor did not significantly influence their probability to make optimal choices after (Table 3). Post-hoc additional analysis also revealed that there was no significant effect of the location on the left or right setup of the human partners on the probability of the subjects of choosing the skillful human partner (full-null model comparison: χ^2 = 0.0005, df = 1, *p* = .982; Supplementary Materials section 7, Table S7).

Table 2

Results of the Subjects' Partner Choices Full Model (Estimates Together with Standard Errors, 95% Confidence Limits, Significance Tests, and the Estimates' Range After Excluding Individuals One at a Time)

Term	Estimate	SE	CL_{Lower}	CL_{Upper}	χ^2	*df*	*p*
(Intercept)	-0.322	0.358	-1.177	0.377			
Session	0.856	0.473	0.083	1.984			
Trial number	0.307	0.334	-0.369	1.150			
Initial preference	0.259	0.275	-0.286	0.951	0.826	1	.364
Session × Trial number	-0.527	0.476	-1.694	0.447	1.245	1	.264

Note. Trial number and Initial preference were z-transformed to a mean of 0 and a standard deviation of 1 (original means and standard deviations were 4.5 ± 2.31 and 0.325 ± 0.13, respectively). Session was dummy coded with test session 1 being the reference level.

Table 3

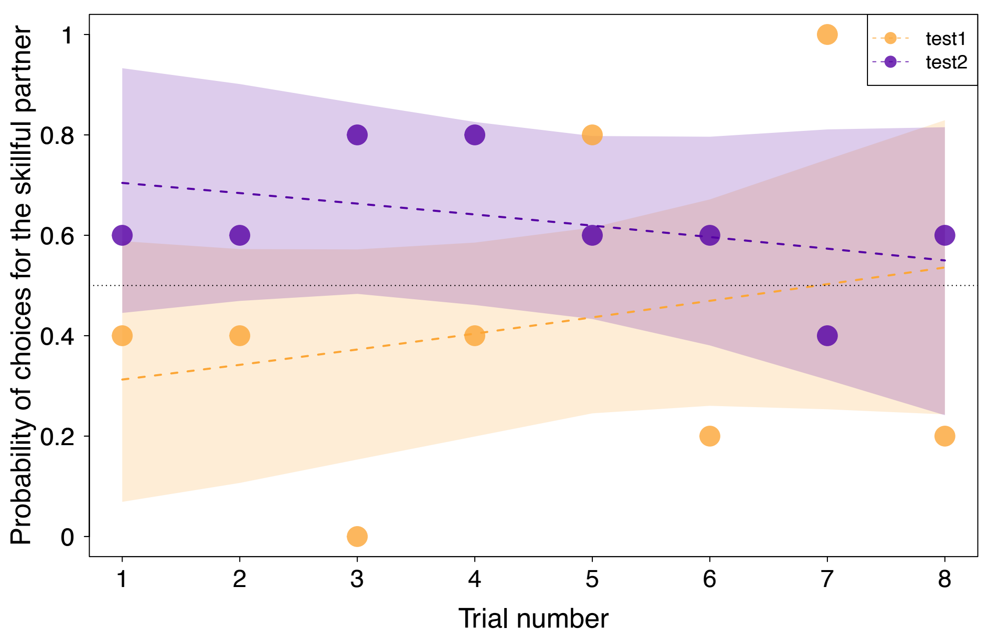
Results of the Subjects' Partner Choices, Reduced Model (Estimates Together with Standard Errors, 95% Confidence Limits, and Significance Tests)

Term	Estimate	SE	CL_{Lower}	CL_{Upper}	χ^2	*df*	*p*
(Intercept)	-0.315	0.352	-1.166	0.437			
Session	0.843	0.468	0.103	2.018	3.304	1	.069
Trial number	0.048	0.234	-0.455	0.577	0.041	1	.839
Initial preference	0.254	0.270	-0.339	0.848	0.826	1	.363

Note. Trial number and Initial preference were z-transformed to a mean of 0 and a standard deviation of 1 (original means and standard deviations were 4.5 ± 2.31 and 0.325 ± 0.13, respectively). Session was dummy coded with test session 1 being the reference level.

Figure 4

Probabilities Of Subjects' Choices For The Skillful Human Partner



Note. The raw data and fitted model and its 95% confidence limits are shown for both sessions for initial preference being at its average. Each data point represents the mean probability of choosing the skillful human partner for the five subjects at each trial. Note that at the first trial, the confidence intervals of both sessions encompass the value of 0.5 indicating that the performance in the first trial did not significantly deviate from chance.

Was the Attention of the Subjects Predictive of their Partner Choices?

During the demonstration phase, subjects looked on average 6.1 ± 2.9 s out of 16.3 s (37.8%) per demonstration trial at the skillful human partner and 11.1 ± 6.9 s out of 26.7 s (40.9%) at the unskilled human partner. In total, they looked on average 145.4 s (37.2%) at the skillful human partner (range across individuals: 125.5 s to 182.9 s, 31.5% to 44.7%), 265.5 s (41.3%) at the unskilled human partner (range across individuals: 209.9 s to 370.4 s, 34.2% to 47.7%), and 410.9 s (39.8%) at both human partners (range across individuals: 335.4 s to 505 s, 33.1% to 46.2%) across all demonstration trials. Overall, the probability to choose the skillful human partner was not significantly influenced by attention and its interaction with session ($\chi^2 = 2.123$, $df = 2$, $p = .346$; Tables S4 and S5 in the Supplementary Materials). This result indicates that neither the attention of the subjects to both human partners during the demonstration phase (i.e., the minimum total looking time per subject to the skillful and unskilled human partner) nor its interaction with test session influenced their partner choices. Therefore, contrary to our expectation, the individuals who observed the human partners more during the demonstration did not make more optimal partner choices in the test sessions.

Discussion

We investigated whether Tonkean macaques can choose optimal human partners for cooperation in two co-action tasks. Subjects had several opportunities to observe a skillful and an unskilled human partner performing individually at a solo version of a task (familiar task) and then had to choose one of them to cooperate with in the familiar task and a novel task from the same domain. During each test trial, the subjects obtained information on the human partners' skills through direct experience – receiving food only if they chose the skillful human partner. Optimal choices from the first trials would then indicate a spontaneous preference for the skillful human partner based solely on observation, suggesting an evaluation of the human partners' skills. Optimal choices increasing with trials would indicate a learning effect based on direct experience with the human partners. Contrasting subjects' partner choices in a novel task condition and a familiar task condition allowed to investigate whether they generalized their social evaluation across different tasks from the same domain.

We found that Tonkean macaques did not prefer to cooperate with a skillful human partner compared to an unskilled human partner. Overall, they did not spontaneously use information acquired by observation prior to the test to choose optimal human partners, indicating no evaluation or use of knowledge on the human partners' skills. Even though they tended to choose the skillful human partner more often in their second test session compared to the first session, this increase in optimal choices across sessions was marginally non-significant. This result indicates no learning by experience with the human partners across the eight or 16 presented test trials. Our subjects' choices emerged irrespective of how much attention they paid to the human partners' actions when they observed the human partners solving solo versions of the task in the demonstration phase. Contrary to our predictions, Tonkean macaques who observed the human partners the most during the demonstration then did not make more optimal choices thereafter.

These findings must be carefully interpreted as resulting from a small sample size and with noticeable inter-individual differences in our subjects' partner choices. Three subjects chose the skillful human partner six out of eight times only in the second test session and two subjects did not make more than four optimal choices in both sessions. One individual stood out by choosing the skillful human partner above chance at both test sessions and from the first trial on. This choice pattern could indicate that she transferred her knowledge from the observation phase to the test phase and the two co-action tasks, but it could also be the result of chance. However, this Tonkean macaque chose again six out of eight times a skillful over an unskilled human actor at opening containers in another experiment (Hirel et al., 2025). This inter-individual variation in performance should be investigated in the future by testing the same individuals in different settings. Our paradigm presents a promising way to address this question, as being suitable to test individuals with different partners pairs (humans as well as conspecifics) differing in their degree of competence or social relationship and with different tasks and in different contexts, like competition (see Keupp & Herrmann, 2024).

