ANTHROZOÖS
https://doi.org/10.1080/08927936.2025.2469400

Barking Up the Wrong Tree: Human Perception of Dog Emotions Is Influenced by Extraneous Factors

Holly G. Molinaro and Clive D. L. Wynne

Psychology Department, Arizona State University, Tempe, AZ, USA

ABSTRACT

Human perception of others' emotions is complex. Specifically, little is known about what other extraneous factors may influence human perception of emotions in animals. This two-experiment study investigated contextual and demographic factors that impact on human perception of dog emotions. A dog was video recorded in putatively positive or negative situations. In experiment 1, participants were shown videos with and without visual background (context). In experiment 2, videos were edited so the dog appeared in mismatched contexts. In both experiments, undergraduates ($n = 383$, 485) rated the videos for valence and arousal, and freely described the dog's emotions. In experiment 1, the presence of context impacted participants' ratings: valence was rated more positive in videos without context and in videos of positive situations, whilst arousal was rated higher specifically for videos of positive situations without context. In experiment 2, videos in which the human was doing something positive were rated higher for levels of valence and lower for levels of arousal regardless of the situation in which the dog had been recorded. We highlight that extraneous factors besides the dog itself are major contributing influences on how humans perceive dogs' emotions. These results serve to expand our comprehension of how we perceive animal emotions, strengthening the basis for providing better welfare to animals under human care. They also contribute to the consideration of what theories of emotion may be best suited to understanding animal emotional expression.

KEYWORDS
Constructed theory of emotion; dog; human–animal interaction; welfare

A better understanding of what animals experience and how people perceive those states offers the promise of improved welfare for animals under human care, but understanding animal emotions is challenging. There is much debate even over how to conceive of emotions (Anderson & Adolphs, 2014; LeDoux, 2012) and what animals may be capable of experiencing (de Vere & Kuczaj, 2016; Kremer et al., 2020).

Prior studies of human perception of dog emotions have focused primarily on how readily people can identify basic emotions in dogs. It has been found that humans consider dog emotions the same as their own (Konok et al., 2015; Kujala et al., 2017; Schirmer

CONTACT Holly G. Molinaro hgmolina@asu.edu Psychology Department, Arizona State University, 950 S McAllister Ave, Tempe, AZ 85281, USA

Supplemental data for this article can be accessed online at https://doi.org/10.1080/08927936.2025.2469400.

et al., 2013). People report that dogs have the same basic emotions as humans, and they therefore perceive these emotions in a similar way as they perceive human emotions (Arahori et al., 2017; Martens et al., 2016; Morris et al., 2008). Facial expressions have been found to be essential for humans in perceiving dog emotions (Correia-Caeiro et al., 2021), and many studies focus on how humans perceive emotions from just the face of a dog (Bloom & Friedman, 2013; Park & Kim, 2020; Trevathan-Minnis et al., 2021).

Most studies on people's perception of dog emotions assume dogs share basic emotions with humans. These studies often involve an expert assigning an emotion to an auditory or visual recording of a dog, with human participants' ratings considered correct if they match the expert's (e.g., Burza et al., 2022; Hantke et al., 2018; Lakestani et al., 2014; Wan et al., 2012). Burza et al. (2022) reported that humans could recognize four basic emotions from a dog's eyes and facial expressions, and Bloom and Friedman (2013) found humans could correctly rate six basic emotions from photographs of a dog's face. Other studies either provide a preset list of emotional or behavioral terms for participants to choose from or allow them to generate their own descriptions, using consensus to infer an accurate understanding of the dog's emotions (Tami & Gallagher, 2009; Walker et al., 2010). However, these studies assess the agreement between humans rather than directly evaluating the perception of dog emotions. A recent review by Correia-Caeiro et al. (2023) highlights strong anthropomorphic biases in the current literature, suggesting a need for alternative approaches to studying how people perceive dog emotions.

Several studies show that recognizing human emotional expressions involves more than just reading facial appearances. Aviezer et al. (2008) found that the same facial expression was perceived differently based on context, such as holding a diaper (disgusted) versus at a grave site (sadness). Calbi et al. (2017) showed that after viewing fearful scenes (e.g., a spider), participants rated neutral faces as more fearful. Additionally, emotional perception can be influenced by body language, context, and voice (de Gelder et al., 2006; Meeren et al., 2005). These findings suggest that people might also be influenced by additional background factors when perceiving animal emotions, especially given that Caeiro et al. (2017) found that dogs and humans display different facial expressions even in similar contexts. This indicates that humans' perceptions of dog emotional expressions may be inaccurate.

Some indication of how broader background factors may influence the perception of dog emotional expressions comes from studies that have found that demographic factors can be important in people's perception of dog emotions. Lakestani (2007) reported that age, culture, gender, and ethnicity influenced children's labeling of dog emotions. Amici et al. (2019) found significant differences in the ability to recognize dog emotions between groups of adults from cultures that did or did not treat dogs as pets, as measured by their ability to match adult-labeled emotions in photographs of dogs. Similarly, Szánthó et al. (2017) found differences between people living in Germany and Hungary in a survey of emotional ratings of dogs. These findings point to the possibility that people's perception of dog emotions may be influenced by more than just the behavior of the dog itself.

Mendl et al. (2010) proposed mitigating anthropomorphic tendencies in animal emotion perception research by refraining from asking individuals to rate an animal's emotions based on the six basic human emotions and instead having them rate the animal's affect. *Affect* describes a general feeling or mood (Russell, 2003) which

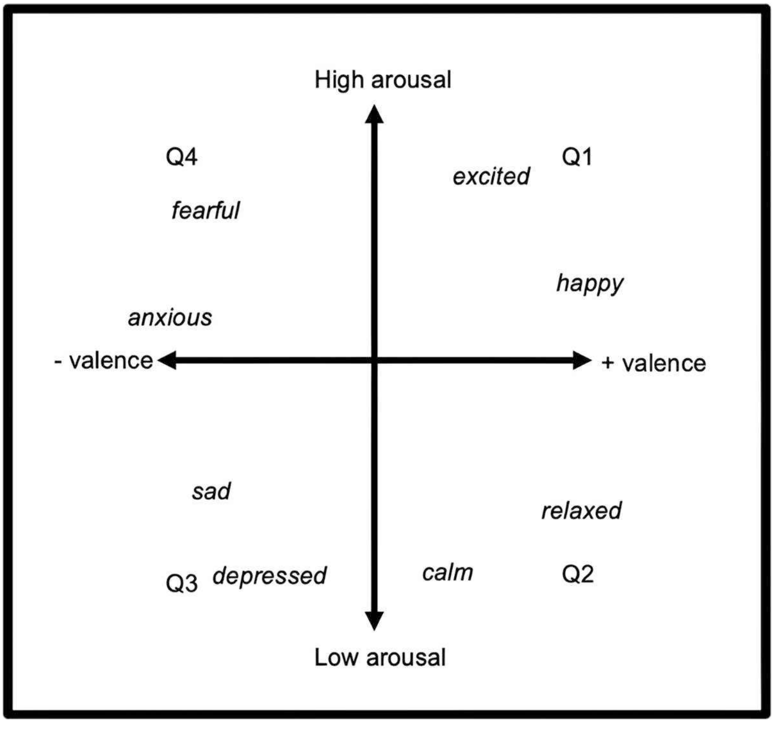


Figure 1. The two-dimensional space that constitutes an emotion, including a range of valence and arousal. Sample emotional words are placed to indicate what each quadrant represents (e.g., medium arousal and high valence indicate a sense of happiness for humans). Modified from Mendl et al. (2010).

can be analyzed in a two-dimensional space (Figure 1). One dimension is *valence:* the extent to which something feels good or bad. Additionally, a feeling can be subdued or excitatory – known as *arousal* (Barrett, 2017).

The current study investigated human perception of dog emotions by manipulating extraneous factors in two experiments. We aimed to elucidate whether humans' perceptions of a dog's emotions were directly influenced by the dog itself or by other aspects of the environment and situation around the dog. Unlike prior studies that focused on specific dog features or used predefined emotion lists, we showed participants videos of a freely behaving dog and asked them to rate the dog's valence and arousal. We also explored how demographic factors and participants' moods impacted their perceptions. We hypothesized that human perception of dog emotions is influenced by other extraneous situational factors, rather than just what the dog is doing, and that observer's demographics and mood would affect these perceptions.

Experiment 1: Context vs. No Context

This experiment examined how the presence or absence of background context (such as a wall, carpet, a person, and other objects present in the room) influenced participants' judgments of a dog's emotional state.

Methods

Filming

A 14-year-old medium-sized pointer/beagle mix (approximately 30 kg) and his owner were filmed in six scenarios. The dog was chosen for his lighter facial coloring, which was easier to see on video (Figure 2). A single dog was used to restrict the number of variables under consideration. Filming occurred at their home to ensure the dog's reactions to the emotion-inducing stimuli were not influenced by any response to a novel environment. All videos were shot against a blank wall using an iPhone 13 Pro Max (Apple Inc., Cupertino, CA).

The dog was presented with stimuli his owners believed, based on their experience with him, were linked to positive or negative emotions, including a treat, praise, play (positive), and a cat, a tape measure, and reprimand (negative). The video recordings captured the dog's entire body and the presentation of stimuli, with multiple iterations (three to five) to provide editing choices.

Video Editing

Video clips, (15–30 s), were edited in iMovie software (Apple Inc., Cupertino, CA) to include only the interaction between the dog and the owner and stimuli. Clips were selected to ensure both the dog and the person were in frame, capturing all of the dog's behavioral actions. Adobe After Effects (Adobe Systems Inc., San Jose, CA) software was used to edit out the background and context, leaving the dog against a black backdrop while preserving the dog's sounds and removing owner and stimulus sounds (example in Figure 2). The final materials comprised 12 videos: six unedited originals and six with context edited out, split evenly between positive and negative scenarios. These resulted in four video categories each containing three videos, one from each of three scenarios: No Context Negative (NCN), Context Negative (CN), No Context Positive (NCP), and Context Positive (CP).

Survey

A survey was created using Qualtrics software (Qualtrics, Provo, Utah), comprised of two blocks of questions. The first block gathered demographic data (age, gender, ethnicity,

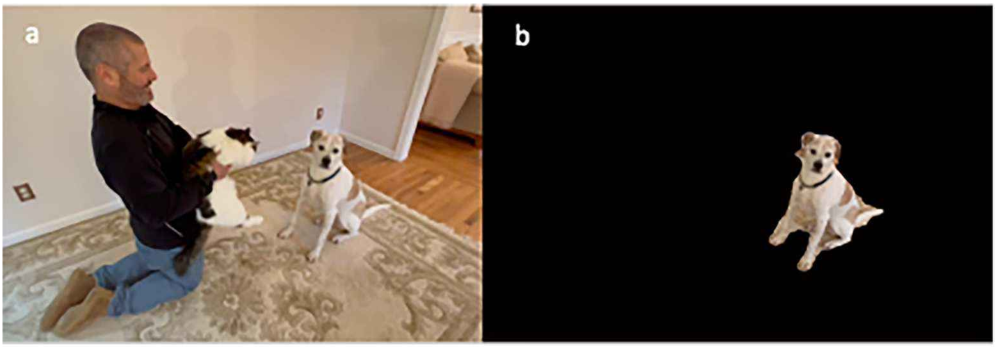


Figure 2. Comparison of frame grabs from the same video with (a) and without (b) context. The owner gave consent for his image to be used.

dog experience, and current mood). In the second block, participants first watched six videos alternately depicting positive and negative scenarios without context. They provided emotional adjectives in a free response question, rated valence from 1 (extremely bad) to 10 (extremely good) and arousal from 1 (extremely calm) to 10 (extremely agitated), and specified cues used to perceive the dog's emotions. All six context videos were presented after those without context to prevent anticipation of the study's purpose. The survey aimed to be completed within 30 min.

Participants

The survey was sent out to Arizona State University (ASU) psychology undergraduates. ASU's Institutional Review Board for human subjects approved this research project on January 24, 2022. The survey was launched on January 31, 2022 and ended on February 26, 2022, once 400 respondents had completed the survey.

