The Neural Basis of Understanding the Expression of the Emotions in Man and Animals

Journal:	*Social Cognitive and Affective Neuroscience*
Manuscript ID	SCAN-16-458.R2
Manuscript Type:	Original Manuscript
Date Submitted by the Author:	11-Oct-2016
Complete List of Authors:	Spunt, Robert; Robert Spunt, Humanities & Social Sciences; Ellsworth, Emily; Caltech, Social Sciences Adolphs, Ralph; Caltech, Social Sciences
Keywords:	anthropomorphism, emotion understanding, face perception, fMRI, social cognition, theory of mind

Running Title: UNDERSTANDING ANIMAL EMOTION

The Neural Basis of Understanding the Expression of the Emotions in Man and Animals

Robert P. Spunt, Emily Ellsworth, Ralph Adolphs

California Institute of Technology

Corresponding Author:

Robert P. Spunt

California Institute of Technology

1200 E. California Blvd.

Pasadena, CA 91125

Email: spunt@caltech.edu

Abstract

Humans cannot help but attribute human emotions to nonhuman animals. While such attributions are often regarded as gratuitous anthropomorphisms and held apart from the attributions humans make about each other's internal states, they may be the product of a general mechanism for flexibly interpreting adaptive behavior. To examine this, we used functional magnetic resonance imaging (fMRI) in humans to compare the neural mechanisms associated with attributing emotions to humans and nonhuman animal behavior. While undergoing fMRI, participants first passively observed the facial displays of human, nonhuman primate, and domestic dogs, and subsequently judged the acceptability of emotional (e.g., "annoyed") and facial descriptions (e.g., "baring teeth") for the same images. For all targets, emotion attributions selectively activated regions in prefrontal and anterior temporal cortices associated with causal explanation in prior studies. These regions were similarly activated by both human and nonhuman targets even during the passive observation task; moreover, the degree of neural similarity was dependent on participants' self-reported beliefs in the mental capacities of nonhuman animals. These results encourage a non-anthropocentric view of emotion understanding, one that treats the idea that animals have emotions as no more gratuitous than the idea that humans other than ourselves do.

Keywords: anthropomorphism, emotion attribution, face perception, social cognition

Introduction

"But man himself cannot express love and humility by external signs, so plain as does a dog, when with drooping ears, hanging lips, flexuous body, and wagging tail, he meets his beloved master. Nor can these movements in the dog be explained by acts of volition or necessary instincts, any more than the beaming eyes and smiling cheeks of a man when he meets an old friend." (Darwin, 1872, pp. 10-11)

In this brief passage from his treatise *The Expression of the Emotions in Man and Animals*, Charles Darwin provides an excellent illustration of the human mind's capacity to see in the behavior of nonhuman animals the same kinds of covert emotional states it sees in the behavior of other humans. From an anthropocentric viewpoint, the attribution of humanlike emotion to nonhuman animals represents a clear case of *anthropomorphism*, the projection of our own attributes onto nonhuman entities (Epley, Waytz, & Cacioppo, 2007), and is not unlike the attribution of beliefs, motives, and intentions to other nonhuman entities, such as moving circles and triangles (Heider & Simmel, 1944), robots (Fussell, Kiesler, Setlock, & Yew, 2008), and hurricanes (Barker & Miller, 1990). Unsurprisingly, then, Darwin's sophisticated attributions of emotion to nonhuman animals are still the subject of intense debate regarding whether and how emotion should be used in the science of animal behavior (de Waal, 2011). Nevertheless, be the beloved family dog or the genetically-similar chimpanzee and bonobo, humans cannot help but attribute human emotions to animals. In the present study, we designed a behavioral task for isolating the cognitive processes that produce reliable attributions of emotional states to the behavior of nonhuman animals. Using this task in conjunction with functional magnetic resonance imaging (fMRI), we for the first time directly compared the neural basis of attributing the same emotions to human and nonhuman animals.

What is involved when one human attributes an emotion to another human? Taking Darwin's example above, the target's overt facial behaviors (e.g., "beaming eyes", "smiling cheeks") are viewed as expressions of a specific covert emotional state presumed to exist in the target. This assumption – that observable behaviors are caused by unobservable states of mind like belief and emotion – is the foundation of what the philosopher Daniel Dennett termed *the intentional stance* (Dennett, 1989), and what in psychology and more recently social neuroscience has come to be known as a *theory of mind* (Gallagher & Frith, 2003; Saxe, Carey, & Kanwisher, 2004). Our proclivity for making such attributions about behavior has been well studied by social and developmental psychologists, yielding several frameworks for characterizing the dimensions on which we make such attributions. For

instance, there are two-dimensional schemes for experience and agency (Gray, Gray, & Wegner, 2007), or for competence and warmth (Fiske, Cuddy, & Glick, 2007). Specifically for faces, there may be dimensions of trustworthiness and dominance (Todorov, Said, Engell, & Oosterhof, 2008). In all these cases, the dimensions on which we represent emotion in humans are dimensions of attributes that cannot be directly observed but are inferred based on information that is already known or directly observable, such as nonverbal behavior. In other words, representations of emotion tend to be conceptually abstract, with an unreliable correspondence with specific sensorimotor events (Vallacher & Wegner, 1987; Spunt, Kemmerer, & Adolphs, 2016).

There is also a literature suggesting that very similar, or identical, psychological dimensions characterize the attributions humans tend to make about a variety of nonhuman entities, ranging from nonhuman animals to robots to Gods (Epley et al., 2007). While such anthropomorphic attributions are typically placed in a special category and often treated as irrational and gratuitous (e.g., *(Wynne, 2004)), they may actually be the rational consequence of fundamental similarities in both the form and function of human and nonhuman animal behaviors (de Waal, 2011). The form of the human body and face – including its capacity for expression - has much in common with other mammalian species (Brecht & Freiwald, 2012), particularly other primates (Sherwood et al., 2003; Parr, Waller, Vick, & Bard, 2007). Not surprisingly, then, untrained human observers evidence an ability to extract affective information from novel dog facial expressions that rivals their ability to do the same for human infant expressions (Schirmer, Seow, & Penney, 2013). In turn, evidence suggests that domesticated dogs have through selective breeding and adaptation to human ecologies evolved an ability to discriminate affective information in human faces (Axelsson et al., 2013; Berns, Brooks, & Spivak, 2012; Cuaya, Hernández-Pérez, & Concha, 2016).

To our knowledge, no prior neuroimaging study has experimentally manipulated the demand to attribute emotions to nonhuman animals in order to identify its neural basis and evaluate its similarity to the well-known neural basis of reasoning about the mental states of another human being. Several studies have compared the neural correlates of passively observing dogs and humans under different conditions; these studies largely demonstrate that regions associated with processing salient features of human behavior - such as facial expression and biological motion - are also activated when observing corresponding features in dogs (Blonder et al., 2004; Franklin et al., 2013). These studies collectively indicate that humans spontaneously deploy similar higher-level visual processes across

human and nonhuman animal targets. However, given that they did not experimentally control the demand to anthropmorphize the nonhuman animal targets, their findings leave open questions regarding its neural basis.

A related line of prior neuroimaging studies demonstrate that individual differences in prior experience and beliefs related to nonhuman animals affect the brain regions humans recruit when observing nonhuman animals. For instance, prior experience with dogs was associated with increased activity in the posterior superior temporal sulcus during the passive viewing of meaningful dog gestures (Kujala, Kujala, Carlson, & Hari, 2012), presumably reflecting increased engagement of cortical regions concerned with processing biological motion. A more recent study compared the neural response to observing domesticated dogs in pet owners and non-pet owners and found that pet owners more strongly activated a set of cortical regions spanning insular, frontal, and occipital cortices (Hayama, Chang, Gumus, King, & Ernst, 2016). Related work examining brain structure suggests that individual differences in anthropomorphism for nonhuman animals (see Waytz, Cacioppo, & Epley, 2010) correlated with gray matter volume in a region of left TPJ thought to be involved in aspects of theory-of-mind (Cullen, Kanai, Bahrami, & Rees, 2014). These studies suggest the importance of distinguishing the *capacity* to appreciate anthropomorphic descriptions of nonhuman animals, and the *tendency* to deploy that capacity spontaneously, in the absence of an explicit stimulus to do so (Keysers & Gazzola, 2014).

Although the neural basis of attributing emotion to animals remains unknown, the neural basis of attributing mental states to other humans is known to be reliably associated with a set of cortical regions commonly referred to as the theory-of-mind (ToM) or mentalizing network (Fletcher et al., 1995; Goel, Grafman, Sadato, & Hallett, 1995; Happé et al., 1996; Gallagher & Frith, 2003; Saxe et al., 2004; Amodio & Frith, 2006). In a series of prior studies of healthy adults, we have shown that attributions about social situations activate an anatomically well-defined network of brain regions (Spunt, Falk, & Lieberman, 2010; Spunt, Satpute, & Lieberman, 2011; Spunt & Lieberman, 2012a, 2012b). In a recent study, we found that, remarkably, this same neural system appears to be engaged when we make causal attributions about completely nonsocial events, such as attributing the sight of water gushing out of a gutter to an unseen rainstorm (Spunt & Adolphs, 2015). In line with prior neuroimaging studies of social and/or nonsocial reasoning (see Van Overwalle, 2011), activation of this inferential process was reliably stronger for social situations. Thus, it appears that social attributions are executed by processes that, though intrinsically domain-

general, may acquire specialization for the social domain over the course of social development.

