PSYCHOLOGICAL SCIENCE

Do marmosets understand others' conversations? A thermography approach

R. K. Brügger*, E. P. Willems, J. M. Burkart

What information animals derive from eavesdropping on interactions between conspecifics, and whether they assign value to it, is difficult to assess because overt behavioral reactions are often lacking. An inside perspective of how observers perceive and process such interactions is thus paramount. Here, we investigate what happens in the mind of marmoset monkeys when they hear playbacks of positive or negative third-party vocal interactions, by combining thermography to assess physiological reactions and behavioral preference measures. The physiological reactions show that playbacks were perceived and processed holistically as interactions rather than as the sum of the separate elements. Subsequently, the animals preferred those individuals who had been simulated to engage in positive, cooperative vocal interactions during the playbacks. By using thermography to disentangle the mechanics of marmoset sociality, we thus find that marmosets eavesdrop on and socially evaluate vocal exchanges and use this information to distinguish between cooperative and noncooperative conspecifics.

INTRODUCTION

How nonhuman primates see the world, and in particular how they see their social world, is a longstanding question. For instance, can they extract information from the mere observation of social interactions between conspecifics, and do they also assign value to these interactions? An increasing number of behavioral studies have been addressing such questions over the last few decades (1), providing a glimpse into the rich social lives of primates. What is typically lacking, however, are reliable measures of what is happening inside the mind of the observer, what we will call the inside perspective—in line with the seminal work of Cheney and Seyfarth (1). This inside perspective is notoriously difficult to quantify. New noninvasive technologies such as thermography have the potential to provide this inside perspective by quantifying even subtle changes in emotional arousal (2). Combining thermography with behavioral experiments could thus be a powerful approach to disentangle the mechanics of complex social behavior. Here, we pioneer this approach to investigate how marmoset monkeys process and evaluate cooperative and noncooperative vocal interactions between third parties.

The ability to extract information from social interactions of conspecifics [social eavesdropping (3)] and assign value to those interactions [social evaluation (4)] has been investigated in several species [fishes, e.g., (5); birds, e.g., (6); mammals, e.g., (7)]. Social evaluation in the context of cooperation is most important for species with a social system in which cooperation plays a substantial role and individuals have to be aware of the cooperativeness of their potential partners. Accordingly, it is ubiquitous in humans who constantly classify individuals as cooperative versus noncooperative interaction partners, merely based on how these individuals interact with third parties (8–10). Social evaluation emerges early in human ontogeny, and a multitude of studies has addressed whether infants [as early as 3 months of age (11)] are able to distinguish between third parties that behave cooperatively versus antisocially, and whether they show

a preference for either of the parties [reviewed in (12, 13)]. Such studies generally present infants with video or live scenarios that show not only various agents in social situations of helping or hindering [climbing the hill scenario, e.g., (14)] but also fairness [allocating goods scenario, e.g., (15)] or benevolence [comforting and threatening scenario (16)]. Infants' reactions are assessed with gaze behavior (15) or directly expressed preferences (i.e., choosing one agent over the other) (11), and typical results show that infants have a preference for agents that they had seen interact cooperatively (i.e., they prefer a helper over a hinderer, someone who behaves fairly over someone who does not, and a comforting agent over a threatening one).

Some of these studies with infants have also been applied to nonhuman animals to investigate the phylogenetic origin of social evaluation [reviewed in (4, 17)]. Analogous to human children, animal subjects can first observe scenarios of social interactions between human actors. Their subsequent preference for one actor over the other is measured by giving the subjects the choice to accept food that is offered simultaneously by both actors. This approach has been applied to several nonhuman primates [great apes (18, 19), capuchin monkeys/squirrel monkeys (20, 21), common marmosets (22, 23), Japanese macaques (23)], dogs (24, 25), and, more recently, even dolphins (26) as well as horses (27). The scenarios varied with regard to content, showing, e.g., helpers versus nonhelpers or reciprocal versus nonreciprocal exchangers, and, in the latter context, whether food or tokens were used in exchange scenarios. The results are mixed, but most of the experiments conclude that the subjects have a bias against the noncooperative actor, i.e., they are less willing to accept food from a human actor who refused to help or did not reciprocate (4, 17).

However, several issues have been raised concerning such studies. First, the scenarios are typically implemented with human actors instead of conspecifics [but see (28)], and there is no a priori reason to assume that animals reason about humans as they do about conspecifics. Second, the scenarios often involve food exchanges, which make the final choice in the preference test vulnerable to alternative explanations. For instance, rather than figuring out the content of the interaction, animals may simply learn which human actor is most likely to give them food. In some studies, subjects could simply form

Department of Anthropology, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.
*Corresponding author. Email: rahelkatharina.bruegger@uzh.ch

an association between a specific actor and food without considering the content of the interaction per se (*24*). Third, some of the results could arguably be mere results of side biases (*25, 29*). In summary, many studies on social evaluation on animals are inconclusive to date.

These studies have the fundamental flaw that they only measure direct behavioral responses of the subjects without taking the inside perspective into account. This is a major limitation because it is well perceivable that an individual correctly perceives and evaluates the scenario, but nevertheless does not show any behavioral reaction. For instance, recent results from a pilot study (see section S1) suggest that common marmosets (*Callithrix jacchus*) may evaluate whether or not other individuals behave prosocially toward conspecifics, but often do not show this in overt behavior.

Common marmosets are cooperative breeders and have been reported to socially evaluate humans with the prevalent paradigm used to test for social evaluation (*22, 23*). They live in a social system with extended reliance on allomaternal care, and being able to distinguish between potential cooperation partners is of high importance. Within their family groups, help is provided by all group members (even if they are unrelated) in the form of infant carrying, provisioning, and shared vigilance (*30*). Food sharing plays an important role in marmoset infant rearing. All group members are known to readily share food with immatures, often using food calls to initiate proactive sharing interactions (*31*). In this pilot study (see section S1), a focal helper, together with the immatures from the group, was temporarily removed from their group. The group then heard a vocal playback that represented either a positive interaction between the helper and the immatures over food (begging calls from the immatures and food offering calls from the absent helper) or a negative one (begging calls from the immatures and agonistic chatter calls from the helper). A playback mimicking a food sharing interaction was expected to be perceived as a cooperative interaction, whereas the refusal to share food was expected to be perceived as a noncooperative interaction. Afterward, the group was reunited to test whether the other group members showed more socio-positive and/or less socio-negative behaviors toward the focal helper after the positive playback and vice versa for the negative playback.

Overall, group members did not systematically show more socio-positive behaviors after the simulation of a cooperative interaction or more socio-negative behaviors after simulating a noncooperative interaction. Nevertheless, in some cases, noncooperative playbacks elicited strong punishment, namely, of two female helpers who were ready for their own reproductive career. These females were therefore not well integrated in the group and may potentially have competed for the breeding position and posed a potential threat to the immatures (*32, 33*). It may thus well be that marmosets can engage in social evaluation and show negative reactions to poor cooperators but do so only in times of group instability and not when all group members are well integrated. Alternatively, however, marmosets may simply lack the ability to process call sequences holistically as meaningful vocal social interactions. Instead, they may rather process these sequences serially as a concatenation of independent events. Here, we use thermography to distinguish between these two possibilities of what happens in the mind of marmoset monkeys when listening to conspecific vocal interactions.
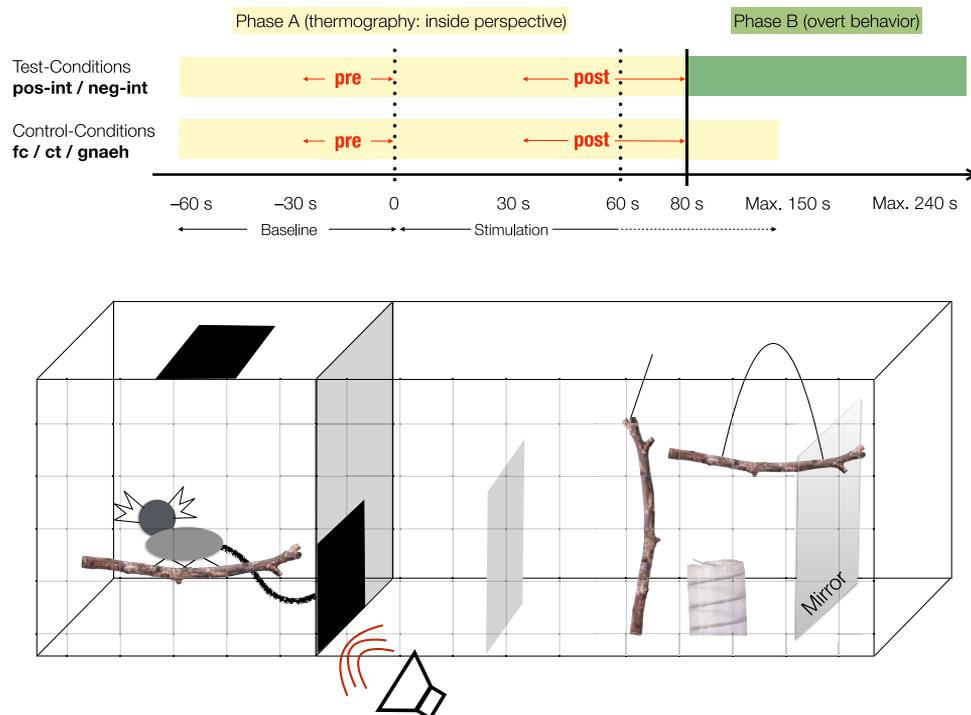