In addition, our subjects were attentive to the human partners' actions less than 50 % of the time. This lack of interest in the human partner performance is in accordance with previous findings showing that long-tailed macaques only paid close attention to conspecific partner performance if it had directly relevant consequences for themselves (Keupp et al., 2019, 2021). Indeed, in our study, the human partners' performance during the demonstration phase did not affect the subjects' food intake, which might have led to a lack of sufficient interest in our subjects. However, we had no predictions on how much time Tonkean macaques should observe the human partners' actions as we still do not know how many seconds of demonstrations primates need to observe to obtain the relevant information about others. This question would have to be assessed in the future with a different study design that systematically varies the amount of information provided.

The particular characteristics of our subjects should also be considered (see STRANGE-related biases; Webster & Rutz, 2020). Our selection protocol may have influenced which Tonkean macaques in the group participated in our experiment and how they behaved during the test. Only individuals who were not anxious of being separated from their group and who came regularly to the experimental rooms could complete the training and be tested as subjects. Dominant individuals also tended to monopolize the access to the experimental rooms while females carrying infants tended to avoid being inside the experimental rooms. Our pool of subjects was then biased toward adults, males, dominants, bold, and highly motivated and experienced individuals to experimental testing with human experimenters. Individual features such as sex, age, social rank, rearing history, as well as personality, have been found to influence the use of social information (e.g., Canteloup et al., 2021; Carter et al., 2014; Kendal et al., 2015; Watson et al., 2018). Consequently, the particular individual characteristics of our subjects may have influenced their performance in the experiment, and our results may not be representative of the species' abilities. The overall lack of optimal partner selectivity in our experiment stands in contrast with previous findings with monkeys (Anderson et al., 2016, 2013, 2013; Kawai et al., 2014). However, contrary to our experiment, these previous studies presented a larger amount of test trials, each including an observation of the human actors' actions and, directly after, a choice to take a food piece from one of the human actors. For example, capuchins and squirrel monkeys had to choose between a reciprocator and a non-reciprocator in twelve sessions of twelve trials each (Anderson et al., 2016, 2013). In addition, their subjects' choices had no effect on their food intake as they always got the same piece of food no matter which human actor they chose. Therefore, these monkeys may not have used associative learning despite the large trial number, but they had 144 opportunities to observe each human actor compared to 24 possible observations in our study. Except for one study indicating no change in capuchins' choices against non-helpers through sessions (Anderson et al., 2013), information on how the choices of these monkeys developed through trials is not provided in the respective papers. Knowing the development of these monkeys' choice pattern would help us understand whether we did not offer sufficient opportunities for Tonkean macaques to acquire relevant knowledge about the human partners' skills, or whether other reasons may explain our contrasting findings. One possible reason for the differences in monkeys' social evaluative performance between our experiment and previous studies may also be the investigation of different domains of competence and contexts. Indeed, these previous studies tested monkeys on prosocial behaviors which relate to a different domain than foraging or cooperative skillfulness and may require different cognitive capacities. The monkeys in previous studies could observe human experimenters interacting with each other prosocially or not and then decide who they will accept food from. In contrast, Tonkean macaques in our experiment observed the human partners individually performing the task and then decided which human partner they wanted to cooperate with. Taking food from an actor's hand might be a less cognitively challenging situation than having to choose a partner for a cooperative activity. Indeed, choosing a cooperative partner can be complex as it may involve taking into account not a single but a broad range of social information as well as interacting and coordinating with another individual (Hermes et al., 2016).

Another possibility is that skillfulness might not be a salient feature at all to assess others' value for cooperative tasks for monkeys as for apes or humans. To our knowledge, evidence of active cooperative foraging in wild monkey species populations has never been described. However, monkeys pay attention to the foraging behaviors of others and adapt their own behaviors toward the foraging experts of their group in non-cooperative contexts (Fruteau et al., 2009; O'Hearn et al., 2025; Ottoni et al., 2005; Stammbach, 1988; Tan et al., 2018). Skillfulness of others might then be a relevant characteristic for monkeys to consider in contexts other than cooperation. For example, being able to evaluate individuals' skills and use it to selectively learn new behaviors from the best experts can be truly advantageous. However, social learning in primates has mainly been studied with a focus on the learning mechanisms (what to learn) or the social aspects (who to learn from), but only taking into account the social relationships between individuals and never the individual characteristics (Fragaszy & Visalberghi, 2004; Kendal, 2021; Laland, 2004). Future research could investigate whether monkeys form and use impressions of others' expertise to choose from whom to learn new behaviors and how this selective learning through social evaluation might affect the evolutionary dynamics of innovation and culture (e.g., see model-based biases in vervet monkeys and chimpanzees: Canteloup et al., 2021; Kendal et al., 2015).

In summary, we investigated Tonkean macaques' ability to choose optimal partners in a cooperative context using evaluation of human partners' skillfulness acquired by observation. Even though our subjects did not prefer to cooperate with a skillful human partner, showing no evidence of social evaluation abilities, the inter-individual differences on our small sample size encourage to further investigate whether and how Tonkean macaques can form impressions about others on different behaviors and contexts. Besides, our methodology represents a promising way to investigate different contexts, the different types of social information involved, and the cognitive mechanisms underlying social evaluation and partner selectivity in nonhuman primates.

Acknowledgements

We are thankful to the management and staff of the Centre de Primatologie – Silabe de l'Université de Strasbourg (https://www.silabe.com) for their authorization, support, and help with the data collection, especially the valuable assistance of the animal keepers, the technical team, and the researchers and students. A very special thanks to Pascal Ancé and Pierre-Henri Moreau for agreeing to play the roles of the partners in the experiment. For their invaluable support with the conception and the manufacturing of the experimental apparatuses, we would like to thank Louis Frank and M. Zippert and his team of the chemistry workshop of the University of Göttingen. We are grateful to Adam Provin, Michele Marziliano, and Rose Riguet for their precious help with the data collection, and Nadja Vöglte for reliability coding. Furthermore, we thank the editor and an anonymous reviewer for helpful comments about an earlier version of this paper. Finally, we thank Julie Desriac for the illustration of the experimental setup.

Author Contributions: Hirel Marie: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. Meunier Hélène: Resources, Writing – review & editing. Mundry Roger: Formal analysis, Writing – review & editing. Rakoczy Hannes: Conceptualization, Funding acquisition, Writing – review & editing. Fischer Julia: Funding acquisition, Writing – review & editing. Keupp Stefanie: Conceptualization, Methodology, Project administration, Supervision, Writing – review & editing.

Funding: This work was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; Project number: 254142454 / GRK 2070 "Understanding Social Relationships") and an individual grant to SK (project number: 425330201).

Conflict of Interest: The authors declare no conflict of interest.