Here, we extend this research to identify the neural mechanisms supporting anthropomorphism during the perception of facial expressions that we see in nonhuman primates and dogs. Specifically, we examined three questions. First, do we use the same mechanisms to attribute emotions to the facial expressions of humans and nonhuman animals, when asked in a task to make such attributions? Second, do we spontaneously recruit these mechanisms to similar degrees when merely passively watching human and nonhuman animal behavior, in the absence of an explicit attribution task? Third, is the level of spontaneous recruitment dependent on individual differences in experience with, and attitudes towards, humans and nonhuman animals?

Method

Participants

Eighteen adults from the Los Angeles metropolitan area participated in the study in exchange for financial compensation. All participants were screened to ensure that they were right-handed, neurologically and psychiatrically healthy, had normal or corrected-to-normal vision, spoke English fluently, had IQ in the normal range (as assessed using the Wechsler Abbreviated Scales of Intelligence), and were not pregnant or taking any psychotropic medications at the time of the study. All participants provided written informed consent according to a protocol approved by the Institutional Review Board of the California Institute of Technology.

Exclusions. For the Explicit Task described below, data from two participants was excluded from the analysis, one due to unusually poor task performance (no response to 46% of trials), and one due to excessive head motion during image acquisition (translation > 8mm). This left 16 participants (8 males, 8 females; mean age = 29.00, age range = 21-46) for the analysis of the Explicit Task. For the Implicit Task described below, data from one participant was excluded from analysis due to unusually poor task performance (no response to 42% of catch trials). This left 17 participants (9 males, 8 females; mean age = 28.71, age range = 21-46) in the analysis of the Implicit Task.

Power Analysis. The open-source MATLAB toolbox fmripower was used to estimate power for detecting effects in the *Emotion* > *Expression* contrast for each of the a priori ROIs used to test our primary hypotheses in the Explicit Task (described below). We used the *Mind* > *Body* contrast from Spunt, Meyer, and Lieberman (2015), which featured an event-

related design similar to that of the present study. For the 4 ROIs used to test our primary hypotheses, 90% detection power could be achieved with an average of 9.25 subjects (SD = 3.95, MIN/MAX = 6/15). We thus aimed to have data from at least 15 subjects, a sample size sufficiently large to test our primary hypotheses, and recruited N=18 with the expectation that a few participants might drop out for technical reasons, as indeed a few did.

Experimental Design

For the study, participants were asked to perform two tasks, each of which was intended to capture distinctive features of emotion attribution. Both tasks featured the same stimulus set containing photographs of human, nonhuman primate, and dog facial expressions. The tasks differed primarily in the overt task that participants had to perform for each stimulus. In the following sections, we describe in detail the components of the experimental design, beginning with the stimulus set.

Stimuli. The experimental stimuli were composed of 30 naturalistic photographs of facial expressions from each of three categories: *Humans*, *Nonhuman Primates*, and *Dogs* (see **Figure 1** for examples, and **Figure S1**, **Figure S2**, and **Figure S3** for the full stimulus sets). For brevity's sake, we henceforth use the term "Primate" to refer to the Nonhuman Primate photographs. Given that three of the Primate photographs show the animal in restricted conditions (e.g., behind the bars of a cage), we clarify here that we describe the photographs as "naturalistic" from the perspective of a human observer rather than the depicted targets. Insofar as opportunities for humans to observe primates often occur while the animal is in captivity (e.g., zoo, laboratory), photographs of animals in restricted conditions are therefore "naturalistic" from the perspective of the human observer.

We acquired stimuli from a variety of online stock photography sources. Each set contained 26 color photographs and four grayscale photographs; these were not distinguished in analyses. Moreover, the three sets were matched on valence, on face area proportionate to the whole image area, and on estimated luminance (all ps > .84 from independent-samples t-tests comparing the means). We used Amazon.com's web service Mechanical Turk to collect normative ratings of photograph valence from approximately 30 native English-speaking U.S. citizens (on a 7-point Likert scale; 1 = Extremely Negative, 4 = Neutral, 7 = Extremely Positive).

Explicit Task. In the Explicit Task, we experimentally manipulated mental state attribution by presenting participants with each stimulus twice, once with the demand to evaluate a description of the target's emotional state (e.g., bored?), and once with the

demand to instead evaluate a description of the physical characteristics of their facial expression (e.g., mouth open?). The Explicit Task can thus be described as a 2 (Cue: Emotion vs. Expression) x 3 (Target: Humans vs. Nonhuman Primates vs. Dogs) factorial design. In numerous published studies, we have shown that conceptually similar manipulations provide a contrast that robustly and selectively modulates activity in the regions of the brain associated with mental state attribution (Spunt et al., 2011; Spunt & Lieberman, 2012a, 2012b; Spunt & Adolphs, 2014).

Table 1 displays the 10 verbal cues featured in the study. Emotion cues regarded the mental state of the focal animal in the photograph (e.g., bored?), while Expression cues regarded an observable motor behavior (e.g., gazing up?). Each cue was paired with four photographs designed to elicit the response 'yes', and two photographs designed to elicit the response 'no'. These pairings were selected based on the responses of an independent sample of Mechanical Turk respondents. For these prior normative responses, each cue-stimulus pairing was evaluated by approximately 30 native English-speaking U.S. citizens. We retained only those pairings that produced the same response (accept or reject) in the majority (>80%) of respondents. In our subsequent analysis of participant performance, this consensus data was used to code responses as normative vs. counternormative. Independent samples t-tests showed that question-photograph consensus did not differ across the three stimulus sets (ps > .60).

Implicit Task. The Explicit Task allowed us to tackle our primary research question directly: Do people activate the same neural regions, and hence presumably engage the same psychological processes, to attribute mental states to the facial expressions of humans and nonhuman animals alike? Given that anthropomorphism is motivated by extrinsic task demands, this limited our ability to address our secondary and tertiary research questions, which ask whether humans commonly attribute human-like emotions to nonhuman animals spontaneously, in the absence of any extrinsic demands to do so. To better capture aspects of this spontaneously expressed motivation to anthropomorphize nonhuman animals, we had participants perform a simple visual 1-back task on all stimuli. To minimize demand characteristics, participants always performed this Implicit Task before being introduced to the Explicit Task. This ensured that the evoked responses to each stimulus category were maximally spontaneous and not primed by the words used in the explicit task.

In the Implicit Task, the 90 stimuli were presented in pseudo-random order in an

event-related fashion along with 30 phase-scrambled images, which were selected by phase-scrambling the entire set of 90 stimuli and determining a subset of 30 that was matched on luminance. Each stimulus appeared onscreen for 2 seconds. Trials were separated by an inter-stimulus interval during which a fixation-cross appeared onscreen. The duration ranged from .5 seconds to 5 seconds, with a mean interval of 1.5 seconds. During stimulus presentation, participants were asked to perform a visual 1-back task, indicating with a right index finger button press whenever a presented image was the same as the one previously shown. Three 1-back "catch" trials were included for each of the four stimulus conditions and excluded from the fMRI analysis. The percentage of 1-backs detected by participants was high (Mean = 96.57%, Min = 75.00%), ensuring that all participants attended to the images.

Post-Task Stimulus Ratings and Personality Measurement. Immediately following their scan, all subjects also rated each of the facial expression images on two scales. To assess the extent to which participants understood the emotional state of each target, participants answered the question “Do you understand what he or she is feeling?” (1-9 scale; 1=Not at all, 9=Completely). To assess the valence of the participant’s emotional reaction to viewing each target, participants answered the question “How does the photograph make you feel?” (1-9 scale; 1=Very bad, 9=Very good).

In addition, participants completed several questionnaires designed to measure personality attributes we predicted would affect their tendency to anthropomorphize animals. These were the The Belief in Animal Mind Questionnaire (BAM; (Hills, 1995)), the Empathy to Animals Scale (ETA; Paul, 2000), the Pet Attitude Scale (PAS; Templer, Salter, Dickey, Baldwin, & Veleber, 1981), and the Individual Differences in Anthropomorphism Questionnaire (IDAQ; Waytz et al., 2010). Due to restriction of range in our small sample in responses to the BAM, ETA, and PAS, we excluded these questionnaires from further analysis. Moreover, given our exclusive focus on the anthropomorphization of nonhuman animals as well as prior work indicating distinct correlates of the living and non-living subscales of the IDAQ (Cullen et al., 2014), we restricted our analysis to the animal anthropomorphization subscale, which demonstrated both excellent internal reliability (α = .90) and range of responses across participants (Scale Limits: 1 - 9; Score Range: 2.2 - 9). Due to subject timing schedules, responses to the BAM, ETA, and IDAQ were available for only 16 participants.

Experimental Procedures

For both tasks, trials were presented to participants in a pseudo-randomized event-related design (**Figure 1**). For the Implicit Task, the order and onsets of trials were optimized to maximize the efficiency of separately estimating the Face > Scramble contrast for each of the three target categories. For the Explicit Task, the order and onsets of trials were optimized to maximize the efficiency of separately estimating the Emotion > Expression contrast for each of the three targets; in addition, the order in which the Emotion/Expression cues appeared was counterbalanced across stimuli within each target. For both tasks, design optimization was achieved by generating the design matrices for one million pseudo-randomly generated designs and for each summing the efficiencies of estimating each contrast of interest. The most efficient design for each task was retained and used for all participants.

Stimulus presentation and response recording used the Psychophysics Toolbox (version 3.0.9; Brainard, 1997) operating in MATLAB. An LCD projector was used to present the task on a screen at the rear of the scanner bore that was visible to participants through a mirror positioned on the head coil. Participants were given a button box and made their responses using their right-hand index and middle fingers. Before the experimental tasks, participants were introduced to the task structure and performed brief practice versions featuring stimuli not included in the experimental task.

