Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/7/6/eabc8790/DC1

Supplementary Materials for

Do marmosets understand others' conversations? A thermography approach

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

*Corresponding author. Email: rahelkatharina.bruegger@uzh.ch

Published 3 February 2021, *Sci. Adv.* **7**, eabc8790 (2021)
DOI: 10.1126/sciadv.abc8790

This PDF file includes:

Section S1. Pilot study: methods and results

Methods

Subjects and housing

15 captive born adult common marmosets (*Callithrix jacchus*) were involved in this pilot study, we tested seven individuals as focals (see Table S1. for detailed information about the subjects and the group compositions). Feeding and housing conditions were the same as in the study presented in the main text (see main text, Methods, Subjects and Housing).

Table S1. Individuals participating in pilot study. *Includes with details about the group composition at the time of testing as well as the order of testing for the different conditions. B = breeder, h = helper, m = male, f = female, i = infant; C(1–3) = control (1–3), Pos = Positive Playback, Neg = Negative Playback*

Group	Group composition (at time of testing)	Individual	Sex	Status	Birth date	Data collection period	Order of conditions
Mina	2 b, 1 fh, 2 mh, 2 i	Mibba	f	helper	17 June 2009	31 Aug.–24 Sept. 2015	C1, Pos, C2, Neg, C3*
		Mars	m	helper	11 July 2012		C1, Neg, C2, Pos, C3
		Merkur	m	helper	11 July 2012		C1, Pos, C2, NA**, NA
Jojoba	2b, 2 fh, 2 mh, 2 i	Jet	m	helper	3 Apr. 2009	25 Feb.–25 Mar. 2016	C1, Pos, C2, Neg, C3
		Jupie	f	helper	29 Aug. 2009		C1, Neg, C2, Pos, C3
		Joyce	f	helper	21 Apr. 2013		C1, Neg, C2, Pos, C3
		James	m	helper	21 Apr. 2013		C1, Pos, C2, Neg, C3

* For C3 Mibba was only tested in ***pre phase*** (aggression escalated, and session needed to be aborted)

** Could not compile negative playback due to a lack of negative calls from this individual.

Procedure and playback stimuli

To test whether marmosets engage in social evaluation, we removed a helper together with the group's offspring from the rest of the family group in an adjacent enclosure with auditory but not visual contact. We then used two types of playbacks to mimic to the rest of the group a socio-positive or socio-negative interaction between the helper and the offspring. After this separation the groups were reunited and we coded how the group members behaved towards the focal helper, specifically if they showed more socio-positive behaviors after we simulated a positive interaction vs. more socio-negative behaviors after simulating the negative interaction.

All calls used for the playback stimuli were recorded prior to the start of the experiment with Avisoft UltraSoundGate 116H (Sampling rate 250000 Hz; FFT size 256) and a Condenser Microphone CM16/CMPA. Recorded calls were visually inspected using Avisoft-SASLab Pro 5.1.17, saved as separate files and then combined to the final playback files using iMovie 10.1.2. Each playback file lasted 11 minutes and contained 20 call sequences with a maximum duration of 15 seconds interspersed with pauses of an average duration of 30 seconds. The playback started with a distinct tone (to help synchronize the videos recorded from different camera angles) followed by one minute of silence, to allow the experimenter to position herself and for the monkeys to settle down.

Positive playbacks contained focal helper food calls and infant begging calls (positive condition), thus mimicking a begging infant and the helper responding to the begging with food sharing (*41*). Negative playbacks were composed of focal helper chatter calls and infant begging calls (negative condition) (*41*), mimicking a helper chasing away a begging infant. To ensure that the simulation of the food sharing or food conflict interaction would sound as naturalistic as possible, we used different sequences of the respective call types and adjusted the volume to a common level.

Each focal individual (except Merkur, for details see Table S1.) participated in 3 control and 2 test sessions alternating on 5 consecutive days. The procedure for each session was:

- 15-minute *baseline-phase*: family group moves freely in their home cage (2.4 m × 1.5 m × 0.8 m)
- 11-minute *playback-phase*: separation of focal helper together with immatures from rest of the family group (only visually) and presentation of playback stimulus (on test days) or just silence (control days).
- 15-minute *post-phase*: family group is reunited and moves freely in their home cage

Individuals always experienced control sessions on test days 1, 3 and 5 and the positive or negative test session counterbalanced aver all individuals on day 2 or 4. We used control sessions to assess what the baseline reaction towards focal helper is after a short-term separation without any playback. For the analysis we lumped all control conditions. When considering the reactions towards the focal helper Mibba, we split up the control conditions to show the carry over effect of the negative conditions that happened on the day before.