Data Accessibility: The supplementary document, videos, data, statistical code associated with this article can be found at: https://osf.io/tnkz2/

References

Anderson, J. R., Bucher, B., Kuroshima, H., & Fujita, K. (2016). Evaluation of third-party reciprocity by squirrel monkeys (*Saimiri sciureus*) and the question of mechanisms. *Animal Cognition*, *19*(4), 813–818. https://doi.org/10.1007/s10071-016-0980-7

Anderson, J. R., Kuroshima, H., Takimoto, A., & Fujita, K. (2013). Third-party social evaluation of humans by monkeys. *Nature Communications*, *4*(1), 1561. https://doi.org/10.1038/ncomms2495

Anderson, J. R., Takimoto, A., Kuroshima, H., & Fujita, K. (2013). Capuchin monkeys judge third-party reciprocity. *Cognition*, *127*(1), 140–146. https://doi.org/10.1016/j.cognition.2012.12.007

Baayen, R. H. (2008). Analyzing Linguistic Data. *Cambridge University Press*. https://doi.org/10.1017/CBO9780511801686

Ballesta, S., & Meunier, H. (2023). Is this worth the trouble? Strategic conflict management in Tonkean macaques. *iScience*, *26*(11), 108176. https://doi.org/10.1016/j.isci.2023.108176

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.48550/arXiv.1406.5823

Borgeaud, C., Waal, E. van de, & Bshary, R. (2013). Third-Party Ranks Knowledge in Wild Vervet Monkeys (Chlorocebus aethiops pygerythrus). *PLoS ONE*, *8*(3), e58562. https://doi.org/10.1371/journal.pone.0058562

Brügger, R. K., Willems, E. P., & Burkart, J. M. (2021). Do marmosets understand others' conversations? A thermography approach. *Science Advances*. https://doi.org/10.1126/sciadv.abc8790

Canteloup, C., Cera, M. B., Barrett, B. J., & van de Waal, E. (2021). Processing of novel food reveals payoff and rank-biased social learning in a wild primate. *Scientific Reports*, *11*(1), 9550. https://doi.org/10.1038/s41598-021-88857-6

Canteloup, C., & Meunier, H. (2017). 'Unwilling' versus 'unable': Tonkean macaques' understanding of human goal-directed actions. *PeerJ*, *5*, e3227. https://doi.org/10.7717/peerj.3227

Canteloup, C., Piraux, E., Poulin, N., & Meunier, H. (2016). Do Tonkean macaques (*Macaca tonkeana*) perceive what conspecifics do and do not see? *PeerJ*, *4*, e1693. https://doi.org/10.7717/peerj.1693

Canteloup, C., Poitrasson, I., Anderson, J. R., Poulin, N., & Meunier, H. (2017). Factors influencing deceptive behaviours in Tonkean macaques (*Macaca tonkeana*). *Behaviour*, *154*(7–8), 765–784. https://doi.org/10.1163/1568539X-00003443

Carter, A. J., Marshall, H. H., Heinsohn, R., & Cowlishaw, G. (2014). Personality predicts the propensity for social learning in a wild primate. *PeerJ*, *2*, e283. https://doi.org/10.7717/peerj.283

Dobson, A. J. (2018). An introduction to generalized linear models. *Chapman and Hall/CRC*, *4th ed.* https://doi.org/10.1201/9781315182780

Duboscq, J., Micheletta, J., Agil, M., Hodges, K., Thierry, B., & Engelhardt, A. (2013). Social Tolerance in Wild Female Crested Macaques (Macaca nigra) in Tangkoko-Batuangus Nature Reserve, Sulawesi, Indonesia. *American Journal of Primatology*, *75*(4), 361–375. https://doi.org/10.1002/ajp.22114

Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, *7*(2), 117–140. https://doi.org/10.1177/001872675400700202

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Fragaszy, D., & Visalberghi, E. (2004). Socially biased learning in monkeys. *Learning & Behavior*, *32*(1), 24–35. https://doi.org/10.3758/BF03196004

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Fruteau, C., Voelkl, B., Damme, E. van, & Noë, R. (2009). Supply and demand determine the market value of food providers in wild vervet monkeys. *Proceedings of the National Academy of Sciences*, *106*(29), 12007–12012. https://doi.org/10.1073/pnas.0812280106

Fusaro, M., Corriveau, K. H., & Harris, P. L. (2011). The good, the strong, and the accurate: Preschoolers' evaluations of informant attributes. *Journal of Experimental Child Psychology*, *110*(4), 561–574. https://doi.org/10.1016/j.jecp.2011.06.008

Hermes, J., Behne, T., & Rakoczy, H. (2015). The role of trait reasoning in young children's selective trust. *Developmental Psychology*, *51*(11), 1574–1587. https://doi.org/10.1037/dev0000042

Hermes, J., Behne, T., & Rakoczy, H. (2018). The development of selective trust: Prospects for a dual-process account. *Child Development Perspectives*, *12*(2), 134–138. https://doi.org/10.1111/cdep.12274

Hermes, J., Behne, T., Studte, K., Zeyen, A.-M., Gräfenhain, M., & Rakoczy, H. (2016). Selective cooperation in early childhood – How to choose models and partners. *PoOS ONE*, *11*(8), e0160881. https://doi.org/10.1371/journal.pone.0160881

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://doi.org/10.1037/a0028929

Heyes, C. (2012). Simple minds: A qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2695–2703. https://doi.org/10.1098/rstb.2012.0217

Hirel, M., Thiriau, C., Roho, I., & Meunier, H. (2020). Are monkeys able to discriminate appearance from reality? *Cognition*, *196*, 104123. https://doi.org/10.1016/j.cognition.2019.104123

Hirel, M., Marziliano, M., Meunier, H., Rakoczy, H., Fischer, J., & Keupp, S. (2025). Social evaluation of skilfulness in Tonkean macaques (*Macaca tonkeana*) and brown capuchins (*Sapajus apella*). BioRxiv. https://doi.org/10.1101/2025.05.13.653791

Horner, V., Proctor, D., Bonnie, K. E., Whiten, A., & Waal, F. B. M. de. (2010). Prestige affects cultural learning in chimpanzees. *PLoS ONE*, *5*(5), e10625. https://doi.org/10.1371/journal.pone.0010625

Kawai, N., Yasue, M., Banno, T., & Ichinohe, N. (2014). Marmoset monkeys evaluate third-party reciprocity. *Biology Letters*, *10*(5), 20140058. https://doi.org/10.1098/rsbl.2014.0058

Kendal, R., Hopper, L. M., Whiten, A., Brosnan, S. F., Lambeth, S. P., Schapiro, S. J., & Hoppitt, W. (2015). Chimpanzees copy dominant and knowledgeable individuals: Implications for cultural diversity. *Evolution and Human Behavior*, *36*(1), 65–72. https://doi.org/10.1016/j.evolhumbehav.2014.09.002

Kendal, R. L. (2021). Social learning and teaching overview. In A. B. Kaufman, J. Call, & J. C. Kaufman (Eds.), *The Cambridge Handbook of Animal Cognition* (pp. 443–471). Cambridge University Press. https://doi.org/10.1017/9781108564113.024

Keupp, S., Abedin, F., Jeanson, L., Kade, C., Kalbitz, J., Titchener, R., Mussweiler, T., Bugnyar, T., & Fischer, J. (2021). Performance-based social comparisons in humans and long-tailed macaques. *Animal Behavior and Cognition*, *8*(3), 325-350. https://doi.org/10.26451/abc.08.03.02.2021

Keupp, S., & Herrmann, E. (2024). Domain-specific inferences about conspecifics' skills by chimpanzees. *Scientific Reports*, *14*(1), 21996. https://doi.org/10.1038/s41598-024-73340-9

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), 20180784. https://doi.org/10.1098/rsbl.2018.0784

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Laland, K. N. (2004). Social learning strategies. *Animal Learning & Behavior*, *32*(1), 4–14. https://doi.org/10.3758/BF03196002

Manrique, H. M., Zeidler, H., Roberts, G., Barclay, P., Walker, M., Samu, F., Fariña, A., Bshary, R., & Raihani, N. (2021). The psychological foundations of reputation-based cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1838), 20200287. https://doi.org/10.1098/rstb.2020.0287

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models.* Routledge. https://doi.org/10.1201/9780203753736

Melis, A. P., Hare, B., & Tomasello, M. (2006). Chimpanzees recruit the best collaborators. *Science*, *311*(5765), 1297–1300. https://doi.org/10.1126/science.1123007

Melis, A. P., Hare, B., & Tomasello, M. (2008). Do chimpanzees reciprocate received favours? *Animal Behaviour*, *76*(3), 951–962. https://doi.org/10.1016/j.anbehav.2008.05.014