Image Acquisition

All imaging data was acquired at the Caltech Brain Imaging Center using a Siemens Trio 3.0 Tesla MRI scanner outfitted with a 32 channel phased-array head coil. We acquired 1330 whole-brain T2*-weighted echoplanar image volumes (EPIs; multi-band acceleration factor=4, slice thickness=2.5 mm, in-plane resolution=2.5 mm x 2.5 mm, 56 slices, TR=1000 ms, TE=30 ms, flip angle=60°, FOV=200 mm) for the two experimental tasks. Participants' in-scan head motion was minimal (max translation = 2.78 mm, max rotation = 1.88°). We also acquired an additional 904 EPI volumes for each participant for use in a separate study. Finally, we acquired a high-resolution anatomical T1-weighted image (1 mm isotropic) and field maps used to estimate and correct for inhomogeneity-induced image distortion.

Image Preprocessing

Images were processed using Statistical Parametric Mapping (SPM12 version 6685, Wellcome Department of Cognitive Neurology, London, UK) operating in MATLAB. Prior to statistical analysis, each participant's images for each task were subjected to the following

preprocessing steps: (1) the first four EPI volumes were discarded to account for T1-equilibration effects; (2) slice-timing correction was applied; (3) the realign and unwarp procedure was used to perform distortion correction and concurrent motion correction; (4) the participants' T1 structural volume was co-registered to the mean of the corrected EPI volumes; (5) the group-wise DARTEL registration method included in SPM12 (Ashburner, 2007) was used to normalize the T1 structural volume to a common group-specific space, with subsequent affine registration to Montreal Neurological Institute (MNI) space; (6) all EPI volumes were normalized to MNI space using the deformation flow fields generated in the previous step, which simultaneously re-sampled volumes (2 mm isotropic) and applied spatial smoothing (Gaussian kernel of 6 mm isotropic, full width at half maximum); and finally, (7) a log-transformation was applied to the EPI timeseries for each task to ensure that the regression weights estimated from our single-subject models were interpretable as percent signal change.

Single-Subject Analysis

For each participant, general linear models were used to estimate a model of the EPI timeseries for each task. Models for both tasks included as covariates of no interest the six motion parameters estimated from image realignment and a predictor for every timepoint where in-brain global signal change (GSC) exceeded 2.5 SDs of the mean GSC or where estimated motion exceeded 0.5 mm of translation or 0.5° of rotation. In addition, the hemodynamic response was modelled using the canonical (double-gamma) response function; high-pass filtered at 1/100 Hz; and estimated using the SPM12 RobustWLS toolbox, which implements the robust weighted least-squares estimation algorithm (Diedrichsen & Shadmehr, 2005).

Explicit Task. The model for the Explicit Task included six covariates of interest corresponding to the six cells created by crossing factors corresponding to the Cue (*Emotion* vs. *Expression*) and Target (*Humans* vs. *Nonhuman Primates* vs. *Dogs*). These covariates excluded foil trials (i.e., trials to which the normative response was to reject) and trials to which the participant gave either no response or the counternormative response. These excluded trials were modeled in a separate covariate of no interest. The neural response to each trial was defined with variable epochs spanning the onset and offset of the target stimulus (Grinband, Wager, Lindquist, Ferrera, & Hirsch, 2008). The onset of the verbal cues preceding each target were also modeled in an additional covariate of no interest. A final covariate of no interest was included which modeled variability in response time (RT) across

all trials included in the covariates of interest.

Implicit Task. The model for the Implicit Task included four covariates of interest corresponding to the timeseries of the four stimulus types presented to participants (*Humans*, *Primates*, *Dogs*, *Scrambles*). These covariates excluded all 1-back "catch" trials, which were modeled in a separate covariate of no interest. Given that stimulus duration was fixed across conditions, we modeled the neural response to each trial using fixed 2-second epochs that spanned the onset and offset of each image.

Group Analysis

Contrasts of Interest. To test for a relationship with emotion attribution across all three targets (humans, primates, dogs), we used paired-samples t-tests to identify those regions that independently demonstrated an association with the Emotion > Expression contrast in the Explicit Task for all three targets. We followed this by testing for target-independent (spontaneous) responses to the Face > Scramble contrast in the Implicit Task for each stimulus category separately.

We interrogated several additional contrasts to identify effects that reliably differed across the human and nonhuman targets. Our primary objective here was to identify those effects where Human targets differed from both Primate and Dog targets. For the Explicit Task, we examined two contrasts, one for the Cue by Target interaction ($Human_{Emotion>Expression}$ > $Nonhuman_{Emotion>Expression}$); and one for the main effect of Target ($Human_{Emotion+Expression}$ > $Nonhuman_{Emotion+Expression}$). For the Implicit Task, we examined the *Human* > *Nonhuman* comparison. Finally, exploratory analyses of the *Primate* > *Dog* comparison are reported in **Table S2**. These analyses indicate that in both the Explicit and Implicit Tasks, responses in regions of interest were not reliably different across the two nonhuman targets.

Regions of Interest. Each effect of interest was first interrogated using a set of independently-defined regions of interest (ROI) based on the group-level Why/How contrasts from Study 1 (N = 29) and Study 3 (N = 21) reported in Spunt and Adolphs (2014). These images are publicly available on NeuroVault (http://neurovault.org/collections/445/). In Spunt and Adolphs (2015), we observed evidence for domain-general responses to the Why > How contrast in four left hemisphere ROIs: Dorsomedial PFC, the Lateral OFC, TPJ, and Anterior STS. Two of these regions - the dorsomedial PFC and lateral OFC - showed a response to the Why > How contrast that was reliably stronger for facial expressions than for nonsocial events. Building on these prior study findings, our ROI analyses here were focused on the

same four ROIs highlighted in that prior study. The peak coordinate and spatial extent of each ROI is provided in Table S1. For each ROI, we tested our hypotheses with t-tests on the extracted average parameter estimate across voxels. For each test, we report p-values corrected for multiple comparisons across ROIs using the false-discovery rate (FDR) procedure described in Benjamini and Yekutieli (2001). Confidence intervals (CIs) for these effects were estimated using the bias corrected and accelerated percentile method (10,000 random samples with replacement; implemented using the BOOTCI function in MATLAB).

Whole-Brain. ROI analyses were complemented by whole-brain analyses. To model the 2 x 3 design of the Explicit Task at the group-level, we entered participants' contrast images for the six cells of the design into a random-effects analysis using the flexible factorial repeated-measures ANOVA module within SPM12 (within-subject factors: Cue, Target; blocking factor: Subject). To examine the one-way design of the Implicit Task, we conducted one-sample t-tests on single-subject contrast images for effects of interest. When interrogating the group-level model for both tasks, we tested the conjunction null for comparisons of interest using the minimum statistic method (Nichols, Brett, Andersson, Wager, & Poline, 2005).

We interrogated the resulting group-level t-statistic images by applying a cluster-forming (voxel-level) threshold of $p < .001$ followed by cluster-level correction for multiple comparisons at a family-wise error (FWE) of .05. Cluster-level correction was achieved for conjunction images by identifying the maximum cluster extent threshold necessary to achieve cluster-wise correction across the individual images entering the conjunction. For visual presentation, thresholded t-statistic maps were overlaid on the average of the participants' T1-weighted anatomical images.

Exploratory Analysis of Individual Differences. We conducted between-subject analyses to explore possible individual differences in the experimentally-unconstrained brain activity observed during the Implicit Task. We specifically explored two interrelated *a priori* hypotheses regarding individual differences in the attribution of human-like emotions to nonhuman animals. The first hypothesis follows on our finding that brain regions associated with anthropomorphism in the Explicit Task were reliably activated by the stimuli in the Implicit Task. If these activations reflect spontaneous anthropomorphism, then they should be strongest in those individuals who, on our post-scanning questionnaire, endorsed the highest levels of understanding the animal stimuli, as well as in those participants who endorsed a disposition to anthropomorphize nonhuman animals in their everyday lives. To

explore this first hypothesis, we extracted signal from each ROI in the Primate > Scramble and Dog > Scramble contrasts and computed the Pearson correlation with the scores derived from the post-task stimulus ratings and personality questionnaires.

The second hypothesis adopts a different approach to interpreting the activity observed in the Implicit Task. Namely, if spontaneous activity in response to nonhuman animals reflects anthropomorphism, then the extent to which that activity is similar to the same individual's spontaneous activity in response to humans should be associated with their tendency to anthropomorphize nonhuman animals. To explore this hypothesis, for each participant we correlated their whole-brain (grey-matter masked) response pattern for human faces to their response pattern for each of the nonhuman animal face conditions. These correlations were computed on the t-statistic images and subsequently Fisher's z-transformed for the between-subject analysis. This produced two measures of human/nonhuman neural similarity for each participant, which we also correlated with the scores derived from the post-task stimulus ratings and personality questionnaires.

We chose to estimate similarity across the whole-brain response because it requires no assumptions about the content of the brain states being compared. Rather, it simply asks to what extent to a participant responds similarly when naturally observing humans and nonhuman animals, and is agnostic regarding *what* specific mental contents and processes are involved. In adopting this strategy, we recognized that within-subject similarities would likely be underestimated due to the admission of functional responses that are irrelevant to anthropomorphism, such as early visual processing. However, we reasoned that this problem in within-subject similarity estimates would likely wash out in our analysis, which is on the between-subject variability.

