The perception of emotional expressions allows animals to evaluate the social intentions and motivations of each other. This usually takes place within species; however, in the case of domestic dogs, it might be advantageous to recognize the emotions of humans as well as other dogs. In this sense, the combination of visual and auditory cues to categorize others’ emotions facilitates the information processing and indicates high-level cognitive representations. Using a cross-modal preferential looking paradigm, we presented dogs with either human or dog faces with different emotional valences (happy/playful versus angry/aggressive) paired with a single vocalization from the same individual with either a positive or negative valence or Brownian noise. Dogs looked significantly longer at the face whose expression was congruent to the valence of vocalization, for both conspecifics and heterospecifics, an ability previously known only in humans. These results demonstrate that dogs can extract and integrate bimodal sensory emotional information, and discriminate between positive and negative emotions from both humans and dogs.

1. Introduction

The recognition of emotional expressions allows animals to evaluate the social intentions and motivations of others [1]. This provides crucial information about how to behave in different situations involving the establishment and maintenance of long-term relationships [2]. Therefore, reading the emotions of others has enormous adaptive value. The ability to recognize and respond appropriately to these cues has biological fitness benefits for both signaller and the receiver [1].

During social interactions, individuals use a range of sensory modalities, such as visual and auditory cues, to express emotion with characteristic changes in both face and vocalization, which together produce a more robust percept [3]. Although facial expressions are recognized as a primary channel for the transmission of affective information in a range of species [2], the perception of emotion through cross-modal sensory integration enables faster, more accurate and more reliable recognition [4]. Cross-modal integration of emotional cues has been observed in some primate species with conspecific stimuli, such as matching a specific facial expression with the corresponding vocalization or call [5–7]. However, there is currently no evidence of emotional recognition of heterospecifics in non-human animals. Understanding heterospecific emotions is of particular importance for animals such as domestic dogs, who live most of their lives in mixed species groups and have developed mechanisms to interact with humans (e.g. [8]). Some work has shown cross-modal capacity in dogs relating to the perception of specific activities (e.g. food-guarding) [9] or individual features (e.g. body size) [10], yet it remains unclear whether this ability extends to the processing of emotional cues, which inform individuals about the internal state of others.
Dogs can discriminate human facial expressions and emotional sounds (e.g. [11–18]); however, there is still no evidence of multimodal emotional integration and these results relating to discrimination could be explained through simple associative processes. They do not demonstrate emotional recognition, which requires the demonstration of categorization rather than differentiation. The integration of congruent signals across sensory inputs requires internal categorical representation [19–22] and so provides a means to demonstrate the representation of emotion.

In this study, we used a cross-modal preferential looking paradigm without familiarization phase to test the hypothesis that dogs can extract and integrate emotional information from visual (facial) and auditory (vocal) inputs. If dogs can cross-modally recognize emotions, they should look longer at facial expressions matching the emotional valence of simultaneously presented vocalizations, as demonstrated by other mammals (e.g. [5–7,21,22]). Owing to previous findings of valence [5], side [22], sex [11,22] and species [12,23] biases in perception studies, we also investigated whether these four main factors would influence the dogs’ response.

2. Material and methods

Seventeen healthy socialized family adult dogs of various breeds were presented simultaneously with two sources of emotional information. Pairs of grey-scale gamma-corrected human or dog face images from the same individual but depicting different expressions (happy/playful versus angry/aggressive) were projected onto two screens at the same time as a sound was played (figure 1a). The sound was a single vocalization (dog barks or human voice in an unfamiliar language) of either positive or negative valence from the same individual, or a neutral sound (Brownian noise). Stimuli (figure 1b) featured one female and one male of both species. Unfamiliar individuals and an unfamiliar language (Brazilian Portuguese) were used to rule out the potential influence of previous experience with model identity and human language.

Experiments took place in a quiet, dimly-lit test room and each dog received two 10-trial sessions, separated by two weeks. Dogs stood in front of two screens and a video camera recorded their spontaneous looking behaviour. A trial consisted of the presentation of a combination of the acoustic and visual stimuli and lasted 5 s (see electronic supplementary material for details). Each trial was considered valid for analyses when the dog looked at the images for at least 2.5 s. The 20 trials presented different stimulus combinations: 4 face-pairs (2 human and 2 dog models) × 2 vocalizations (positive and negative valence) × 2 face positions (left and right), in addition to 4 control trials (4 face-pairs with neutral auditory stimulus). Therefore, each subject saw each possible combination once.

We calculated a congruence index = (C – I)/T, where C and I represent the amount of time the dog looked at the congruent (facial expression matching emotional vocalization, C) and incongruent faces (I), and T represents total looking time (looking left + looking right + looking at the centre) for the given trial, to measure the dog’s sensitivity to audio-visual emotional cues delivered simultaneously. We analysed the congruence index across all trials using a general linear mixed model (GLMM) with individual dog included in the model as a random effect. Only emotion valence, stimulus sex, stimulus species and presentation position (left versus right) were included as the fixed effects in the final analysis because first- and second-order interactions were not significant. The means were compared to zero and confidence intervals were presented for all the main factors in this model. A backward selection process was applied to identify the significant factors. The normality assumption was verified by visually inspecting plots of residuals with no important deviation from normality identified. To verify a possible interaction between the sex of subjects and stimuli, we used a separate GLMM taking into account these factors. We also tested whether dogs preferentially looked at a particular valence throughout trials and at a particular face in the control trials (see the electronic supplementary material for details of index calculation).