Mielke, A., Preis, A., Samuni, L., Gogarten, J. F., Wittig, R. M., & Crockford, C. (2018). Flexible decision-making in grooming partner choice in sooty mangabeys and chimpanzees. *Royal Society Open Science*, *5*(7), 172143. https://doi.org/10.1098/rsos.172143

Miss, F. M., Meunier, H., & Burkart, J. M. (2022). Primate origins of corepresentation and cooperative flexibility: A comparative study with common marmosets (*Callithrix jacchus*), brown capuchins (*Sapajus apella*), and Tonkean macaques (*Macaca tonkeana*). *Journal of Comparative Psychology*, *136*(3), 199–212. https://doi.org/10.1037/com0000315

Molesti, S., & Majolo, B. (2016). Cooperation in wild Barbary macaques: Factors affecting free partner choice. *Animal Cognition*, *19*(1), 133–146. https://doi.org/10.1007/s10071-015-0919-4

Mundry, R. (2023). *Some R functions* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7670524

O'Hearn, W. J., Beckmann, J., Von Fersen, L., Dal Pesco, F., Mundry, R., Keupp, S., Diakhate, N., Niederbremer, C., & Fischer, J. (2025). Increased female competition for males with enhanced foraging skills in Guinea baboons. *Proceedings of the Royal Society B: Biological Sciences*, *292*(2042), 20242925. https://doi.org/10.1098/rspb.2024.2925

Ottoni, E. B., de Resende, B. D., & Izar, P. (2005). Watching the best nutcrackers: What capuchin monkeys (*Cebus apella*) know about others' tool-using skills. *Animal Cognition*, *8*(4), 215–219. https://doi.org/10.1007/s10071-004-0245-8

Perry, S., Barrett, H. C., & Manson, J. H. (2004). White-faced capuchin monkeys show triadic awareness in their choice of allies. *Animal Behaviour*, *67*(1), 165–170. https://doi.org/10.1016/j.anbehav.2003.04.005

Petit, O., Desportes, C., & Thierry, B. (1992). Differential Probability of "Coproduction" in Two Species of Macaque (*Macaca tonkeana, M. mulatta*). *Ethology*, *90*(2), 107–120. https://doi.org/10.1111/j.1439-0310.1992.tb00825.x

Placì, S., Padberg, M., Rakoczy, H., & Fischer, J. (2019). Long-tailed macaques extract statistical information from repeated types of events to make rational decisions under uncertainty. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-48543-0

R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.r-project.org/

Resende, B. D. de, Izar, P., & Ottoni, E. B. (2004). Social play and spatial tolerance in tufted capuchin monkeys (*Cebus apella*). *Revista de Etologia*, *6*(1), 55–61. http://pepsic.bvsalud.org/scielo.php?script=sci_abstract&pid=S1517-28052004000100008&lng=pt&nrm=iso&tlng=en

Riley, E. P. (2010). The endemic seven: Four decades of research on the Sulawesi macaques. *Evolutionary Anthropology: Issues, News, and Reviews*, *19*(1), 22–36. https://doi.org/10.1002/evan.20246

Russell, Y. I., Call, J., & Dunbar, R. I. M. (2008). Image scoring in great apes. *Behavioural Processes*, *78*(1), 108–111. https://doi.org/10.1016/j.beproc.2007.10.009

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Silk, J. B. (1999). Male bonnet macaques use information about third-party rank relationships to recruit allies. *Animal Behaviour*, *58*(1), 45–51. https://doi.org/10.1006/anbe.1999.1129

Silk, J. B., Alberts, S. C., & Altmann, J. (2004). Patterns of coalition formation by adult female baboons in Amboseli, Kenya. *Animal Behaviour*, *67*(3), 573–582. https://doi.org/10.1016/j.anbehav.2003.07.001

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104*(43), 17228–17233. https://doi.org/10.1073/pnas.0706741104

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., & Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences*, *104*(44), 17435–17440. https://doi.org/10.1073/pnas.0704598104

Stammbach, E. (1988). Group responses to specially skilled individuals in a *Macaca fascicularis* group. *Behaviour*, *107*(3–4), 241–266. https://doi.org/10.1163/156853988X00368

Subiaul, F., Vonk, J., Okamoto-Barth, S., & Barth, J. (2008). Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal Cognition*, *11*(4), 611–623. https://doi.org/10.1007/s10071-008-0151-6

Suchak, M., Eppley, T. M., Campbell, M. W., & Waal, F. B. M. de. (2014). Ape duos and trios: Spontaneous cooperation with free partner choice in chimpanzees. *PeerJ*, *2*, e417. https://doi.org/10.7717/peerj.417

Tan, A., Hemelrijk, C., Malaivijitnond, S., & Gumert, M. (2018). Young macaques (*Macaca fascicularis*) preferentially bias attention towards closer, older, and better tool users. *Animal Cognition*, *21*(4), 551–563. https://doi.org/10.1007/s10071-018-1188-9

Thierry, B., Anderson, J. R., Demaria, C., Desportes, C., & Petit, O. (1994). Tonkean macaque behaviour from the perspective of the evolution of Sulawesi macaques. In J.J. Roeder, B. Thierry, J.R. Anderson, N. Herrenschmidt (Eds.), *Current Primatology Vol.2: Social development, learning and behaviour* (pp. 103–117). Université Louis Pasteur.

Thierry, B., Bynum, E. L., Baker, S., Kinnaird, M. F., Matsumura, S., Muroyama, Y., O'Brien, T. G., Petit, O., & Watanabe, K. (2000). The social repertoire of Sulawesi macaques. *Primate Research*, *16*(3), 203–226. https://doi.org/10.2354/psj.16.203

Titchener, R., Hermes, J., Fischer, J., Rakoczy, H., & Keupp, S. (2023). *Social evaluation of co-action partners: Children's partner recruitment in cooperation and competition tasks*. *PsyArXiv*. https://doi.org/10.31234/osf.io/c3gwu

Watson, S. K., Vale, G. L., Hopper, L. M., Dean, L. G., Kendal, R. L., Price, E. E., Wood, L. A., Davis, S. J., Schapiro, S. J., Lambeth, S. P., & Whiten, A. (2018). Chimpanzees demonstrate individual differences in social information use. *Animal Cognition*, *21*(5), 639–650. https://doi.org/10.1007/s10071-018-1198-7

Webster, M. M., & Rutz, C. (2020). How STRANGE are your study animals? *Nature*, *582*(7812), 337–340. https://doi.org/10.1038/d41586-020-01751-5

Whitehouse, J., & Meunier, H. (2020). An understanding of third-party friendships in a tolerant macaque. *Scientific Reports*, *10*(1), 9777. https://doi.org/10.1038/s41598-020-66407-w

Wittig, R. M., Crockford, C., Langergraber, K. E., & Zuberbühler, K. (2014). Triadic social interactions operate across time: A field experiment with wild chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1779), 20133155. https://doi.org/10.1098/rspb.2013.3155

Wittig, R. M., Crockford, C., Seyfarth, R. M., & Cheney, D. L. (2007). Vocal alliances in Chacma baboons (*Papio hamadryas ursinus*). *Behavioral Ecology and Sociobiology*, *61*(6), 899–909. https://doi.org/10.1007/s00265-006-0319-5

Wu, J., Balliet, D., & Van Lange, P. A. M. (2016). Reputation, gossip, and human cooperation. *Social and Personality Psychology Compass*, *10*(6), 350–364. https://doi.org/10.1111/spc3.12255

Young, C., Majolo, B., Schülke, O., & Ostner, J. (2014). Male social bonds and rank predict supporter selection in cooperative aggression in wild Barbary macaques. *Animal Behaviour*, *95*, 23–32. https://doi.org/10.1016/j.anbehav.2014.06.007.