Results

Behavioral Outcomes

Table 2 summarizes cue acceptance and RT data from the Explicit Task and post-task stimulus ratings of emotion understanding and emotional valence. Due to ceiling effects on cue acceptance which produced highly non-normal distributions, we did not subject cue acceptance data to statistical tests and excluded counternormative attributions from our analysis of the both the remaining behavioral outcomes (RT and post-task ratings) and the estimated neural response (as described in the *Method*).

Explicit Task Performance. We used a series of repeated measures analysis of variance (ANOVA) to examine the remaining behavioral outcomes in the Explicit Task.

Complete results are presented in **Table S2**. When examining the simultaneous effects of Cue and Target on Acceptance RT, we observed a significant main effect of both Cue and Target but no interaction. The main effect of Cue - namely, that RT to acceptance for Emotion cues were longer than for Expression cues – parallels a reliable behavioral effects observed in the Why/How Task from which the present study's task was adapted (Spunt & Adolphs, 2014; Spunt & Adolphs, 2015). Critically, the absence of an interaction effect indicates that this behavioral effect was of a similar magnitude in all three targets. Post-hoc t-tests indicated that the RT difference across Emotion and Expression cues revealed reliable above-zero effects for all three targets (all *p*s < .0001). We additionally observed no evidence that the magnitude of this RT effect differed in any of the pairwise comparisons (all *p*s > .20). Post-hoc t-tests demonstrate that the main effect of Target on RT to acceptance was driven by reliably longer RTs for judgments of nonhuman targets when compared to the same judgments for human targets (all *p*s < .0001).

We emphasize here our belief that the main effect of Target on RT does not pose a significant impediment to interpreting our fMRI-derived measures of brain activity. As described above, our regression model of the fMRI timeseries for the Explicit Task additionally included a parametric covariate modelling RT variability in response amplitude across all included trials. As we have shown now in multiple published papers (Spunt & Adolphs, 2014; Spunt et al., 2016), RT variability does not sufficiently explain the univariate response in regions associated with attentional manipulations akin to the Emotion/Expression manipulation used here.

Post-Task Ratings. A one-way ANOVA revealed a significant main effect of Target on participants' post-task ratings of emotion understanding. Post-hoc contrasts revealed that this effect was driven by higher levels of understanding for human targets compared to each nonhuman targets, with no reliable difference across the two nonhuman targets. Finally, a one-way ANOVA revealed no evidence for an effect of Target on participants' post-task ratings of valence.

Target-Independent Effects

As listed in **Table 3** and displayed in **Figure 3**, every ROI except the left TPJ showed an independently significant association with the *Emotion* > *Expression* contrast for all three targets. These were the dorsomedial prefrontal cortex (dmPFC), the lateral orbitofrontal cortex (LOFC), and the anterior superior temporal sulcus (aSTS). Notably, left TPJ failed to demonstrate an association with the *Emotion* > *Expression* contrast for *any* of the three

targets, including Humans. As displayed in Figure 2a and listed in **Table S3**, a whole-brain analysis of the *Emotion* > *Expression* conjunction across targets demonstrated target-independent responses in regions of the dmPFC and LOFC similar in location to our *a priori* ROIs for those regions. These findings strongly suggest that explicitly attributing emotions (as opposed to merely describing expressions) to nonhuman animal faces relies on the same core inferential mechanisms supporting the attribution of emotion to the faces of other humans.

We next tested for target-general responses in the Implicit Task. As also listed in **Table 3**, every ROI - including the left TPJ - showed an independently significant association with the comparison to Scrambled images for all three targets. As displayed in **Figure 2** and listed in **Table S3**, a whole-brain analysis of the Face > Scramble conjunction across targets demonstrated target-independent responses in a distributed set of cortical regions spanning the fusiform and inferior occipital gyri bilaterally, the amygdala/parahippocampal gyrus bilaterally, the precuneus, and both the vmPFC and dmPFC. These findings expand on those observed in the Explicit Task by demonstrating that nonhuman animal facial expressions elicit similar regional responses in the brain even in the absence of an explicit emotion attribution task, and that these regional responses include those associated with emotion attribution in the Explicit Task as well as additional regions associated more so with face perception and social attention in prior studies (Blonder et al., 2004; Kujala et al., 2012; Franklin et al., 2013; Hayama et al., 2016).

Target-Dependent Effects

As listed in **Table 4** and displayed in **Figure 4**, we observed scant evidence for univariate responses in any ROI that reliably discriminated the human from the nonhuman stimuli. When examining each of the direct, pairwise comparisons of the three targets, only the left aSTS and left TPJ demonstrated reliable effects. The left aSTS showed greater activation for human compared to primate faces in both the Explicit and Implicit Tasks, and for dog compared to primate faces in the Explicit Task. The left TPJ showed greater activation for human compared to primate faces in only the Explicit Task. While of potential interest, these effects are more likely to be interpreted in terms of factors other than the human vs. nonhuman distinction, such as prior experience and conceptual familiarity, which would naturally be elevated for facial expressions of dogs and humans when compared to nonhuman primates.

Exploratory Analysis of Individual Differences

Table 5 summarizes the results of exploratory between-subject analyses examining the relationship between brain activity during the Implicit Task and participants' post-task ratings of emotion understanding and the nonhuman animal subscale of the IDAQ. Univariate contrast in our ROIs did not show consistent effects on ratings of emotion understanding but did show preliminary evidence for a positive association with the IDAQ subscale. Multivariate response similarity across human and each nonhuman target revealed stronger evidence for an association between "human-like" brain activity and both higher levels of emotion understanding and dispositional anthropomorphism towards animals.

Discussion

We identified the brain regions underlying a basic form of anthropomorphism, namely, the attribution of human-like emotion to the facial expressions of nonhuman animals. Using an adapted version of the Yes/No Why/How Task (Spunt & Adolphs, 2014), we were able to independently identify the neural basis of attributing the same emotional states to other humans, to nonhuman primates, and to dogs. In both region-of-interest and whole-brain analyses, we found no evidence for a uniquely human neural substrate for the attribution of emotion. Instead, we found that attributions of emotion to each of the three targets draws on a shared neural mechanism spanning dorsomedial and lateral orbitofrontal prefrontal cortices.

Thus, to answer our first research question: *Yes*, attributions of emotion to both humans and to nonhuman animals draw on the same neural mechanism. Rather than viewing this result as the misapplication of an inferential mechanism for attributing mental states to other humans, we instead view this as the rational consequence of fundamental continuities in both the form and function of human and many nonhuman animal behaviors (de Waal, 2011; Brecht & Freiwald, 2012). As noted earlier, the human body much in common with other mammalian species, making it natural that many nonhuman animals would behave in ways that physically resemble human emotional behaviors. A distinct yet related view follows from the observation that, due to common descent and/or to common ecology, many nonhuman animal species produce behaviors with functions that resemble those ascribed to human emotional behaviors, for instance, social communication. And, while it may be the case that we cannot know what it feels like to be a dog or a chimpanzee, it is just as much the case that we cannot know what it feels like to be *any* human other than

ourselves.

Our second research question attempted to distinguish the *ability* to attribute emotion to nonhuman animal behavior from the *tendency* to recruit that ability spontaneously, in the absence of the kinds of explicit verbal cues used to manipulate emotion attribution in the Explicit Task. Thus, before introducing participants to the Explicit Task, we asked them to observe the experimental stimuli while performing a minimally demanding 1-back task. In this Implicit Task, we found that human and nonhuman facial expressions both elicited activity in a distributed network of brain regions including those regions associated with emotion attribution in the subsequent Explicit Task. Of course, this does not permit the reverse inference that participants were making emotion attributions while performing the 1-back task. Rather, it reinforces the conclusion that the similar functional responses observed for human and nonhuman targets in the Explicit Task were not merely the product of the strong demand characteristics imposed by experimental protocol. Thus, to answer our second research question: *Yes*, when observing nonhuman animal facial expressions, participants activate this mechanism spontaneously even in the absence of explicit verbal cues to attribute emotion.

Our final research question built on extant research demonstrating that the tendency to attribute emotion to nonhuman animals is a measurable trait that shows considerable variability in the general population (Waytz et al., 2010). In a set of exploratory analyses of both univariate contrast in our a priori ROIs and multivariate response pattern similarity across the whole-brain, we found evidence consistent with the proposition that individuals who are more dispositionally prone to attribute mental states to nonhuman animals will be more similar in their neural responses to humans and nonhuman animals. Thus, to provide a preliminary answer to our third and final research question: *Yes*, the extent to which humans and nonhuman animals spontaneously produce similar neural responses appears to be somewhat related to individual differences in beliefs about the mental capacities of nonhuman animal species.

Future studies with larger sample sizes should further investigate individual differences in animal emotion attribution as they appear in both typically developing populations and in psychiatric disorders. Such studies would profit by considering other idiographic variables that influence how individuals perceive and think about nonhuman animals, such as adhering to a vegetarian or vegan diet for ethical reasons. Doing so may shed light on the mechanisms by which human-animal relationships, in particular, pet

ownership, can have positive effects on mental health and symptom improvement in a wide variety of disorders (Matchock, 2015).

Finally, it is well known that fMRI can distinguish separable neural processes, but cannot provide definitive conclusions regarding similar neural processes. It is thus possible that emotion attribution for human faces, nonhuman primate faces, and dog faces, recruit different neural mechanisms. While we consider this possibility unlikely, it could arise from separate but intermingled populations of neurons within the same brain regions. Future fMRI studies using multivoxel analyses or adaptation protocols could further test this possibility, as could single-unit recordings from neurosurgical patients (although with the exception of dmPFC, the regions we found are rarely implanted with electrodes).

