Current Biology

Supplemental Information

Dogs Can Discriminate Emotional Expressions of Human Faces

Corsin A. Müller, Kira Schmitt, Anjuli L.A. Barber, and Ludwig Huber

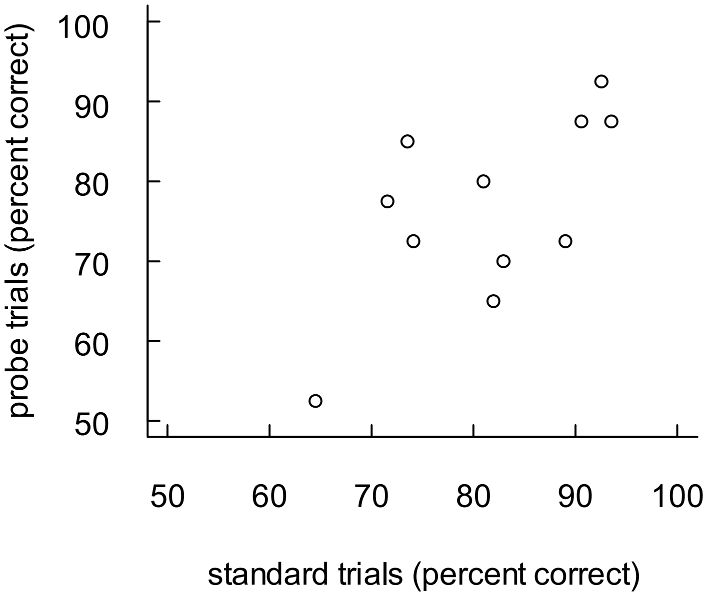

Figure S1. Correlation between proportion of correct choices in standard trials and probe trials in the test phase. Relates to Figure 4.

Table S1: Subjects, experimental groups and training performances [a]

ID	Breed	Sex	Age (yrs)	Pre-training: No. Sessions [b]	Training Set [c]	Rewarded Expression	Training: No. Sessions [b]
1	Border Collie	female	8	**16**	1a	Happy	**11**
2	Border Collie	male	6	**8**	1a	Happy	**6**
3	Border Collie	female	7	**29**	1b	Happy	26
4 [d]	Mongrel	male	9	**31**	1b	Happy	35
5	Fox Terrier	male	8	**13**	1b	Happy	**9**
6	Border Collie	female	4	**13**	2a	Happy	**10**
7	Border Collie	male	9	**6**	2a	Happy	**6**
8	Border Collie	male	10	**14**	2b	Happy	**17**
9	Border Collie	female	9	**7**	2c	Happy	**4**
10 [d]	Border Collie	female	7	**11**	1a	Angry	**28**
11	Border Collie	female	5	**6**	1a	Angry	30
12	Border Collie	female	5	**12**	1b	Angry	33
13 [d]	Border Collie	female	5	**16**	1b	Angry	**19**
14	Golden Retriever	female	10	**9**	1c	Angry	31
15 [d]	Border Collie	female	4	**8**	2a	Angry	**24**
16	Sheltie	female	1	**7**	2b	Angry	**31**
17	Border Collie	female	2	**4**	2b	Angry	35
18	Border Collie	female	5	**14**	2c	Angry	26
19 [d]	Border Collie	male	5	**20**	1c	Happy	5
20 [d]	German Shepherd	female	7	30	2b	Happy	NA
21 [e]	Border Collie	female	6	**5**	2a	Angry	8
22	Mongrel	male	7	30	2c	Angry	NA
23	Border Collie	male	9	0	2c	Happy	NA
24 [e]	Border Collie	male	9	12	1c	Angry	NA

[a] subjects below the bold line were not included in the analysis of the respective phase, either because they did not reach the learning criterion in the preceding phase or because they dropped out for other reasons (see Supplemental Experimental Procedures below for details).

[b] bold: reached learning criterion

[c] 1: lower halves, 2: upper halves; a: stimulus pairs 1-15; b: stimulus pairs 6-20; c: stimulus pairs 11-25

[e] dogs that, prior to the present study, had participated in a task in which they were rewarded for discriminating between two unfamiliar faces with neutral expressions, or elements of the two faces, presented on a touch-screen monitor [S1].