Behavioral coding and data analysis

All sessions were video recorded, yielding around 17 hours of video footage. Social behaviors directed at or including the focal helper were coded with INTERACT (Mangold GmbH, V 15.0) and are listed below:

Socio-positive behaviors

- *Rest huddle:* resting in direct body contact with focal helper, in % of total time
- *Receive grooming:* getting picked at the fur or skin with hands or mouth, in % of total time

Socio-negative behaviors

- *Receive attack:* getting hit, lunged at or jumped on aggressively, in frequency/10 min
- *Receive chatter:* having aggressive chatter calls directed at, in frequency/10 min
- *Submission:* facial or whole-body submission, flattened ear tufts, in frequency/10 min
- *Avoid/replace:* leaving current place as soon as conspecific approaches, in frequency/10 min

The statistical analysis was performed in R (version 3.5.3).
As a first step we applied an unrotated Principal Component Analysis (PCA) on the six z-transformed behavioral variables capturing the socio-positive and socio-negative behaviors directed towards and including the focal. We then performed a parallel analysis to decide which factors we would retain. To do so we generated 10000 random data sets with equal sample size and dimensions as our empirical data and extracted only components with Eigenvalues greater than the 95 % quantile value obtained from the random sample. We then performed a rotated PCA and retained the two factors as revealed in the parallel analysis, for further analysis. The first factor corresponded to socionegative behaviors (***RPC1 – socio-negativie factor***: *variance explained:* 52 %, *Eigenvalue:* 3.13, *factor loadings*: rest huddle: –0.04, receive grooming: –0.04, receive attack: 0.95, receive chatter: 0.93, submission: 0.93, avoid/replace: 0.70;) the second to sociopositive behaviors (***RPC2 – socio-positive factor***: *variance explained:* 23 %, *Eigenvalue:* 1.40, *factor loadings*: rest huddle: 0.83, receive grooming: 0.82, receive attack: 0.03, receive chatter: –0.02, submission:–0.01, avoid/replace: –0.21); The Kaiser-Meyer-Oklin test verified that sampling adequacy was “good” (overall KMO = 0.73 and no individual KMO value below 0.5). One outlier was eliminated with the help of the Grubbs Test ($G = 3.71250$, $U = 0.54105$, *p*-value = 0.0003; highest value 5.014 (value of Merkur control 3 was considered an outlier).

To obtain the response variable, namely the behavioral reaction towards the focal animal, we first calculated the difference between the two retained rotated PC-Factors per session and per individual; a positive difference thus reflects receiving more socio-positive behaviors, a negative difference receiving more socio-negative behaviors. To quantify the behavioral reactions of group members after the playback in comparison to the baseline phase, we subtracted in each test session the values of the baseline-phase from the values of the post-phase, this resulting value is called *"behavioral reaction score"*.
Lastly, we used a linear mixed model to investigate how the behavioral reaction score varied in response to the different test conditions. We calculated a linear mixed-effects model (function "lme" from package "nlme") and compared the full model to the null model with a likelihood-ratio test (function "anova"). We assessed the effects of the condition and sex of focal helper and their interaction on the behavioral reaction towards the focal helper. To control for the effect of the group we included the group membership as a fixed effect (only 2 groups, thus not suitable as random effect) and to control for repeated measures of individuals we included individual (the respective focal helper) as a random intercept. Model assumptions were tested by inspecting the residual plot and cook's distances. We assessed multicollinearity with the fixed effect correlation matrix of the model (values < 0.7).

Results

The full model including condition, focal sex and their interaction as well as family group was not significantly different from the null model including only individual as a random intercept (likelihood ratio test: N = 31, $\chi^2_{(6)} = 4.607$, $p = 0.593$, Table S2.), indicating that group members did not behave differently when helpers were reunited after a positive vs. a negative playback. Indeed, when calculating estimated marginal means for the behavioral reaction scores of the control condition (EMM (SE) = 0.249 (0.274)) the negative condition (EMM (SE) = 0.420 (0.384)) and the positive condition (EMM (SE) = 0.550 (0.360)) were very similar, even though (as expected) the score was highest after the positive condition; this effect did not differ from the reactions after the control or the negative conditions. Fig. 1A shows that there is already quite large variation amongst the reactions towards the focal helper even after the control conditions, but in general individuals received a bit more socio-positive behaviors after the separation, even if there was no playback stimulus. Further, it is apparent that the distribution of behavioral reactions is widest after the positive playback and the highest positive values of reaction scores occur after positive playback. In contrast, after the negative playback the reactions seemed to be less biased against the positive end.
A striking outlier from this general pattern was the female helper "Mibba" who experienced strong negative behavioral reactions after the negative playback. These aggressive behaviors were mainly exhibited by males of the group (*54*). Furthermore, "Mibba" was the only helper where we had to abort a session during the baseline-phase following the negative playback because the aggression towards her escalated and she had to be separated from the rest of the group. This happened in the control condition 3 which "Mibba" experienced the day after the negative condition. Fig. 1B shows an overview of the behavioral reaction scores for "Mibba". Luckily "Mibba" could be reintroduced to her family group after some days. Because we only had data for the baseline-phase of the control session 3 for Mibba, this data could not be included in the main analysis (full model, Table S2.).