We used a novel adaptation of the Yes/No Why/How Task (Spunt & Adolphs, 2014) to examine the neural basis of a common form of anthropomorphism, namely, the attribution of complex emotional states to the facial expressions of nonhuman animals. By comparing such attributions to those made about the facial expressions of humans, we were able to show that the attribution of emotion to nonhuman animals relies on the same executive processes already known to be critical for understanding the behavior other humans in terms of mental states (Spunt et al., 2011; Spunt & Lieberman, 2012a), and is consistent with our recent work suggesting that these processes can be flexibly deployed to understand the reasons for phenomena in nonsocial domains (Spunt & Adolphs, 2015). Thus, attribution of emotion to nonhuman animals may represent one of the many possible expansions humans have made on their capacity to reason about the causes of human behavior. More broadly and in the spirit of Darwin's seminal treatise on human and nonhuman animal emotion, our findings here further encourage a non-anthropocentric view of emotion understanding, one that treats the idea that animals have emotions as no more gratuitous than the idea that humans other than ourselves do.

Acknowledgments

The Authors would like to acknowledge Mike Tyszka and the Caltech Brain Imaging Center for help with the neuroimaging; the Della Martin Foundation for postdoctoral fellowship support to R.P.S.; Samuel P. and Frances Krown for sponsoring a Summer Undergraduate Research Fellowship to E. E; and three anonymous reviewers for helpful comments. This work was supported in part by the National Institutes of Health (R01 MH080721-03 to R.A.). Additional funding was provided by the Caltech Conte Center for Social Decision-Making.

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. Nat Rev Neurosci, 7(4), 268-277. doi:10.1038/nrn1884

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. Neuroimage, 38(1), 95-113. doi:10.1016/j.neuroimage.2007.07.007

Axelsson, E., Ratnakumar, A., Arendt, M. L., Maqbool, K., Webster, M. T., Perloski, M., . . . Lindblad-Toh, K. (2013). The genomic signature of dog domestication reveals adaptation to a starch-rich diet. Nature, 495(7441), 360-364. doi:10.1038/nature11837

Barker, D., & Miller, D. (1990). Hurricane gilbert: Anthropomorphising a natural disaster. Area, 22(2), 107-116. Retrieved from http://www.jstor.org/stable/20002812

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of statistics, 1165-1188. Retrieved from http://www.jstor.org/stable/2674075

Berns, G. S., Brooks, A. M., & Spivak, M. (2012). Functional MRI in awake unrestrained dogs. PLoS One, 7(5), e38027. doi:10.1371/journal.pone.0038027

Blonder, L. X., Smith, C. D., Davis, C. E., Kesler-West, M. L., Garrity, T. F., Avison, M. J., & Andersen, A. H. (2004). Regional brain response to faces of humans and dogs. Brain Res Cogn Brain Res, 20(3), 384-394. doi:10.1016/j.cogbrainres.2004.03.020

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433-436. doi:10.1163/156856897X00357

Brecht, M., & Freiwald, W. A. (2012). The many facets of facial interactions in mammals. Curr Opin Neurobiol, 22(2), 259-266. doi:10.1016/j.conb.2011.12.003

Cuaya, L. V., Hernández-Pérez, R., & Concha, L. (2016). Our faces in the dog's brain: Functional imaging reveals temporal cortex activation during perception of human faces. PLoS One, 11(3), e0149431. doi:10.1371/journal.pone.0149431

Cullen, H., Kanai, R., Bahrami, B., & Rees, G. (2014). Individual differences in anthropomorphic attributions and human brain structure. Soc Cogn Affect Neurosci, 9, 1276-1280. doi:10.1093/scan/nst109

Darwin, C. (1872). The expression of the emotions in man and animals. London: J. Murray.

de Waal, F. B. (2011). What is an animal emotion. Ann N Y Acad Sci, 1224, 191-206. doi:10.1111/j.1749-6632.2010.05912.x

Dennett, D. C. (1989). The intentional stance. Cambridge: The MIT Press.

Diedrichsen, J., & Shadmehr, R. (2005). Detecting and adjusting for artifacts in fmri time series data. Neuroimage, 27(3), 624-634. doi:10.1016/j.neuroimage.2005.04.039

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. Psychol Rev, 114(4), 864-886. doi:10.1037/0033-295X.114.4.864

Fiske, S. T., Cuddy, A. J., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. Trends Cogn Sci, 11(2), 77-83. doi:10.1016/j.tics.2006.11.005

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S. J., & Frith,

C. D. (1995). Other minds in the brain: A functional imaging study of “theory of mind” in story comprehension. Cognition, 57(2), 109-128. doi:10.1016/0010-0277(95)00692-R

Franklin, R. G., Nelson, A. J., Baker, M., Beeney, J. E., Vescio, T. K., Lenz-Watson, A., & Adams, R. B. (2013). Neural responses to perceiving suffering in humans and animals. Soc Neurosci, 8(3), 217-227. doi:10.1080/17470919.2013.763852

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. Proceedings from Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ‘theory of mind’. Trends Cogn Sci, 7(2), 77-83. doi:10.1016/S1364-6613(02)00025-6

Goel, V., Grafman, J., Sadato, N., & Hallett, M. (1995). Modeling other minds. NeuroReport, 6(13), 1741-1746. doi:10.1097/00001756-199509000-00009

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. Science, 315(5812), 619. doi:10.1126/science.1134475

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fmri designs. Neuroimage, 43(3), 509-520. doi:10.1016/j.neuroimage.2008.07.065

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., . . . Frith, C. (1996). ‘theory of mind’ in the brain. Evidence from a pet scan study of asperger syndrome. NeuroReport, 8(1), 197-201. doi:10.1097/00001756-199612200-00040

Hayama, S., Chang, L., Gumus, K., King, G. R., & Ernst, T. (2016). Neural correlates for perception of companion animal photographs. Neuropsychologia, 85, 278-286. doi:10.1016/j.neuropsychologia.2016.03.018

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. The American Journal of Psychology, 57(2), 243. doi:10.2307/1416950

Hills, A. M. (1995). Empathy and belief in the mental experience of animals. Anthroz Jour Inter Peo Ani, 8(3), 132-142. doi:10.2752/089279395787156347

Keysers, C., & Gazzola, V. (2014). Dissociating the ability and propensity for empathy. Trends Cogn Sci, 18(4), 163-166. doi:10.1016/j.tics.2013.12.011

Kujala, M. V., Kujala, J., Carlson, S., & Hari, R. (2012). Dog experts’ brains distinguish socially relevant body postures similarly in dogs and humans. PLoS One, 7(6), e39145. doi:10.1371/journal.pone.0039145

Matchock, R. L. (2015). Pet ownership and physical health. Curr Opin Psychiatry, 28(5), 386-392. doi:10.1097/YCO.0000000000000183

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. Neuroimage, 25(3), 653-660. doi:10.1016/j.neuroimage.2004.12.005

Parr, L. A., Waller, B. M., Vick, S. J., & Bard, K. A. (2007). Classifying chimpanzee facial expressions using muscle action. Emotion, 7(1), 172-181. doi:10.1037/1528-3542.7.1.172

Paul, E. S. (2000). Empathy with animals and with humans: Are they linked. Anthroz Jour Inter Peo Ani, 13(4), 194-202. doi:10.2752/089279300786999699

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. Annu Rev Psychol, 55, 87-124. doi:10.1146/annurev.psych.55.090902.142044

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

Sherwood, C. C., Holloway, R. L., Gannon, P. J., Semendeferi, K., Erwin, J. M., Zilles, K., & Hof, P. R. (2003). Neuroanatomical basis of facial expression in monkeys, apes, and humans. Ann N Y Acad Sci, 1000, 99-103. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14766625

Spunt, R. P., & Adolphs, R. (2014). Validating the why/how contrast for functional mri studies of theory of mind. Neuroimage, 99, 301-311. doi:10.1016/j.neuroimage.2014.05.023

Spunt, R. P., & Adolphs, R. (2015). Folk explanations of behavior: A specialized use of a domain-general mechanism. Psychol Sci, 26(6), 724-736. doi:10.1177/0956797615569002

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. Psychol Sci, 21(11), 1593-1598. doi:10.1177/0956797610386618

Spunt, R. P., Kemmerer, D., & Adolphs, R. (2016). The neural basis of conceptualizing the same action at different levels of abstraction. Soc Cogn Affect Neurosci, 11(7), 1141-1151. doi:10.1093/scan/nsv084

Spunt, R. P., & Lieberman, M. D. (2012a). An integrative model of the neural systems supporting the comprehension of observed emotional behavior. NeuroImage, 59(3), 3050-3059. doi:10.1016/j.neuroimage.2011.10.005

Spunt, R. P., & Lieberman, M. D. (2012b). Dissociating modality-specific and supramodal neural systems for action understanding. J Neurosci, 32(10), 3575-3583. doi:10.1523/JNEUROSCI.5715-11.2012

Spunt, R. P., Meyer, M. L., & Lieberman, M. D. (2015). The default mode of human brain function primes the intentional stance. J Cogn Neurosci, 27(6), 1116-1124. doi:10.1162/jocn_a_00785

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: An fmri study of mentalizing and mechanizing during action observation. J Cogn Neurosci, 23(1), 63-74. doi:10.1162/jocn.2010.21446

Templer, D. I., Salter, C. A., Dickey, S., Baldwin, R., & Veleber, D. M. (1981). The construction of a pet attitude scale. Psycholog Record, 31(3), 343-348. Retrieved from http://psycnet.apa.org/psycinfo/1982-06859-001

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. Trends Cogn Sci, 12(12), 455-460. doi:10.1016/j.tics.2008.10.001

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. Psychol Rev, 94(1), 3-15. doi:10.1037/0033-295X.94.1.3

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. NeuroImage, 54(2), 1589-1599. doi:10.1016/j.neuroimage.2010.09.043

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. Perspect Psychol Sci, 5(3), 219-232. doi:10.1177/1745691610369336

Wynne, C. D. (2004). The perils of anthropomorphism. Nature, 428(6983), 606. doi:10.1038/428606a

Table 1

Verbal cues used in the Explicit Task to experimentally control the incidence of emotion attribution when observing the facial expressions of Humans, Nonhuman Primates, and Dogs. Emotion cues prompted participants to evaluate the emotional state implied by the target expression, while Expression cues prompted participants to evaluate a factual statement about the target expression itself. Each cue was paired with six targets from each stimulus category. Every target appeared twice during the Explicit Task, once paired with an Emotion cue and once with a Verbal cue. Thus, the Emotion > Expression contrast is attentional.