Fig. 1. Experimental setup. Schematic representation of the experimental setup with the phases, periods, and subphases of the experiment. During phase A (yellow), marmosets were sitting on a perch in the front of the compartment on the left-hand side. During phase B (green, only after test conditions), marmosets could choose to explore the compartment on the right-hand side or go back to the home enclosure (via the black sliding doors that were opened after phase A). The baseline period lasted for 60 s before the onset of the playback stimulus. The playback stimulus started at time point 0 and lasted for 60 s (dotted lines on timeline). Temperature values were only extracted from subphases pre and post (red arrows on timeline: 30 to 80 s after stimulus onset).

Our goal was to investigate social evaluation in common marmosets, taking both the inside perspective and the overt behavioral responses into account. We developed a paradigm that would not require the use of human actors or food and that was not prone to produce location biases. To do so, we presented playbacks of vocal cooperative or noncooperative interactions to stranger conspecifics, which produced a strict third-party context not confounded by within-group social relationships (*34*). In the first phase of the study (phase A—thermography: inside perspective), we used thermography to investigate how the subjects would process the presented scenarios. In the second phase (phase B—overt behavior), we assessed whether the subjects would have a preference for the cooperative over the noncooperative individuals that they had heard in the playbacks.

The goal of phase A was to quantify whether marmosets perceived the playbacks holistically as social interactions and would thus extract more information from the exchange of calls than would be available if the calls occurred as isolated signals (*35*). We therefore measured the thermal change in response not only to the positive and negative interaction playback (consisting of begging calls from immatures as well as food offering calls or aggressive chatter calls from adults) but also to the separate calls (i.e., playback of only begging calls, only chatter calls, only food calls). If the marmosets perceived the interaction playback as a social interaction rather than as a concatenation of separate independent events, the response to the interaction playbacks should be systematically different compared to the mere additive responses to the separate calls. For instance, if immature begging calls and adult food calls have a relaxing effect, but the combination of begging calls and food calls (positive interaction playback) has a strong arousing effect, then the reaction to the interaction playback can clearly not be the result of a simple additive effect of the separate calls. This would suggest that marmosets process and perceive the interaction playbacks holistically as conversations rather than as the sum of the single elements.

Thermography allows very precise measurements of changes in physiological arousal via infrared radiation of the skin not only in humans but also in animals, including common marmosets (*2*) [humans, e.g., (*36*); other primates, e.g., (*37*, *38*)]. It is fundamentally based on the principle that changes in emotional states are co-occurring with changes in the autonomic nervous system (ANS). The ANS controls cutaneous blood flow during sympathetic activation (fight or flight response) and results in a reduction of body surface temperature in regions where blood flow is reduced (*37*, *39*). To answer whether marmosets understood the playback stimuli holistically, which is as an interaction, we compared whether the thermal reaction shown while witnessing the interaction playbacks was different from the additive effect of arousal levels measured while witnessing the corresponding control playbacks separately. To additionally validate the thermal measurements, we compared the patterns to independent but less sensitive behavioral markers of arousal and also controlled for potentially confounding effects of activity (*2*).

The goal of phase B was to behaviorally quantify whether the marmosets socially evaluated the opposite sex stranger simulated with the playback. We predicted that if they socially evaluated the playbacks, they would more likely approach an interlocutor involved in a cooperative interaction compared to a noncooperative one. After the presentation of the interaction playbacks, the marmosets were therefore given the possibility to enter the compartment from where

the playback was broadcast. In this compartment, a mirror partially covered by various enrichment materials was used to give the marmosets the illusion that a conspecific was in this compartment [marmosets do not recognize their mirror image but react to their reflection with social behaviors; see (*40*)]. We measured whether the marmosets would preferentially enter this compartment and approach the mirror after having heard cooperative versus noncooperative playbacks. We expected that they would preferentially do so after the cooperative playbacks, which would indicate a preference for cooperative over noncooperative strangers.

MATERIALS AND METHODS
Subjects and housing
We tested 21 captive-born adult common marmosets (*C. jacchus*; four female breeders, three male breeders, seven female helpers, and seven male helpers; see table S3 for detailed information about subjects' age and respective group compositions). They were housed as family groups or pairs in heated indoor enclosures equipped with various climbing materials (branches, ropes, tubes, and platforms), a sleeping box, and floors covered by bark mulch. The animals had regular access to outdoor enclosures and a separate testing room via a semitransparent tube system.

Feeding of all animals occurred at least twice a day, once in the morning with a vitamin-enriched mash and at around noon with fresh fruit and vegetables. During the afternoon, animals received various additional protein sources such as insects or nuts as well as gum. Water was always available ad libitum.

Ethics statement
All the experiments were in accordance with the Swiss legislation and licensed by the Kantonales Veterinäramt Zürich (license number: ZH223/16, degree of severity: 0).

Procedure and playback stimuli
General procedure
The experiment consisted of two phases. During the first phase (phase A—thermography: inside perspective), the marmosets heard one of five playback stimuli, i.e., two test stimuli and three control stimuli. The test stimuli were each composed of two different call types to simulate a cooperative or negative interaction over food. The control stimuli were composed of each of the separate call types used for the test stimuli and were added to examine whether the simulated interactions were perceived as interactions rather than as the sum of the parts of the playback. For the second phase (phase B—overt behavior), where we wanted to link the test stimuli with the behavioral preferences for the simulated individuals of the playback, the experimenter opened the two doors of the testing compartment (two doors of the compartment on the left, indicated in black; Fig. 1), and the animals could either decide to explore the additional compartment with the putative caller or return to their home enclosure via the door on top. Each individual experienced only one condition per day in a randomized order (with the constraint that half of the individuals experienced pos-int first and the second half experienced neg-int first; for details, see table S3).
Phase A: Playback stimuli
To assemble the playback stimuli, we used recordings of the vocalizations of unrelated strangers: two adult animals, a male and a female helper, and immatures. Each focal individual thus experienced the

playback of one and the same opposite sex outgroup adult and the same outgroup immature (in the interaction playbacks).

The five playback stimuli each lasted for 1 min and consisted of six call sequences starting every 10 s with a duration of 1 to 8 s, mimicking naturalistic vocal interactions as closely as possible. During test conditions, we simulated either a cooperative interaction involving food with call sequences of infant begging calls and adult food calls (pos-int) or a negative interaction over food with call sequences of infant begging calls and adult chatter calls (neg-int) (*41*). During noninteraction control conditions, we played back stimuli that consisted of sequences of one call type only out of the calls used in the test stimuli, namely, food calls only (fc), chatter calls only (ct), and infant begging calls only (gnaeh), thus simulating the presence of a single individual. Playback files were assembled using iMovie 10.1.8, and the loudness of individual calls was adjusted to a common level.

Phase A: Procedure

The experiments were conducted individually in a separate experimental room inside a testing compartment (60 cm × 50 cm × 50 cm; see Fig. 1, compartment on the left) and only started after each individual was trained to follow the experimental protocol (see section S2). Phase A was recorded with an infrared thermography camera (FLIR T620; temperature sensitivity: 0.04°C, resolution: 640 × 480 pixels, sampling rate: 30 frames per second). For additional settings, see section S2.