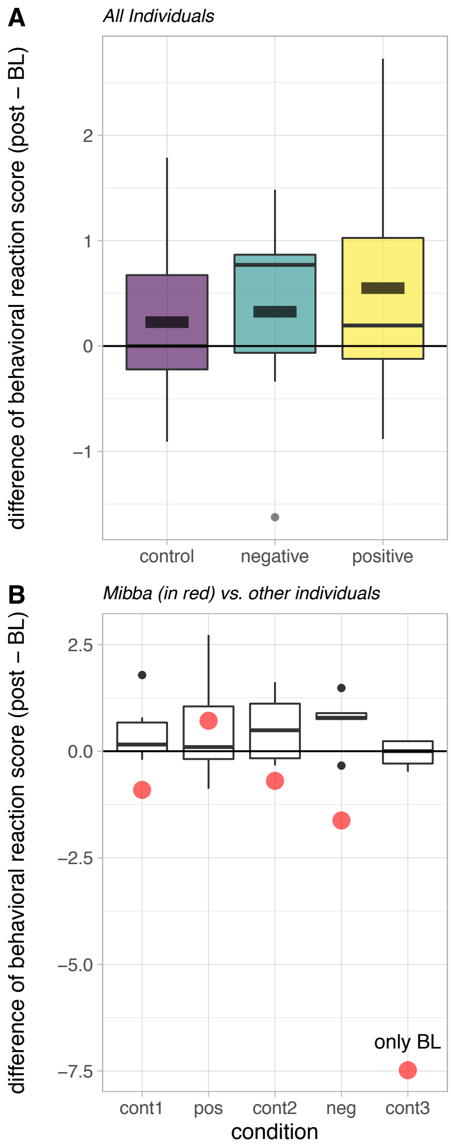

Table S2. LMM on behavioral reaction score. *Full model on the behavioral reaction score.*

behavioral reaction score				
Fixed factors	β	SE	*t*-value	*p*-value
intercept	−0.088	0.394	−0.223	0.826
condition				
control vs. positive & negative	−0.053	0.149	−0.353	0.728
positive vs. negative	−0.014	0.337	−0.041	0.968
sex (male)	0.673	0.493	1.363	0.245
family group (Mina)	0.297	0.478	0.620	0.569
condition*sex interaction				
control vs. positive & negative*sex (male)	0.237	0.203	1.169	0.256
positive vs. negative*sex (male)	0.101	0.465	0.216	0.831

Fig. S1. Behavioral reaction score by condition. *(A) Boxplot (median, interquartile range) showing the difference in behavioral reaction scores (controlled for baseline-phase) by condition for all individuals. Values above zero indicate relatively more socio-positive behaviors, values below zero indicate relatively more socio-negative behaviors directed towards the focal individual. Bold black lines indicate the estimated marginal means by condition. (B) Behavioral reaction scores for female helper "Mibba" visualized as red points (note that for control session 3 the point shows the behavioral reaction score of the baseline-phase only (session had to be aborted). Behavioral reaction scores of all other individuals are visualized as outlined boxplots (median, interquartile range).*

In a follow up experiment of this pilot study, we tested three different groups of common marmosets (N = 8) with the same result: no systematic increase in socio-negative behaviors nor socio-positive behaviors towards the separated helpers, but again in a specific case (female helper Jandira) the individuals reacted very negatively after the simulation of the socio-negative interaction (*55*).

Section S2. Main study: procedure & playback stimuli

Training Procedure. Prior to starting the experiment, all animals were trained to enter the experimental room with the experimenter voluntarily (first with the whole family group and then in a second step alone) and sit on the perch provided for them in the compartment on the left. They were also occasionally allowed access to the compartment on the right but without having the mirror inside. This should ensure that animals knew of this additional compartment and were familiar with the layout of it. Only animals showing no visible signs of arousal (e.g. escape attempts, running in compartment) while being separated in the experimental compartment in the presence of the experimenter were included in the experiment.

Phase A. Thermal camera settings. The uncooled camera was placed in the experimental room at least 15 minutes before the experiment started and the animals were filmed with the camera mounted on a tripod from a one meter distance. The camera was connected to a laptop to directly save the data to an external hard drive. We used the program FLIR Tools+ to control the camera and record the videos. Emissivity was constantly set to 0.98 and the distance to the focal plane to 1 meter. We recorded room temperature and humidity periodically during each session and adjusted the camera settings accordingly. We additionally video recorded both phases of the experiment to be able to code behavioral data.