Attentional Focus	
Emotion	*Expression*
annoyed?	baring teeth?
bored?	gazing up?
confident?	looking at the camera?
excited?	mouth closed?
reflective?	mouth open?

Table 2

Means and standard deviations (parenthetically) summarizing the frequency and response time (RT) with which participants accepted normative Emotion and Expression cues for each of the three targets in the Explicit Task, as well as their ratings of each stimulus target collected post-task. Participants indicated the extent to which they understood what each target was feeling (*Understanding*), and the extent to which each target made them feel good versus bad (*Valence*). See the main text further details, and **Table S2** for statistical analysis of these outcomes.

	Acceptance (%)		Acceptance RT (ms)		Stimulus Ratings (Post-Task)	
	Emotion	*Expression*	*Emotion*	*Expression*	*Understanding*	*Valence*
Human	97.19 (4.46)	98.12 (3.10)	914 (126)	789 (99)	6.93 (1.03)	5.31 (.48)
Primate	94.95 (5.20)	89.95 (6.82)	960 (112)	849 (108)	5.83 (.91)	5.07 (.62)
Dog	84.80 (12.18)	95.24 (6.98)	948 (114)	848 (94)	6.05 (1.11)	5.30 (.47)

Table 3

Outcomes of region of interest (ROI) tests on the within-stimulus comparisons from the Explicit and Implicit Tasks. ROIs that showed an effect across all stimulus categories are marked with an asterisk. P-values after adjusting for multiple ROIs using a false-discovery rate procedure. Further details on the ROIs used are provided in the main text and **Table S1**. L = Left; LOFC = Lateral Orbitofrontal Cortex; TPJ = Temporoparietal Junction; aSTS = Anterior Superior Temporal Sulcus; dmPFC = Dorsomedial Prefrontal Cortex.

Contrast Name	Humans				Nonhuman Primates				Dogs			
ROI Label	t-stat	*pFDR*	95% CI		t-stat	*pFDR*	95% CI		t-stat	*pFDR*	95% CI	
Explicit Task: Emotion > Expression												
LOFC*	5.415	<.001	0.34	0.68	5.357	<.001	0.25	0.51	7.531	<.001	0.41	0.69
TPJ	0.203	1.000	-0.17	0.24	1.896	0.161	-0.01	0.29	1.958	0.144	0.01	0.30
aSTS	3.532	0.013	0.09	0.31	2.086	0.151	0.00	0.21	3.168	0.018	0.07	0.26
dmPFC*	3.192	0.017	0.13	0.52	5.668	<.001	0.32	0.63	6.913	<.001	0.36	0.64
Implicit Task: Face > Scramble												
LOFC*	2.911	0.021	0.12	0.51	2.788	0.027	0.11	0.54	2.720	0.032	0.05	0.29
TPJ*	4.171	0.002	0.26	0.68	2.798	0.027	0.10	0.48	2.878	0.030	0.13	0.53
aSTS*	5.777	<.001	0.20	0.39	3.734	0.008	0.08	0.25	4.821	0.002	0.09	0.21
dmPFC*	5.620	<.001	0.42	0.86	4.364	0.004	0.35	0.88	2.980	0.030	0.13	0.61

Table 4

Outcomes of region of interest (ROI) tests examining contrasts of human to nonhuman targets in the Explicit and Implicit Tasks. ROIs showing effects in both nonhuman targets are marked with an asterisk. P-values are adjusted for multiple ROIs using a false-discovery rate procedure. Further details on ROIs are provided in the main text and **Table S1**. L = Left; LOFC = Lateral Orbitofrontal Cortex; TPJ = Temporoparietal Junction; aSTS = Anterior Superior Temporal Sulcus; dmPFC = Dorsomedial Prefrontal Cortex.

Contrast Name	Human > Primate				Human > Dog			
ROI Label	t-stat	*pFDR*	95% CI		t-stat	*pFDR*	95% CI	
Explicit Task: $Human_{Emotion>Expression}$ > $Nonhuman_{Emotion>Expression}$								
LOFC	-0.156	1.000	-0.22	0.17	-1.038	0.877	-0.42	0.12
TPJ	-1.862	0.343	-0.27	-0.01	-1.102	0.877	-0.35	0.10
aSTS	2.013	0.343	0.01	0.17	0.629	1.000	-0.08	0.16
dmPFC	-1.157	0.741	-0.25	0.05	-1.401	0.877	-0.44	0.07
Explicit Task: $Human_{Emotion+Expression}$ > $Nonhuman_{Emotion+Expression}$								
LOFC	0.380	1.477	-0.08	0.14	0.543	1.335	-0.06	0.10
TPJ	4.469	0.002	0.11	0.29	0.851	1.335	-0.04	0.12
aSTS*	5.352	0.001	0.12	0.24	3.659	0.021	0.06	0.20
dmPFC	2.971	0.026	0.05	0.22	0.476	1.335	-0.07	0.12
Implicit Task: Human > Nonhuman								
LOFC	-0.256	1.000	-0.08	0.06	1.077	0.620	-0.04	0.11
TPJ	2.257	0.160	0.01	0.11	3.002	0.070	0.03	0.13
aSTS	4.334	0.004	0.03	0.08	2.667	0.070	0.01	0.11
dmPFC	0.372	1.000	-0.03	0.06	1.789	0.257	-0.02	0.13

Table 5

Outcomes of exploratory between-subject ROI analyses of variability in spontaneous activation in the Implicit Task, and in the similarity of each participant's whole-brain (gray-matter masked) response pattern for human faces to their response pattern for each of the nonhuman animal face conditions. We extracted signal from each ROI in the Primate > Scramble and Dog > Scramble contrasts and computed the Pearson correlation with the scores derived from the post-task emotion understanding ratings and scales on the animal-specific subscale of the Individual Differences in Anthropomorphism Questionnaire (IDAQ). Response pattern similarity was calculated using the Pearson correlation of the t-statistic images representing the response pattern for each Target in the Implicit Task, and then subsequently Fisher's z-transformed for the between-subject analysis. This produced two measures of human/nonhuman neural similarity for each participant, which we also correlated with the collected post-task measures. L = Left; OFC = Lateral Orbitofrontal Cortex; TPJ = Temporoparietal Junction; aSTS = Anterior Superior Temporal Sulcus; dmPFC = Dorsomedial Prefrontal Cortex.

		Primate > Scramble				*Dog > Scramble*				*Similarity to Human*	
		LOFC	TPJ	aSTS	dmPFC	LOFC	TPJ	aSTS	dmPFC	Primate	Dog
Target-Specific Emotion Understanding	*Humans*	-.19	.03	-.25	-.31	-.37	-.41	-.32	-.31	.46	.45
	Primates	.14	.03	.15	.14	-.25	-.09	-.03	-.43	.65**	.56*
	Dogs	.15	.22	-.10	-.05	-.07	-.41	-.20	-.25	.60*	.53*
IDAQ (Animals Subscale)		.37	.23	.10	.54*	.32	.18	.49	.11	.54*	.41

Note: *p < .05, **p < .01

Figure 1

Experimental design. **(a)** In the Implicit Task, participants perform a visual 1-back on a series of naturalistic images showing human, nonhuman primate, and dog facial displays. Images appear in an event-related design intermixed with a phase-scrambled subset of the same images which provided a baseline for univariate contrasts. We refer to the task as "implicit" only to designate that at the time of performing the task, participants were not explicitly directed to attend to or think about the images in a particular way, and were naïve to the fact that in a subsequent task they would be asked to consider the emotional states of each target. **(b)** The sequence of screens from an example trial in the Explicit Task, which participants learn about only after completing the Implicit Task. The task features the same images used in the Implicit Task (excluding scrambles). Each image is presented twice, once preceded by a verbal cue directing participants to accept or reject an emotion attribution about the target, and once by a verbal cue directing participants to accept or reject an expression attribution about the target (all verbal cues presented in **Table 1**). Once the image appears, participants have 1750 ms to commit a *Yes*/*No* manual response. Every cue preceded an equal number of images from each target type and elicited a response of either *Yes* (2/3 of trials) or *No* (1/3 of trials) in the majority of respondents in an independent sample.