Each experimental session started when the focal animal was sitting quietly on a perch. Phase A consisted of a baseline period (see Fig. 1) lasting for 60 s, where the subject was given a small food reward every 20 s to ensure that the focal animal's attention was kept toward the camera. Then, a 60-s stimulation period (see Fig. 1) followed, where the focal animal was exposed to one of the five playback stimuli. During this time, the experimenter turned her back to the animal and observed the situation via the thermal camera recording on the laptop. Thermal data were extracted for up to a maximum of 80 s. No additional food was given during this period. To additionally validate the thermal measurements, we videorecorded the animals to quantify independent behavioral markers of arousal and activity.

Phase B: Procedure

After the test stimuli of phase A, we assessed whether the subjects showed a preference for the cooperative (mimicked in pos-int) versus noncooperative (mimicked in neg-int) interaction partner. To do so, the experimenter simultaneously opened two sliding doors, leaving the monkey with the choices of returning to their home enclosure or exploring an additional compartment from where the playback had been broadcast (right side; Fig. 1). On the far right of the compartment (at a maximum distance from the compartment door), a mirror was installed that produced the illusion of a conspecific in that compartment [common marmosets do not recognize their mirror image but treat it as a conspecific; see (*40*)]. This mirror was only visible once the focal animal passed the visual barrier that was placed in the first third of the compartment (gray shield; Fig. 1). In front of the mirror, the compartment was equipped with some branches and other familiar materials to enhance the illusion of encountering the previously simulated opposite sex outgroup individual. This preference test was conducted only twice, after the test stimuli, to avoid that animals would quickly learn that the playback was not real, and no other individual was present in the adjacent compartment. A buffer period of 20 s (i.e., the period after the dotted line at 60 s in Fig. 1) allowed the experimenter to open the doors, and phase B lasted for a maximum of 180 s (for a detailed overview of the duration of phase B for each

focal animal, see table S4) or until the individuals decided to leave the experimental room and enter the tube to the home enclosure.

Data coding
Thermal data: Data coding

Thermal data were extracted with a MATLAB (R2018b) script that allowed to manually mark the region of interest (ROI), in our case, the nose with an ellipse and save the minimal temperature value of this region to a .csv file to always extract the coldest point on the nasal tip. Because marmosets were allowed to move freely in the experimental compartment, we placed ellipses on the ROI on all frames that satisfied the following four strict criteria: (i) minimal distance of the animal's ROI to the mesh to ensure that the nose was in the focal plane of the camera. (ii) ROI was not covered or partially covered by mesh. (iii) The animal's head was oriented straight toward the camera (maximum tilt angle in all directions about 45°). (iv) The animal's ROI was not blurred because of movement of the subject or from adjusting the camera. According to the insights of a previous study (*2*), we coded the following time periods: 30 s before the stimulus onset (subphase pre; see Fig. 1, red arrows) and 30 s after the stimulus onset until phase B of the experiment started or a maximum of 80 s after the stimulus onset (subphase post; see Fig. 1, red arrows). With this data extraction protocol, we were most likely to record the flight or fight response and were still able to gather enough appropriate frames. Before including them in the final dataset, we visually inspected the extracted temperature values for obvious outliers. We only included sessions with at least 10 frames per subphase to ensure robust high-quality data for each session (*2*). We assessed interrater reliability by analyzing 20% of all thermography videos by a second rater and found an interclass correlation coefficient (ICC 3) of 0.94, 95% confidence interval (CI) [0.88, 0.97].

Thermal data: Data preparation

We extracted raw minimum nasal temperatures from a total of 9912 video frames collected on 21 individuals over 90 sessions. Because we were interested in changes of nasal temperature, we centered all values around their respective session's baseline value, i.e., the mean minimum temperature over $t = -30$ to 0 s.

As individuals moved unpredictably, the number of extracted frames varied substantially across sessions (means ± SD = 110 ± 54.4 frames). Therefore, to reduce random measurement error and variation (e.g., due to the animal's breathing), we aggregated data over 1-s intervals to amplify the signal-to-noise ratio. This resulted in a dataset comprising a total of 1711 temperature values (means ± SD per session = 19 ± 7.38).

Behavioral data

Behavioral data were coded from the videos with the software INTERACT (Mangold GmbH, version 18.0.1.10). For phase A, we coded both piloerection of the tail and occurrences of high arousal calls to be able to verify that nasal temperature change was correlated with behavioral measures of arousal. We additionally coded activity levels to control for this potential confound. For phase B, we quantified the marmosets' latency to looking into the mirror. Detailed definitions of these variables can be found in section S3.

We assessed interrater reliability for all behaviors used in phase A by analyzing 30% of all videos by a second rater. We reached the following interclass correlation coefficient (ICC 3) for piloerection: 0.99, 95% CI [0.97, 0.99]; high arousal calls: 1, 95% CI [1, 1]; activity: 1, 95% CI [1, 1]. For phase B, we analyzed 20% of all videos by a second rater and reached interclass correlation coefficient (ICC 3) for latency to look into mirror: 0.94, 95% CI [0.71, 0.99].

Statistical analysis

All statistical analyses were performed in R (version 3.5.3). We used linear mixed-effects models (lme), generalized linear mixed-effects models (glmms), and cox proportional hazard mixed-effects models (function "lme," package "nlme"; function "glmer," package "lme4"; function "coxme," package "coxme") and always compared the full model with all fixed effects of interest to the null model only including the random intercept using likelihood ratio tests (LRt; function "Anova," package "car"). Model assumptions were checked with residual histograms and qq-plots of residuals as well as plots of residuals against fitted values. We checked the presence of influential cases with Cooks distance (function "CookD," package "predictmeans"). Multicollinearity was assessed with the fixed effect correlation matrix of the model (values < 0.7). Post hoc comparisons were conducted on the full model (function "emmeans," package "emmeans").

Overall thermal reaction

To assess whether and how temperature changed in response to the five different playback stimuli, we calculated an LMM (model 1) with centered nasal temperature as dependent variable. We were interested in the effects of condition (pos-int, neg-int, fc, ct, gnaeh), subphase (pre and post), as well as sexstatus [coded as breeders (including male and female breeders) = b, male helpers = mh, and female helpers = fh] and their two- and three-way interactions. If the individuals reacted differently to the playback, we would expect an interaction effect between condition and subphase. We controlled for dependencies within our data, by including session number, nested within subject ID and family group as random intercepts, while correcting for heteroscedasticity by specifying separate variance functions for each condition-sexstatus combination. To quantify the thermal changes within each condition, we compared estimated marginal means of the two subphases across condition separately for the sexstatus classes.

Individual thermal reactions

For each individual and session we separately determined whether the temperature change from subphase pre to subphase post in response to the playbacks was significant with a Wilcoxon rank sum test (function "wilcox.test," package "stats").

Additive effect

Next, we tested whether the changes in nasal temperature in response to the interaction playbacks were different from a mere additive effect of the changes in response to the calls separately. To do so, we estimated additive effects for each individual and both test conditions. We used all post temperature measurements from the corresponding control conditions (i.e., ct and gnaeh for the neg-int and fc and gnaeh for the pos-int) and used the function "crossing" (package "tidyr") to generate a dataset containing two columns with all possible combinations of temperature values measured in the two control conditions (Cartesian product). We then summed up each row resulting in the measure for the negative additive effect (neg-add) and the positive additive effect (pos-add). The number of elements of the Cartesian product of two finite sets corresponds to the product of the number of elements in both sets (for example, if condition fc contained 10 measurements and gnaeh contained 5 measurements, then the Cartesian product would contain 50 measurements). Because we wanted to compare the additive measurements to the measurements of the interaction conditions, we randomly selected the same number of data points from the newly calculated additive measurements as the corresponding post subphase of the interac-

tion condition contained. The resulting data (Cartesian product dataset) thus contained all the measurements of the subphase post of the interaction playback condition and the calculated additive effect measurements.

With this dataset, we calculated a linear mixed model (model 2a) and examined the effects of condition (neg-int, pos-int, neg-add, and pos-add), sexstatus (coded as breeders, fh, and mh), and their interaction on the temperature measurements. Random intercepts were set for individual ID, nested within family group. We compared the estimated marginal means between the additive effect and the reaction after the interaction playback for the sexstatus classes separately by setting custom contrasts (positive contrast: comparing the positive additive effect to the reaction to the positive interaction playback; negative contrast: comparing the negative additive effect to the reaction to the negative interaction playback).

