Hear them roar: A comparison of black-capped chickadee (*Poecile atricapillus*) and human (*Homo sapiens*) perception of arousal in vocalizations across all classes of terrestrial vertebrates

Jenna V. Congdon[ac], Allison H. Hahn[d]*, Piera Filippi[efgh]*, Kimberley A. Campbell[ac]*, John Hoang[c], Erin N. Scully[ac], Daniel L. Bowling[h], Stephan A. Reber[i], & Christopher B. Sturdy[abc]

Department of Psychology[a], Neuroscience and Mental Health Institute[b], University of Alberta[c]; Department of Psychology, St. Norbert College[d]; Institute of Language, Communication and the Brain, Aix-en-Provence[e]; Laboratoire Parole et Langage LPL UMR 7309, CNRS, Aix-Marseille University[f]; Laboratoire de Psychologie Cognitive LPC UMR 7290, CNRS, Aix-Marseille University[g]; Department of Cognitive Biology, University of Vienna[h]; Department of Philosophy, Lund University[i]

Corresponding author: Christopher B. Sturdy, csturdy@ualberta.ca

Recently, evidence for acoustic universals in vocal communication was found by demonstrating that humans can identify levels of arousal in vocalizations produced by species across three biological classes (Filippi et al., 2017). Here, we extend this work by testing whether two vocal learning species, humans and chickadees, can discriminate vocalizations of high and low arousal using operant discrimination go/no-go tasks. Stimuli included vocalizations from nine species: Giant panda, American alligator, common raven, hourglass treefrog, African elephant, Barbary macaque, domestic pig, black-capped chickadee, and human. Subjects were trained to respond to high or low arousal vocalizations, then tested with additional high and low arousal vocalizations produced by each species. Chickadees (Experiment 1) and humans (Experiment 2) learned to discriminate between high and low arousal stimuli and significantly transferred the discrimination to additional panda, human, and chickadee vocalizations. Finally, we conducted discriminant function analyses using four acoustic measures, finding evidence suggesting that fundamental frequency played a role in responding during the task. However, these analyses also suggest roles for other acoustic factors as well as familiarity. In sum, the results from these studies provide evidence that chickadees and humans are capable of perceiving arousal in vocalizations produced by multiple species.

Keywords: arousal; vocalizations; black-capped chickadee; human

Introduction

In *The Descent of Man*, Darwin proposed that our earliest terrestrial ancestors expressed emotion through their vocalizations (Darwin, 1871). This century-old hypothesis still underlies many theories of emotional communication today, highlighting the importance of understanding the specific information contained within vocalizations and the evolutionary origins of a trait critical to our survival (i.e., shared mechanisms for vocal emotional expression). Organisms produce acoustic signals in response to specific stimuli to convey information (Aubin & Jouventin, 2002). For example, signals may include the production of aggressive calls by hourglass treefrogs (*Dendrosophys ebraccatus*) during competitive signaling interactions with other males (Reichert, 2011), mobbing calls produced by black-capped chickadees (*Poecile atricapillus*) when they encounter predators (Templeton, Greene, & Davis, 2005), or distress calls produced by domestic pigs (*Sus scrofa domesticus*), all of which comprise information about an individual’s emotional state and potential intentions (e.g., mating intentions, predator threat, distress; Linhart, Ratcliffe, Reby, & Spinka, 2015). Acoustic signals can vary based on different environmental contexts and adopt a wide range of functions in mate choice, potential threat perception, and species recognition in a broad range of taxa (Gerhardt, Humfeld, & Marshall, 2007). In diverse species, including insects, amphibians, birds, and mammals, acoustic signals convey a variety of cues: caller identity, location, social context, emotional state, current condition, and developmental state (Maynard-Smith & Harper, 2004). Since the vocalizations of many taxa share similar functions, vocal characteristics are especially well-suited to investigate both within and between species similarities (i.e., do animals produce, and perceive signals that share overall meaning within and across species?). Moreover, some phylogenetic comparative studies support the idea that acoustic traits can predict patterns of diversification across genera by indicating lineage-specific acoustic differences (Seddon, Merrill, & Tobias, 2008). These studies suggest that organisms can identify the vocal signals not only within their own species, but also among others due to distinct characteristics. Overall, the capacity to perceive the information in acoustic signals is evolutionarily important because it allows individuals to respond appropriately in the face of specific environmental challenges (e.g., the presence of a predator), ultimately increasing their fitness.