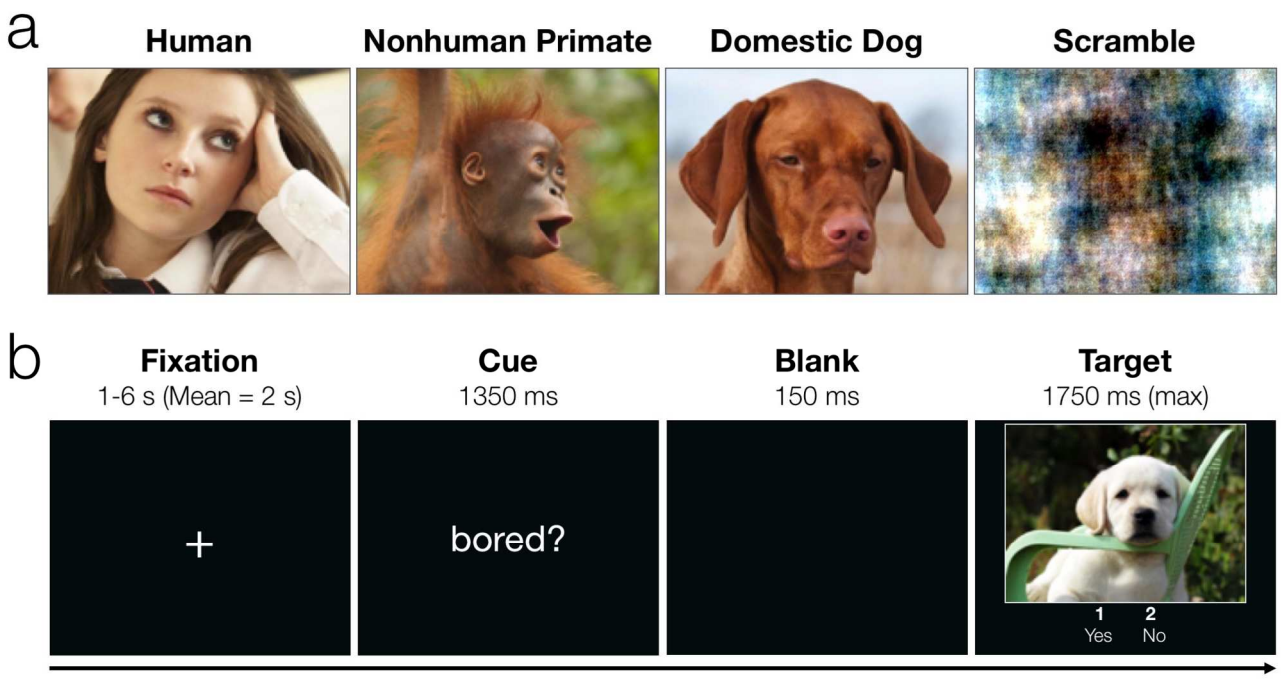


Figure 2

Percent signal change in *a priori* regions of interest (ROI). For each ROI, the leftmost set of bars represent the mean response across voxels (relative to fixation baseline) to the six conditions of the Explicit Task; the rightmost set of bars represent the mean response across voxels to each target in the Implicit Task (relative to the response to the scramble stimulus condition). For further details on the ROIs, see the main text and **Table S1**. Statistical tests corresponding to the data plotted here are presented in **Table 3 and Table 4**. OFC = Orbitofrontal Cortex; TPJ = Temporoparietal Junction; aSTS = Anterior Superior Temporal Sulcus; dmPFC = Dorsomedial Prefrontal Cortex.

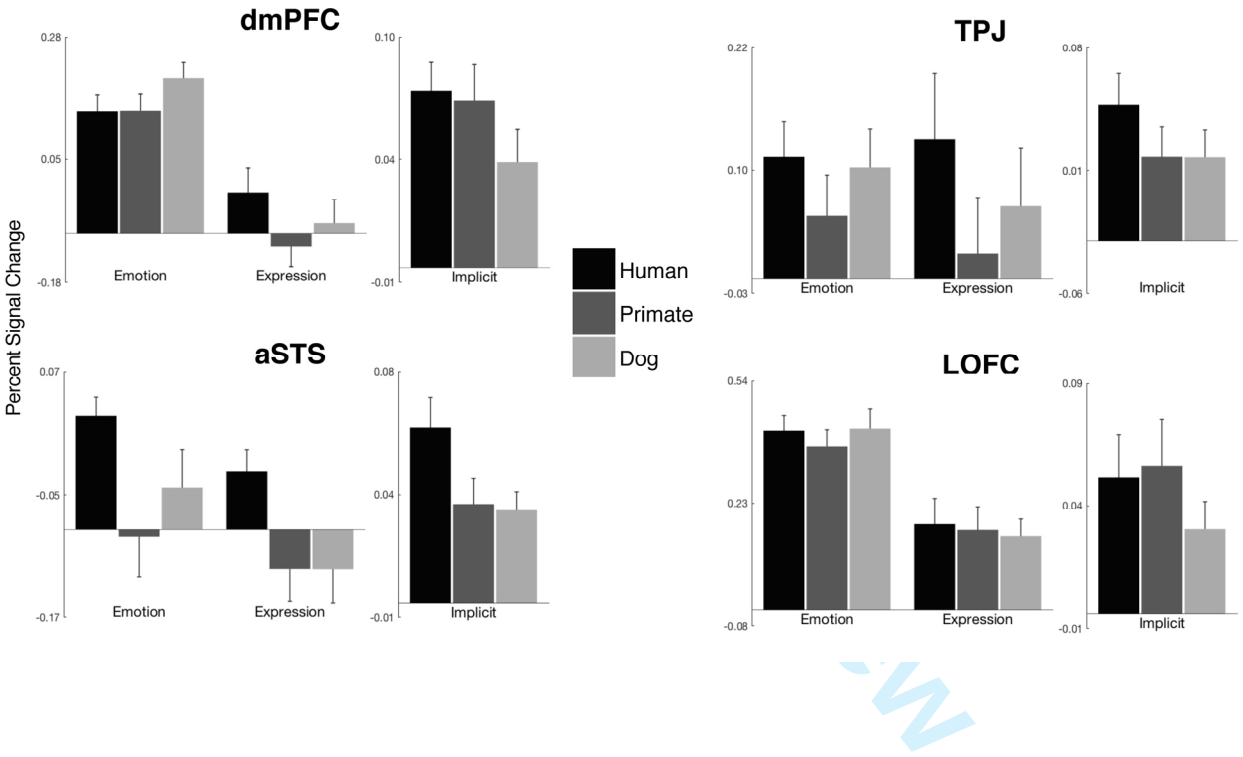


Figure 3

Whole-brain analysis of target-independent effects. Statistical maps are cluster-level corrected at a familywise error rate of 0.05. **(a)** *Explicit Task*. Regions showing significantly positive or negative responses in the *Emotion* > *Expression* contrast for all targets. (b) *Implicit Task*. Regions showing a significantly positive response in the *Face* > *Scramble* contrast for all targets. **(c)** *Explicit/Implicit Task Conjunction*. The region of dorsomedial prefrontal cortex showing a target independent-effect in both tasks.

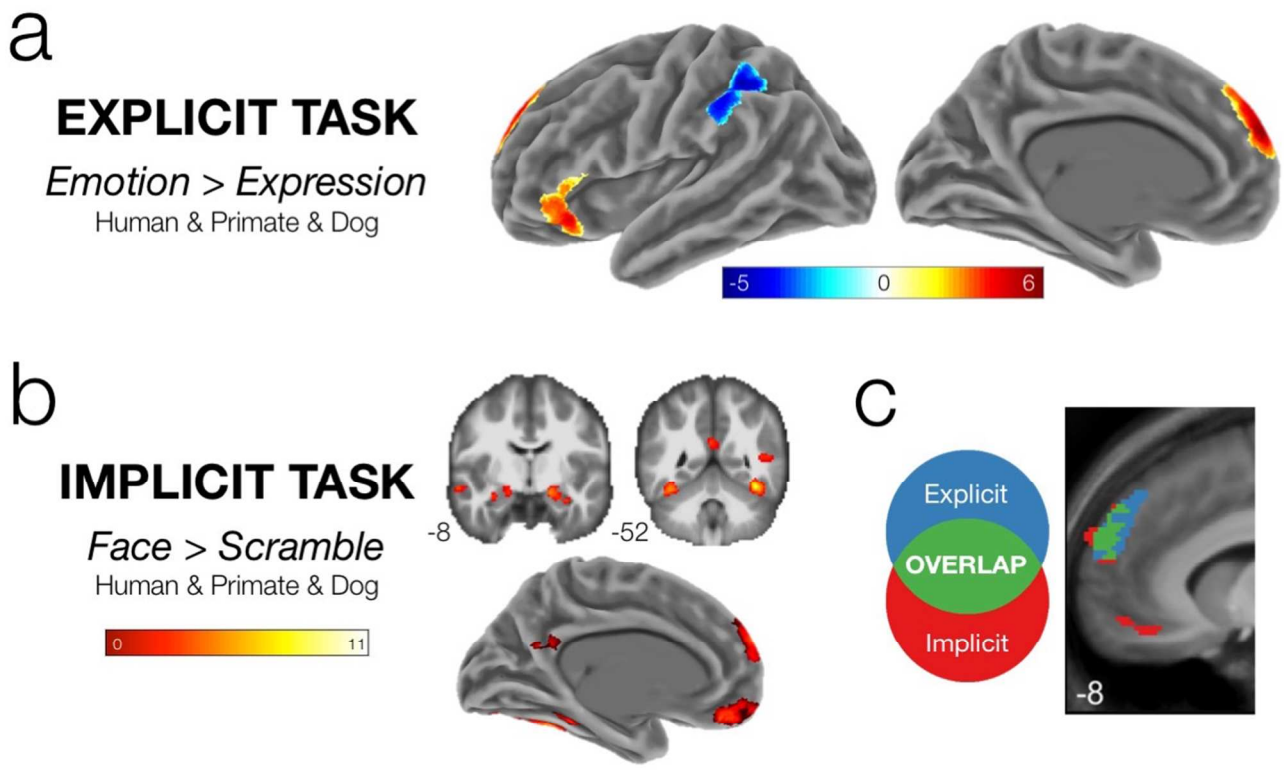
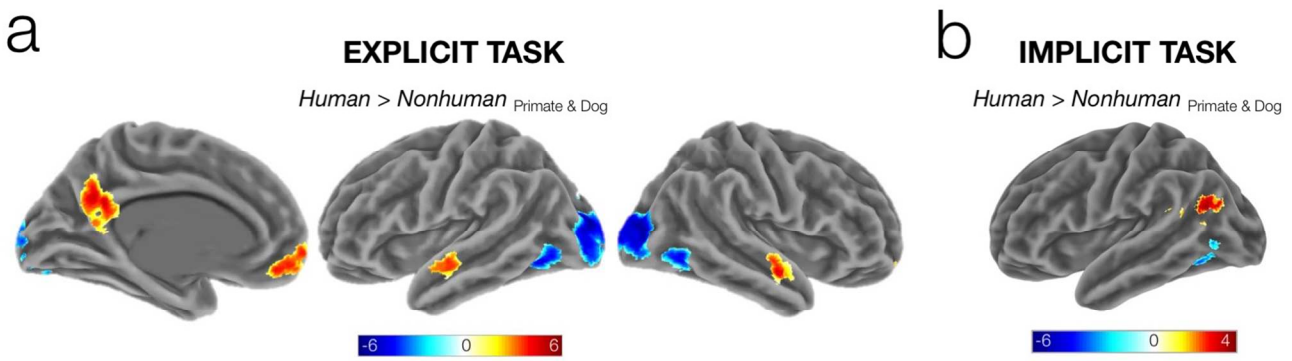