Supplementary Material

Supplemental Videos S1-S4

All four can be found at: https://osf.io/tnkz2/

Section 1: Information about Subjects and Experimental Rooms

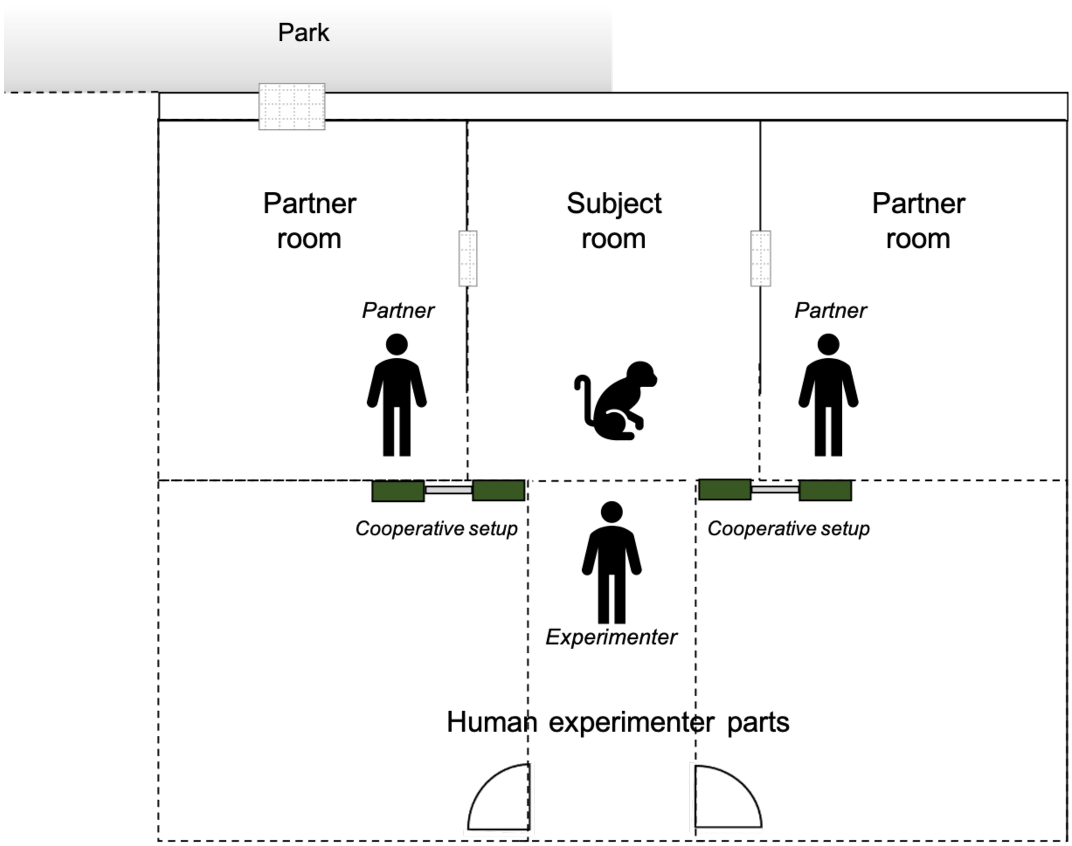
Table S1

Demographic Information and Training Progress of the 15 Tonkean Macaques who Started the Training Phase

Name	Sex	Age (Years) at the Start of Training	Training Step Reached	Other Relevant Information
Abricot	M	9	All steps validated	Removed from the group during training
Alaryc	M	9	All steps validated	Removed from the group during training
Barnabé	M	8	All steps validated	Removed from the group during training
César	M	7	All steps validated	Removed from the group during training
Néréis	F	23	All steps validated	Dominant female of the group
Eric	M	5	All steps validated	Injury before experiment
Ficelle	F	4	All steps validated	Gave birth during training
Dory	F	6	Step 7	Gave birth during training
Patchouli	M	11	Step 7	Dominant male of the group, no regular attendance
Bérénice	F	8	Step 3	Gave birth during training
Lassa	F	11	Step 1	Gave birth during training
Nema	F	11	Step 1	Gave birth during training
Gandhi	M	3	Step 1	no regular attendance
Horus	M	2	Step 1	no regular attendance
Yin	F	13	Step 1	no regular attendance

Figure S1

Configuration of the Experimental Rooms n°2



Note. Experimental rooms where one subject has been tested and the first training steps of the four other subjects took place. The subject was in the middle room ('Subject room'), with all trapdoors closed. The two cooperative setups were fixed on the human side of the mesh panels. The subject had access to the boxes at the bottom of each cooperative setup and could operate them through the mesh. The two human partners were situated in the lateral rooms ('Partner room'), each on one side. They could only access and operate the top boxes from each cooperative setup. The subject room was separated from the partner rooms with mesh for the first one and a half meter in front of the cooperative setups. In this way, the subject could perfectly see the human partners and observe them manipulating the boxes.

Training and familiarization were carried out individually in experimental rooms situated next to their outdoor enclosure (Figure S1). One subject was also tested in these experimental rooms. Due to animal husbandry management decisions, the four other subjects were moved into a smaller enclosure midway through the training. Consequently, these individuals received the last steps of training and familiarization as well as the test sessions in this different enclosure (see Figure 2).

Section 2: Training Phase Procedure

1. How to operate the tasks. All Tonkean macaques were individually trained to perform each of the tasks with a single setup (one box apparatus). During this training step, a single puzzle box was presented to the Tonkean macaques in one of the four possible locations (top left, top right, bottom left, bottom right). In this way, Tonkean macaques experienced the different locations, including the slots where the partners will be working later on (i.e., the respective top boxes). They received one piece of fruit (mango, apple, or two or three raisins) for each successful apparatus manipulation. We considered them to have understood the tasks when they were able to successfully perform at each task for at least six balls in a row, and when they had experienced at least two sessions in each of the four locations. Ten out of the 15 Tonkean macaques who started this step passed the criterion and moved on to the next step.

2. Introduction to the co-action setups. The Tonkean macaques were individually familiarized with the cooperative setups (*i.e.*, two boxes connected with the transparent hose) to make them understand that the ball rolls from one box to the other and reward is obtained only once the ball reaches the aluminum tray of the bottom box. For this and the following training steps, only the Slider task was used, unless otherwise indicated. In this step, one cooperative setup was mounted and the Tonkean macaques had the opportunity to access and interact with both boxes. The ball was inserted by the experimenter into the top box. To be rewarded, the Tonkean macaques had to successfully move the ball down inside the top box first, and then inside the bottom box. We repeated this procedure at both locations. The Tonkean macaques had to move from one box to the other and operated successfully both boxes without being prompted by the experimenter, for four trials in a row on both locations, to go to the next training step. All ten Tonkean macaques passed the criterion and moved on to the next step.

3. Introduction to the human partner. One cooperative setup was mounted and one human actor (different from the human partners in the actual test sessions) stood in front of the double setup and played the role of the partner. The top box was made inaccessible to the Tonkean macaques either by fixing a transparent panel on the mesh so that they could not operate the top box anymore or by closing the trapdoor allowing access to it. We conducted sessions of eight trials, in which both the Tonkean macaque and the human partner failed once. The idea was to familiarize the Tonkean macaques that they can fail as well as the human partner, and the outcome was the same in any case (no reward for any of them). Tonkean macaques performed alternatively the top and the bottom boxes on each cooperative setup's location. Two sessions per location per box were conducted for each Tonkean macaque (i.e., eight sessions in total). No success criteria were used for this step. We just wanted our subjects to experience both roles (perform on the top box or on the bottom box) with success and failures to maximize the chances for them to understand the cooperativeness of the puzzle boxes.