Free Response Coding

Each free response was coded for the following five dimensions:

1. Valence (on a scale from –2 to +2)
2. Arousal (on a scale from –2 to +2).

   These first two were chosen to replicate the valence and arousal scale questions.
3. Presence or absence of anthropomorphism (1,0)
4. Presence or absence of the mental state terms (1,0)
5. Presence or absence of action state terms (1,0).

To establish the coding paradigm, three research assistants blind to the study question initially reviewed a subset of 150 free responses and collectively determined scoring for the five dimensions through consensus coding (Richards & Hemphill, 2018). They documented their definitions and revisited the subset, adjusting definitions until unanimous agreement was reached. Once a consensus codebook was established, a larger subset (20% of responses) was reviewed to ensure consistency, with any discrepancies resolved through collaborative discussion. This process yielded the primary codebook (see online supplemental material). Subsequently, the assistants independently coded another 20% subset of responses. Cronbach's alpha was used for interrater reliability. Reliability between the three coders was high for each of the subcodes (valence, $\alpha = 0.962$; arousal, $\alpha = 0.927$; anthropomorphism, $\alpha = 0.933$; mentalizing, $\alpha = 0.912$; action, $\alpha = 0.934$). The three coders then individually coded all 4,596 free responses for all five dimensions listed above.

Data Analysis

IBM SPSS Statistics (Version 28.0.1.0, IBM Corp., Armonk, NY) was used to analyze all data. Data were visually assessed and were approximately normally distributed (with the exception of the free response data) and displayed homogeneity of variance. Results were considered significant at an alpha level of $p < 0.05$.

Two-way repeated measures ANOVAs were used to analyze the effects of context on valence and arousal responses separately. There were two within-subjects' factors: the situation (positive or negative) and context (present or absent). For both valence and arousal, responses were averaged across the positive, as well as the negative, situations.

Free responses were analyzed using nonparametric tests owing to their non-normal distribution. For valence and arousal, coder values were averaged per video, then across participants for each category (NCN, CN, NCP, CP), resulting in one valence and arousal value within a −2 to 2 range. Friedman's two-way analysis of variance by ranks was applied to all video categories, with Wilcoxon signed-rank tests for pairwise comparisons using Bonferroni correction. Effect sizes for valence and arousal were determined using matched pairs rank biserial correlations, which range from −1 (all values of the second sample are larger than all values of the first) to +1 (all values of the second sample are smaller than all values of the first). Anthropomorphism, mentalizing, and action values were averaged per video across coders, with discrepancies resolved by the first author. These values were then averaged across participants for each category. Cochran's *Q* test was used to assess differences across all videos, followed by McNamar's post hoc pairwise comparisons with Bonferroni correction.

Three-way ANOVAs were used to examine the impact of demographic factors (age class, gender, and ethnicity; the fixed independent variable factors) on averaged valence and arousal responses. Age class was divided into two groups (18–21 and 22+), gender into three (Female, Male, Other), and ethnicity into five (Asian, Black, Hispanic, White, Other) due to small sample sizes in certain categories.

One-way ANOVAs were used to analyze the effects of dog experience on both averaged valence and arousal scores. The "extremely familiar" and "very familiar" groups were combined owing to small sample sizes, so that there were five categories of dog experience (none, somewhat, familiar, familiar currently, very/extremely).

A simple linear regression was used to assess the influence of pre-video mood on valence and arousal scores. Given the Likert scale's approximation of a continuous variable, we treated both pre-mood and averaged valence and arousal as continuous variables (Johnson & Creech, 1983; Norman, 2010; Sullivan & Artino, 2013). Additionally, a paired *t*-test was used to compare pre-survey and post-survey moods.

Two chi-square tests were used to assess response proportions for what participants used to determine the dog's emotions: the first for videos with the dog on a black background, and the second for videos with the full context/background.

Results

Response Inclusion Criteria and Demographics

The median time to complete the survey was 755 s. All survey responses in the 25% or 75% quartile of durations were checked for inadequate attention to the videos or unreliable answers, leaving 383 responses. Demographics are in the online supplemental materials.

Effect of Context and Positive or Negative Situation on Valence and Arousal Scores

Valence responses were greater (more positive) without context ($F_{(1,381)} = 6.60$, $p < 0.05$, $\eta_p^2 = 0.02$) and were also greater in positive situations ($F_{(1,381)} = 286.50$, $p < 0.05$, $\eta_p^2 =$

0.43), with a significant interaction ($F_{(1,381)} = 497.32$, $p < 0.05$, $\eta_p^2 = 0.57$). Valence was rated higher (i.e., dogs were considered more positive) in positive videos with context than without context but was rated higher in negative videos without context (Figure 3(a)).

Ratings of the arousal of dogs in the videos did not differ based on the presence or absence of context ($F_{(1,380)} = 2.22$, $\eta_p^2 = 0.01$) but were higher (dogs were considered more aroused) in positive situations ($F_{(1,380)} = 135.07$, $p < 0.05$, $\eta_p^2 = 0.26$). The interaction was significant ($F_{(1,380)} = 150.11$, $p < 0.05$, $\eta_p^2 = 0.28$), indicating that videos in the negative situations were viewed as less arousing without context, but videos in the positive situation were viewed as less arousing with context (Figure 3(b)).

Free Responses

Across the 12 videos, 4,596 free responses were analyzed (383 free response answers per video). Valence-coded free responses differed among all video categories ($\chi^2_{(3)} =$ 1071.62, $p < 0.001$). Wilcoxon signed-rank tests were significant for all pairwise comparisons and the effect sizes were large for all pairs except NCP/NCN (online supplemental materials). CP videos had the highest levels of valence (most positive) in free responses, followed by NCP, NCN, and CN (Figure 4(a)). Arousal-coded free responses also differed among all video categories ($\chi^2_{(3)} = 1091.72$, $p < 0.001$), Wilcoxon signed-rank tests were significant for all pairwise comparisons, and the effect sizes were large for all pairs (online supplemental materials). NCP videos had the highest level of arousal throughout free responses, followed by CP, CN, and NCN (Figure 4(b)). Anthropomorphism-coded free responses differed among all video categories ($\chi^2_{(3)} =$

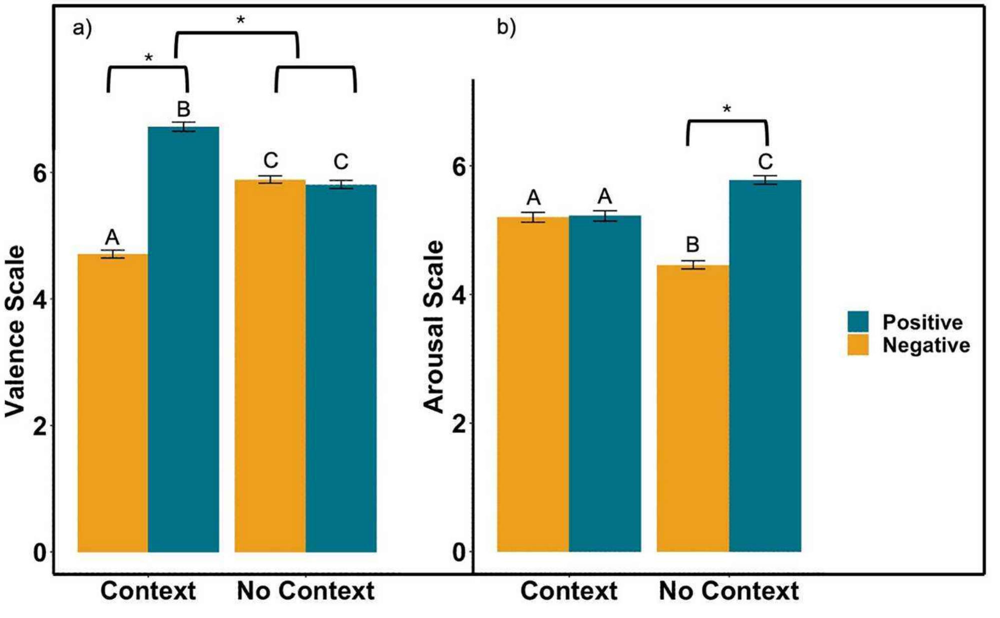


Figure 3. Average valence responses (a) and arousal responses (b) for positive and negative situations, with and without context from experiment 1. Means marked with different letters are significantly different. Asterisks (*) denote significant differences. Error bars show standard error.

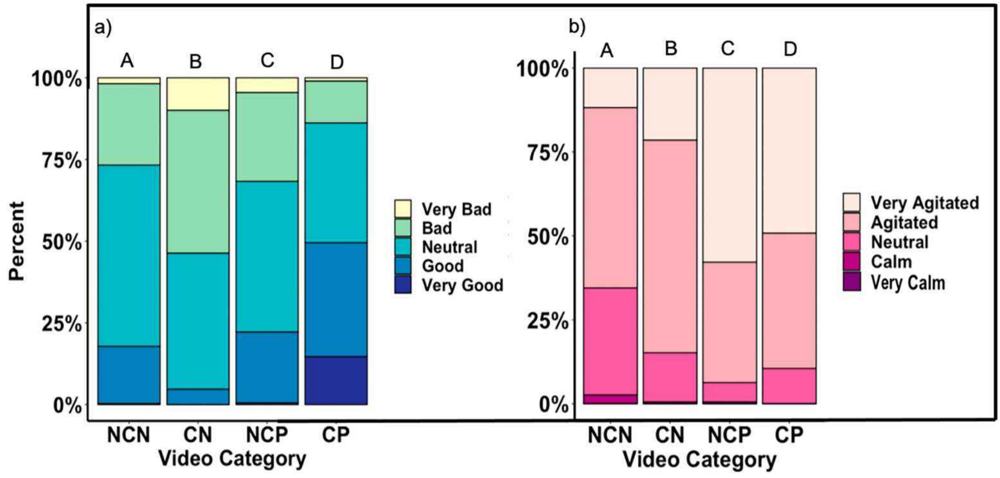


Figure 4. Free responses coded for valence (a) and arousal (b) in experiment 1. Different letters denote significant differences (Bonferroni-corrected Wilcoxon signed-rank tests).

70.98, $p < 0.001$). McNamar's post hoc pairwise comparisons adjusted with Bonferroni correction indicated significant differences between NCN vs. CP, NCP, and CN (Figure 5(a); online supplemental materials).

Free responses coded for mentalizing differed among all video categories ($\chi^2_{(3)} = 32.57$, $p < 0.001$). McNamar's post hoc pairwise comparisons adjusted with Bonferroni correction indicated significant differences between NCN vs. CP, NCP, and CN (Figure 5(b); online supplemental materials). Action-coded free responses differed among all video categories ($\chi^2_{(3)} = 203.91$, $p < 0.001$). McNamar's post hoc pairwise comparisons adjusted with Bonferroni correction indicated all videos were statistically different from one another except CN vs. CP (Figure 5(c); online supplemental materials).

Effect of Demographics on Valence and Arousal Scores

Age class had a significant main effect on valence response, averaging over gender and ethnicity ($F_{(1,382)} = 4.84$, $p < 0.05$, $\eta^2_p = 0.013$). Respondents 18–21 years of age gave significantly higher valence scores overall than those 22 years or higher. There was no significant main effect of gender or ethnicity, nor were there any significant interactions among the variables. Age class also had a significant effect on arousal responses ($F_{(1,382)} = 4.62$, $p < 0.05$, $\eta^2_p = 0.013$), with those 18–21 years of age reporting higher arousal scores than those 22 years of age or older. There were no significant effects of the gender or ethnicity variables, nor any interactions, on the arousal responses.

Effect of Dog Experience on Valence and Arousal Scores

Dog experience significantly impacted averaged valence scores ($F_{(4,381)} = 5.58$, $p < 0.05$, $\eta^2_p = 0.056$). Tukey's post hoc test revealed that the "very or extremely familiar" group had significantly higher valence ratings than the "somewhat familiar" group. There were no significant effects of dog experience on averaged arousal scores.