To further corroborate the findings of this analysis with an even more conservative approach, we additionally calculated an analogous model to model 2a but summarized all the temperature measurements as means per session (model 2b). Last, we calculated the additive effect for each individual separately with Wilcoxon rank sum tests (function "wilcox.test," package "stats").

Independent measures of arousal and the effect of activity

Because thermography is a novel approach to assess arousal in marmosets, we corroborated our findings with independent measures of arousal. To do so, we examined the link between nasal temperature and the two independently assessed measures of arousal (piloerection and the frequency of high arousal calls). We calculated two LMMs with mean difference in nasal temperature (post–pre) as a dependent variable and set either piloerection (model 3) or frequency of high arousal calls as predictor (model 4). We controlled for the repeated temperature measures within a session by adding session nested in subject and family group as a random intercept. To examine the link between activity and nasal temperature, we conducted five different LMMs (model 5a to 5e) by splitting up the original dataset by condition and only using the measurements taken in subphase post to control for collinearity between the factors condition and activity (2). We thus examined the main effect of activity (during the subphase post) on nasal temperature of the subphase post while controlling for individual nested within family group (random intercept).

Preference for cooperative individual

In phase B, to investigate the preference for a cooperative versus noncooperative interaction partner, we analyzed whether the probability to look into the mirror, and do so earlier, was higher after the positive versus negative interaction playback. We therefore fitted a cox proportional hazards mixed-effects model (model 6) on the latency to look into the mirror (in additional compartment; see Fig. 1, right side) after hearing the interaction playbacks (test conditions: pos-int and neg-int). We assessed the effects of condition and direction of thermal change (as a proxy for the arousal level of phase A). The variable direction of thermal change was used as a factor with three levels, "increase," "decrease," or "none," depending on the results of the Wilcoxon tests for each individual (see the "Individual thermal reactions" section). The Wilcoxon test needed to show an absolute effect size of >0.3 (and a P value of ≤0.05) to be considered an increase/decrease, in all other cases, the variable was set to none. Although we counterbalanced the order of the conditions over all the individuals, we added order, as well as the interaction between order and condition to the model, to assess the influence of the

order in which the individual experienced the conditions. We accounted for the hierarchical structure of the data by including individual nested in family group as random intercepts. To control for the different lengths of phase B, we further included the logarithm of the duration as an offset term into the model. Results are reported as hazard ratios (HR; HR > 1: increased likelihood of looking into the mirror, HR < 1: decreased likelihood of looking into mirror). The Kaplan-Meier survival curve for the latencies to enter the compartment were produced using the packages "survival" and "survminer."

RESULTS

Overall thermal reaction

First, we analyzed the changes in arousal level to the five different stimuli. The full model that included the fixed effects condition, subphase as well as sexstatus, the two-way interactions with subphase and condition, as well as sexstatus and their three-way interaction effect (model 1; Table 1) explained the data significantly better than the null model [likelihood ratio test: $N_{total} = 1711$, $N_{individuals} = 21$, $N_{sessions} = 90$; pseudo-$R^2_c = 0.512$; $\chi^2_{(29)} = 384.619$, $P < 0.0001$]. As predicted, we found that the different stimuli elicited a change in arousal level from the baseline to the stimulation phase, but the three-way interaction between condition, subphase, and sexstatus was significant, indicating that the different classes of animals showed varying reactions to the stimuli. Thus, to investigate this interaction effect further and to compare the thermal reaction from subphase pre to post split up by condition and sexstatus, we used pairwise comparisons of estimated marginal means (see table S5 and Fig. 2).

Most playback stimuli elicited significant arousal changes from the baseline to the stimulation period. Breeders (males and females together; see the left panel in Fig. 2) showed significant changes in nasal temperatures in three of five conditions. The neg-int playback and the fc playback led to a decrease in nasal temperature, indicative of an increase in arousal, and the ct playback led to an increase in nasal temperature, indicating a decrease in arousal. Female helpers (see the middle panel in Fig. 2) showed the strongest decreases in nasal temperature, especially after the playbacks simulating interactions. Infants being alone also led to an increase in arousal but to a lesser extent. Male helpers (see the right panel in Fig. 2), on the other hand, only showed temperature increases after hearing the opposite sex outgroup individual playback, indicating a decrease in arousal even after the simulated negative interaction or the individual emitting chatter calls. The simulated positive interaction did not elicit a significant change in arousal levels, as did simulating an immature being alone. The highest temperature increase in male helpers occurred after hearing an individual being alone emitting food calls. For individual results comparing the baseline to the post phase with Wilcoxon rank sum test, see table S6.

Additive effect

Next, we investigated whether the change in arousal in the interaction playbacks could be simply explained as a mere additive effect of the single stimuli. The full model including condition, sexstatus, and their two-way interaction explained nasal temperature significantly better than the null model [likelihood ratio test: $N_{total} = 434$, $N_{individuals} = 17$, $N_{sessions} = 54$, pseudo-$R^2_c = 0.610$; $\chi^2_{(11)} = 110.782$, $P < 0.0001$]. We found a significant two-way interaction between condition and sexstatus (model 2a; Table 1). Thus, to compare the additive effect to the reaction to the simulated interaction playback (pos-int and neg-int) by sexstatus class, we compared the estimated marginal means with the relevant contrasts (table S5). We found significant differences between the additive effect and the reaction to the interaction playback for all classes of animals and all conditions,

Table 1. Type II analyses of deviance tables for models 1 to 2b. Bold values indicate $P < 0.05$. Only the highest-order (interaction) terms warrant biological interpretation.

| Fixed factors	χ^2	df	P
Model 1: Centered nasal temperature			
Condition (pos-int, neg-int, fc, ct, gnaeh)	9.402	4	0.0518
Subphase (pre, post)	5.344	1	**0.0208**
Sexstatus (breeders, fh, mh)	19.255	2	**<0.0001**
Condition*subphase	87.073	4	**<0.0001**
Condition*sexstatus	7.432	8	0.4908
Subphase*sexstatus	263.889	2	**<0.0001**
Condition*subphase*sexstatus	157.382	8	**<0.0001**
Model 2a: Nasal temperature (Cartesian product dataset)			
Condition (pos-int, pos-add, neg-int, neg-add)	87.550	3	**<0.0001**
Sexstatus (breeders, fh, mh)	11.461	2	**0.003**
Condition*sexstatus	28.468	6	**<0.0001**
Model 2b: Nasal temperature (Cartesian product dataset with means per session)			
Condition (pos-int, pos-add, neg-int, neg-add)	13.1326	3	**0.004**
Sexstatus (breeders, fh, mh)	15.4435	2	**0.0004**
Condition*sexstatus	4.1416	6	0.658
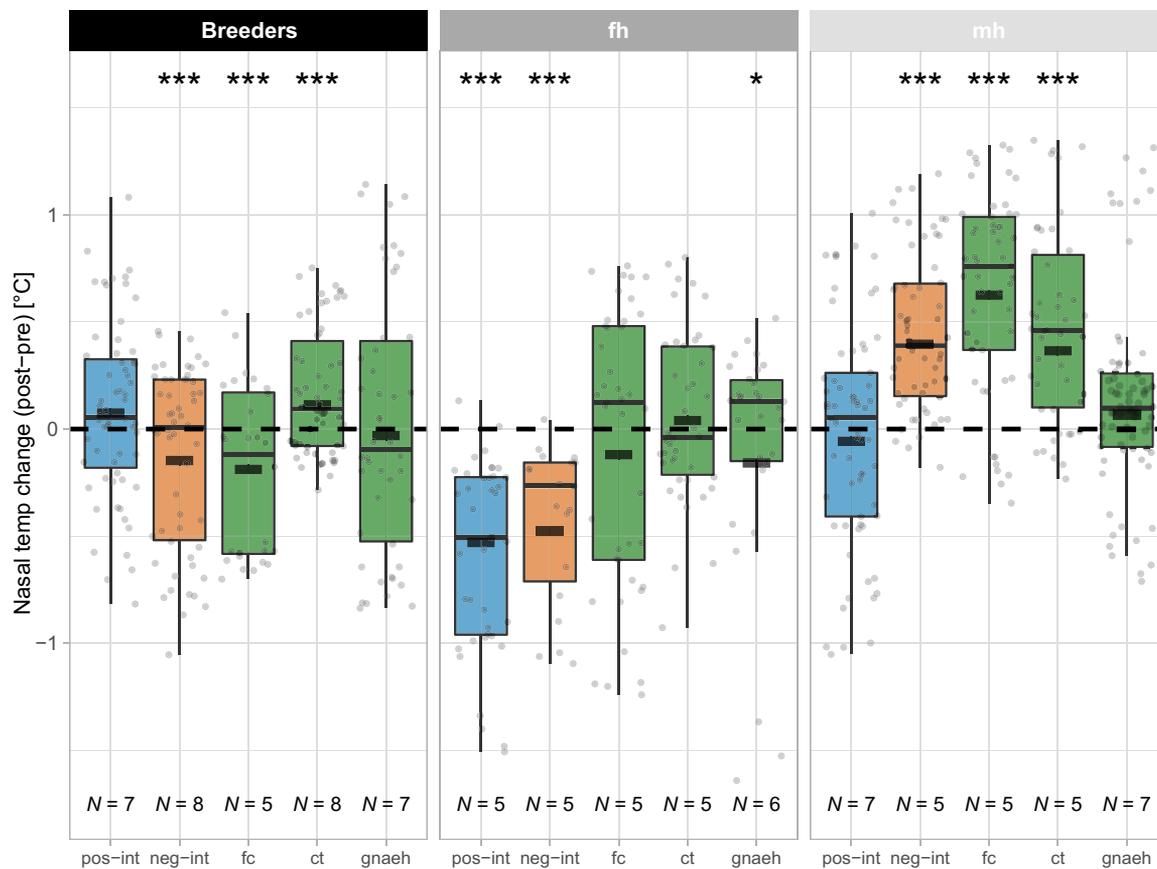
Fig. 2. Overall thermal reaction. Changes in arousal in response to the playbacks. Boxplots showing temperature changes relative to baseline (i.e., session-specific mean minimum nasal temperature) by condition and sexstatus [breeders, female helpers (fh), and male helpers (mh)]. Bold black bars indicate estimated marginal means based on model 1 (see table S5), and gray points represent centered data points. Note that negative values represent a decrease in nasal temperature and thus an increase in arousal. $*P \leq 0.05$, $***P \leq 0.001$.