4. How to indicate a choice by touching a target. All Tonkean macaques first learned to touch a target to be rewarded, using the clicker training method. They received one raisin for each touch and moved on to the next step when they were able to touch the target several times in a row. The two targets were two wooden sticks with a yellow square shape at the end, and otherwise completely identical. Then, we wanted them to learn to make a choice between two targets resulting in different outcomes. The experimenter presented two food options (one raisin on one side vs four raisins on the other side) out of reach of the Tonkean macaque and then simultaneously presented the targets (one above each food option). By touching one of the targets, the Tonkean macaque could obtain the food option of the respective side. The optimal choices for the Tonkean macaques were then to touch the target corresponding to the side with four raisins. Sessions of eight trials were conducted, with a criterion of 100% success to go to the next training step. The position of the food options was pseudo-randomized within sessions (i.e., equally often on both sides and no more than twice in a row at the same one). One individual did not return to the experimental rooms regularly to continue. Then, nine out of the ten Tonkean macaques passed the criterion and moved on to the next step.

5. How to choose a setup. This training step we conducted for the Tonkean macaques to understand how to choose between the two setups and experience the consequences of their choices on their food intake. Only single puzzle boxes were used, one on each side of the experimental room (at the bottom positions). Experimenter 1 called the Tonkean macaque in the middle of the room, at equidistance from the single boxes, and presented two different food options (one vs three raisins) to the Tonkean macaque. After ensuring the Tonkean macaque had looked at both food options, Experimenter 1 placed them on each side in front of each single box before presenting the targets. After the Tonkean macaque had touched one target, Experimenter 2 went in front of the chosen box and Experimenter 1 gave the chosen food option to Experimenter 2. Experimenter 2 inserted the ball into the box and gave the food reward once the Tonkean macaque successfully navigated the ball inside the box. The location of the food options was pseudo-randomized within sessions (i.e., equally often on each location but not more than twice in a row at the same location). One session consisted of eight trials. The Tonkean macaques had to choose the optimal option (i.e., the larger food option) in all trials of a session, or at least seven out of eight trials in two consecutive sessions. All nine Tonkean macaques passed the criterion and moved on to the next step.

6. Introduction of the two human partners. This training step was conducted to familiarize the Tonkean macaques with the possibility to choose between two human partners to perform with at the cooperative setups. Two cooperative setups were mounted, one on each side. Two humans (different from the human partners used in the experiment) simultaneously moved to stand in front of the cooperative setups and acted as skillful partners. Human partners were both skillful as we did not want the subjects to learn anything about these actors or to begin making any association with the testing conditions with success or failure. The experimenter presented the targets, and once the Tonkean macaque had touched one of them, placed the ball inside the top box of the chosen setup. Choosing one side/partner or the other led to the same food outcomes: one piece of fruit for the Tonkean macaque and the human partner. One session of four trials was conducted, with the only criterion that the Tonkean macaques must choose both human partners and sides at least once. If a Tonkean macaque always chose the same side or human partner, two additional trials were conducted on the unchosen location and/or with the unchosen human partner. As these Tonkean macaques had shown to quickly develop side biases in previous and other ongoing studies (personal communications), we wanted them to successfully experience both sides and human partners at this step.

7. Comprehension check in the cooperative context. The Tonkean macaques had to make a choice between the two cooperative setups, one with a human partner (‘partner setup’) and one without a partner (‘empty setup’). The Tonkean macaques had to choose the partner setup to obtain food (optimal choice), otherwise they went empty-handed. When a Tonkean macaque chose the empty setup, the experimenter inserted the ball inside the top box of the empty setup and waited for 15 seconds before retrieving the ball. As no human partner was there to operate the top box, the Tonkean macaque was unable to operate the ball inside the bottom box. The side of the human partner was pseudo-randomized across trials. When Tonkean macaques demonstrated at least 80% success (i.e., choosing the partner setup) in two consecutive sessions of eight trials, we considered them to have sufficient comprehension of the test procedure. Individuals who reached this criterion were allowed to participate as subjects in the experiment. Two individuals did not return in the experimental rooms regularly to continue. Then, seven out of the nine Tonkean macaques passed the criterion and moved on to the next step.

8. Preparatory step to the testing conditions. The subjects performed two sessions of four trials with the exact same procedure as training step 6, one session with the Slider task and one with the Cog task. This last step was to familiarize the subjects with the testing conditions before the start of the experiment, and remind them that they can make a choice between two human partners to perform at these co-action tasks.

At the end of this training procedure, seven out of 15 Tonkean macaques managed to pass the criteria and showed enough motivation to come regularly to the experimental rooms. However, a few days before the experiment, one individual was injured and another never wanted to return to the experimental rooms. These two individuals could not be tested, giving a total number of five subjects for the experiment.

Section 3: Video Coding

We coded the videos using the Behavioral Observation Research Interactive Software (“Boris”; Friard & Gamba, 2016). We coded the following behaviors: (a) the partner choices of the subjects during initial preference assessment and test sessions; and (b) the looking behaviors of the subjects (i.e., head orientation, or eye orientation when visible) directed toward an actor or the experimenter (face, body, or hands) during the demonstration phase. The action of the human partner (box manipulation, food handling, or other) when observed by the subjects was also coded (see Table S2 and S3 for details). During the demonstration phase, only one human partner was present at a time. We only coded subjects’ looking time when the human partner was performing the task, not when it was the subject’s turn.

Inter-coder reliability was calculated using Cohen’s kappa coefficient for partner choices, and intraclass correlation coefficients (Koo & Li, 2016) for looking behaviors based on a single rating ($k = 2$ ratters), two-way mixed-effects model and consistency (using the function icc from the package irr (version 0.84.1) in R (version 4.3.2; R Core Team, 2022). Due to technical problems with the cameras during Test session 2 of the subject Barnabé, two trials of the session were not recorded and could not be coded. These two trials were included in the data analysis but not in the reliability coding. Hence, the inter-coder reliability test for the partner choices was based on 118 measurements (instead of 120 trials in total). For the looking behaviors, the inter-coder reliability was based on 240 measurements.

The subjects were given a maximum of one minute to make a choice, otherwise the trial was aborted. In this case, the same trial (with the same human partners’ side configuration) was repeated. If a subject did not make a choice three times in a row, the session was aborted. Only one subject did not touch a target within the minute twice in row for the third trial of the first session, but then made a choice at the third attempt so we were able to continue the session. The reason why this subject did not touch the target is unknown but she conspicuously moved and waited in front of one of the cooperative setups both times instead. The experimenter then considered this “moving and waiting in front of one cooperative setup” as a clear behavior indicating a choice and decided to accept it as a partner choice for the next trials. This subject did not touch a target but moved and waited in front of one cooperative setup twice after that, on the fifth and eighth trial of the first test session.

Table S2

Definitions of Behaviors Coded for the Initial Preference and Test Phase

Behavior	Definition
Start of trial	When the subject was in the middle, in front of the experimenter, and the two human partners stood just behind the experimenter (before moving on each side).
Decision	When the subject touched one of the targets or took a piece of food from a human, who was then considered as the chosen actor by the experimenter and coded on videos later. For the subject in experimental room n°2: when they touched the target or moved to a position in front of one cooperative setup.
End of trial	For the initial preference sessions: when the subject took the food from the actor. For the test sessions: for an unsuccessful trial (i.e., choice for the unskilled human partner), when the experimenter took the ball back from the box; for a successful trial (i.e., choice for the skillful human partner), when the experimenter placed the food bowl back on the ground after giving a food reward to the subject and the human partner.