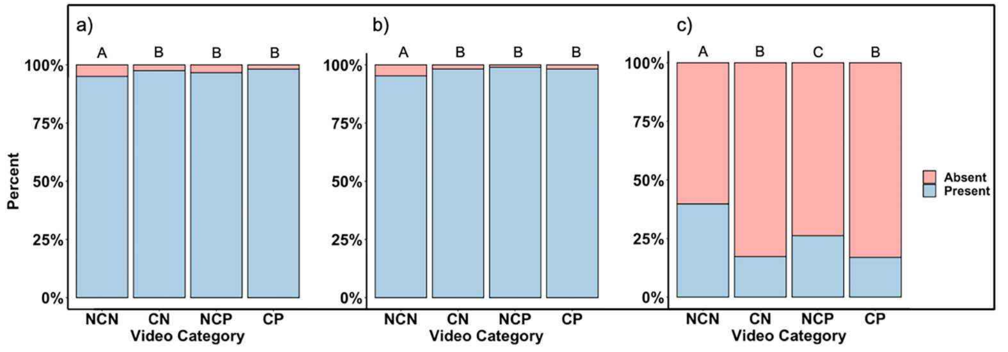


Figure 5. Free responses coded for anthropomorphism (a), mentalizing (b), and action (c) in experiment 1. Different letters denote significant differences for each graph (Bonferroni-corrected McNamar's post hoc pairwise comparisons).

Mood Pre- and Post-Survey

Higher pre-survey mood was predicative of higher valence scores ($F_{(1,376)} = 22.63$, $p < 0.05$, $R^2 = 0.057$; Figure 6) but not average arousal scores ($F_{(1,376)} = 0.05$, $p = 0.83$). Mood significantly increased from the beginning to the end of the survey ($t_{(363)} = -5.41$, $p < 0.001$).

Participants' Responses Regarding Perception of Dog Emotions

Participants' responses on how they perceived dog emotions in context-free videos differed significantly among choices ($\chi^2 = 246.30$, $p < 0.05$). Furthermore, their response proportions changed after viewing videos with context ($\chi^2 = 299.00$, $p < 0.05$). Participants initially relied on the dog's tail and behavior to perceive emotions, but when context was reintroduced, they used the context/background instead (Figure 7).

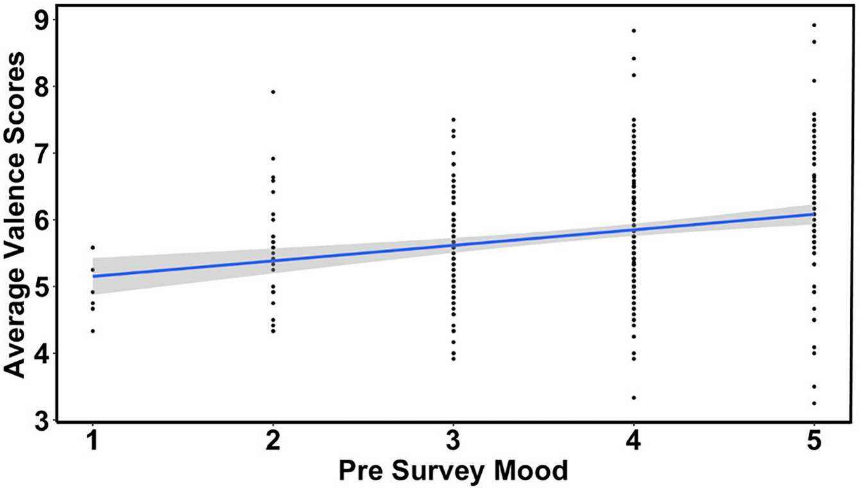


Figure 6. Mood effects on averaged valence scores in experiment 1. The blue line shows linear regression; the shaded area indicates standard error.

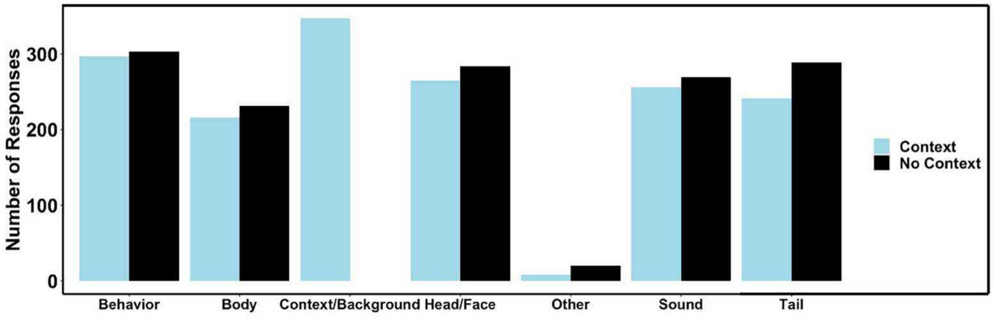


Figure 7. Frequencies of participant responses regarding what they reported relying on to perceive the dog's emotions. Black bars indicate the question was asked after first viewing the videos without context. Light blue bars indicate the question was asked after videos with context were presented with the addition of the option to select "context/background" as a possible cue to the dog's emotions.

Experiment 2: Mismatched Contexts

Prior studies in the human domain indicate how the situational context and its extraneous factors can influence perception. Thus, the aim of experiment 2 was to elucidate how perception would be altered if the situational context was misleading.

Methods

The methods were identical to those in experiment 1, except in the following details.

Filming and Video Editing

Different mismatched video components were merged to create video clips with the human and any stimuli on the left and the dog on the right side of the screen (Figure 8, online supplemental material). The positive situations used were toy, leash, and praise. The negative situations used were reprimand, vacuum, and tape measure. Again, the stimuli were selected based on both the owners' impressions of how the dog responded to different objects and scenarios. There were six video categories with three videos per category for a total of 18 videos: original context with the human doing something positive and the dog receiving something positive (Original +/+), original context with the human doing something negative and the dog receiving something negative (Original −/−), mismatched context with the human doing something negative and the dog receiving something positive (Mismatched −/+), mismatched context with the human doing something positive and the dog receiving something negative (Mismatched +/−), swapped context with the human doing something positive and the dog receiving something positive (Swapped +/+), and swapped context with the human doing something negative and the dog receiving something negative (Swapped −/−).

Survey

After the demographic questions, participants were presented with the 18 videos and asked to rate valence and arousal as in experiment 1. There was an attentional check

Figure 8. Two screenshots from two videos used in experiment 2: mismatched contexts. Panel (a) shows a mismatched video, with the vacuum as a negative stimulus, but the dog is actually responding to seeing a leash. Panel (b) shows the original leash content video, with the human showing a leash and the dog reacting to that positive stimulus. Thus, the right side of both panels is the same video of the dog. However, the context changes on the left side. All videos, whether original content or mismatched, had this black background with the human and stimuli on the left side and the dog on the right side.

question halfway through. They were also asked the same free response questions. Order was randomized subject to the constraint that the same human actions and stimuli were never shown successively.

Participants

The survey was sent out to ASU psychology undergraduates through the Sona Systems (https://www.sona-systems.com/) online research credit website as for experiment 1. Five hundred responses were requested. Participants from experiment 1 were ruled ineligible to participate in experiment 2. ASU's Institutional Review Board for human subjects approved this research project on March 29, 2022, and the survey ran from April 1, 2022 until April 29, 2022.

Free Response Coding

The Cronbach's alpha, with the same coders as experiment 1, was high for each of the subcodes (valence, $\alpha = 0.987$; arousal, $\alpha = 0.976$; anthropomorphism, $\alpha = 0.734$; mentalizing, $\alpha = 0.935$; action, $\alpha = 0.900$).

Results

Response Inclusion Criteria and Demographics

Out of 513 recorded responses, 485 were included for analysis after removing inaccurately answered, rapid, inconsistent, or repeated responses. For analysis, age was combined into two groups: 18–21 and 22 years and over. Gender was combined into three groups (male, female, other). Ethnicity was combined into five groups (White, Hispanic, Black, Asian, Other). Finally, dog experience was combined into four groups (None/Somewhat, Familiar, Familiar Currently, and Very/Extremely Familiar; online supplemental materials).

Effect of Context on Valence and Arousal Scores

There were significant differences in valence responses between the video categories ($F_{(5,480)} = 269.48$, $p < 0.01$, $\eta_p^2 = 0.73$). Post hoc contrasts revealed significant differences between all video categories, with the exception of Mismatched +/- and Swapped +/+ (Figure 9).

There was also a significant difference between all video categories in terms of arousal ($F_{(5,480)} = 107.73$, $p < 0.01$, $\eta_p^2 = 0.53$). Post hoc contrasts revealed significant differences between all video categories except Original −/− and Swapped −/− (Figure 10).

Free Responses

Over the 18 videos there were 8,730 free responses (485 per video). Valence-coded free responses differed among all video categories ($\chi^2_{(5)} = 2143.36$, $p < 0.001$.) Wilcoxon signed-rank tests were significant for all pairwise comparisons (online supplemental materials). Original +/+ had the highest valence coded in the free responses, followed by Mismatched +/−, Swapped +/+, Swapped −/−, Original −/−, Mismatched −/+ (Figure 11(a)). Effect sizes ranged widely but Original +/+ and Mismatched −/+ had an effect size of −1, meaning every single respondent indicated a higher level of valence in their free response for Original +/+ videos compared with Mismatched -/+ videos, even though in both cases the dog had been recorded in a positive situation.

Arousal-coded free responses differed among all video categories ($\chi^2_{(5)} = 2098.54$, $p < 0.001$). Wilcoxon signed-rank tests were significant for all pairs (online supplemental materials). Original +/+ had the highest arousal coded in the free responses, followed by Mismatched −/+, Swapped +/+, Mismatched +/−, Swapped −/−, Original −/− (Figure 11(b)).

Anthropomorphism-coded free responses were different among all video categories ($\chi^2_{(5)} = 69.07$, $p < 0.001$). McNamar's post hoc pairwise comparisons adjusted with Bonferroni correction indicated some videos had significantly different responses. Original +/+ had the highest average response, followed by Swapped +/+, Mismatched +/−, Original −/−, Mismatched −/+, Swapped −/− (Figure 12(a); online supplemental materials).

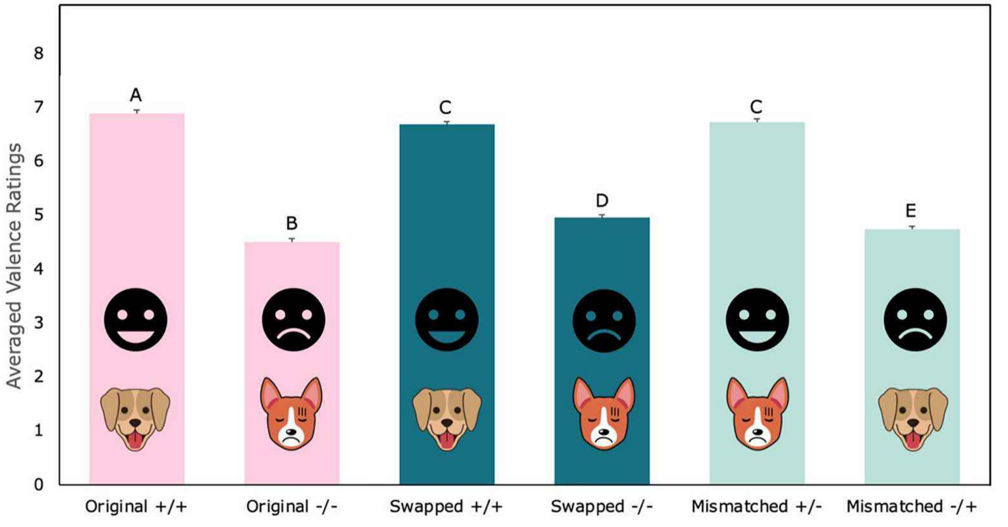


Figure 9. Averaged valence scores across the six different video categories for experiment 2: mismatched contexts. Error bars show standard error; different letters denote significant differences. Smiling and frowning human faces indicate the positive or negative situational context, while smiling and frowning cartoon dog faces indicate the situation the dog was actually in during the recording.