Acoustic signals are produced in many contexts of intraspecies (i.e., conspecific) and interspecies (i.e., heterospecific) communication, which can be beneficial for the receiver as well as the signaler. For example, groups consisting of individuals of different species can benefit from alarm calls that warn of predators (Caro, 2005), as well as contact calls that maintain group cohesion by notifying nearby individuals of their location via the production of brief, soft short-range calls (Marler, 2004; e.g., *tseet* call of the black-capped chickadee: Smith, 1991). Black-capped chickadee mobbing calls have been observed to recruit conspecifics and other avian species (i.e., heterospecifics) to attack and harass nearby predators with mobbing behaviour (Templeton, Greene, & Davis, 2005), thus providing an example of both intra- and interspecies communication. There are many cases of species eavesdropping on heterospecifics' vocalizations; grey squirrels (*Sciurus carolinensis*) attend to the calls of blue jays (*Cyanocitta cristata*) in order to assess the risk of food caches being pilfered (Schmidt & Ostefeld, 2008), and migrating passerines assess the acoustic sounds of local species as an indicator of habitat quality (Mukhin et al., 2008).

Acoustic signals can reflect the signalers' physiological states, which result from the context of vocal production (Morton, 1977; Rendall et al., 2009). In particular, acoustic signals can contain information about an organism's state of arousal, which is the degree of physiological alertness or attentiveness (Russell, 1980); arousal plays a critical role in regulating attention, information processing, motivation, and emotional reactivity (Nesse, 1990). The ability for the receiver to discriminate arousal in vocalizations is especially important. While acoustic signals can provide details about the caller's species, sex, age, motivational state, dominance status, etc., cues expressing high arousal level convey immediately relevant information, such as the presence of a threat to the life of an organism (Stoeger, Baotic, Li, & Charlton, 2012). Arousal-based changes in physiology influence acoustic features of the voice such as duration and harmonic-to-noise ratio (HNR; i.e., harshness of a sound). For instance, infant African elephant roars primarily function to signal arousal levels, with duration and HNR being of particular importance (Stoeger, Charlton, Kratochvil, & Fitch, 2011). Detecting level of threat, or arousal, from another organism's vocalizations (conspecific or heterospecific) is perhaps one of the most important evolutionary functions for survival (e.g., red-breasted nuthatches, *Sitta canadensis*, eavesdrop on black-capped chickadees mobbing calls indicating a nearby predator; Templeton & Greene, 2007).

Recently, Filippi and colleagues (2017) found that humans (*Homo sapiens*) can identify levels of arousal in vocalizations produced by multiple species across three taxonomic classes: Amphibia, Mammalia, and Reptilia (including Aves). Thus, humans appear to have the ability to identify the emotional content contained in both conspecific and heterospecific vocalizations. So far, tests of this ability across such a wide range of species have only been made in humans. In order to better understand both the presence of and the ability to perceive these acoustic universals, we need to look beyond humans. Here, we seek to evaluate this ability in songbirds because their method of vocal acquisition parallels that of humans, suggesting that they may perceive or process vocalizations in a comparable manner produced through convergent evolution.

In Experiment 1, we tested black-capped chickadees, a non-migratory, North American songbird. Chickadees have been used as a comparative model as both songbirds and humans are vocal learners, meaning that the young of these animal groups learn their species-specific vocalizations, critical to their survival, from an adult model. These species belong to a relatively small group (i.e., humans, songbirds, hummingbirds, parrots, bats, elephants, and cetaceans; Jarvis, 2006) that are known to possess this ability. Therefore, we believe that there could be similarities in how these two evolutionarily-distinct species perceive and respond to auditory stimuli, specifically having similar abilities in perceiving arousal-based vocal cues communicated by the signaler. Thus, we investigated if black-capped chickadees and humans would respond similarly on a comparative task discriminating between high and low arousal vocalizations. Filippi and colleagues (2017) demonstrated that humans were capable of detecting arousal in vocalizations by directly asking participants to simultaneously differentiate between high and low arousal vocalizations (pairwise presentations), and their results provided evidence for acoustic universals. In the wild, animals would typically not hear two vocalizations (one of high arousal and one of low arousal) and then need to decide how to respond. As it is not natural to have to decide, from two vocalizations, which is higher arousal, it is more likely that animals would hear vocalizations and then have to judge if they are of high or low arousal to determine how they should respond to their surroundings (e.g., nearby predator). Considering how vocalizations would typically be heard in the wild, in the current study we presented vocalizations consecutively, such that subjects had to assess a single vocalization before hearing the next.