Section S3. Main study: behavioral coding

Phase A – playback

Independent measures of arousal. We assessed piloerection of the tail, which is an established measure for high levels of arousal (*56–58*) by coding the level of piloerection every 10 seconds for the entire duration of phase A and assigning the following levels: 0 when the tail showed no piloerection; 0.5 when either upper or lower part of the tail showed piloerection; 1 when the whole tail showed strong piloerection. If at the respective timepoint the animal's piloerection of the tail was not fully or clearly visible, we extended the time interval for a maximum five seconds and if it was still not visible, we coded it as not available (N.A.).
We then calculated a mean piloerection score (with a range from 0–1) for each period (baseline, see main text, Fig. 2, –60–0 s and stimulation, main text, Fig. 2, 0–max 150 s) of phase A and subtracted the mean score for the baseline period from the stimulation period which resulted in a mean score for the change in extent of piloerection from the baseline to stimulation period.
Additionally, we coded all occurrences of the following calls: tsik, tsik-egg and egg calls during the stimulation period of phase A (see main text, Fig. 2, time 0–max. 150 s). We counted each unambiguously recognizable call from the audio track of the video camera and divided it by the total duration of the stimulation period (in minutes) resulting in a call frequency per minute. Tsik, tsik-egg and egg calls are calls known to be uttered in high arousal situations (*41, 59*).

Activity. The activity level of animals was coded as number of position changes occurring during the stimulation period of phase A of the experiment. A change in position was defined by the animal moving all four limbs at least one body length from their current position. Movements on the perch were not included in this count as this perch was rather short (33 cm) and therefore the physical activity highly restricted. If animals jumped/moved from one position to the next via a third position that they only occupied for less than 10 frames (0.4 s), it was only recorded as a single change in position. We divided the number of position changes by the length of the stimulation period (in minutes, 0–max 150 s, see main text, Fig. 2) which resulted in a frequency of position changes per minute for the stimulation period.

Phase B – mirror experience

Latency to looking into the mirror. To quantify marmoset's willingness to approach the mirror after experiencing the positive or negative interaction playback stimulus (during stimulation period of phase A) we coded the latency to looking into the mirror. Phase B started when both trap doors were completely open. Looking into mirror was only possible if marmosets decided to enter the additional compartment on the right via the sliding door and by passing by the visual barrier placed in the first third of the additional compartment (main text, Fig. 2, right side). Because duration of Phase B was not identical for all focal individuals, we recorded the total duration of this phase to be included as an offset in the model in the model.

Section S4. Main Study: validation of thermography data

First, we replicated the findings from Ermatinger et al. 2019 (*2*), by showing that in this independent dataset a decrease in nasal temperature was correlated with an independent measure of arousal, i.e. piloerection of the tail. Second, we used the frequency of known high arousal calls such as tsik-, tsik-egg- as well as egg-calls (*41, 59*). We again found a negative correlation for this independent measure of arousal, further corroborating the link between nasal temperature changes and emotional arousal. Lastly, we replicated the finding on the possible confounding effect of activity and could show that similar to a study with chimpanzees (*38*), decreases in nasal temperature can occur without physical activity and furthermore that the relationship between physical activity and nasal temperature was not the same across conditions, indicating that nasal temperature is not solely caused by physical activity. Another and related possible confounder when using nasal temperature is heavy breathing. An increase in breathing maybe even resulting from higher physical activity due to an increase in arousal could cause a cooling effect of the nose (*60*). But evidence of this effect is inconclusive, as a study by Ioannou et al. (2016) (*61*) showed that nasal temperature actually increases in situations with increased deep breathing patterns, indicating that the breathing might not be able to mask all changes in cutaneous blood flow (see also *39*). Regardless, compared to most studies using thermography in animals, the stimuli used in the present study are expected to elicit only mild changes in arousal and thus effects of heavy breathing and activity should be negligible.