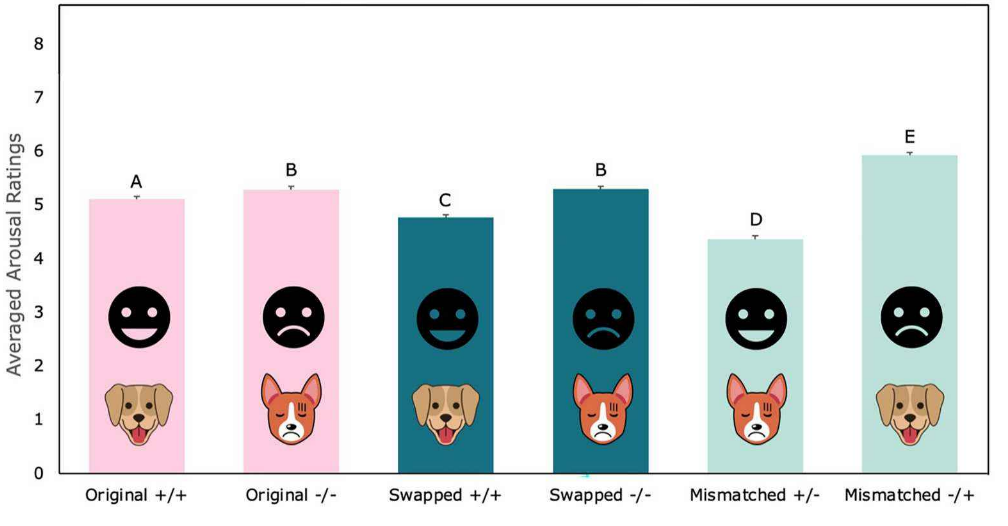


Figure 10. Averaged arousal ratings across video categories in experiment 2. Other details are the same as in Figure 9.

Mentalizing-coded free responses were different among all video categories ($\chi^2_{(5)} = 41.49$, $p < 0.001$). McNamar's post hoc pairwise comparisons adjusted with Bonferroni correction indicated some videos had significantly different responses: Original +/+ had the

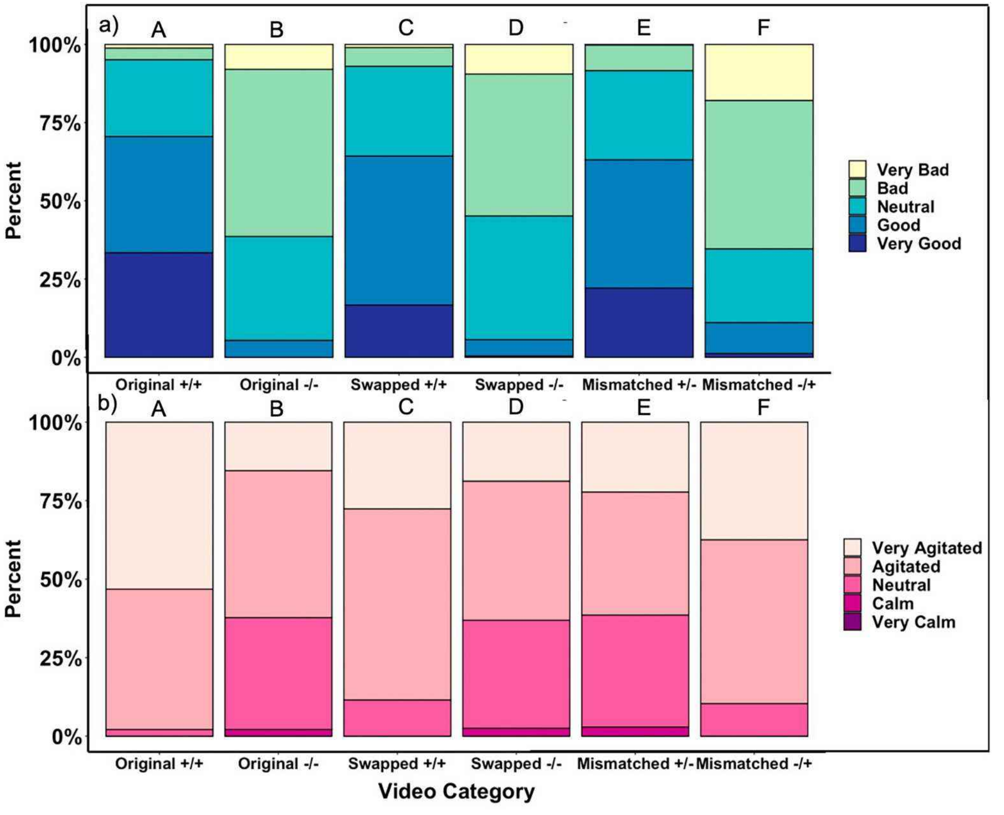


Figure 11. Free responses coded for valence (a) and arousal (b) for experiment 2: mismatched contexts. Different letters denote significant differences for each graph.

highest, followed by Swapped +/+, Mismatched −/+, Mismatched +/−, Original −/−, Swapped −/− (Figure 12(b); online supplemental materials).

Action-coded free responses were different among all video categories ($\chi^2_{(5)} = 166.06$, $p < 0.001$). Bonferroni-corrected McNamar's post hoc pairwise comparisons indicated some videos had significantly different responses: Mismatched −/+ had the highest level, followed by Swapped −/−, Swapped +/+, Mismatched +/−, Original +/+, Original −/− (Figure 12(c); online supplemental materials).

Effect of Demographics on Valence and Arousal Scores

Participants aged 22 years and older had higher valence ratings than those 18–21 ($F_{(1,484)} = 18.21$, $p < 0.01$, $\eta^2_p = 0.038$). Neither ethnicity nor gender had any significant effects, nor were there any significant interactions between the three variables on valence scores.

There were no main effects of age class, gender, or ethnicity on arousal but there was a significant interaction between ethnicity and age class ($F_{(4,484)} = 3.70$, $p = 0.006$). Post hoc tests indicated that African Americans over 22 years of age were an outlier, with significantly higher arousal scores. There were no other significant interactions among the variables on arousal scores.

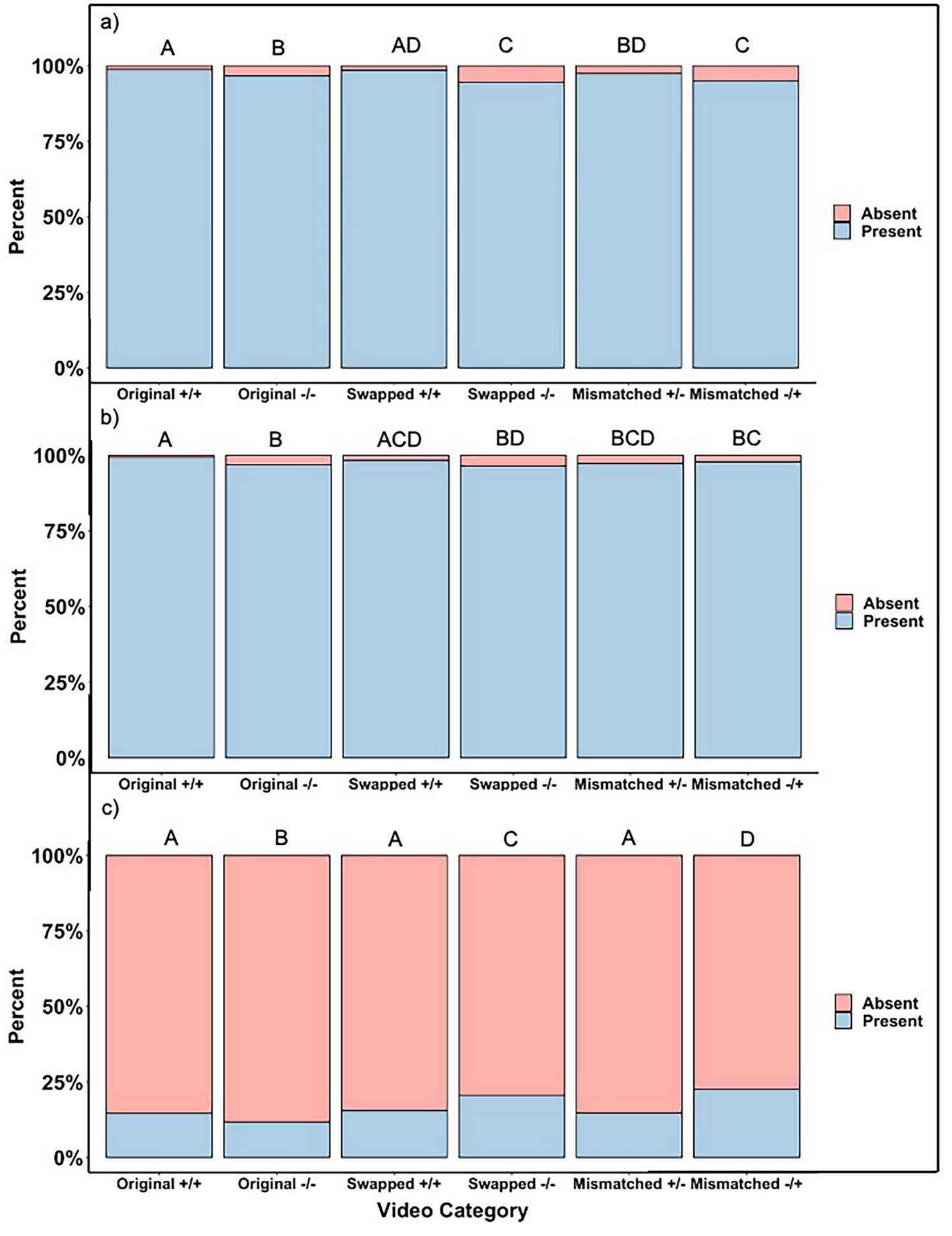


Figure 12. Free responses coded for anthropomorphism (a), mentalizing (b), and action (c) for experiment 2. Other details are the same as Figure 11.

Effect of Dog Experience on Valence and Arousal Scores

Dog experience had a significant effect on averaged valence scores ($F_{(3,484)} = 5.89$, $p < 0.01$, $\eta_p^2 = 0.035$). Tukey's post hoc tests revealed significant differences between those very/extremely familiar with dogs and those with no dog experience and those familiar currently. The "very/extremely familiar with dogs" group reported higher levels of

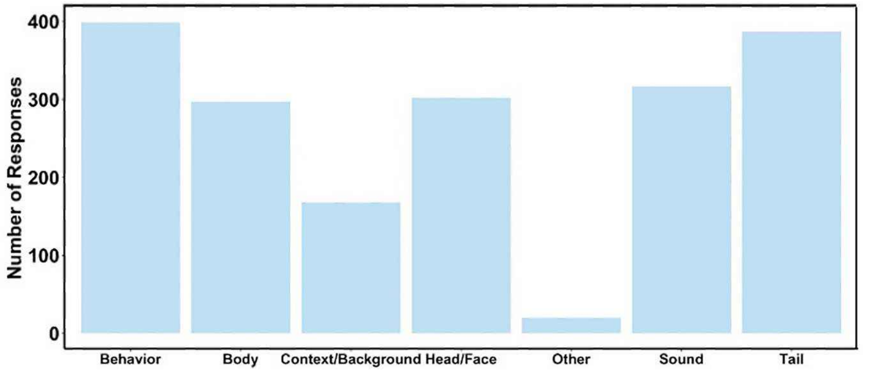


Figure 13. Counts of what participants reported they were using to perceive and understand the emotion of the dog in the videos from experiment 2. Options were not all chosen equally ($\chi^2 = 397.02$, $p < 0.01$).

valence then the two other groups. There was no difference in averaged arousal scores based on dog experience.