Table S3

Definitions of Behaviors Coded for the Demonstration Phase

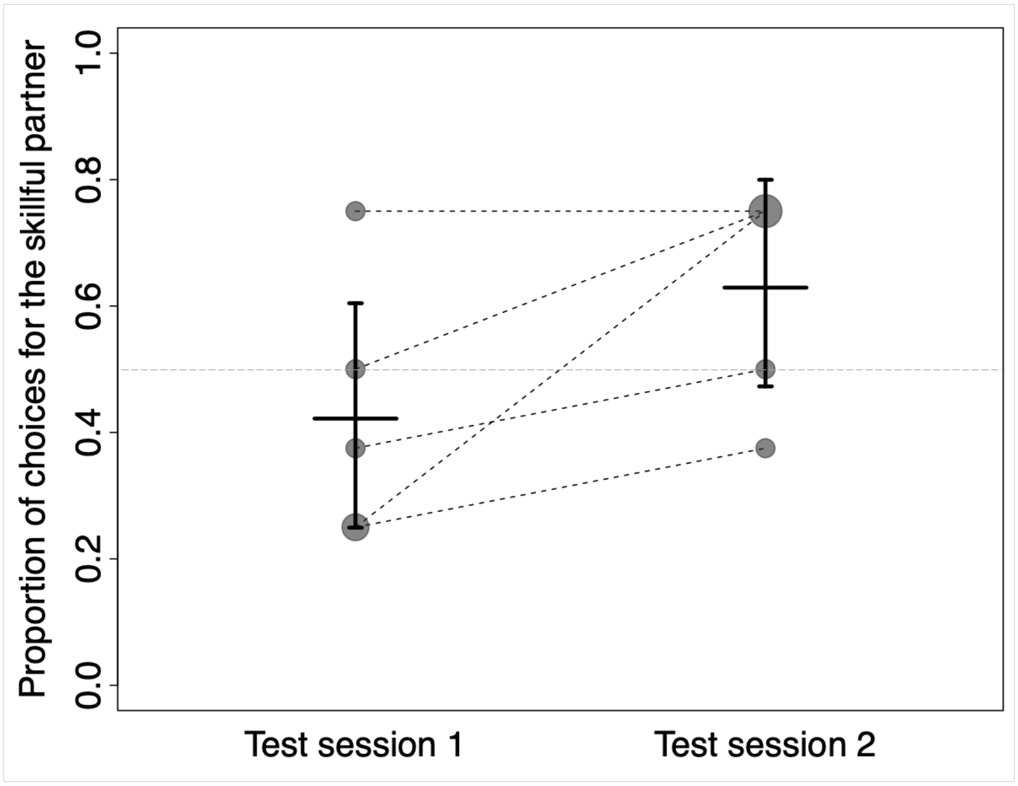
Behavior	Definition
Start of trial	When the human actor was in place, standing in front of the setup, and the experimenter had inserted the ball into the box.
Looking	Head (or eyes when possible) directed toward the human actor's face, body, or hands. The action of the human actors while the subject was looking could fall into one of three categories: manipulation of the box, food handling, or other context.
End of box manipulation	For successful trials: when the experimenter took the ball back from the ball collector. For unsuccessful trials: when the human actor stopped manipulating the box.
End of trial	When the experimenter put the food bowl back on the ground or touched the box to take the ball back.

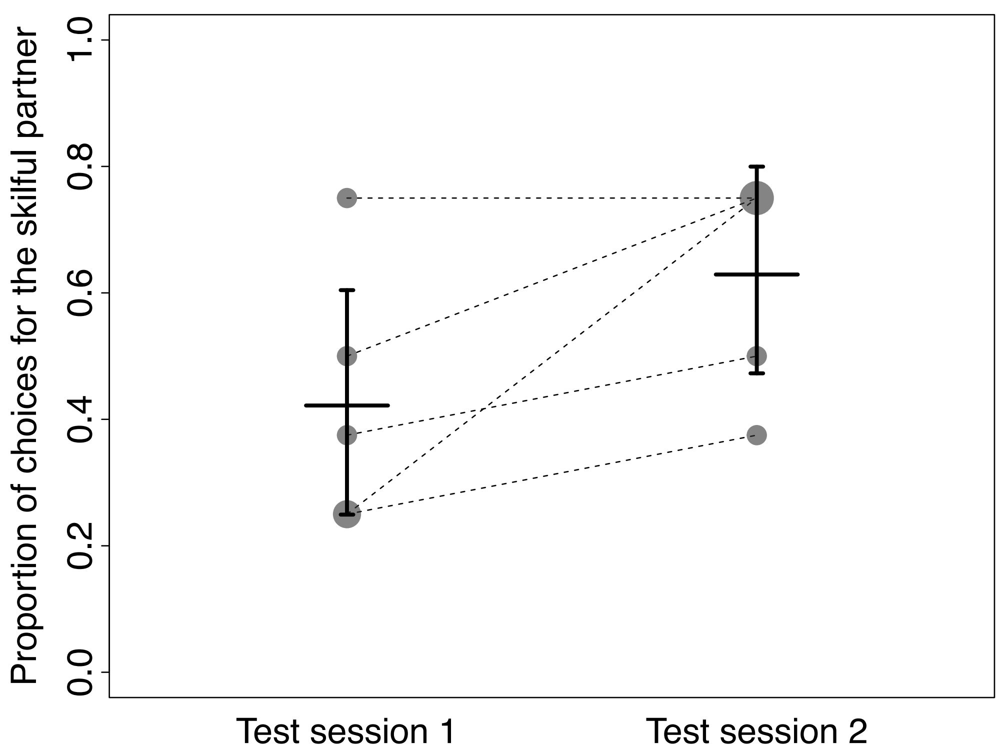
Section 4: Did the Subjects Prefer the Skillful Human Partner over the Unskilled Human Partner?

With this model, we wanted to estimate the effect of session (test session 1, test session 2) on the number of choices of the skillful human partner. We included subjects' initial partner preference, session number, and trial number as fixed effects, with trial number in an interaction with session. In order to avoid pseudo-replication and overconfident model estimates and to keep type *I* error rate at the nominal level of 5% (Barr et al., 2013; Schielzeth & Forstmeier, 2009), we included subject ID as a random intercepts effect, and random slopes of session (manually coded and then centered) and trial number within subject into the model. Prior to fitting the model, we z-transformed trial number and initial partner preference to a mean of zero and a standard deviation of one, to ease interpretation of the model estimates (Schielzeth, 2010) and model convergence. Originally, we also included estimates of the correlations between the random intercept and slopes into the model. We compared the log-likelihoods of the full model with one lacking the correlations in the random structure. As the removal of the correlations lead only to a small decrease of the maximum likelihood in the model fit (log-likelihoods; full model including the correlation parameters: -52.30, *df* = 11; full model lacking the correlation parameters: -52.41, *df* = 8), we decided to remove them from our full model. We assessed model stability by dropping individuals from the data, one at a time, fitting the full model to each of the subsets, and finally comparing the range of model estimates, with the model estimates we obtained for the full data set. To this end we used a function written by RM (Mundry, 2023). The model revealed to be of acceptable stability.

Figure S2

Comparison of the Proportion of Choices of the Subjects for the Skillful Human Partner in the Two Test Sessions





Note. The raw data and the fitted reduced model and its 95% confidence interval are shown for both sessions. Five subjects have been tested at each session, which represent ten data points in total.

Section 5: Was the Attention of the Subjects Predictive of their Partner Choices?

With this model, we wanted to estimate the effect of attention during the demonstration on the number of choices for the skillful human partner in the test phase. To potentially obtain the relevant information on the human partners' skills, subjects needed to observe both human partners during the demonstration phase. As a measure of the subjects' attention to both human partners, we hence chose the minimum looking time per subject to the skillful and unskilled human partner, i.e., the time during which each subject observed each human partner, and used this for the following analysis. Into the model, we included the attention during the demonstration, test session (test session 1, test session 2), and their interaction. In order to avoid pseudo-replication and overconfident model estimates and to keep type I error rate at the nominal level of 5% (Barr et al., 2013; Schielzeth & Forstmeier, 2009), we included subject ID as a random intercept effect and test session (manually coded and then centered) as a random slope within subjects. Prior to fitting the model, we z-transformed attention to a mean of zero and a standard deviation of one, to ease interpretation of the model estimates (Schielzeth, 2010) and model convergence. Originally, we also included estimates of the correlations between the random intercept and slope into the model. We compared the log-likelihoods of the full model with one lacking the correlation in the random structure. As the removal of the correlation led to a very minor change of the maximum likelihood in model fit (log-likelihoods; full model including the correlation parameter: -52.402, *df* = 7; full model lacking the correlation parameter: -52.405, *df* = 6), we decided to remove it from our full model. We assessed model stability as described above. The model revealed to be of acceptable stability.