Table S3. Individuals participating in main study. *List of all individuals participating in the main study with details about the group composition at the time of testing and in which conditions individuals were participating in the main study by order of testing. Only sessions used in the final dataset are shown. Unnamed conditions indicate that individuals have not participated in these conditions or no data from this condition was usable. B = breeder, h = helper, m = male, f = female, pos-int = positive interaction (food calls and gnaehs), neg-int = neg interaction (chatter calls and gnaehs), fc = food calls only, ct = chatter calls only, gnaeh = gnaehs only*

Data collection period		25 Feb.–22. Mar. 2018				
Group	Group composition (at time of testing)	Individual	Sex	Status	Birth date	Conditions (in order of testing)
Lancia	2 b, 1 fh, 1 mh	Lancia	f	breeder	26 Aug. 2002	ct, pos-int, fc; gnaeh, neg-int
		Lexus	m	breeder	19 July 2003	ct, gnaeh, neg-int, pos-int
Lea	2 b, 1 fh	Lea	f	breeder	18 Aug. 2007	gnaeh, ct, pos-int, fc
		Kyros	m	breeder	03 Jan. 2008	gnaeh, ct, pos-int, neg-int
Mina	2 b, 2 fh, 1 mh	Mina	f	breeder	10 July 2006	neg-int, ct, gnaeh, pos-int, fc
		Merkur	m	helper	11 July 2012	gnaeh, neg-int, fc, ct, pos-int
		Mojita	f	helper	04 May 2015	fc, pos-int, ct, gnaeh, neg-int
		Manuka	f	helper	04 May 2015	pos-int
Nikita	2 b, 1 fh, 3 mh	Ninja	m	helper	12 Apr. 2015	ct, neg-int, fc, pos-int, gnaeh
		Nunchaku	m	helper	12 Apr. 2015	pos-int, gnaeh, fc, ct, neg-int
		Nori	m	helper	28 June 2016	neg-int, gnaeh, pos-int
Nina	2 b, 4 fh, 1 mh	Nina	f	breeder	13 July 2006	gnaeh, fc, ct, pos-int, neg-int
		Lex	m	breeder	12 July 2006	neg-int, gnaeh, ct, pos-int
		Nebula	f	helper	19 Dec. 2012	neg-int, gnaeh, fc, pos-int
		Nougat	f	helper	15 Apr. 2015	fc, ct, gnaeh
		Nutella	f	helper	15 Apr. 2015	gnaeh, neg-it, pos-int, ct
		Nux	f	helper	4 July 2016	ct, fc, gnaeh, pos-int
		Nox	m	helper	4 July 2016	ct, gnaeh, pos-int, fc
Tabor	2 mh/2 b (was paired during testing period)	Tabor	m	helper/ breeder	30 Oct. 2008	pos-int, gnaeh, ct, fc
Vesta	1 fh, 1 mh	Vesta	f	helper	5 Oct. 2004	ct, fc, gnaeh, neg-int, pos-int
		Vito	m	helper	30 May 2006	fc, pos-int, gnaeh, neg-int, ct

Table S4. Durations of phase B. *Duration of phase B for each individual and condition. These values have been added to model 6 as an offset term.*

Individual	**Total duration of phase B [s]**	
	POS-INT	***NEG-INT***
Lancia	93	80
Lexus	93	90
Lea	1 (leaves experimental room)	82
Kyros	69	68
Mina	91	69
Merkur	80	130
Mojita	130	78
Manuka	132	NA
Ninja	134	81
Nunchaku	133	107
Nori	118	127
Nina	68	78
Lex	71	81
Nebula	74	98
Nutella	109	175
Nux	77	NA
Nox	110	NA
Tabor	77	92
Vesta	85	98
Vito	73	103

Table S5. Estimated marginal means pairwise comparison based on model 1 & 2a. *For model 1 the comparison occurs between the subphases pre and post, split up by sexstatus and condition. For model 2a we planned contrast comparisons based on model between positive additive effect and the positive interaction as well as the negative additive effect and the negative interaction split up by sexstatus (coded as breeders, female helper, male helper). Significant differences indicate that the marmosets' response to the interaction playbacks does not correspond to an additive effect, whereas a non-significant difference may result from a pure additive understanding of the interaction. Bold values indicate p < 0.05.*