Mood Pre- and Post-Survey

Pre-survey mood did not significantly affect valence or arousal scores ($F_{(1,484)} = 1.40$; $F_{(1,484)} = 0.40$), and mood did not change significantly from the beginning to the end of the survey ($t_{(484)} = -1.68$, $p = 0.10$).

Participants' Responses on How They Were Perceiving Dog Emotions

Participants' responses regarding what they thought they were using to perceive the dog's emotions were not all chosen equally ($\chi^2 = 397.02$, $p < 0.01$), with the most reported being the dog's overall behavior and tail. Only 10% reported using context and background (Figure 13).

General Discussion

Two experiments demonstrated that people's perception of dog emotions is influenced by factors beyond the dog itself. In experiment 1, videos of a dog in putatively positive and negative scenarios significantly impacted valence scores as expected, confirming appropriate labeling. When context was removed, participants' valence ratings did not distinguish between positive and negative scenarios. When rating the dog's arousal, there was no difference between the scenarios when there was context present, but when the context was removed, positive videos of the dog were rated as more aroused. Free responses coded for valence and arousal supported scaled responses, revealing high levels of anthropomorphism and mentalizing regardless of video type.

Experiment 2 showed that people's perception of a dog's emotional state in a video was influenced more by environmental and situational context than by the dog's behavior itself. This was particularly true for valence: in all videos where the context was positive, participants rated the dog's valence as also positive, even if the dog was actually in a negative situation. Arousal was also affected by situational context, with positive contexts leading to lower arousal regardless of the dog's actions. Free responses coded for valence also indicated that situational context was impactful. However, participants' arousal ratings were uncorrelated with valence ratings in experiment 2, with higher arousal reported in the dog's positive scenarios and lower arousal in the dog's negative scenarios. Participants also showed high levels of anthropomorphism and mentalizing, with slightly less anthropomorphism in negative contexts.

In our experiments, the background context included an emoting human actor, the stimuli being used, as well as floor and wall coverings. While further experiments are needed to pinpoint exactly what participants were using to infer the dog's emotions, some conclusions can be drawn by comparison of certain scenarios in the current study. All conditions included floor and wall coverings as well as essential stimuli such as the tape measure and vacuum cleaner. However, there were some conditions that did not include the human actor. Thus, the impact of the human can be explored by comparing this condition with the others which included the actor.

In one scenario in experiment 2, the dog was shown alongside a vacuum cleaner without the human actor being visible. The vacuum cleaner was also shown to participants spliced together with the dog's reactions to being shown his leash (again, without the human actor visible). The mean valence score for the original vacuum cleaner video (one of the videos in the Original −/− condition) was 3.83 ($SE = 0.086$). The mean valence score for the original leash video (Original +/+ condition) was 7.57 ($SE = 0.095$). As can be seen in Figure 9, these are typical levels of response to the Original −/− and Original +/+ conditions. When participants were shown the vacuum cleaner (without the human) alongside the dog reacting to seeing his leash (Mismatched −/+ condition), their mean valence was 4.31 ($SE = 0.094$). This value is typical of the entire Mismatched −/+ category (Figure 9). This implies that participants viewing these videos were rating the dog's emotional valence using other aspects of the context than the human actor's behavior.

Our finding that people rely heavily on extraneous information around an individual dog when forming an impression of the animal's emotional state is inconsistent with *Basic and Discrete Emotional Theory's* contention that emotions are universal and can be recognized across species based on fixed physiological or behavioral markers (e.g., Anderson & Adolphs, 2014; Tomkins, 1962). The present results are more consistent with an approach to animal emotion derived from the *Constructed Theory of Emotion* (CTE: Barrett, 2017; Barrett & Russell, 2014; Gendron et al., 2020). CTE argues that emotional experience and perception derive from multiple factors including context, culture and past experiences. CTE also predicts that prior mood should influence emotion perception (Barrett, 2017), but we observed only small and contradictory effects in this study. Future studies should be designed to explicitly manipulate prior mood in order to identify whether more robust effects can be observed.

Our finding of over 90% anthropomorphized responses aligns with literature showing that anthropomorphism is common when attributing emotions to dogs and other animals. Mentalizing often occurs alongside anthropomorphism (Bahlig-Pieren & Turner, 1999; Gomez-Melara et al., 2021; Konok et al., 2015). One approach to account for these findings is the SEEK theory of anthropomorphism (Epley et al., 2007). According to this theory, complexity and understanding influence the tendency toward anthropomorphism. If the positive situations were more difficult to understand, this would result, according to SEEK theory, in more anthropomorphism, while if the negative situations were easier to understand, this would result in less anthropomorphism. Future studies should manipulate scenario difficulty to test this hypothesis.

Individuals who reported very high levels of familiarity with dogs perceived the negative scenarios as less negative and thus reported higher valence overall. While familiarity generally increases expertise (Krueger, 1975), this may not be the case for dog emotions. Some prior studies suggest experience with dogs may positively affect the perception of dog emotions (Bloom & Friedman, 2013; Wan et al., 2012); however, others report the opposite (Pongrácz et al., 2006; Tami & Gallagher, 2009). Our results showcase that owners more familiar with dogs could actually be missing negative cues from dogs. However, it could also be that familiarity with dogs leads one to ascribe more positive emotions onto every dog one sees. Clearly this issue requires further investigation to identify the confounding variables that lead to such diverse findings, especially if from an applied welfare perspective experience with dogs does lead to owners missing more negative emotional cues from their pets.

Limitations

There were several limitations to these experiments. One limitation is that, due to ethical constraints, the negative-emotion-inducing stimuli were relatively mild. People's perception of stronger negative dog emotions might be less influenced by contextual factors. Future research could also include a more diverse participant sample and dogs of different morphologies and temperaments.

Another limitation is the potential for multiple interpretations of the videos as they include various stimuli that could influence participants' ratings. Future studies should improve stimulus materials by isolating key contextual elements, such as the human face or background factors, that may affect perception. We recognize that our videos, with their rich and varied environments, may contain conflicting elements (e.g., a smiling person holding a negative stimulus like a cat). While more control during filming would have reduced these issues, it could also introduce artificiality. Our goal was to capture real-world effects, but we encourage future research to adopt a more controlled approach to identify which contextual factors most influence perceptions of dog emotions.

Another limitation involves the scales used. Post hoc analyses revealed that the valence and arousal scores of both experiment 1 and experiment 2 were very slightly but significantly correlated with one another (Experiment 1: $R^2 = 0.034$, $p < 0.001$, Experiment 2: $R^2 = 0.039$, $p < 0.001$). We believe that these slightly (around 3% of shared variance) correlated scales still form a superior method to asking participants to choose "human emotion" words from a list (i.e., happy, sad, etc.). Given that research indicates that perceived emotional intensity can be affected by the connotations of the terms used to anchor valence and arousal scales (Estes & Adelman, 2008; Kuperman et al., 2014), we understand that the words that anchored the scales we used (e.g., calm, agitated) might have influenced participants' responses. Furthermore, demographic factors may affect the interpretation of terms like "valence" and "arousal" (Mohammad, 2018). However, the observed correspondence between the scaled and free responses for valence and arousal indicates that the participants did indeed understand how to use the scales.

Conclusions

These results have important implications for animal welfare. Our findings suggest that people often make errors in judging a dog's emotional state: they are influenced by extraneous factors such as the context in which the dog is seen or the presence of other stimuli. These errors can have significant welfare implications, potentially causing us to overlook or misinterpret a dog's actual needs. Conversely, when people do make correct judgments, particularly when they align with the dog's true emotional state, it highlights a potential for better-informed welfare decisions. However, the challenge lies in discerning when our interpretations are correct and when they are clouded by human biases. By acknowledging these biases and striving to understand dog emotions through their specific behavioral and physiological needs, rather than through a human emotional lens, we can improve welfare standards.

Acknowledgements

We thank Rich Molinaro for his excellent acting, Oliver the dog for his emotionally expressive behaviors, and Saffron the cat for being a negative stimulus. Both Oliver's and Saffron's memory will live on in this research. Thanks to Holly Miller at General Mills for continued support with this project. Thanks go to dogs Bailey, Velvet, and Xephos for aiding filming enhancements, and special appreciation to Beatriz Algibez Flores, Nicole Fabela, and Alyssa Samson for their diligent free response coding.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

Amici, F., Waterman, J., Kellermann, C. M., Karimullah, K., & Bräuer, J. (2019). The ability to recognize dog emotions depends on the cultural milieu in which we grow up. *Scientific Reports*, *9*(1), 16414. https://doi.org/10.1038/s41598-019-52938-4

Anderson, D. J., & Adolphs, R. (2014). A framework for studying emotions across species. *Cell*, *157*(1), 187–200. https://doi.org/10.1016/j.cell.2014.03.003

Arahori, M., Kuroshima, H., Hori, Y., Takagi, S., Chijiiwa, H., & Fujita, K. (2017). Owners' view of their pets' emotions, intellect, and mutual relationship: Cats and dogs compared. *Behavioural Processes*, *141*, 316–321. https://doi.org/10.1016/j.beproc.2017.02.007

Aviezer, H., Hassin, R. R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M., & Bentin, S. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*(7), 724–732. https://doi.org/10.1111/j.1467-9280.2008.02148.x

Bahlig-Pieren, Z., & Turner, D. C. (1999). Anthropomorphic interpretations and ethological descriptions of dog and cat behavior by lay people. *Anthrozoös*, *12*(4), 205–210. https://doi.org/10.2752/089279399787000075

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Macmillan.
Barrett, L. F., & Russell, J. A. (Eds.). (2015). *The psychological construction of emotion.* The Guilford Press.
Bloom, T., & Friedman, H. (2013). Classifying dogs' (*Canis familiaris*) facial expressions from photographs. *Behavioural Processes*, *96*, 1–10. https://doi.org/10.1016/j.beproc.2013.02.010
Burza, L. B., Bloom, T., Trindade, P. H. E., Friedman, H., & Otta, E. (2022). Reading emotions in dogs' eyes and dogs' faces. *Behavioural Processes*, *202*, 104752. https://doi.org/10.1016/j.beproc.2022.104752
Caeiro, C., Guo, K., & Mills, D. (2017). Dogs and humans respond to emotionally competent stimuli by producing different facial actions. *Scientific Reports*, *7*(1), 15525. https://doi.org/10.1038/s41598-017-15091-4
Calbi, M., Heimann, K., Barratt, D., Siri, F., Umiltà, M. A., & Gallese, V. (2017). How context influences our perception of emotional faces: A behavioral study on the Kuleshov Effect. *Frontiers in Psychology*, *8*, 1684. https://doi.org/10.3389/fpsyg.2017.01684
Correia-Caeiro, C., Guo, K., & Mills, D. (2021). Bodily emotional expressions are a primary source of information for dogs, but not for humans. *Animal Cognition*, *24*(2), 267–279. https://doi.org/10.1007/s10071-021-01471-x
Correia-Caeiro, C., Guo, K., & Mills, D. S. (2023). Visual perception of emotion cues in dogs: A critical review of methodologies. *Animal Cognition*, *26*(3), 727–754. https://doi.org/10.1007/s10071-023-01762-5
de Gelder, B., Meeren, H. K., Righart, R., van den Stock, J., van de Riet, W. A., & Tamietto, M. (2006). Chapter 3. Beyond the face: Exploring rapid influences of context on face processing. *Progress in Brain Research*, *155 B*, 37–48. https://doi.org/10.1016/S0079-6123(06)55003-4
de Vere, A. J., & Kuczaj, S. A. (2016). Where are we in the study of animal emotions? *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(5), 354–362. https://doi.org/10.1002/wcs.1399
Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864
Estes, Z., & Adelman, J. S. (2008). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion*, *8*(4), 441–444. https://doi.org/10.1037/1528-3542.8.4.441
Gendron, M., Hoemann, K., Crittenden, A. N., Mangola, S. M., Ruark, G. A., & Barrett, L. F. (2020). Emotion perception in Hadza hunter-gatherers. *Scientific Reports*, *10*(1), 3867. https://doi.org/10.1038/s41598-020-60257-2
Gomez-Melara, J. L., Acosta-Naranjo, R., Izar, P., Sah, S. A. M., Pladevall, J., Maulany, R. I., Ngakan, P. O., Majolo, B., Romero, T., & Amici, F. (2021). A cross-cultural comparison of the link between modernization, anthropomorphism and attitude to wildlife. *Sustainability*, *13*(23), 13095. https://doi.org/10.3390/su132313095
Hantke, S., Cummins, N., & Schuller, B. (2018). What is my dog trying to tell me? The automatic recognition of the context and perceived emotion of dog barks. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5134–5138). IEEE. https://doi.org/10.1109/ICASSP.2018.8461757
Johnson, D. R., & Creech, J. C. (1983). Ordinal measures in multiple indicator models: A simulation study of categorization error. *American Sociological Review*, *48*(3), 398–407. https://doi.org/10.2307/2095231
Konok, V., Nagy, K., & Miklósi, Á. (2015). How do humans represent the emotions of dogs? The resemblance between the human representation of the canine and the human affective space. *Applied Animal Behaviour Science*, *162*, 37–46. https://doi.org/10.1016/j.applanim.2014.11.003
Kremer, L., Klein Holkenborg, S. E. J., Reimert, I., Bolhuis, J. E., & Webb, L. E. (2020). The nuts and bolts of animal emotion. *Neuroscience and Biobehavioral Reviews*, *113*, 273–286. https://doi.org/10.1016/j.neubiorev.2020.01.028
Krueger, L. E. (1975). Familiarity effects in visual information processing. *Psychological Bulletin*, *82*(6), 949–974. https://doi.org/10.1037/0033-2909.82.6.949