with the exception of the positive contrast in the breeders (comparing the positive interaction playback to the positive additive effect) (Fig. 3). Even in the more conservative analysis, when summarizing the data with a mean per session and calculating the effects of condition, sexstatus, and their interaction, the full model was still significantly different from the null model [likelihood ratio test: $N_{total} = 54$, $N_{individuals} = 17$, $N_{sessions} = 54$, pseudo-$R^2_c = 0.499$; $\chi^2_{(11)} = 24.214$, $P < 0.012$] and showed both a significant effect of condition and sexstatus, but not their interaction (model 2b; Table 1). Looking more closely at the estimated marginal means to compare the positive and negative contrast, we found that the difference between the negative interaction playback and the negative additive remained significant [EMM (SE) = 0.488 (0.184), 95% CI = [0.051, 0.924], $t = 2.646$, $P = 0.026$]. The difference between the positive interaction playback and the positive additive effect was not as strong and no longer reached significance in this additional analysis [EMM (SE) = 0.320 (0.181), 95% CI = [−0.109, 0.748], $t = 1.766$, $P = 0.176$]. Last, on an individual level, when comparing the negative interaction playback to respective additive effect, 73% of individuals showed a significant difference (with 10 individuals showing $r > 0.5$ and 1 individual with $r > 0.3$). For the positive interaction playback, 64% of individuals showed a thermal reaction that was significantly different from the positive additive effect (with seven individuals showing $r > 0.5$ and two individuals with $r > 0.3$; see table S7).

Independent measures of arousal and the effect of activity

To validate the assumption that a decrease in nasal temperature is indicative of an increase in arousal levels, we examined the relationship between nasal temperature and two independent measures of arousal, namely, piloerection and the frequency of high arousal calls. Both variables significantly predicted nasal temperature and showed a negative relationship with temperature changes [model 3: piloerection β (SE) = −0.305 (0.122), df = 64, $t = −2.507$, $P = 0.014$; model 4: high arousal calls β (SE) = −0.085 (0.030), df = 64, $t = −2.885$, $P = 0.005$]. Thus, higher levels of piloerection and higher frequencies of high arousal calls were associated with a decrease in nasal temperature (fig. S2, A and B). Likelihood ratio tests confirmed that both models were significantly different from the null model [model 3: $N_{total} = 1665$, $N_{individuals} = 21$, $N_{sessions} = 87$, pseudo-$R^2_c = 0.359$; $\chi^2_{(1)} = 5.921$, $P = 0.015$; model 4: $N_{total} = 87$, $N_{individuals} = 21$, $N_{sessions} = 87$, pseudo-$R^2_c = 0.354$; $\chi^2_{(1)} = 7.556$, $P = 0.006$]. We additionally investigated the influence of activity on nasal temperature and replicated the finding from Ermatinger *et al.* (*2*) that activity per se cannot explain changes in nasal temperature. Activity only significantly predicted nasal temperature in the positive-interaction (pos-int) and the food call condition (table S8, model 5a to 5e, and fig. S2C). Higher physical activity correlated with a stronger decrease in nasal temperature. In the three other
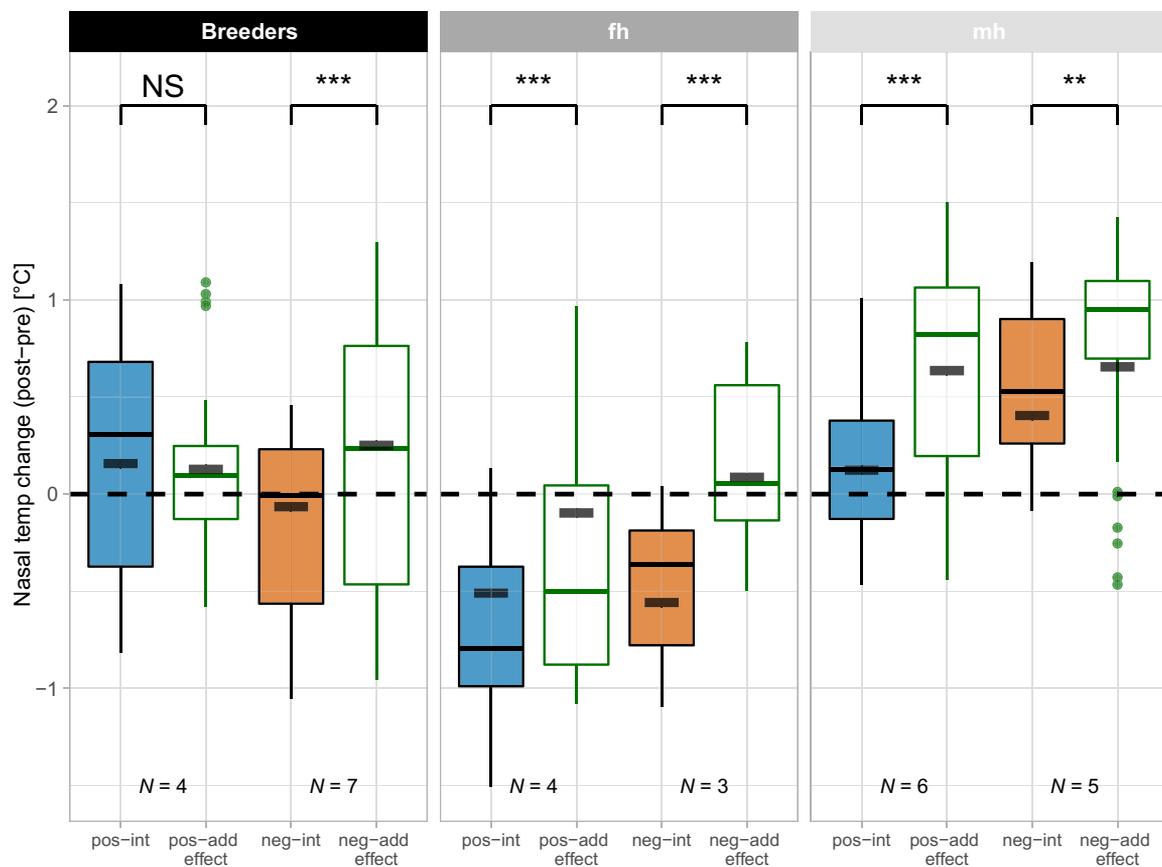
Fig. 3. Additive effect. Boxplots comparing the simulated additive effect (green outlined boxplots) and the measured reaction after the interaction playback (black outlined boxplots) for the positive and negative condition split up by sexstatus classes [breeders, female helpers (fh), and male helpers (mh)]. Bold black bars indicate estimated marginal means based on model 2a (see table S5). **$P \leq 0.01$, ***$P \leq 0.001$; NS, not significant.

conditions, the relationship between physical activity and nasal temperature was not consistent.