In summary, the current study used an operant conditioning go/no-go discrimination task, in which we could test chickadees (Experiment 1) and humans (Experiment 2) in an analogous manner (i.e., consecutive presentations with no instructions about the nature of the task) in order to determine the extent to which both species can discriminate vocalizations based on arousal levels. Considering that we tested humans again with this new paradigm, direct comparison to Filippi et al. (2017) is less critical. Overall, using a different task to ask a similar question also informs us about how robust the results are. In addition to testing, in order to investigate the mechanisms underpinning arousal perception in animal vocalizations, we explored which acoustic parameters predict category of vocalization (high or low arousal) across and within the vocalizing species included in this study.

Experiment 1: Black-capped Chickadees

Methods

Subjects

Twelve black-capped chickadees (eight males and four females, identified by DNA analysis; Griffiths, Double, Orr, & Dawson, 1998) were tested between September 2015 and April 2016. Birds at least one year of age (determined by examining the colour and shape of their outer tail retrices; Pyle, 1997) were captured in Edmonton (North Saskatchewan River Valley, 53.53˚N, 113.53˚W, Mill Creek Ravine, 53.52˚N, 113.47˚W), or Stony Plain (53.46˚N, 114.01˚W), Alberta, Canada between December 2010 and January 2015.

Prior to the experiment, birds were individually housed in Jupiter Parakeet cages (30 × 40 × 40 cm; Rolf C. Hagen, Inc., Montreal, QB, Canada) in colony rooms containing several other black-capped chickadees. Birds had visual and auditory, but not physical, contact with one another. Birds had *ad libitum* access to food (Mazuri Small Bird Maintenance Diet; Mazuri, St

Louis, MO, USA), water (vitamin supplemented on alternating days; Prime vitamin supplement; Hagen, Inc.), grit, and cuttlebone. Birds were given three to five sunflower seeds daily, one superworm (*Zophobas morio*) three times a week, and a mixture of greens (spinach or parsley) and eggs twice a week. Birds were maintained on a light:dark cycle that mimicked the natural light cycle for Edmonton, Alberta, Canada.

Throughout the experiment, birds were housed individually in operant chambers (see apparatus below), maintained on the natural light cycle for Edmonton, Alberta, and had *ad libitum* access to water (vitamin supplemented on alternate days), grit, and cuttlebone. Birds were given two superworms daily (one in the morning and one in the afternoon). Food (i.e., Mazuri) was only available as a reward for correct responding during the operant discrimination task. Birds had previous experience discriminating synthetic tones, musical chords, *fee-bee* songs, and/or black-capped chickadee call notes (McMillan et al., 2015; Hahn et al., 2016; Guillette, Hahn, Hoeschele, Przyslupski, & Sturdy, 2015; respectively), but no operant experience with the stimuli used in this experiment.

Ethical Note

Throughout Experiment 1, birds remained in the testing apparatus to minimize the transport and handling of each bird. Following the experiment, birds were returned to the colony room for use in future experiments. Birds remained healthy during the experiment. All animal studies were conducted in accordance with the Canadian Council on Animal Care Guidelines and Policies and with approval from the Animal Care and Use Committee for Biosciences for the University of Alberta (“Neuroethology of Songbird Acoustic Communication: Laboratory Studies”, AUP00000107), and the University of Calgary Life and Environmental Sciences Animal Care Committee. Birds were captured and research was conducted under an

Environment Canada Canadian Wildlife Service Scientific permit (#13-AB-SC004), Alberta Fish and Wildlife Capture and Research permits (#56076 and #56077), and City of Edmonton Parks Permit.

Apparatus

During the experiment, birds were housed individually in modified colony room cages (30 × 40 × 40 cm) placed inside a ventilated, sound-attenuating chamber. The chambers were illuminated by a 9-W, full spectrum fluorescent bulb. Each cage contained three perches, a water bottle, and a grit cup. An opening on the side of the cage (11 × 16 cm) provided each bird access to a motor-driven feeder (see Njegovan, Hilhorst, Ferguson, & Weisman, 1994). Infrared cells in the feeder and the request perch (perch closest to the feeder) monitored the position of the bird. A personal computer connected to a single-board computer (Palya & Walter, 2001) scheduled trials and recorded responses to stimuli. Stimuli were played from the personal computer hard drive, through either a Cambridge A300 Integrated Amplifier, Cambridge Azur 640A Integrated Amplifier (Cambridge Audio, London, England), or an NAD310 Integrated Amplifier (NAD Electronics, London, England) and through a Fostex FE108 Σ or Fostex FE108E Σ full-range speaker (Fostex Corp., Japan; frequency response range 80-18,000 Hz) located beside the feeder. See Sturdy and Weisman (2006) for a detailed description of the apparatus.