Model 1: nasal temperature ~ condition*subphase*sexstatus (random effect: family group(individual(session)))						
contrast	**sexstatus**	**condition**	**estimate (SE)**	**95 % CI**	**t-ratio**	***p*-value**
subphase pre (baseline) vs. post	breeders	pos-int	0.0749 (0.041)	–0.006; 0.156	1.820	0.0689
		neg-int	–0.147 (0.042)	–0.229; –0.066	–3.543	**0.0004**
		fc	–0.188 (0.051)	–0.288; –0.088	–3.685	**0.0002**
		ct	0.113 (0.032)	0.051; 0.174	3.574	**0.0004**
		gnaeh	–0.031 (0.063)	–0.154; 0.092	–0.494	0.6216
	female helpers	pos-int	–0.530 (0.050)	–0.628; –0.433	–10.694	**< 0.0001**
		neg-int	–0.476 (0.062)	–0.598; –0.354	–7.654	**< 0.0001**
		fc	–0.120 (0.073)	–0.262; 0.022	–1.654	0.0982
		ct	0.040 (0.053)	–0.064; 0.144	0.757	0.4493
		gnaeh	–0.158 (0.075)	–0.305; –0.010	–2.096	**0.0362**
	male helpers	pos-int	–0.057 (0.051)	–0.156; 0.042	–1.132	0.2576
		neg-int	0.396 (0.040)	0.317; 0.475	9.809	**< 0.0001**
		fc	0.625 (0.056)	0.515; 0.734	11.186	**< 0.0001**
		ct	0.365 (0.059)	0.250; 0.481	6.228	**< 0.0001**
		gnaeh	0.067 (0.041)	–0.014; 0.148	1.628	0.1038
Model 2a: nasal temperature (cartesian dataset) ~ condition*sexstatus (random effect: family group(individual))						
	sexstatus	**condition**	**estimate (SE)**	**95 % CI**	**t-ratio**	***p*-value**
	breeders	positive contrast	0.029 (0.106)	–0.209; 0.268	0.277	1.0000
		negative contrast	–0.316 (0.081)	–0.498; –0.134	–3.909	**0.0002**
	female helpers	positive contrast	–0.413 (0.104)	–0.647; –0.179	–3.966	**0.0002**
		negative contrast	–0.645 (0.136)	–0.950; –0.339	–4.743	**< 0.0001**
	male helpers	positive contrast	–0.512 (0.078)	–0.687; –0.337	–6.593	**< 0.0001**
		negative contrast	–0.251 (0.085)	–0.443; –0.059	–2.935	**0.0071**

Table S6. Wilcoxon tests on thermal change (post vs. pre) by individual. *Results for each Wilcoxon rank sum test comparing the subphases pre to post for each session separately.*

sexstatus	individual		gnaeh-fc	gnaeh-ct	fc	ct	gnaeh
fb	Lancia	*p*-value	< 0.0001	0.003	0.1087	0.0221	0.3429
		r	0.8327	0.742	0.5072	0.6345	0.3354
		direction	decrease	increase	none	increase	none
	Lea	*p*-value	0.1714	0.1508	0.1625	0.4	0.0121
		r	0.4325	0.4543	0.4417	0.3181	0.7564
		direction	none	none	none	none	increase
	Mina	*p*-value	0.0081	0.0069	0.0141	< 0.0001	0.4896
		r	0.5919	0.5296	0.6807	0.8721	0.141
		direction	increase	decrease	increase	increase	none
	Nina	*p*-value	< 0.0001	0.9433	0.2698	0.0721	< 0.0001
		r	1	0.0197	0.2676	0.375	0.7573
		direction	increase	none	none	none	decrease
mb	Kyros	*p*-value	0.7626	< 0.0001	NA	0.4181	0.0028
		r	0.0526	0.6871	NA	0.1653	0.8291
		direction	none	increase	NA	none	increase
	Lex	*p*-value	0.0202	0.0374	NA	0.0079	< 0.0001
		r	0.5633	0.434	NA	0.6443	0.9819
		direction	decrease	decrease	NA	decrease	decrease
	Lexus	*p*-value	0.3874	< 0.0001	NA	0.3969	0.3053
		r	0.1634	0.942	NA	0.2187	0.2292
		direction	none	decrease	NA	none	none
fh	Manuka	*p*-value	< 0.0001	NA	NA	NA	NA
		r	0.7878	NA	NA	NA	NA
		direction	decrease	NA	NA	NA	NA
	Mojita	*p*-value	< 0.0001	0.0357	< 0.0001	< 0.0001	0.0471
		r	1	0.7425	0.955	1	0.4816
		direction	decrease	decrease	decrease	increase	increase
	Nebula	*p*-value	0.0021	0.381	< 0.0001	NA	0.1922
		r	0.7068	0.3312	0.9826	NA	0.3074
		direction	decrease	none	decrease	NA	none
	Nougat	*p*-value	NA	NA	0.5	0.1429	0.0028
		r	NA	NA	0.3372	0.3916	0.8291
		direction	NA	NA	none	none	decrease
	Nutella	*p*-value	0.1818	0.0055	NA	0.4557	0.0357
		r	0.3854	0.6732	NA	0.1994	0.7425
		direction	none	decrease	NA	none	increase
	Nux	*p*-value	0.0121	NA	0.0339	0.2977	0.2286
		r	0.7564	NA	0.463	0.2783	0.4551
		direction	decrease	NA	increase	none	none
	Vesta	*p*-value	0.0089	< 0.0001	< 0.0001	< 0.0001	0.6354
		r	0.585	0.856	1	0.711	0.1185
		direction	decrease	decrease	increase	decrease	none
mh	Merkur	*p*-value	< 0.0001	0.0084	< 0.0001	< 0.0001	0.0616
		r	0.9886	0.5497	1	0.8721	0.347
		direction	decrease	increase	increase	increase	none
	Ninja	*p*-value	0.0674	0.9452	0.0045	0.0023	< 0.0001
		r	0.4089	0.0191	0.7327	0.6641	0.7628
		direction	none	none	increase	increase	decrease
	Nori	*p*-value	< 0.0001	0.0111	NA	NA	< 0.0001
		r	0.9409	0.4354	NA	NA	0.7089
		direction	decrease	increase	NA	NA	increase
	Nox	*p*-value	0.3677	NA	0.1778	0	0.5237
		r	0.26	NA	0.4262	0.9409	0.1769
		direction	none	NA	none	increase	none
	Nunchaku	*p*-value	0.027	< 0.0001	0.0939	0.0485	< 0.0001
		r	0.4338	1	0.4326	0.3664	0.9382
		direction	increase	increase	none	decrease	increase
	Tabor	*p*-value	< 0.0001	0.1246	< 0.0001	0.0727	< 0.0001
		r	0.9052	0.3619	0.8377	0.518	0.6659
		direction	decrease	none	decrease	none	decrease
	Vito	*p*-value	< 0.0001	< 0.0001	< 0.0001	0.0012	0.242
		r	1	0.9945	1	0.9006	0.1924
		direction	increase	increase	increase	increase	none