[d] dogs that, prior to the present study, had participated in a task, in which they were rewarded for discriminating between two familiar faces, with neutral expression, projected in live size unto a screen [S2].

Table S2. Binomial models comparing performance in the four test conditions to chance level.

Condition [a]	Estimate [b]	Standard error	t_{10}	p-value
NFSH	1.39	0.32	4.40	0.001
TFOH	1.03	0.25	4.06	0.002
NFOH	0.94	0.21	4.48	0.001
TFLH	1.44	0.30	4.90	<0.001

[a] NFSH: novel face, same half; TFOH: training face, other half; NFOH: novel face, other half; TFLH: training face, left half

[b] in binomial models with logit link, as used here, an estimate of zero corresponds to a choice probability of 50%

Supplemental Experimental Procedures

Subjects

All 24 subjects were adult pet dogs (*Canis familiaris*) whose owners volunteered to participate in the study and gave written consent (for details on breed and age see Table S1). All but one of the dog owners were female. Two of the 24 subjects did not reach the pre-training criterion within 30 sessions and thus did not proceed to the training phase. Another subject developed an aversive response to the noise of the automatic feeder in the first session. Pre-training with this subject was therefore discontinued. A further seven subjects dropped out of the study at different stages due to limitations in owner availability. Of these, three dropped out during the pre-training phase or after less than 10 sessions in the training phase and where excluded from the analyses. The other four subjects completed at least 25 sessions in the training phase and were included in the analysis of the latter. The resulting sample size for the analysis of the training phase results was therefore 18, of which 11 reached the training criterion and thus proceeded to the test phase.

All subjects had prior experience with the touch-screen apparatus and had at least learned to solve a two-choice task that required them to discriminate two geometrical figures presented simultaneously on the computer screen (a square and a circle). Six of the subjects had previously participated in a study in which they were rewarded for discriminating between two unfamiliar faces, or elements of the two faces, presented on the touch-screen monitor [S1]. Two of the subjects had previously participated in a task, in which they were rewarded for discriminating between two familiar faces projected in life-size unto a screen [S2]. The eight subjects with pre-experience in discriminating faces did not perform better in the pre-training (median 16 sessions to criterion) than the subjects without such pre-experience (median 12 sessions to criterion; Mann-Whitney U-test: $W = 67.5$, $p = 0.31$). No corresponding analysis was run for the training phase, as only four subjects with experience completed it. Inspection of the data gives no evidence that experimental experience influenced the performance in the training phase either:

among the subjects with prior experience with a face discrimination task, the subject that was rewarded for touching the happy stimulus did not reach the learning criterion (Subject 4 in Table S1) whereas one of the three subjects rewarded for touching the angry stimulus reached the criterion quickest in this group (Subject 13 in Table S1).

Material

The test apparatus consisted of a 15-inch touch-screen TFT monitor (resolution: 1024 x 768 pixels) connected to a CPU and an automatic feeder (for further technical details see ref [S3]). The monitor was visually shielded on both sides and on top by 40-cm long panels to ensure that the dogs' choices could not be influenced by the owner or the experimenter. The monitor could be shifted vertically and was adjusted to the nose height of the different subjects.

As stimuli for the pre-training, we used photographs of the face (with a neutral expression) and of the back of the head of ten female lab members. These pictures were taken with a Canon EOS 6D and a Walimex Pro Studioset VC-200 under standardized conditions against a white background. As stimuli for the training and test phase we used photographs of 25 Caucasian women with a happy or an angry expression, which were obtained from validated databases [S4–S6]. We used only photographs of women as stimuli in this study because most of our dog owners were female and we assumed that the subjects would have the best chance of solving the task for stimuli resembling the face they are most familiar with. All photographs were cropped and resized to 8 x 12 cm. For the training phase, either the top 6.5 cm or the bottom 5.5 cm of the emotional faces was covered with a white rectangle. Covering areas of slightly unequal size was necessary to ensure that in all pictures the complete eye region (including the lower eyelids) was visible when the upper halves of the faces were presented. In the test phase, we additionally used emotional faces with the right half covered (we showed the left halves rather than the right halves of the pictures due to the documented left-gaze bias of dogs when looking at pictures of human faces [S7, S8]).