Kujala, M. V., Somppi, S., Jokela, M., Vainio, O., & Parkkonen, L. (2017). Human empathy, personality and experience affect the emotion ratings of dog and human facial expressions. *PLoS ONE*, *12*(1), e0170730. https://doi.org/10.1371/journal.pone.0170730

Kuperman, V., Estes, Z., Brysbaert, M., & Warriner, A. B. (2014). Emotion and language: Valence and arousal affect word recognition. *Journal of Experimental Psychology: General*, *143*(3), 1065–1081. https://doi.org/10.1037/a0035669

Lakestani, N. N. (2007). Children and dogs: How do children interpret dog behaviour. *GSAVA Annual Congress*. Berlin.

Lakestani, N. N., Donaldson, M. L., & Waran, N. (2014). Interpretation of dog behavior by children and young adults. *Anthrozoös*, *27*(1), 65–80. https://doi.org/10.2752/175303714X13837396326413

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, *73*(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004

Martens, P., Enders-Slegers, M. J., & Walker, J. K. (2016). The emotional lives of companion animals: Attachment and subjective claims by owners of cats and dogs. *Anthrozoös*, *29*(1), 73–88. https://doi.org/10.1080/08927936.2015.1075299

Meeren, H. K. M., Van Heijnsbergen, C. C. R. J., & De Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16518–16523. https://doi.org/10.1073/pnas.0507650102

Mendl, M., Burman, O. H. P., & Paul, E. S. (2010). An integrative and functional framework for the study of animal emotion and mood. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1696), 2895–2904. https://doi.org/10.1098/rspb.2010.0303

Mohammad, S. M. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In I. Gurevych & M. Yusuke (Eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 174–184). Association for Computational Linguistics. https://doi.org/10.18653/v1/p18-1017

Morris, P., Doe, C., & Godsell, E. (2008). Secondary emotions in non-primate species? Behavioural reports and subjective claims by animal owners. *Cognition and Emotion*, *22*(1), 3–20. https://doi.org/10.1080/02699930701273716

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632. https://doi.org/10.1007/s10459-010-9222-y

Park, J., & Kim, A. (2020). A dog doesn't smile: Effects of a dog's facial expressions and gaze on pet product evaluation. *Journal of Product and Brand Management*, *30*(5), 641–655. https://doi.org/10.1108/JPBM-04-2019-2335

Pongrácz, P., Molnár, C., & Miklósi, Á. (2006). Acoustic parameters of dog barks carry emotional information for humans. *Applied Animal Behaviour Science*, *100*(3–4), 228–240. https://doi.org/10.1016/j.applanim.2005.12.004

Richards, K. A. R., & Hemphill, M. A. (2018). A practical guide to collaborative qualitative data analysis. *Journal of Teaching in Physical Education*, *37*(2), 225–231. https://doi.org/10.1123/jtpe.2017-0084

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Schirmer, A., Seow, C. S., & Penney, T. B. (2013). Humans process dog and human facial affect in similar ways. *PLoS ONE*, *8*(9), e74591. https://doi.org/10.1371/journal.pone.0074591

Sullivan, G. M., & Artino, A. R. (2013). Analyzing and interpreting data from Likert-type scales. *Journal of Graduate Medical Education*, *5*(4), 541–542. https://doi.org/10.4300/jgme-5-4-18

Szánthó, F., Miklósi, Á, & Kubinyi, E. (2017). Is your dog empathic? Developing a dog emotional reactivity survey. *PLoS ONE*, *12*(2), e0170397. https://doi.org/10.1371/journal.pone.0170397

Tami, G., & Gallagher, A. (2009). Description of the behaviour of domestic dog (*Canis familiaris*) by experienced and inexperienced people. *Applied Animal Behaviour Science*, *120*(3–4), 159–169. https://doi.org/10.1016/j.applanim.2009.06.009

Tomkins, S. S. (1962). *Affect imagery consciousness: Volume I: The positive affects*. Springer.

Trevathan-Minnis, M., Bloom, T., Atlas, N., Macdonald, D. A., & Friedman, H. L. (2021). Using dog facial-emotion recognizability to explore the putative black dog syndrome. *The Humanistic Psychologist*, *49*(4), 616–629. https://doi.org/10.1037/hum0000252

Walker, J., Dale, A., Waran, N., Clarke, N., Farnworth, M., & Wemelsfelder, F. (2010). The assessment of emotional expression in dogs using a free choice profiling methodology. *Animal Welfare*, *19*(1), 75–84. https://doi.org/10.1017/S0962728600001196

Wan, M., Bolger, N., & Champagne, F. A. (2012). Human perception of fear in dogs varies according to experience with dogs. *PLoS ONE*, *7*(12), e51775. https://doi.org/10.1371/journal.pone.0051775

<u>Supplementary Materials</u>

Demographics for Experiment 1 and 2

		Experiment 1	Experiment 2
Age	18-21 years 22 years +	95% 5%	95% 5%
Gender			
	Female	57%	43%
	Male	41%	56%
	Other	2%	1%
Ethnicity			
	Caucasian	50%	61%
	Asian	22%	16%
	Hispanic	17%	15%
	African American	6%	4%
	American Indian, Middle Eastern, Native Hawaiian, or Other	5%	4%
Dog Experience			
	Very/Extremely Familiar	49%	56%
	Familiar	20%	18%
	Somewhat Familiar	17%	13%
	Familiar Currently	12%	11%
	None	3%	1%

Note. For dog experience, very/extremely familiar means that either their daily job involved working with dogs, or they have always lived with pet dogs. Familiar means they had a dog before but did not have one now while 'familiar currently' means they own a dog now but have not lived with a dog in the past. Somewhat familiar means they never had a dog as a pet. None means that they never had a dog as a pet in addition to having no experience with dogs whatsoever.

Tables

Valence coded free response results from Experiment 1: context vs no context.

Comparison	Valence Coded Free Response Post Hoc Results		
	Wilcoxon Signed Rank Test Z Score	Wilcoxon Signed Rank Test P Value	Matched Pairs Rank Biserial Correlation (Effect Size)
CN - NCN	-17.07	< 0.001*	-1.00
NCP - NCN	-4.70	< 0.001*	-0.357
NCN – CP	-16.83	< 0.001*	-0.999
NCP - CN	-16.99	< 0.001*	1.00
CP – CN	-16.98	0.001*	1.00
CP - NCP	-17.02	< 0.001*	1.00

Note. Asterisks (*) indicate significant pairwise differences with a Bonferroni correction for multiple tests. CN = Context Negative, CP = Context Positive, NCN = No Context Negative, NCP = No Context Positive.

Arousal coded free response results from Experiment 1: context vs no context.

Comparison	Arousal Coded Free Response Post Hoc Results		
	Wilcoxon Signed Rank Test Z Score	Wilcoxon Signed Rank Test P Value	Matched Pairs Rank Biserial Correlation (Effect Size)
CN - NCN	-17.12	< 0.001*	0.999
NCP - NCN	-17.15	< 0.001*	0.999
NCN – CP	-17.05	< 0.001*	-1.00
NCP - CN	-17.02	< 0.001*	0.998
CP – CN	-16.49	< 0.001*	0.997
CP - NCP	-14.55	< 0.001*	-0.972

Note. Asterisks (*) indicate significant pairwise differences with a Bonferroni correction for multiple tests. CN = Context Negative, CP = Context Positive, NCN = No Context Negative, NCP = No Context Positive.

McNemar posthoc pairwise comparisons for anthropomorphism, mentalizing, action coded free response from Experiment 1: context vs no context.

	McNemar Test P Value Coded Free Response Post Hoc Results		
	Anthropomorphism	Mentalizing	Action
CN - NCN	< 0.001*	< 0.001*	< 0.001*
NCP - NCN	< 0.001*	< 0.001*	< 0.001*
NCN – CP	< 0.001*	< 0.001*	< 0.001*
NCP - CN	0.125	0.250	< 0.001*
CP – CN	0.250	1.00	1.00
CP - NCP	1.00	0.250	< 0.001*

Note. Asterisks (*) indicate significant pairwise differences with a Bonferroni correction for multiple tests. CN = Context Negative, CP = Context Positive, NCN = No Context Negative, NCP = No Context Positive.

Valence coded free response results from Experiment 2: mismatched contexts.

	Valence Coded Free Response Post Hoc Results		
	Wilcoxon Signed Rank Test Z Score	Wilcoxon Signed Rank Test P Value	Matched Pairs Rank Biserial Correlation (Effect Size)
OCNN - OCPP	-19.091	< 0.001*	-1.00
MCNP - OCPP	-19.124	< 0.001*	-1.00
MCPN - OCPP	-13.547	< 0.001*	-0.780
SWPP - OCPP	-16.34	< 0.001*	-0.943
SWNN - OCPP	-19.13	< 0.001*	-1.00
MCNP - OCNN	-7.96	< 0.001*	-0.045
MCPN - OCNN	-19.15	< 0.001*	1.00
SWPP - OCNN	-19.14	< 0.001*	1.00
SWNN – OCNN	-10.59	< 0.001*	0.717
MCPN - MCNP	-10.13	< 0.001*	1.00
SWPP - MCNP	-19.12	< 0.001*	1.00
SWNN - MCNP	-11.74	< 0.001*	0.668
SWPP - MCPN	-4.21	< 0.001*	-0.36
SWNN - MCPN	-19.15	< 0.001*	-1.00
SWNN - SWPP	-19.18	< 0.001*	-1.00

Note. Asterisks (*) indicate significant pairwise differences with a Bonferroni correction for multiple tests. OCNN = Original Context, Negative/Negative, OCPP = Original Context, Positive/Positive, MCNP = Mismatched Context, Negative/Positive, MCPN = Mismatched Context, Positive/Negative, SWPP = Swapped Context, Positive/Positive, SWNN = Swapped Context, Negative/Negative.

Arousal coded free response results from Experiment 2: mismatched contexts.