3. Results

Dogs showed a clear preference for the congruent face in 67% of the trials (n = 188). The mean congruence index was 0.19 ± 0.03 across all test trials and was significantly greater than zero (t_{16} = 5.53, p < 0.0001), indicating dogs looked significantly longer at the face whose expression matched the valence of vocalization. Moreover, we found a consistent congruent looking preference regardless of the stimulus species (dog: t_{16} = 5.39, p < 0.0001; human: t_{16} = 2.48, p = 0.01; figure 2a), emotional valence (negative: t_{16} = 5.01, p < 0.0001; positive: t_{16} = 2.88, p = 0.005; figure 2b), stimulus gender (female: t_{16} = 4.42, p < 0.0001; male: t_{16} = 3.45, p < 0.0001; figure 2c) and stimulus position (left side: t_{16} = 2.74, p < 0.01; right side: t_{16} = 5.14, p < 0.0001; figure 2d). When a backwards selection procedure was applied to the model with the four main factors, the final model included only stimulus species. The congruence index for this model was significantly higher for viewing dog rather than human faces (dog: 0.26 ± 0.05, human: 0.12 ± 0.05, F_{1,170} = 4.42; p = 0.04, figure 2a), indicating that dogs demonstrated greater sensitivity towards conspecific cues. In a separate model, we observed no significant interaction between subject sex and stimulus sex (F_{1,169} = 1.33, p = 0.25) or main effects (subject sex: F_{1,169} = 0.17, p = 0.68; subject stimulus: F_{1,169} = 0.56, p = 0.45).

Dogs did not preferentially look at either of the facial expressions in control conditions when the vocalization was the neutral sound (mean: 0.04 ± 0.07; t_{16} = 0.56; p = 0.58). The mean preferential looking index was −0.05 ± 0.03, which was not significantly different from zero (t_{16} = −1.6, p = 0.13), indicating that there was no difference in the proportion of viewing time between positive and negative facial expressions across trials.

4. Discussion

The findings are, we believe, the first evidence of the integration of heterospecific emotional expressions in a species other than humans, and extend beyond primates the demonstration of cross-modal integration of conspecific emotional expressions. These results show that domestic dogs can obtain dog and human emotional information from both auditory and visual inputs, and integrate them into a coherent perception of emotion [21]. Therefore, it is likely that dogs possess at least the mental prototypes for emotional categorization (positive versus negative affect) and can recognize the emotional content of these expressions. Moreover, dogs performed in this way without any training or familiarization with the models, suggesting
that these emotional signals are intrinsically important. This is consistent with this ability conferring important adaptive advantages [24].

Our study shows that dogs possess a similar ability to some non-human primates in being able to match auditory and visual emotional information [5], but also demonstrates an important advance. In our study, there was not a strict temporal correlation between the recording of visual and auditory cues (e.g. relaxed dog face with open mouth paired with playful bark), unlike the earlier research on primates (e.g. [5]). Thus the relationship between the modalities was not temporally contiguous, reducing the likelihood of learned associations accounting for the results. This suggests the existence of a robust categorical emotion representation.

Although dogs showed the ability to recognize both conspecific and heterospecific emotional cues, we found that they responded significantly more strongly towards dog stimuli. This could be explained by a more refined mechanism for the categorization of emotional information from conspecifics, which is corroborated by the recent findings of dogs showing a greater sensitivity to conspecifics’ facial expressions [12] and a preference for dog over human images [23]. The ability to recognize emotions through visual and auditory cues may be a particularly advantageous social tool in a highly social species such as dogs and might have been exapted for the establishment and maintenance of long-term relationships with humans. It is possible that during domestication, such features could have been retained and potentially selected for, albeit unconsciously. Nonetheless, the communicative value of emotion is one of the core components of the process and even less-social domestic species, such as cats, express affective states such as pain in their faces [25].

**Figure 1.** (a) Schematic apparatus. R2: researcher, C: camera, S: screens, L: loudspeakers, P: projectors, R1: researcher. (b) Examples of stimuli used in the study: faces (human happy versus angry, dog playful versus aggressive) and their correspondent vocalizations.
There has been a long-standing debate as to whether dogs can recognize human emotions. Studies using either visual or auditory stimuli have observed that dogs can show differential behavioural responses to single modality sensory inputs with different emotional valences (e.g. [12,16]). For example, Müller et al. [13] found that dogs could selectively respond to happy or angry human facial expressions; when trained with only the top (or bottom) half of unfamiliar faces they generalized the learned discrimination to the other half of the face. However, these human-expression-modulated behavioural responses could be attributed solely to learning of contiguous visual features. In this sense, dogs could be discriminating human facial expressions without recognizing the information being transmitted.

Our subjects needed to be able to extract the emotional information from one modality and activate the corresponding emotion category template for the other modality. This indicates that domestic dogs interpret faces and vocalizations using more than simple discriminative processes; they obtain emotionally significant semantic content from relevant audio and visual stimuli that may aid communication and social interaction. Moreover, the use of unfamiliar Portuguese words controlled for potential artefacts induced by a dog’s previous experience with specific words. The ability to form emotional representations that include more than one sensory modality suggests cognitive capacities not previously demonstrated outside of primates. Further, the ability of dogs to extract and integrate such information from an unfamiliar human stimulus demonstrates cognitive abilities not known to exist beyond humans. These abilities may be fundamental to a functional relationship within the mixed species social groups in which dogs often live. Moreover, our results may indicate a more widespread distribution of the ability to spontaneously integrate multimodal cues among non-human mammals, which may be key to understanding the evolution of social cognition.

Figure 2. Dogs’ viewing behaviour (calculated as congruence index). (a) Species of stimulus; (b) valence of stimulus; (c) sex of stimulus; (d) side of stimulus presentation. *p < 0.05, **p < 0.01, ***p < 0.001.

## References