Procedure

In the pre-training phase as well as in the training and test phases, the dogs received sessions of 30 trials. Training and test sessions lasted between 3 and 12 minutes (median: 5 min) and were separated by breaks of at least 2 minutes. Test sessions (3-8 min, median 4 min) did not last longer than the last five training sessions of the dogs that reached the training criterion (3-10 min, median 5 min). Depending on the duration of the sessions and the motivation of the dogs, between 2 and 8 sessions (median: 4 sessions) were conducted per visit. The subjects were never forced to participate. If a dog did not approach the apparatus voluntarily after a break, training/testing was continued on a later day. Visits were typically separated by between 1 and 14 days, though on one occasion, two training visits were separated by 42 days. This did not seem to affect the performance as the respective dog reached the first half of the learning criterion before and the second half immediately after the long break.

In each trial a stimulus pair of photographs of the same person was shown with the rewarded stimulus presented randomly on the right or on the left side. If the subject chose the correct stimulus by touching it with the nose, a 600 Hz tone was played and a piece of dry dog food was dispensed. If the subject chose the incorrect stimulus, a 200 Hz tone was played and a red screen was presented for 3 seconds. Incorrect choices were followed by correction trials, presenting the same stimulus pair again, until the dog chose correctly (the correction trials were excluded from the dataset prior to the analyses). For dogs with poor impulse control, judged by the experimenter in the first session of the pre-training, the dog was pushed one step back by the experimenter after each trial (this was possible without looking at the stimuli herself, cf. Movie S1).

In the pre-training phase, the face was the rewarded stimulus and the back of the head was the unrewarded stimulus for all subjects. The set of ten stimulus pairs was shown three times in each session and the order of the stimulus pairs was randomized within the sets. The pre-training served to confirm that the subjects do not suffer from substantially impaired eye sight, to ensure that all subjects are familiar with the two-choice procedure on the touch-screen apparatus and that a possibly aversive effect of unfamiliar human faces does not interfere with the required approach and touch response. In the training phase, the stimulus showing the happy expression was the rewarded stimulus for half of the subjects, whereas for the other half of the dogs the stimulus with the angry expression was rewarded. We used a subset of 15 of the 25 stimulus pairs in the training phase, the remaining stimuli were saved for the test phase. The subset of 15 stimulus pairs was shown twice in each training session and the order of the stimulus pairs was randomized within the subsets. We used three different subsets of the 25 stimulus pairs so that not each dog saw the same set of stimuli during training and in the probe trials (cf. Table S1). This served to ensure that single stimulus pairs could not have an unduly large effect on our results.

In the test phase, five probe trials were interspersed semi-randomly within sessions of standard trials (trials identical to those presented in the training phase). Successive probe trials were separated by at least three and at most ten standard trials. In probe trials, the dogs were rewarded for touching either of the two stimuli presented. The probe trials of the three conditions showing the upper or lower halves of the faces (cf. Fig. 2) were presented alternately in the first six test sessions. The probe trials showing the left halves of the faces were presented in the last two test sessions to ensure that these probe trials could not facilitate transfer of the learned contingency from the lower halves to the upper halves of the faces (or vice versa) in the other probe conditions.

Subjects proceeded from the pre-training to the training phase once they had reached a learning criterion of at least 24 correct choices in three consecutive sessions. Pre-training was discontinued if a subject did not reach this criterion within 30 sessions (900 trials) though for one subject, which had reached the required level of 80% correct choices in the 29th and 30th session, one extra pre-training session was added (for discussion of the pre-training results, see

Supplemental Results below). Subjects proceeded from the training phase to the test phase if they reached a learning criterion of at least 23 correct choices in 4 out of 5 consecutive sessions (11 of 18 subjects reached this criterion). Training was discontinued if a subject did not reach this criterion within 35 sessions.

Analyses

As not all subjects reached the learning criterion in the training phase, we compared the rate at which subjects reached the criterion between the groups of subjects rewarded for touching the happy or the angry stimulus, and between the groups of subjects shown the upper or the lower halves of the faces, with Cox proportional hazards models (R-package "survival" [S9, S10]). Performance in the different conditions of the test phase was compared using a generalized linear mixed effects model (GLMM, R-package "lme4" [S11]) with proportion of correct choices in the numerator and number of trials in the denominator of the binomial response term. Dog identity was included as a random factor to account for the repeated measures structure in the dataset. Performance in each condition was compared to the chance level of 50% using binomial generalized linear models (GLMs) with the intercept as the only predictor.