Preference for cooperative individuals

We examined the preference for the cooperative versus noncooperative interlocutor by measuring the latency to enter the additional compartment from where the playback was broadcast, and thus the propensity of individuals to approach the previously simulated strangers. As predicted, we found that the likelihood to enter the compartment in phase B and look into the mirror was lower after hearing the negative interaction playback (see Table 2, model 6, and Fig. 4). The full cox proportional hazards mixed-effects model explained a significantly larger proportion of the variation than the null model [likelihood ratio test: $N_{total} = 37$, $N_{individuals} = 20$, $N_{sessions} = 37$; $\chi^2_{(5)} = 16.664$, $P = 0.005$] and revealed no effect of the arousal level during phase A (playback), order of the condition, or the interaction between condition and order.

DISCUSSION

In this study, we used marmoset monkeys to pioneer a promising approach to link physiology and sociality. To investigate social evaluation of vocal interactions between third parties, we used vocal playbacks of conspecific outgroup individuals to eliminate the involvement of human actors, food, and potential side biases and combined

a thermography approach to evaluate how marmosets process these vocal interactions (phase A) with behavioral observations to quantify their preference for cooperative versus noncooperative interlocutors (phase B). As predicted, thermal measurements revealed significant changes in the marmosets' arousal levels after experiencing the playback stimuli. The reaction to the interaction playbacks were not merely the sum of the reactions when they experienced the constitutive parts of the vocal interaction separately (phase A). This suggests that the marmosets perceived and processed the interaction playbacks holistically as "conversations" rather than as the sum of the single elements. Subsequently (phase B), the marmosets preferentially entered the compartment from where the playback was broadcast after they had heard the cooperative interaction playback. This indicates that the marmosets not only processed the vocal interactions holistically but also used this information to evaluate the interactions by showing a preference for a cooperative stranger.

The temperature changes after experiencing the different playback stimuli varied among the breeders and male and female helpers, and not all playbacks elicited significant changes from the baseline, indicating that (classes of) individuals did not react uniformly to our test stimuli. This variation in responses appears consistent with the natural history of these animals. The biggest changes in temperature, suggesting the strongest emotional responses, were shown by female and male helpers, but in markedly opposite directions.

Table 2. Summary table model 6. Bold values indicate $P < 0.05$. Cox proportional hazards model with random intercept on latency to look into mirror.

Model 6: Latency to looking into mirror

Fixed factors	β	SE	HR [95% CI]	z	P
Condition (neg-int)	−2.866	1.058	0.057 [0.007, 0.453]	−2.71	**0.007**
Direction of thermal change					
Decrease and increase versus none	−0.304	0.218	0.738 [0.481, 1.132]	−1.39	0.160
Decrease versus increase	−0.624	0.412	0.536 [0.239, 1.200]	−1.52	0.130
Order (2)	−0.126	0.975	0.882 [0.130, 5.964]	−0.13	0.900
Condition*order [Neg-int*order (2)]	3.295	1.876	26.988 [0.683, 1066.107]	1.76	0.079
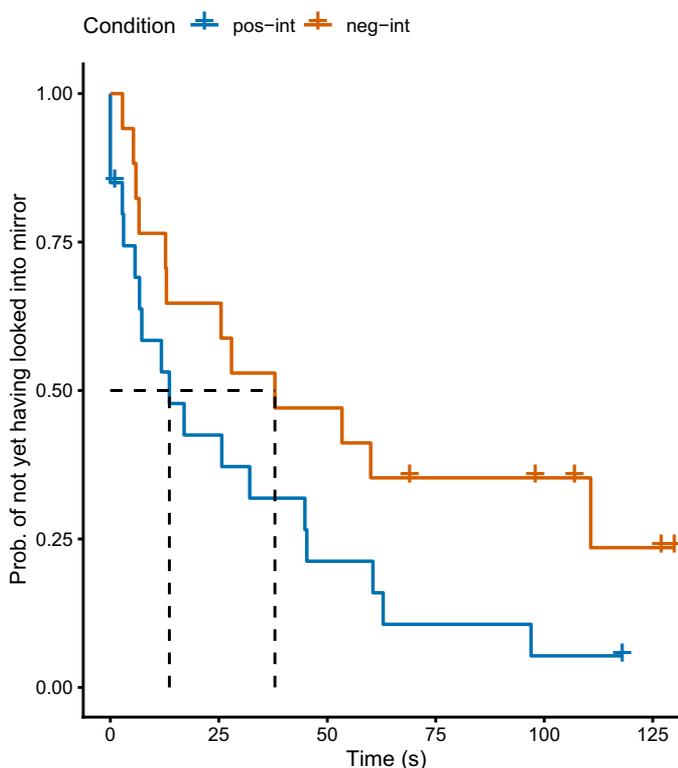
Fig. 4. Preference for cooperative individuals. The probability of not yet having looked into the mirror after the positive (blue) and negative (orange) playback (Kaplan-Meier curves). $N = 37$ with 30 events of looking into the mirror. Dashed lines indicate median survival pointers and show that the median latency to look into the mirror was 13.6 s after the positive playback but 37.9 s after the negative playback.

Female helpers showed strong temperature drops and thus arousal after the simulation of both positive and negative interactions between a male stranger and an immature. They were thus always highly aroused when they perceived an outgroup male with an immature, which is indicative of a neighboring group with young immatures. In nature, when female helpers try to immigrate into a new group, periods with small infants in the new group are arguably most difficult because this is when female-female competition is particularly high (*42, 32, 33*). It may thus well be that the female helpers'

reaction reflected high alertness to the potential presence of highly competitive female breeders. Unfamiliar male strangers without immatures, on the other hand, were apparently considered less threatening and potentially even seen as a mating partner, as evident in the lack of arousal after hearing vocalizations of male strangers alone.

Male helpers showed strong increases in nasal temperature, indicating a decrease of arousal and thus relaxation. In humans, an increase in the nasal, periorbital, and mouth region is associated with sexual arousal, likely due to increased skin perfusion rates to raise the sensitivity of the respective organs (*43*). Male helpers showed an increase in nasal temperature after witnessing the playback of an adult female stranger alone as well as after the simulated aggressive interaction between the female and the immature. All these stimuli that elicited an increase in nasal temperature represent potential mating opportunities: either a stranger female who is alone or interacting aggressively with an immature and thus is likely neither the mother nor well integrated in her own group. The stimuli that did not lead to a change in arousal represented a female who is well integrated in her group and therefore unlikely a potential mate or an immature alone.

Breeders, finally, showed the least pronounced changes in nasal temperature (about 0.1°C change after chatter calls to almost 0.2°C change after food calls). The increase of arousal after food calls might indicate an anticipatory excitement toward a potential food source. This is especially likely for female breeders, who are known to be very food motivated (*44, 2, 45*). The slight decrease in nasal temperature in reaction to the negative interaction between a stranger and the immature appears mostly driven by male breeders. They are often considered primary caretakers in marmoset groups. Their arousal in response to the situation where the begging immature is aggressively denied food could thus well be an expression of their concern for the well-being of immatures in general (*46*). Although the playback simulates outgroup individuals, it is known from reports of captive individuals that adoption of immatures up to a certain age is readily possible (*47*). A larger sample of breeders will be necessary to systematically address sex differences in the breeders.

These changes in arousal consistent with the natural history of the animals are a first indication that the marmosets correctly understood the playbacks. However, a possible alternative is that rather than understanding the interaction playbacks as social interactions, the subjects independently, but simultaneously, reacted to the separate

elements of the interaction. To exclude this possibility, we scrutinized whether the reactions to the interaction playbacks were different from such a mere additive effect. This analysis of the additive effect revealed that changes in arousal elicited by the interaction playbacks differed from the reaction that would theoretically be expected when the interaction would only be perceived as a simple concatenation of the separate calls used in the playback. We found a significant difference between this simulated additive effect in female and male helpers for both comparisons: the reaction to the negative interaction compared to the theoretically expected additive effect and the positive interaction compared to the additive effect. For breeders, this difference was only found for the negative comparison. This may suggest that they did not extract more information from the positive interaction than from the separate elements, but it may also be an artifact of the overall lower responses and that male and female breeders were lumped in one single category due to sample size.