Table S4

Results of the Full Model Using Subjects' Choices as the Response Variable and the Attention as the Main Predictor (Estimates Together with Standard Errors, Confidence Limits, Significance Tests and Range of Estimates when Dropping Individuals One at a Time)

Term	Estimate	SE	CL_{Lower}	CL_{Upper}	χ^2	*df*	*p*	min	max
(Intercept)	-0.311	0.356	-1.151	0.351				-0.752	-0.009
Session	0.824	0.469	-0.034	2.026				0.572	1.098
Attention	0.516	0.362	-0.202	1.496				-0.242	0.698
Session × Attention	-0.503	0.476	-1.722	0.496	1.151	1	.283	-0.748	-0.273

Note. Attention was z-transformed to a mean of zero and a standard deviation of one; mean and sd of the original attention were 145.4 s and 22.3 s, respectively. Session was dummy coded with test session 1 being the reference level.

Table S5

Results of the Reduced Model Using Subjects' Choices as the Response Variable and the Attention as the Main Predictor (Estimates Together with Standard Errors, Confidence Limits, and Significance Tests when Dropping Individuals One at a Time)

Term	Estimate	SE	CL_{Lower}	CL_{Upper}	χ^2	*df*	*p*
(Intercept)	-0.310	0.351	-1.034	0.412			
Session	0.839	0.468	0.000	1.859	3.305	1	.069
Attention	0.276	0.272	-0.273	0.902	0.972	1	.324

Note. Attention was z-transformed to a mean of zero and a standard deviation of one; mean and sd of the original attention were 145.4 s and 22.3 s, respectively. Session was dummy coded with test session 1 being the reference level.

Section 6: Did the Choices on the First Trial Influence the Probability of Choosing the Skillful Partner During the Rest of the Test Session?

Following a relevant suggestion from a reviewer, we investigated whether subjects who chose the skillful human partner on the first trial were more likely to stick with their choice and choose the skillful human partner more often during the rest of the session. We fitted a similar model to estimate the effect of the choice on the first trial in each session on the number of choices for the skillful human partner. Into the model, we included the choice on the first trial (skillful, unskilled), test session (1, 2), and trial number (2 to 8). To avoid the choices on the first trial being present in the response and the predictor, we removed trials number 1 from the dataset, which then included the choices from trials 2 to 8. In order to avoid pseudo-replication and overconfident model estimates and to keep type I error rate at the nominal level of 5% (Barr et al., 2013; Schielzeth & Forstmeier, 2009), we included subject ID as a random intercept effect and test session (manually coded and then centered) and trial number as random slopes within subjects.

Prior to fitting the model, we z-transformed trial number to a mean of zero and a standard deviation of one, to ease interpretation of the model estimates (Schielzeth, 2010) and model convergence. Originally, we also included estimates of the correlations between the random intercept and slope into the model. We compared the log-likelihoods of the full model with one lacking the correlation in the random structure. As the removal of the correlation led to a very minor change of the maximum likelihood in model fit (log-likelihoods; full model including the correlation parameter: -46.472, *df* = 10; full model lacking the correlation parameter: -46.582, *df* = 7), we decided to remove it from our full model. We compared this full model with a null model lacking the effect of choice at first trial in the fixed effects part but being otherwise identical. We obtained confidence intervals of model estimates and fitted values by means of a parametric bootstrap (N = 1,000 bootstraps; function bootMer of the package lme4). We assessed model stability as described above. The model revealed to be of acceptable stability.

Overall, the probability of the subjects to choose the skillful human partner was not significantly influenced by their choices on the first trial in each test session (full-null model comparison: $\chi^2 = 0.802$, *df* = 1, *p* = .371; see Table S6).

Table S6

Results of the Full Model for the Effect of the Choice on First Trial (Estimates Together with Standard Errors, 95% Confidence Limits, Significance Tests, and the Estimates' Range After Excluding Individuals One at a Time)

Term	Estimate	SE	CL_{Lower}	CL_{Upper}	χ^2	*df*	*p*
(Intercept)	-0.477	0.406	-1.531	0.268			
Session	0.734	0.498	-0.313	2.002	2.187	1	.139
Trial number	0.035	0.283	-0.603	0.658	0.015	1	.902
Choice first trial	0.446	0.498	-0.662	1.725	0.802	1	.371

Note. Trial number was z-transformed to a mean of 0 and a standard deviation of 1 (original mean and standard deviation was 5 ± 2.01). Session was dummy coded with test session 1 being the reference level. The reference level for the choice on the first trial was the skillful human partner.

Section 7: Were the Subjects' Choices Influenced by the Location of the Human Partners?

Following a relevant suggestion from a reviewer, we examined whether the location of the human partners on the left or right setup influenced the probability of choosing the skillful human partner. We fitted a similar model except that we included the location of the skillful human partner (left, right), test session (1, 2), and trial number (1 to 8) as fixed effects. In order to avoid pseudo-replication and overconfident model estimates and to keep type I error rate at the nominal level of 5% (Barr et al., 2013; Schielzeth & Forstmeier, 2009), we included subject ID as a random intercept effect and test session and location (manually coded and then centered) and trial number as random slopes within subjects.

Prior to fitting the model, we z-transformed trial number to a mean of zero and a standard deviation of one, to ease interpretation of the model estimates (Schielzeth, 2010) and model convergence. Originally, we also included estimates of the correlations between the random intercept and slope into the model. We compared the log-likelihoods of the full model with one lacking the correlation in the random structure. As the removal of the correlation led to a very minor change of the maximum likelihood in model fit (log-likelihoods; full model including the correlation parameter: -51.219, *df* = 14; full model lacking the correlation parameter: -53.280, *df* = 8), we decided to remove it from our full model. We compared this full model with a null model lacking the effect of choice at first trial in the fixed effects part but being otherwise identical. We obtained confidence intervals of model estimates and fitted values by means of a parametric bootstrap (N = 1,000 bootstraps; function bootMer of the package lme4). We assessed model stability as described above. The model revealed to be of acceptable stability.

Overall, the probability of the subjects to choose the skillful human partner was not significantly influenced by the location on the left or right side of the human partners (full-null model comparison: χ^2 = 0.0005, *df* = 1, *p* = .982; Table S7).

Table S7

Results of the Full Model for the Effect of the Location of the Human Partners (Estimates Together with Standard Errors, 95% Confidence Limits, Significance Tests, and the Estimates' Range After Excluding Individuals One at a Time)

Term	Estimate	SE	CL_{Lower}	CL_{Upper}	χ^2	*df*	*p*
(Intercept)	-0.330	0.472	-1.454	0.607			
Session	0.868	0.478	-0.005	2.096	3.394	1	.065
Trial number	0.062	0.241	-0.451	0.613	0.065	1	.799
Side skillful human partner	0.014	0.564	-1.233	1.193	0	1	.982

Note. Trial number was z-transformed to a mean of 0 and a standard deviation of 1 (original mean and standard deviation was 5 ± 2.01). Session was dummy coded with test session 1 being the reference level. The reference level for the location of the skillful human partner was the left.

Supplementary References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Mundry, R. (2023). *Some R functions* [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7670524

R Core Team. (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.r-project.org/

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution*, *1*(2), 103–113. https://doi.org/10.1111/j.2041-210X.2010.00012.x

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145