Supplemental Results

Pre-training

As mentioned above, two of the 24 subjects did not reach the pre-training criterion and another two dropped out of the study before doing so. The remaining 20 subjects needed between 4 and 31 sessions to reach the pre-training criterion, which is considerably longer than for the subjects in the study of Nagasawa and colleagues [S12]. This discrepancy can be explained by two methodological differences between the two studies: First, rather than a single stimulus pair, we used ten different stimulus pairs for pre-training, which likely made learning more difficult. Second, we did not use pictures of the owner as stimuli but pictures of people that were unfamiliar to most of the subjects and, for some dogs, the face of an unfamiliar person may have been a slightly aversive stimulus that at least initially inhibited the necessary approach and touch behavior.

Test Phase

For the eleven subjects that reached the test phase, the proportion of correct choices in the probe trials (the four conditions pooled) was correlated with the proportion of correct choices in standard trials during the test phase (Figure S1, binomial GLM: $\beta = 0.04$, $t_9 = 2.48$, $p = 0.035$), which is to a large extent carried by the subject with the worst performance in both standard and probe trials. This result is also consistent with our interpretation that the subjects transferred the contingency they had learned for standard trials to the four types of probe trials presenting novel stimuli but with the same two emotional expressions.

Performance in the probe trials did not differ between subjects that were rewarded for touching the happy stimulus (N = 7) and subjects rewarded for touching the angry stimulus (N = 4; binomial GLMM with likelihood ratio test: $\chi^2_{(1)} = 0.18$, $p = 0.67$).

The one dog that chose correctly in less than 65% of the probe trials (in 21 of the 40 trials) was the only study participant with a male owner. This outlier is in line with the suggestion of Nagasawa and colleagues [S12] that discrimination of facial expressions is easier for dogs when they are shown faces of the same gender as their owner, rather than faces of the opposite gender.

Supplemental References

S1. Pitteri, E., Mongillo, P., Carnier, P., Marinelli, L., and Huber, L. (2014). Part-based and configural processing of owner's face in dogs. PLoS One *9*, e108176.

S2. Huber, L., Racca, A., Scaf, B., Virányi, Z., and Range, F. (2013). Discrimination of familiar human faces in dogs (*Canis familiaris*). Learn. Motiv. *44*, 258–269.

S3. Steurer, M. M., Aust, U., and Huber, L. (2012). The Vienna comparative cognition technology (VCCT): an innovative operant conditioning system for various species and experimental procedures. Behav. Res. Methods *44*, 909–918.

S4. Lundqvist, D., Flykt, A., and Öhman, A. (1998). The Karolinska Directed Emotional Faces - KDEF. Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet. Available at: http://www.emotionlab.se/resources/kdef.

S5. Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., and van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. Cogn. Emot. *24*, 1377–1388.

S6. Van der Schalk, J., Hawk, S. T., Fischer, A. H., and Doosje, B. (2011). Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES). Emotion *11*, 907–920.

S7. Guo, K., Meints, K., Hall, C., Hall, S., and Mills, D. (2009). Left gaze bias in humans, rhesus monkeys and domestic dogs. Anim. Cogn. *12*, 409–418.

S8. Racca, A., Guo, K., Meints, K., and Mills, D. S. (2012). Reading faces: differential lateral gaze bias in processing canine and human facial expressions in dogs and 4-year-old children. PLoS One *7*, e36076.

S9. Therneau, T. M., and Grambsch, P. M. (2000). Modeling Survival Data: Extending the Cox Model. (New York: Springer).

S10. Therneau, T. (2014). A package for survival analysis in S. R package version 2.37-7. http://CRAN.R-project.org/package=survival.

S11. Bates, D., Maechler, M., and Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0. http://cran.r-project.org/package=lme4.

S12. Nagasawa, M., Murai, K., Mogi, K., and Kikusui, T. (2011). Dogs can discriminate human smiling faces from blank expressions. Anim. Cogn. *14*, 525–533.