	Arousal Coded Free Response Post Hoc Results		
	Wilcoxon Signed Rank Test Z Score	Wilcoxon Signed Rank Test P Value	Matched Pairs Rank Biserial Correlation (Effect Size)
OCNN - OCPP	-18.99	< 0.001*	-1.00
MCNP - OCPP	-17.63	< 0.001*	-0.997
MCPN - OCPP	-18.56	< 0.001*	-0.998
SWPP - OCPP	-18.34	< 0.001*	-1.00
SWNN - OCPP	-18.71	< 0.001*	-1.00
MCNP - OCNN	-19.09	< 0.001*	1.00
MCPN - OCNN	-18.94	< 0.001*	0.752
SWPP - OCNN	-11.24	< 0.001*	1.00
SWNN – OCNN	-9.04	< 0.001*	0.679
MCPN - MCNP	-18.05	< 0.001*	-0.998
SWPP - MCNP	-14.19	< 0.001*	-1.00
SWNN - MCNP	-18.75	< 0.001*	-1.00
SWPP - MCPN	-17.06	< 0.001*	-0.989
SWNN - MCPN	-7.39	< 0.001*	0.586
SWNN - SWPP	-18.23	< 0.001*	-0.999

Note. Asterisks (*) indicate significant pairwise differences with a Bonferroni correction for multiple tests. OCNN = Original Context, Negative/Negative, OCPP = Original Context, Positive/Positive, MCNP = Mismatched Context, Negative/Positive, MCPN = Mismatched Context, Positive/Negative, SWPP = Swapped Context, Positive/Positive, SWNN = Swapped Context, Negative/Negative.

McNemar posthoc pairwise comparisons for anthropomorphism, mentalizing, action coded free response from Experiment 2: mismatched contexts.

	McNemar Test P Value Coded Free Response Post Hoc Results		
	Anthropomorphism	Mentalizing	Action
OCNN - OCPP	0.002*	< 0.001*	< 0.001*
MCNP - OCPP	< 0.001*	0.016	< 0.001*
MCPN - OCPP	0.031	0.002*	1.00
SWPP - OCPP	1.00	0.063	0.125
SWNN - OCPP	< 0.001*	< 0.001*	< 0.001*
MCNP – OCNN	0.008	0.063	< 0.001*
MCPN - OCNN	0.125	0.500	< 0.001*
SWPP - OCNN	0.004	0.016	< 0.001*
SWNN – OCNN	0.002*	0.500	< 0.001*
MCPN - MCNP	< 0.001*	0.250	< 0.001*
SWPP - MCNP	< 0.001*	0.500	< 0.001*
SWNN - MCNP	0.500	0.016	0.004
SWPP - MCPN	0.063	0.063	0.250
SWNN - MCPN	< 0.001*	0.125	< 0.001*
SWNN - SWPP	< 0.001*	0.004	< 0.001*

Note. Asterisks (*) indicate significant pairwise differences with a Bonferroni correction for multiple tests. OCNN = Original Context, Negative/Negative, OCPP = Original Context, Positive/Positive, MCNP = Mismatched Context, Negative/Positive, MCPN = Mismatched Context, Positive/Negative, SWPP = Swapped Context, Positive/Positive, SWNN = Swapped Context, Negative/Negative.

The 18 videos in the order they appeared in Experiment 2.

Video Number in Survey Order	**Human Situation**	**Positive or Negative Situation (Human Side)**	**Dog Situation**	**Positive or Negative Situation (Dog Side)**	**Original, Swapped, or Mismatched**	**Video Label**
1	Toy	Positive	Toy	Positive	Original	Original +/+
2	Tape Measure	Negative	Reprimand	Negative	Swapped	Swapped -/-
3	Leash	Positive	Leash	Positive	Original	Original +/+
4	Reprimand	Negative	Vacuum	Negative	Swapped	Swapped -/-
5	Praise	Positive	Toy	Positive	Swapped	Swapped +/+
6	Tape Measure	Negative	Praise	Positive	Mismatched	Mismatched -/+
7	Reprimand	Negative	Toy	Positive	Mismatched	Mismatched -/+
8	Vacuum	Negative	Vacuum	Negative	Original	Original -/-
9	Praise	Positive	Praise	Positive	Original	Original +/+
10	Reprimand	Negative	Reprimand	Negative	Original	Original -/-
11	Vacuum	Negative	Tape Measure	Negative	Swapped	Swapped -/-
12	Toy	Positive	Vacuum	Negative	Mismatched	Mismatched +/-
13	Leash	Positive	Tape Measure	Negative	Mismatched	Mismatched +/-
14	Praise	Positive	Reprimand	Negative	Mismatched	Mismatched +/-
15	Toy	Positive	Leash	Positive	Swapped	Swapped +/+
16	Tape Measure	Negative	Tape Measure	Negative	Original	Original -/-
17	Vacuum	Negative	Leash	Positive	Mismatched	Mismatched -/+
18	Leash	Positive	Praise	Positive	Swapped	Swapped +/+

Note. This table indicates whether the video has original, swapped or mismatched content, whether each side of the video represented a positive or negative situation, and the final label of each video.

Demographic Questions from Experiments 1 and 2

Demographics

What is your age?

- ○ 18-21
- ○ 22-25
- ○ 26-29
- ○ 30-40
- ○ 40-50
- ○ 50-60
- ○ 60+

What is your gender?

- ○ Female
- ○ Non-binary / third gender
- ○ Male
- ○ Prefer not to say

Which best describes you?

- ○ White (Eg: German, Irish, English, Italian, Polish, French, etc)
- ○ Hispanic, Latino or Spanish origin (Eg: Mexican or Mexican American, Puerto Rican, Cuban, Salvadoran, Dominican, Colombian, etc)
- ○ Black or African American (Eg: African American, Jamaican, Haitian, Nigerian, Ethiopian, Somalian, etc)
- ○ Asian (Eg: Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, etc)
- ○ American Indian or Alaska Native(Eg: Navajo nation, Blackfeet tribe, Mayan, Aztec, Native Village or Barrow Inupiat Traditional Government, Nome Eskimo Community, etc)

https://asu.co1.qualtrics.com/Q/EditSection/Blocks/Ajax/GetSurveyPrintPreview?ContextSurveyID=SV_2sm5khHOm8WZ1fU&ContextLibraryID=UR_71KOx4E... 2/18

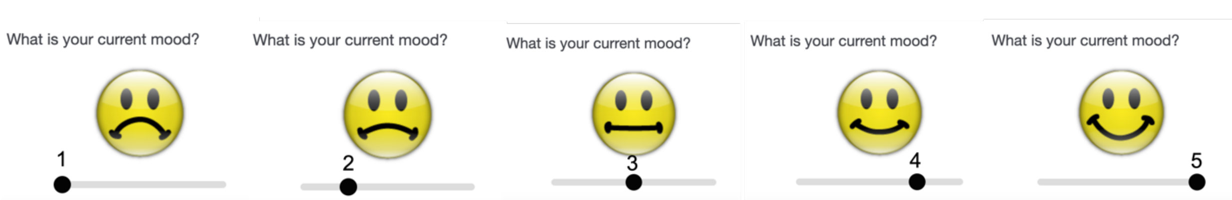
1/20/22, 1:12 PM Qualtrics Survey Software

- ○ Middle Eastern or North African (Eg: Lebanese, Iranian, Egyptian, Syrian, Moroccan, Algerian, etc)
- ○ Native Hawaiian or Other Pacific Islander (Eg: Native Hawaiian, Samoan, Chamorro, Tongan, Fijian, etc)
- ○ ________ One not listed (please specify)

What is your level of experience with dogs?

- ○ None
- ○ Somewhat familiar with dogs but never had one as a pet
- ○ Familiar with dogs in general, owned dog(s) previously but do not own one now
- ○ Familiar with dogs currently, own a dog now but have not lived with dog(s) in the past
- ○ Very familiar, always lived with a pet dog(s)
- ○ Extremely familiar, my daily job includes working with dogs

What is your current mood?



Note: Mood was measured on a sliding 5-point scale, using the face and different facial expressions as a marker for each point.

Free Response Coding Methodology

We based our free response coding methodology on Richards & Hemphill (2018)'s guidelines for qualitatively coding open-ended questions. While various methods exist, a collaborative approach, as employed here, provides the most structured, rigorous, and reliable means to analyze qualitative data like free responses (Braun & Clarke, 2006; Gendron et al., 2020).

The three coders for the two experiments were undergraduates in psychology at ASU. They interviewed by the first author for positions in the Canine Science Collaboratory. Prior experience in qualitative coding was not required, but applicants needed experience in psychological research and the ability to work well in a group setting. The selected undergraduates were unaware of the study questions or hypotheses. They were only informed that participants watched videos of dogs and provided free responses describing what the dog might be feeling.

The themes of valence, arousal, and anthropomorphism were chosen by the authors to determine whether the free responses presented a different narrative than the 1-10- scaled survey responses, and because previous research indicates that perceiving emotions in animals can be anthropomorphic. The themes of mental and action states were derived from Gendron et al. (2020)**,** as they offer an intriguing perspective on how the perception of emotion is influenced by the language we use.

Codebook For Analyzing Free Response Data

Valence (-2,-1,0,1,2)

Valence is the level of positivity or negativity, goodness, or badness. We used a scale of 2 to -2 to rate how good or bad a word's meaning was. Negative words fell below zero and were classified according to the intensity of negativity while positive words were classified above 0. Words that did not apply, had indifferent connotations, or did not imply good or bad intentions were rated as 0.

2 = Extremely positive

1 = Somewhat positive

0 = Indifferent, passive state for dog, not related to good or bad, could be good or bad depending on perspective, projecting "wanting," hunger

-1 = Somewhat negative

-2 = Extreme discomfort/negativity

When there are multiple words in the same answer rated on the same level, the rating for that response remains on that same level

1 (Playful) + **1** (Happy)= **1**

2 (Energetic) + **2** (Excited) = **2**

-2 (Agitated) + **-2** (Hurt) = **-2**

Responses that have opposite ratings cancel each other out to zero

-2 (Agitated) + **2** (Excited) = **0**

1 (Happy) + **-1** (Sad) = **0**

When there are multiple words in a response, add or subtract the levels for each word

1 (Playful) + **-1** (Annoyed) = **0**

1 (Happy) + **2** (Excited) = **2**

1 (Happy) + **-2** (Agitated) = **-1**

0 (Calm) + **2** (Excited) = **2**

0 (Tired) + **-1** (Annoyed) = **-1**

2 (Excited) + **-1** (Sad) = **1**

-1 (Sad) + **-1** (Anxious) + **1** (Happy) = **-1**

Some Examples

Excited: 2, Eager: 2, Happy: 1, Playful: 1, Aware: 0, Calm: 0, Sad: -1, Worried: -1, Agitated: -2, Angry: -2

Arousal (-2,-1,0,1,2)

Arousal is the degree of alertness or stimulation during the waking state. We used a scale of 2 to -2 to rate the degree of arousal. Words with low arousal were classified below 0 based on the intensity of activity and arousal. Words with high arousal were classified above 0 depending on the intensity of the activity.

2 = Vocal/talkative, high anxiety/agitation, highly active

1 = Wanting, trying, happy, playful, interested

0 = Irrelevant/not applicable, opposites, bare minimum

-1 = Relaxed but not lethargic

-2 = Tired, sleepy, very low arousal

When there are multiple words in the same answer rated on the same level, the rating for that response remains on that same level

1 (Happy) + **1** (Playful) = **1**

2 (Excited) + **2** (Agitated) = **2**

-1 (Calm) + **-1** (Relaxed) = **-2**

Responses that have opposite ratings cancel each other out to zero

1 (Curious) + **-1** (Calm) = **0**

2 (Excited) + **-2** (Tired) = **0**

When there are multiple words in a response, add or subtract the levels for each word

0 (Chillin) + **1** (Playing) = **1**

1 (Happy) + **1** (Playful) = **1**

0 (Obedient) + **1** (Impatient) = **1**

-2 (Tired) + **0** (Lonely)= **-2**

-1 (Calm) + **2** (Excited) = **1**

1 (Curious)+ **-1** (Calm) + **-1** (Relaxed) = **-1**

Some Examples

Angry: 2, Stressed: 2, Vocal: 2, Anxious: 1, Happy: 1, Wanting: 1, Bored: 0, Waiting: , Obedient: 0, Calm/Relaxed: -1, Lazy: -1, Tired: -2, Lethargic: -2

Anthropomorphism (0,1)

Anthropomorphism refers to the assignment of human characteristics to nonhuman objects or animals. Anthropomorphism encompasses the inference of emotional states from actions and the inference of thoughts. This includes responses such as *"tired," "happy,"* or *"sad."* An animal may feel positive or negative emotions but we have no way of knowing whether this is the same as happiness or sadness in humans. Additionally, an animal may lay down but we have no way to know if this is because it is truly tired.