Research on social eavesdropping has mostly been focused on showing behavioral changes in the reactions toward the third-party individuals that have been observed. With a few exceptions (48–50), experimental setups and studies done in the wild were not able to implement a "ghost" control condition, where the observed third-party individuals exhibit the same behavior as during the interaction but they have no interaction partner. This control is important to exclude the possibility that individuals observing this interaction only react to cues that are inadvertently present due to the mere presence of the participants themselves (35, 51). The advantage of our paradigm is that we can not only implement such control conditions but also quantify the subjects' emotional reactions to them.

Two different independent measures could validate that the temperature changes reflect changes in arousal and additionally are not just an artifact of activity (for a more in-depth discussion of this validation, see section S4).

Crucially, marmosets not only process vocal interactions of playbacks holistically, but subsequently, their behavioral reactions show that they also evaluate these social interactions: They show a preference for agents who interact cooperatively with a third party. In phase B of the study, a simple free choice trial, we asked whether the marmosets assigned value to the interactions and thus engaged in social evaluation (4, 17). After the interaction playbacks of phase A, the subjects could choose to either return to their home enclosure or enter the compartment from where the playbacks were broadcast. We hypothesized that as cooperative breeders who critically depend on the cooperativeness of group members (30, 31, 52, 53), they would show social evaluation in cooperative contexts and thus preferentially approach an individual that they had heard interacting cooperatively with a third party. We found that individuals approached the compartment with the speaker earlier if they experienced a playback simulating a cooperative interaction compared to the noncooperative interaction.

Overall, these results are in line with our findings from the pilot study. They show that marmosets can engage in social evaluation of third-party interactions, although they may only act accordingly when necessary. In the pilot study, noncooperative vocal interactions similar to the ones used here led to punishment only in the case where helpers were not well integrated in their own group. In well-established, stable groups composed of highly cooperative individuals, occasional noncooperative behaviors appear to be tolerated and not lead to punishment, whereas cooperative behaviors are expected from others and thus do not elicit attention or even appraisal. Thus, despite being able to perceive, process, and evaluate third-party inter-

actions, this does most of the time not lead to an overt behavioral change. In the current study, the response was stronger, and a preference for a cooperative individual was highly significant at the level of the entire sample. Moreover, they showed this preference toward complete strangers. Most likely, this is because the cost of the behavioral response in the current study, i.e., daring a glimpse behind the wall of the adjacent compartment, was rather low, in particular compared to engaging in punishment behavior as in the within-group context of the pilot study. Together, this suggests that a preference for cooperative individuals is rather general in marmosets, but its behavioral expression is very context specific and sensitive to the costs that are involved with it. Such a general sensitivity may also explain why marmosets socially evaluate even nonconspecifics, namely, humans (22, 23).

Our study adds to the growing evidence that many animals are not only passive observers of third-party interactions and shows how thermography can contribute to unveil how such interactions are perceived by nonverbal subjects. We find that marmosets can engage in social evaluation even in the context of cooperation where direct interactions with individuals are much less costly compared to contexts where social evaluation has traditionally been studied [such as the fighting or mating context; see (35)]. Nevertheless, this ability does not systematically lead to overt behavioral reactions or even punishment of noncooperative group members (see results from pilot study, section S1). Rather, this seems to occur only in unstable social situations. It thus appears that social evaluation can be used flexibly in marmosets, in that they become more vigilant to monitor others' cooperative intentions when necessary but do not do that all the time when all members are well integrated in the social group.

SUPPLEMENTARY MATERIALS
Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/6/eabc8790/DC1

REFERENCES AND NOTES
1. D. L. Cheney, R. M. Seyfarth, *How Monkeys See the World: Inside the Mind of Another Species* (University of Chicago Press, 2018).
2. F. A. Ermatinger, R. K. Brügger, J. M. Burkart, The use of infrared thermography to investigate emotions in common marmosets. *Physiol. Behav.* **211**, 112672 (2019).
3. P. K. McGregor, T. M. Peake, Communication networks: Social environments for receiving and signalling behaviour. *Acta Ethologica* **2**, 71–81 (2000).
4. J. Abdai, Á. Miklósi, The origin of social evaluation, social eavesdropping, reputation formation, image scoring or what you will. *Front. Psychol.* **7**, 1772 (2016).
5. T. M. Peake, P. K. McGregor, Information and aggression in fishes. *Anim. Learn. Behav.* **32**, 114–121 (2004).
6. M. Amy, G. Leboucher, Male canaries can visually eavesdrop on conspecific food interactions. *Anim. Behav.* **74**, 57–62 (2007).
7. W.-S. Lai, W.-C. Yu, C.-Y. Liu, M.-T. Kuo, C.-H. Huang, A new method for studying social eavesdropping using male golden hamsters. *Physiol. Behav.* **128**, 202–211 (2014).
8. C. Wedekind, M. Manfred, Cooperation through image scoring in humans. *Science* **288**, 850–852 (2000).
9. E. Fehr, U. Fischbacher, The nature of human altruism. *Nature* **425**, 785–791 (2003).
10. G. Grimalda, A. Pondorfer, D. P. Tracer, Social image concerns promote cooperation more than altruistic punishment. *Nat. Commun.* **7**, 12288 (2016).
11. J. K. Hamlin, K. Wynn, P. Bloom, Three-month-olds show a negativity bias in their social. *Dev. Sci.* **13**, 923–929 (2010).
12. C. Holvoet, C. Scola, T. Arciszewski, D. Picard, Infants' preference for prosocial behaviors: A literature review. *Infant Behav. Dev.* **45**, 125–139 (2016).
13. F. Margoni, L. Surian, Infants' evaluation of prosocial and antisocial agents: A meta-analysis. *Dev. Psychol.* **54**, 1445–1455 (2018).
14. V. Kuhlmeier, K. Wynn, P. Bloom, Attribution of dispositional states by 12-month-olds. *Psychol. Sci.* **14**, 402–408 (2003).

15. A. Geraci, L. Surian, The developmental roots of fairness: Infants' reactions to equal and unequal distributions of resources. *Dev. Sci.* **14**, 1012–1020 (2011).

16. M. Buon, P. Jacob, S. Margules, I. Brunet, M. Dutat, D. Cabrol, E. Dupoux, Friend or foe? Early social evaluation of human interactions. *PLOS ONE* **9**, e88612 (2014).

17. J. R. Anderson, B. Bucher, H. Chijiiwa, H. Kuroshima, A. Takimoto, K. Fujita, Third-party social evaluations of humans by monkeys and dogs. *Neurosci. Biobehav. Rev.* **82**, 95–109 (2017).

18. C. Krupenye, B. Hare, Bonobos prefer individuals that hinder others over those that help. *Curr. Biol.* **28**, 280–286.e5 (2018).

19. E. Herrmann, S. Keupp, B. Hare, A. Vaish, M. Tomasello, Direct and indirect reputation formation in nonhuman great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus*) and human children (*Homo sapiens*). *J. Comp. Psychol.* **127**, 63–75 (2013).

20. J. R. Anderson, H. Kuroshima, A. Takimoto, K. Fujita, Third-party social evaluation of humans by monkeys. *Nat. Commun.* **4**, 1561 (2013).

21. J. R. Anderson, B. Bucher, H. Kuroshima, K. Fujita, Evaluation of third-party reciprocity by squirrel monkeys (*Saimiri sciureus*) and the question of mechanisms. *Anim. Cogn.* **19**, 813–818 (2016).

22. N. Kawai, M. Yasue, T. Banno, N. Ichinohe, Marmoset monkeys evaluate third-party reciprocity. *Biol. Lett.* **10**, 20140058 (2014).

23. N. Kawai, A. Nakagami, M. Yasue, H. Koda, N. Ichinohe, Common marmosets (Callithrix jacchus) evaluate third-party social interactions of human actors but Japanese monkeys (Macaca fuscata) do not. *J. Comp. Psychol.* **133**, 488–495 (2019).

24. S. M. A. Kundey, A. De Los Reyes, E. Royer, S. Molina, B. Monnier, R. German, A. Coshun, Reputation-like inference in domestic dogs (Canis familiaris). *Anim. Cogn.* **14**, 291–302 (2011).

25. M. Nitzschner, J. Kaminski, A. Melis, M. Tomasello, Side matters: Potential mechanisms underlying dogs' performance in a social eavesdropping paradigm. *Anim. Behav.* **90**, 263–271 (2014).