Table S7. Wilcoxon test on additive effect by individual. *Results for each Wilcoxon rank sum test conducted on comparison between either the positive interaction playback and the positive additive effect or the negative interaction playback and the negative additive effect. Bold values indicate p < 0.05.*

sexstatus	individual	Pos-int vs. pos-add	Neg-int vs. neg-add
fb	Lancia	W = 81, *p* < **0.0001** r = 0.967 N = 18	W = 19, *p* = 0.937 r = 0.023 N = 12
	Lea	W = 16, *p* = **0.03** r = 0.774 N = 8	W = 25, *p* = **0.008** r = 0.840 N = 10
	Mina	W = 6, *p* = 0.7 r = 0.157 N = 6	W = 36, *p* = **0.002** r = 0.885 N = 12
	Nina	W = 0, *p* < **0.0001** r = 1 N = 24	W = 0, *p* = **0.008** r = 0.840 N = 10
mb	Kyros	NA	W = 49, *p* = **0.0006** r = 0.919 N = 14
	Lex	NA	W = 1, *p* < **0.0001** r = 0.983 N = 24
	Lexus	NA	W = 49, *p* = **0.0006** r = 0.919 N = 14
fh	Manuka	NA	NA
	Mojita	W = 198, *p* = **0.007** r = 0.473 N = 32	W = 9, *p* = 0.1 r = –0.672 N = 6
	Nebula	W = 3, *p* = 0.7 r = 0.157 N = 6	NA
	Nutella	NA	W = 56, *p* = 0.190 r = 0.309 N = 18
	Nux	W = 9, *p* = 0.1 r = 0.672 N = 6	NA
	Vesta	W = 49, *p* = **0.002** r = 0.820 N = 14	W = 25, *p* = **0.008** r = 0.840 N = 10
mh	Merkur	W = 144, *p* < **0.0001** r = 1 N = 24	W = 87, *p* = **0.004** r = 0.646 N = 20
	Ninja	W = 35. *p* = 0.1014 r = 0.349 N = 22	W = 6, *p* = 0.065 r = 0.533 N = 12
	Nori	NA	NA
	Nox	W = 10, *p* = 0.686 r = 0.143 N = 8	NA
	Nunchaku	W = 196, *p* < **0.0001** r = 1 N = 28	W = 209, *p* = **0.03** r = 0.382 N = 34
	Tabor	W = 0, *p* = **0.002** r = 0.885 N = 12	W = 0, *p* = **0.002** r = 0.885 N = 12
	Vito	W = 93, *p* = **0.034** r = 0.453 N = 22	W = 98, p < **0.0001** r = 0.914 N = 20
		9 of 14 individuals with significant change (NA = 6)	11 of 15 individuals with significant change (NA = 5)

Table S8. Models of effects of activity on nasal temperature (full model). *Summary table of LMMs investigating the effect of activity on nasal temperature. Bold values indicate p < 0.05.*