0 = No anthropomorphic words included in response, observable, objective statements only

1 = At least one anthropomorphic word included in the response

Some Examples

Agitated: 0, Curious: 1, Observant: 0, Playful: 1, Restless: 0, Scared: 1

Mentalizing (0,1)

Mentalizing refers to the assignment of mental attributes or a state of mind to a dog. Mental state is something that the human is inferring that the dog might feel. An example of this is *Waiting* vs *Not Moving*. Both are somewhat describing the same thing but *Not Moving* is an observable trait that the dog is exhibiting whereas *Waiting* is inferred. *Waiting* implies the dog is actively thinking about doing or not doing something while Not Moving implies that it are physically not doing something. One must ask oneself can this trait be observed or is it trying to infer what they are feeling / doing.

1 = Mentalizing, referring to a mental or emotional state

0 = Not mentalizing, observable actions, covert behaviors, nothing in box at all

Some Examples

Chillin: 1, Confused: 1, Anxious: 1, Attentive: 1, Excited (Assume emotionally unless specified otherwise): 1, Happy: 1, Hungry: 1, On-edge: 1, Playful: 1, Waiting: , Uncomfortable/Comfortable: 1, Playing: 0, Hurt(assume physically hurt): 0, Talkative/Vocal/Expressive: 0, Aroused: 0, Agitated: 0, Alert: 0, Calm: 0, Describing any action: 0, On guard: 0

Action State (0,1)

An action state is when a physical and/ or observable action is performed by the dog itself. If something is being done to the dog or it is feeling an emotion, that does not qualify as an action state. For example, *playful* is an emotion that a dog could be feeling, but for the response to qualify as an observable action state the dog must be physically *playing.* If the response is inferring the dog's intentions behind an action, like being *protective* (subjective) vs. the dog simply being *agitated* (objective), it does not qualify as an action state. Any body movement or vocalization, like barking, are considered action states as well.

1 = Action state, physically observable action, vocalizing

0 = Not action state, blank, mental state

Some examples

Tired: 0 vs Sleeping: 1

Anticipating: 0 vs Waiting: 1

Hesitant: 0 vs Hesitating: 1

Playful: 0 vs Playing: 1

On edge: 0 vs On guard: 1

Tense: 1, Jumping: 1, Begging: 1, Energetic: 1, Responsive: 1, Alert: 1, Relaxed/Calm: 1, Chilling: 0, Wanting: 0, Scared: 0, Focused: 0, Motivated: 0

Top 10 Words in Free Response Answers from All Videos Across the Two Experiments

Experiment 1: Context vs No Context – Top 10 Words Used in Free Responses				
	Context		**No Context**	
Videos	*Negative Scenarios*			
Cat	1. Curious	6. Uncomfortable	1. Calm	6. Confused
	2. Confused	7. Dog	2. Curious	7. Relaxed
	3. Annoyed	8. Cat	3. Bored	8. Content
	4. Calm	9. Nervous	4. Sad	9. Something
	5. Interested	10. Scared	5. Dog	10. Tired
Tape Measure	1. Playful	6. Curious	1. Playful	6. Dog
	2. Annoyed	7. Excited	2. Excited	7. Something
	3. Confused	8. Alert	3. Alert	8. Seems
	4. Agitated	9. Dog	4. Agitated	9. Protective
	5. Scared	10. Happy	5. Happy	10. Curious
Reprimand	1. Confused	6. Guilty	1. Hungry	6. Dog
	2. Sad	7. Dog	2. Calm	7. Alert
	3. Curious	8. Calm	3. Curious	8. Patient
	4. Upset	9. Scared	4. Attentive	9. Interested
	5. Attentive	10. Alert	5. Focused	10. Waiting
	Positive Scenarios			
Play	1. Playful	6. Tired	1. Playful	6. Annoyed
	2. Happy	7. Agitated	2. Happy	7. Attention
	3. Excited	8. Dog	3. Excited	8. Bored
	4. Annoyed	9. Lazy	4. Agitated	9. Relaxed
	5. Confused	10. Play	5. Dog	10. Play
Praise	1. Playful	6. Agitated	1. Angry	6. Upset
	2. Happy	7. Angry	2. Agitated	7. Something
	3. Annoyed	8. Obedient	3. Alert	8. Annoyed
	4. Dog	9. Confused	4. Protective	9. Scared
	5. Excited	10. Owner	5. Dog	10. Attention
Treat	1. Hungry	6. Anxious	1. Excited	6. Waiting
	2. Excited	7. Treat	2. Hungry	7. Happy
	3. Happy	8. Dog	3. Curious	8. Dog
	4. Patient	9. Calm	4. Calm	9. Attentive
	5. Focused	10. Curious	5. Focused	10. Something

Note. Top 10 words for each video are ranked by frequency used. The number one word is the most frequently used word.

Experiment 2: Mismatched Contexts – Top 10 Words Used in Free Responses					
Original -/-		**Mismatched -/+**		**Mismatched -/-**	
Reprimand Original		*Reprimand and Toy*		*Reprimand and Vacuum*	
1. Confused	6. Obedient	1. Confused	6. Sad	1. Playful	6. Excited
2. Sad	7. Dog	2. Agitated	7. Dog	2. Angry	7. Dog
3. Upset	8. Guilty	3. Upset	8. Mad	3. Agitated	8. Mad
4. Calm	9. Attentive	4. Playful	9. Excited	4. Upset	9. Annoyed
5. Scared	10. Curious	5. Angry	10. Annoyed	5. Confused	10. Defensive
Vacuum Original		*Vacuum and Leash*		*Vacuum and Tape Measure*	
1. Scared	6. Angry	1. Agitated	6. Confused	1. Tired	6. Unbothered
2. Playful	7. Annoyed	2. Playful	7. Angry	2. Playful	7. Bored
3. Confused	8. Protective	3. Scared	8. Excited	3. Scared	8. Annoyed
4. Agitated	9. Dog	4. Annoyed	9. Upset	4. Confused	9. Agitated
5. Defensive	10. Threatened	5. Defensive	10. Dog	5. Calm	10. Dog
Tape Measure Original		*Tape Measure and Praise*		*Tape Measure and Reprimand*	
1. Tired	6. Uninterested	1. Confused	6. Relaxed	1. Confused	6. Focused
2. Playful	7. Annoyed	2. Calm	7. Focused	2. Calm	7. Nervous
3. Confused	8. Excited	3. Curious	8. Dog	3. Scared	8. Bored
4. Bored	9. Happy	4. Scared	9. Seems	4. Curious	9. Annoyed
5. Calm	10. Dog	5. Confusion	10. Nervous	5. Dog	10. Alert
Original +/+		**Mismatched +/-**		**Mismatched +/+**	
Toy Original		*Toy and Vacuum*		*Toy and Leash*	
1. Excited	6. Excitement	1. Playful	6. Anxious	1. Excited	6. Ready
2. Playful	7. Confused	2. Excited	7. Toy	2. Playful	7. Dog
3. Happy	8. Anxious	3. Happy	8. Eager	3. Happy	8. Playful
4. Frustrated	9. Toy	4. Agitated	9. Dog	4. Agitated	9. Toy
5. Dog	10. Curious	5. Frustrated	10. Play	5. Energetic	10. Angry
Leash Original		*Leash and Tape Measure*		*Leash and Praise*	
1. Excited	6. Anxious	1. Tired	6. Dog	1. Confused	6. Uninterested
2. Happy	7. Ready	2. Excited	7. Bored	2. Calm	7. Intrigued
3. Playful	8. Dog	3. Playful	8. Calm	3. Curious	8. Bored
4. Eager	9. Go	4. Happy	9. Sleepy	4. Excited	9. Happy
5. Agitated	10. Walk	5. Uninterested	10. Confused	5. Interested	10. Dog
Praise Original		*Praise and Reprimand*		*Praise and Toy*	
1. Confused	6. Curious	1. Happy	6. Curious	1. Happy	6. Curious
2. Calm	7. Bored	2. Calm	7. Interested	2. Confused	7. Calm
3. Happy	8. Relaxed	3. Confused	8. Focused	3. Excited	8. Loved
4. Dog	9. Content	4. Attentive	9. Intrigued	4. Dog	9. Proud
5. Uninterested	10. Seems	5. Dog	10. Excited	5. Playful	10. Seems

Note. Top 10 words for each video are ranked by frequency used. The number one word is the most frequently used word

Experiment 2 Filming and Video Editing Methodology

In Experiment 2 mismatched contexts within videos was carried out in Adobe After Effects (Adobe Systems Inc., San Jose, CA) software. This software allows different video clips to be spliced together to create one video. Objects were clipped out of a video and then placed on another background or video image. First, the dog was spliced out of the video before the human or stimuli was also spliced out. Next, those two different video image layers were added back together on a black background. This was done both for the original videos (where the dog and human stimuli had been cut out of the same video and were just put back together on a black background), and for mismatched or swapped videos (where the dog from one video was added together with a human or stimuli from another).

All Videos from Experiment 1 and 2

https://osf.io/bpu3r/

Valence and Arousal for Each Video from Experiment 1 and 2

Experiment 1: Context vs No Context

Video	Valence		Arousal	
	M	**SE**	**M**	**SE**
Video 1a cat (no context)	5.251	0.077	3.570	0.098
Video 1b cat (context)	4.905	0.093	4.551	0.118
Video 2a play (no context)	6.482	0.123	5.553	0.112
Video 2b play (context)	7.013	0.109	4.787	0.106
Video 3a tape measure (no context)	5.560	0.102	6.335	0.105
Video 3b tape measure (context)	5.199	0.100	6.504	0.098
Video 4a praise (no context)	3.924	0.079	7.526	0.095
Video 4b praise (context)	5.763	0.103	6.301	0.109
Video 5a reprimand (no context)	6.851	0.093	3.412	0.104
Video 5b reprimand (context)	4.013	0.098	4.485	0.107
Video 6a treat (no context)	7.016	0.095	4.197	0.103
Video 6b treat (context)	7.380	0.100	4.522	0.122

Experiment 2: Mismatched Contexts

Video	Valence		Arousal	
	M	SE	M	SE
Video 1a - Reprimand Original (Video #10)	4.064	0.092	3.901	0.090
Video 1b - Reprimand and Toy (Video #7)	4.386	0.075	6.431	0.072
Video 1c - Reprimand and Vacuum (Video #4)	4.414	0.084	7.270	0.075
Video 2a - Toy Original (Video #1)	6.915	0.081	6.177	0.082
Video 2b - Toy and Vacuum (Video #12)	6.907	0.092	6.507	0.089
Video 2c - Toy and Leash (Video #15)	6.986	0.093	6.753	0.091
Video 3a - Vacuum Original (Video #8)	3.831	0.086	7.938	0.073
Video 3b - Vacuum and Leash (Video #17)	4.305	0.094	7.670	0.082
Video 3c - Vacuum and Tape Measure (Video #11)	5.175	0.091	4.338	0.108
Video 4a - Leash Original (Video #3)	7.573	0.095	6.443	0.099
Video 4b - Leash and Tape Measure (Video #13)	6.225	0.084	3.732	0.091
Video 4c - Leash and Praise (Video #18)	6.078	0.080	2.866	0.080
Video 5a - Tape Measure Original (Video #16)	5.612	0.075	4.021	0.089
Video 5b - Tape Measure and Praise (Video #6)	5.518	0.079	3.660	0.092
Video 5c - Tape Measure and Reprimand (Video #2)	5.268	0.067	4.262	0.103
Video 6a - Praise Original (Video #9)	6.179	0.086	2.680	0.076
Video 6b - Praise and Reprimand (Video #14)	7.031	0.085	2.852	0.079
Video 6c - Praise and Toy (Video #5)	6.967	0.091	4.678	0.086