Acoustic Stimuli

A total of 180 vocalizations produced by nine species were used as stimuli in the current experiment (originally collected and utilized in Filippi et al., 2017): infant giant panda (*Ailuropoda melanoleuca*), infant American alligator (*Alligator mississippiensis*), adult common raven (*Corvus corax*), adult hourglass treefrog, adult human (*Homo sapiens*; language: Tamil), infant African elephant (*Loxodonta africana*), adult Barbary macaque (*Macaca sylvanus*), adult black-capped chickadee, and infant domestic pig. We obtained 20 vocalizations per species, consisting of ten that were of high arousal and ten that were of low-arousal, in which any given individual produced both high and low arousal vocalizations (for spectrograms of each species' vocalizations, see Figure 1). These vocalizations were produced by male and female signalers, the same individual within a species, recorded and validated for arousal level in previous studies, and unpublished stimuli arousal levels were classified as high or low arousal according to the criteria presented in the original studies from which they were taken (giant panda, Stoeger et al., 2012; hourglass treefrog, Reichert, 2011, 2013; human, Bowling et al., 2013; African elephant, Stoeger et al., 2011; Barbary macaque, Fischer, 1995; domestic pig, Linhart et al., 2015; and unpublished works: American alligator recorded by S.A.R.; common raven recorded by A. Pašukonis; black-capped chickadee recorded by M. Avey and edited by J.V.C. and J.H.; see Table 1 for a summary of the nature of each species' vocalizations and published references). These same vocalizations were used in Filippi et al. (2017).

Below are descriptions of the conditions under which high- and low-arousal vocalizations are produced for each species. Infant giant pandas produce distress calls under multiple contexts (e.g., spontaneous agitation, feeding) which results in differing levels of motor activity (e.g., abrupt head movements) where higher levels of motor activity are associated with higher levels of arousal (i.e., >10 movements per 30s; Stoeger et al., 2012). Infant American alligators produce distress calls when pursued by a predator, and convey increases or decreases in the level of threat posed with the same call type; if they are grabbed, the peak frequency and overall energy of the high arousal calls (uttered with the palatal valve open) shift to significantly higher levels compared to low arousal calls (the same call uttered with the palatal valve closed; Britton, 2001). Common ravens produce defensive calls when physically confronted by a dominant conspecific;

the calls vary in arousal depending on the degree of threat posed (e.g., attack or not; Massen et al., 2014; Stocker et al., 2016). Hourglass treefrogs demonstrate an escalating level of competition during sexual advertisement by producing both aggressive calls (in the context of close-range inter-male agonistic interactions; i.e., high arousal) and advertisement calls (in the context of mate attraction; i.e., low arousal); for use as acoustic stimuli in the current study, these two call types were produced in close temporal proximity to ensure that they were recorded in a similar behavioural context (Reichart, 2011; Reichart, 2013). In a laboratory setting, human speakers were instructed to express emotions of anger (high-arousal) or sadness (low-arousal) through sentences spoken in Tamil (Bowling et al., 2012). Infant African elephants produce distress calls during disturbances in social context which result in greater physiological responses (e.g., larger quantities of secretions from the temporal glands, ears, head, and tail movements) in conditions of high arousal compared to low arousal (Stoeger et al., 2011). Barbary macaques produce disturbance calls when there are disturbances in the surroundings (e.g., a nearby predator), resulting in looking behaviour (low arousal) or increased physical distance from the source of the disturbance (i.e., playback speaker) to a point in which there is an assessed lack of danger (i.e., escape; high arousal; Fischer et al., 1995). Black-capped chickadees produce mobbing calls in response to nearby predators according to the threat level posed (Templeton et al., 2005), and neural activity in some brain regions has been shown to be greater in response to high-threat predator and conspecific calls compared to low-threat calls (arousal level dependent on the degree of threat posed; Avey et al., 2011). Infant domestic pigs produce distress calls when mobility is restricted, increasing motor activity with immobility, where higher levels of motor activity (attempted escape) are associated with higher levels of arousal compared to lower levels of motor activity (relaxed; Linhart et al., 2015). In summary, high arousal vocalizations were obtained from individuals under stressful conditions (e.g., threatening environments, agonistic interactions), while low arousal vocalizations were obtained from less adverse conditions, and despite the differences in indicators all stimuli were correlates of high- or low-arousal and negatively valenced.