Fixed factors	*B*	SE	*t*-value	*p*-value
Model 5a: nasal temperature (only *pos-int*-dataset) Likelihood ratio test (full model vs. null model): N_{total} = 168, $N_{individuals}$ = 20, $N_{sessions}$ = 20, pseudo-R^2_c = 0.869; $\chi^2_{(1)}$ = 7.479, *p* = **0.006**				
intercept	0.262	0.160	1.631	0.110
Activity	–0.070	0.024	–2.993	**0.012**
Model 5b: nasal temperature (only *neg-int*-dataset) Likelihood ratio test (full model vs. null model): N_{total} = 131, $N_{individuals}$ = 17, $N_{sessions}$ = 17, pseudo-R^2_c = 0.; $\chi^2_{(1)}$ = 2.816, *p* = 0.093				
intercept	0.166	0.157	1.058	0.292
Activity (stimulation phase)	–0.044	0.025	–1.736	0.117
Model 5c: nasal temperature (only *fc*-dataset) Likelihood ratio test (full model vs. null model): N_{total} = 120, $N_{individuals}$ = 14, $N_{sessions}$ =14, pseudo-R^2_c = 0.911; $\chi^2_{(1)}$ = 12.568, *p* < **0.0001**				
intercept	0.685	0.160	4.294	**0.000**
Activity (stimulation phase)	–0.098	0.022	–4.525	**0.004**
Model 5d: nasal temperature (only *ct*-dataset): Likelihood ratio test (full model vs. null model): N_{total} = 133, $N_{individuals}$ = 17, $N_{sessions}$ =1 7, pseudo-R^2_c = 0.723; $\chi^2_{(1)}$ = 0.129, *p* = 0.719				
intercept	0.253	0.149	1.692	0.093
Activity (stimulation phase)	–0.009	0.024	–0.378	0.719
Model 5e: nasal temperature (only *gnaeh*-dataset) Likelihood ratio test (full model vs. null model): N_{total} = 138, $N_{individuals}$ = 19, $N_{sessions}$ = 19, pseudo-R^2_c = 0.908; $\chi^2_{(1)}$ = 0.002, *p* = 0.962				
intercept	0.083	0.248	0.333	0.740
Activity (stimulation phase)	–0.002	0.040	–0.049	0.962

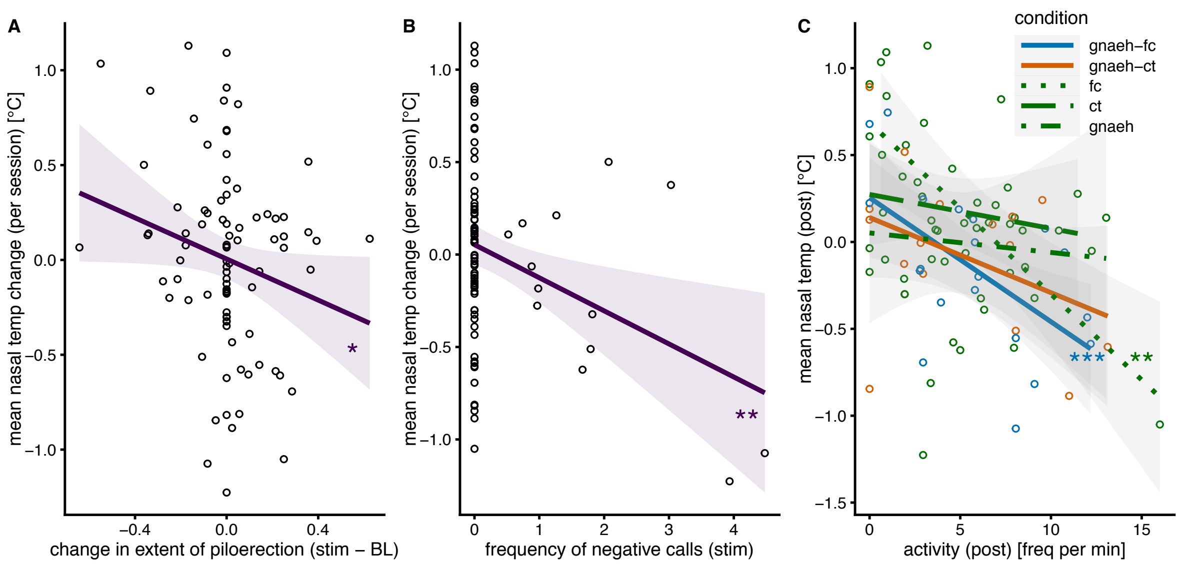

Fig. S2. Independent measures of arousal and the effect of activity. *A) Relationship between nasal temperature and the extent of piloerection (calculated as the difference between baseline and stimulation period). Each point indicates a mean value per session. The shaded area indicates the 95 % confidence intervals of the regression line. (B) Relationship between nasal temperature and the frequency of negative calls (stimulation period). Each point indicates a mean value per session. (C) Relationships between nasal temperature change (as a mean per session) and activity during subphase post. Shaded areas represent the 95 % confidence intervals of the regression line. (significance is indicated with * p ≤ 0.05, ** p ≤ 0.01, *** p ≤ 0.001).*

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. St. 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).