26. C. M. Johnson, J. Sullivan, J. Jensen, C. Buck, J. Trexel, J. S. Leger, Prosocial predictions by bottlenose dolphins (*Tursiops* spp.) based on motion patterns in visual stimuli. *Psychol. Sci.* **29**, 1405–1413 (2018).

27. M. Trösch, S. Pellon, F. Cuzol, C. Parias, R. Nowak, L. Calandreau, L. Lansade, Horses feel emotions when they watch positive and negative horse–human interactions in a video and transpose what they saw to real life. *Anim. Cogn.* **23**, 643–653 (2020).

28. R. Bshary, A. S. Grutter, Image scoring and cooperation in a cleaner fish mutualism. *Nature* **441**, 975–978 (2006).

29. S. F. Brosnan, F. B. M. de Waal, Cebus apella tolerate intermittent unreliability in human experimenters. *Int. J. Primatol.* **30**, 663–674 (2009).

30. W. M. Erb, L. M. Porter, Mother's little helpers: What we know (and don't know) about cooperative infant care in callitrichines. *Evol. Anthropol.* **26**, 25–37 (2017).

31. E. M. Guerreiro Martins, A. C. de A Moura, C. Finkenwirth, M. Griesser, J. M. Burkart, Food sharing patterns in three species of callitrichid monkeys (Callithrix jacchus, Leontopithecus chrysomelas, Saguinus midas): Individual and species differences. *J. Comp. Psychol.* **133**, 474–487 (2019).

32. L. J. Digby, S. F. Ferrari, W. Saltzman, Callitrichines: The role of competition in cooperatively breeding species, in *Primates in Perspective*, C. Campbell, A. Fuentes, K. MacKinnon, S. Bearder, R. Stumpf, Eds. (Oxford Univ. Press, 2007), pp. 85–106.

33. M. E. Yamamoto, A. Araujo, M. de Fatima Arruda, A. K. M. Lima, J. de Oliveira Siqueira, W. T. Hattori, Male and female breeding strategies in a cooperative primate. *Behav. Processes* **109**, 27–33 (2014).

34. C. Rudolf von Rohr, J. M. Burkart, C. P. van Schaik, Evolutionary precursors of social norms in chimpanzees: A new approach. *Biol. Philos.* **26**, 1–30 (2011).

35. K. E. Bonnie, R. L. Earley, Expanding the scope for social information use. *Anim. Behav.* **74**, 171–181 (2007).

36. E. Salazar-López, E. Domínguez, V. Juárez Ramos, J. de la Fuente, A. Meins, O. Iborra, G. Gálvez, M. A. Rodríguez-Artacho, E. Gómez-Milán, The mental and subjective skin: Emotion, empathy, feelings and thermography. *Conscious. Cogn.* **34**, 149–162 (2015).

37. K. Nakayama, S. Goto, K. Kuraoka, K. Nakamura, Decrease in nasal temperature of rhesus monkeys (Macaca mulatta) in negative emotional state. *Physiol. Behav.* **84**, 783–790 (2005).

38. F. Kano, S. Hirata, T. Deschner, V. Behringer, J. Call, Nasal temperature drop in response to a playback of conspecific fights in chimpanzees: A thermo-imaging study. *Physiol. Behav.* **155**, 83–94 (2016).

39. S. Ioannou, V. Gallese, A. Merla, Thermal infrared imaging in psychophysiology: Potentialities and limits: Thermal infrared imaging in psychophysiology. *Psychophysiology* **51**, 951–963 (2014).

40. A. Heschl, J. Burkart, A new mark test for mirror self-recognition in non-human primates. *Primates* **47**, 187–198 (2006).

41. B. M. Bezerra, A. Souto, Structure and usage of the vocal repertoire of Callithrix jacchus. *Int. J. Primatol.* **29**, 671–701 (2008).

42. M. F. Arruda, A. Araújo, M. B. C. Sousa, F. S. Albuquerque, A. C. S. R. Albuquerque, M. E. Yamamoto, Two breeding females within free-living groups may not always indicate polygyny: Alternative subordinate female strategies in common marmosets (Callithrix jacchus). *Folia Primatol.* **76**, 10–20 (2005).

43. A. Merla, G. L. Romani, Thermal signatures of emotional arousal: A functional infrared imaging study, in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, 2007), pp. 247–249.

44. A. M. Michels, Sex differences in food acquisition and aggression in captive common marmosets (Callithrix jacchus). *Primates* **39**, 549–556 (1998).

45. M. F. De la Fuente, N. Schiel, J. C. Bicca-Marques, C. B. Caselli, A. Souto, P. A. Garber, Balancing contest competition, scramble competition, and social tolerance at feeding sites in wild common marmosets (Callithrix jacchus). *Am. J. Primatol.* **81**, e22964 (2019).

46. A. Koenig, Group size, composition, and reproductive success in wild common marmosets (Callithrix jacchus). *Am. J. Primatol.* **35**, 311–317 (1995).

47. D. Maestripieri, Is there mother–infant bonding in primates? *Dev. Rev.* **21**, 93–120 (2001).

48. T. M. Peake, A. M. R. Terry, P. K. McGregor, T. Dabelsteen, Male great tits eavesdrop on simulated male-to-male vocal interactions. *Proc. Biol. Sci. B* **268**, 1183–1187 (2001).

49. T. M. Peake, R. J. Matos, P. K. McGregor, Effects of manipulated aggressive "interactions" on bystanding male fighting fish, *Betta splendens*. *Anim. Behav.* **72**, 1013–1020 (2006).

50. P. K. McGregor, T. M. Peake, H. M. Lampe, Fighting fish Betta splendens extract relative information from apparent interactions: What happens when what you see is not what you get. *Anim. Behav.* **62**, 1059–1065 (2001).

51. S. Marshall-Pescini, C. Passalacqua, A. Ferrario, P. Valsecchi, E. Prato-Previde, Social eavesdropping in the domestic dog. *Anim. Behav.* **81**, 1177–1183 (2011).

52. A. D. Gosselin-Ildari, A. Koenig, The effects of group size and reproductive status on vigilance in captive *Callithrix jacchus*. *Am. J. Primatol.* **74**, 613–621 (2012).

53. M. E. Yamamoto, H. O. Box, The role of non-reproductive helpers in infant care in captive Callithrix jacchus. *Ethology* **103**, 760–771 (1997).

54. R. K. Brügger, Master thesis, University of Zurich (2016).

55. T. Kappeler-Schmalzriedt, thesis, University of Zurich (2017).

56. A. Dettling, C. R. Pryce, R. D. Martin, M. Döbeli, Physiological responses to parental separation and a strange situation are related to parental care received in juvenile Goeldi's monkeys (Callimico goeldii). *Dev. Psychobiol.* **33**, 21–31 (1998).

57. A. C. Dettling, J. Feldon, C. R. Pryce, Repeated parental deprivation in the infant common marmoset (Callithrix jacchus, primates) and analysis of its effects on early development. *Biol. Psychiatry* **52**, 1037–1046 (2002).

58. M. N. Schubiger, F. L. Wüstholz, A. Wunder, J. M. Burkart, High emotional reactivity toward an experimenter affects participation, but not performance, in cognitive tests with common marmosets (*Callithrix jacchus*). *Anim. Cogn.* **18**, 701–712 (2015).

59. Y. Kato, H. Gokan, A. Oh-Nishi, T. Suhara, S. Watanabe, T. Minamimoto, Vocalizations associated with anxiety and fear in the common marmoset (Callithrix jacchus). *Behav. Brain Res.* **275**, 43–52 (2014).

60. I. Pavlidis, J. Levine, P. Baukol, Thermal image analysis for anxiety detection, in *Proceedings 2001 International Conference on Image Processing* (IEEE, 2001), vol. 2, pp. 315–318.

61. S. Ioannou, P. Morris, S. Terry, M. Baker, V. Gallese, V. Reddy, Sympathy crying: Insights from infrared thermal imaging on a female sample. *PLOS ONE* **11**, e0162749 (2016).

Science Advances

Do marmosets understand others' conversations? A thermography approach

R. K. Brügger, E. P. Willems and J. M. Burkart

ARTICLE TOOLS	http://advances.sciencemag.org/content/7/6/eabc8790
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2021/02/01/7.6.eabc8790.DC1
REFERENCES	This article cites 55 articles, 1 of which you can access for free http://advances.sciencemag.org/content/7/6/eabc8790#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service