Figure 4

Whole-brain analysis of target-dependent effects. Statistical maps are cluster-level corrected at a familywise error rate of 0.05. **(a)** *Explicit Task*. Regions showing a differential response to Human targets relative to both Nonhuman targets (collapsing the *Emotion/Expression* factor). **(b)** *Implicit Task.* Regions showing a differential response to Human targets relative to both Nonhuman targets.



Supplementary Materials for

The Neural Basis of Understanding the Expression of the Emotions in Man and Animals

Robert P. Spunt, Emily Ellsworth, Ralph Adolphs
California Institute of Technology

Corresponding Author:
Robert P. Spunt
California Institute of Technology
1200 E. California Blvd.
Pasadena, CA 91125
Email: spunt@caltech.edu

Table S1

Details for the set of left hemisphere regions of interest (ROIs) used to test our primary hypotheses.

		Peak MNI Coordinates		
ROI Name	*Extent*	*x*	*y*	*z*
Lateral Orbitofrontal Cortex (LOFC)	516	-48	27	-6
Temporoparietal Junction (TPJ)	521	-48	-66	30
Anterior Superior Temporal Sulcus (aSTS)	650	-57	-9	-18
Dorsomedial Prefrontal Cortex (dmPFC)	1000	-6	57	36

Table S2

Repeated measures analysis of variance for behavioral outcomes of interest. See **Table 2** in the main text for descriptive data summarizing the outcomes.

Outcome	Effect	SS	df	MS	F	p	*Partial* η^2	95% CI	
Acceptance Rate									
	Within-Subjects								
	- Question	0.008	1	0.008	0.039	0.847	0.003	0.00	0.36
	- Target	0.453	2	0.226	2.426	0.106	0.139	0.01	0.45
	- Question*Target	0.222	2	0.111	1.354	0.274	0.083	0.01	0.35
	Within-Cells	12.137	90	0.135					
	- Subjects	3.935	15	0.262					
	- Question*Subj	2.942	15	0.196					
	- Target*Subj	2.800	30	0.093					
	- Question*Target*Subj	2.460	30	0.082					
	Total	12.819	95	0.135					
Acceptance Response Time									
	Within-Subjects								
	- Question	0.300	1	0.300	50.714	0.000	0.772	0.68	0.88
	- Target	0.053	2	0.027	11.031	0.000	0.424	0.24	0.71
	- Question*Target	0.003	2	0.001	0.776	0.469	0.049	0.01	0.29
	Within-Cells	1.074	90	0.012					
	- Subjects	0.858	15	0.057					
	- Question*Subj	0.089	15	0.006					
	- Target*Subj	0.072	30	0.002					
	- Question*Target*Subj	0.055	30	0.002					
	Total	1.430	95	0.015					
Post-Task Emotion Understanding Ratings									
	Within-Subjects	14.626	2	7.313	22.824	0.000	0.603	0.41	0.79
	- Human > Primate	11.915	1	11.915	37.187	0.000	0.553	0.36	0.74
	- Human > Dog	9.933	1	9.933	31.003	0.000	0.508	0.28	0.75
	- Primate > Dog	0.090	1	0.090	0.281	0.600	0.009	0.00	0.15
	Within-Groups	44.854	45	0.997					
	- Subj	35.242	15	2.349					
	- Group X Subj	9.612	30	0.320					
	Total	59.479	47	1.266					
Post-Task Valence Ratings									

Within-Subjects	0.626	2	0.313	1.434	0.254	0.087	0.01	0.35
- Human > Primate	0.606	1	0.606	2.780	0.106	0.085	0.00	0.33
- Human > Dog	0.260	1	0.260	1.192	0.284	0.038	0.00	0.22
- Primate > Dog	0.072	1	0.072	0.331	0.569	0.011	0.00	0.22
Within-Groups	13.515	45	0.300					
- Subj	6.971	15	0.465					
- Group X Subj	6.544	30	0.218					
Total	14.141	47	0.301					

Table S3

Whole-brain results for contrasts used to test the conjunction hypotheses for both the Explicit and Implicit tasks. Clusters were identified using a cluster-forming threshold of $p < .001$ (uncorrected) and a cluster-level family-wise error rate of 0.05. Within each cluster, we report a maximum of 3 local maxima separated by at least 20 mm. x, y, and z = Montreal Neurological Institute (MNI) coordinates in the left-right, anterior-posterior, and inferior-superior dimensions, respectively. L = Left; R = Right; LOFC = Lateral Orbitofrontal Cortex; aSTS = Anterior Superior Temporal Sulcus; PFC = Prefrontal Cortex; PCC = Posterior Cingulate Cortex.

Contrast Name			MNI Coordinates		
Cluster Label	Extent	t-value	x	y	z
$Human_{Emotion>Expression}$ & $Primate_{Emotion>Expression}$ & $Dog_{Emotion>Expression}$					
L Dorsomedial PFC	481	6.256	-8	56	28
L Lateral Orbitofrontal Cortex	357	5.745	-46	26	-14
$Human_{Expression>Emotion}$ & $Primate_{Expression>Emotion}$ & $Dog_{Expression>Emotion}$					
L rostral Inferior Parietal Lobule	214	5.896	-32	-44	40
L SupraMarginal Gyrus	236	4.732	-58	-34	46
$Human_{Emotion>Expression}$ > $Primate_{Emotion>Expression}$ & $Human_{Emotion>Expression}$ > $Dog_{Emotion>Expression}$					
No suprathreshold clusters					
$Primate_{Emotion>Expression}$ > $Human_{Emotion>Expression}$ & $Dog_{Emotion>Expression}$ > $Human_{Emotion>Expression}$					
No suprathreshold clusters					
$Human_{Emotion+Expression}$ > $Primate_{Emotion+Expression}$ & $Human_{Emotion+Expression}$ > $Dog_{Emotion+Expression}$					
L PCC	632	6.003	-4	-54	26
R aSTS	228	5.177	52	-8	-16
R Ventromedial PFC	473	4.906	8	56	-12
L aSTS	149	4.070	-52	-12	-16
$Primate_{Emotion+Expression}$ > $Human_{Emotion+Expression}$ & $Dog_{Emotion+Expression}$> $Human_{Emotion+Expression}$					
R Middle Occipital Gyrus	1008	6.166	34	-88	10
L Middle Occipital Gyrus	1072	6.082	-28	-92	12
L Fusiform Gyrus	145	5.106	-40	-64	-8

R Inferior Temporal Gyrus	187	5.064	48	-60	-6
$Human_{Face>Scramble}$ & $Primate_{Face>Scramble}$ & $Dog_{Face>Scramble}$					
R Fusiform Gyrus	2019	11.790	38	-48	-16
L Inferior Occipital Gyrus	1989	8.440	-36	-82	-4
L Ventromedial PFC	366	6.928	-4	44	-18
R Amygdala/ParaHippocampal Gyrus	188	6.357	18	-8	-14
L Dorsomedial PFC	162	5.994	-8	60	26
L Amygdala/ParaHippocampal Gyrus	140	5.718	-22	-12	-12
L Precuneus	88	5.159	-2	-52	22
$Human_{Face}$ > $Primate_{Face}$ & $Human_{Face}$ > Dog_{Face}					
No suprathreshold clusters					
$Primate_{Face}$ > $Human_{Face}$ & Dog_{Face} > $Human_{Face}$					
L Fusiform Gyrus	134	6.382	-30	-64	-4

Figure S1

The set of human facial expressions used in both the Explicit and Implicit tasks.

Figure S2

The set of non-human primate facial expressions used in both the Explicit and Implicit tasks.

Figure S3

The set of dog facial expressions used in both the Explicit and Implicit tasks.