All vocalizations were of high quality (i.e., no audible interference and low background noise when viewed on a spectrogram with amplitude cutoffs of -35 to 0 dB relative to vocalization peak amplitude) and were bandpass filtered (outside the frequency range of each vocalization type) using GoldWave version 5.58 (GoldWave, Inc., St. John's, NL, Canada) to reduce any background noise. For each stimulus, 5 ms of silence was added to the leading and trailing portion of the vocalization and tapered to remove transients, and amplitude was equalized using SIGNAL 5.10.24 software (Engineering Design, Berkeley, CA, USA).

During the experiment, stimuli were presented at approximately 75 dB as measured by a Brüel & Kjær Type 2239 (Brüel & Kjær Sound & Vibration Measurement A/S, Nærum, Denmark) decibel meter (A-weighting, slow response) at the approximate height and position of a bird's head when on the request perch.

Procedure

Pretraining. Pretraining began once the bird learned to use the request perch and feeder to obtain food. During Pretraining, birds received food for responding to all stimuli (future S+, S-, and transfer stimuli). A trial began when the bird landed on the request perch and remained for between 900-1100 ms. A randomly-selected stimulus played without replacement until all 180 stimuli had been heard. If the bird left the request perch before a stimulus finished playing, the trial was considered interrupted, resulting in a 30-s time out with the houselight turned off. If the bird entered the feeder within 1 s after the entire stimulus played, it was given 1 s access to food, followed by a 30-s intertrial interval, during which the houselight remained on. If a bird remained on the request perch during the stimulus presentation and the 1 s following the completion of the stimulus it received a 60-s intertrial interval with the houselight on, but this intertrial interval was terminated if the bird left the request perch. This was to encourage a high level of responding on all trials. Birds continued on Pretraining until they completed six 540-trial blocks of ≥ 60% responding on average to all stimuli, at least four 540-trial blocks ≤ 3% difference in responding to future S+ and S- stimuli, at least four 540-trial blocks in which the bird had ≤ 3% difference in responding to future high and low arousal transfer stimuli, and at least four 540-trial blocks in which the bird had ≤ 3% difference in responding to short and long stimuli to ensure that birds did not display stimuli preferences. Following a day of free feed, birds completed a second round of Pretraining in which they completed one 540-trial block of ≥ 60% responding on average to all stimuli, ≤ 3% difference in responding to future S+ and S- stimuli, ≤ 3% difference in responding to future high and low arousal transfer stimuli, and ≤ 3% difference in responding to short and long stimuli to confirm that each bird continued to not display preferences following the break.

Due to a low feeding percentage, one bird was put on a modified Pretraining criteria of twelve 540-trial blocks of ≥ 30% responding on average to all stimuli; all other criteria remained the same. The second round of Pretraining included two 540-trial blocks of ≥ 30% responding on average to all stimuli. The bird remained healthy throughout the experiment. There was a negligible statistical difference in whether or not this bird was included when analyzing overall performance on Discrimination Training. *See 'Trials to Criterion'*.

One other bird met Pretraining criteria, but was moved on to Discrimination Training before completing the second round. However, there was no statistical difference in whether or not this bird was included when analyzing overall performance on Discrimination Training. See ‘Trials to Criterion’.

Discrimination Training. The procedure was the same as during Pretraining; however, only the 108 training vocalizations were presented (with the remaining 72 withheld for use during Transfer testing), and responding to half of these vocalizations were then punished with a 30-s intertrial interval with the houselight off. As during Pretraining, responses to rewarded (S+) vocalizations resulted in 1 s access to food. Discrimination training continued until birds completed six 540-trial blocks with a discrimination ratio (DR) ≥ 0.80 with the last two blocks being consecutive. For DR calculations see Response Measures, below.

Birds were randomly assigned to either a True category discrimination group (*n* = 6) or Pseudo category discrimination group (*n* = 6). Black-capped chickadees in the True category discrimination group were divided into two subgroups (High S+ Group; two males and one female subject; Low S+ Group; two males and one female subject). The Pseudo category discrimination group was also divided into two subgroups (Subgroup 1: two males and one female subject; Subgroup 2: two males and one female subject). The purpose of the Pseudo group was to include a control in which subjects were not trained to categorize according to arousal level, investigating if True group acquisition is due to category learning (significantly fewer trials than the Pseudo groups) or simply rote memorization (similar number of trials compared to the Pseudo groups). See Figure 2 for how stimuli were divided into True and Pseudo subgroups.

Discrimination 85. This phase was identical to Discrimination training, except that the S+ vocalizations were rewarded with a reduced probability (i.e., *P* = 0.85). On unrewarded S+ trials, entering the feeder after the stimulus finished playing resulted in a 30-s intertrial interval, during which the houselight remained on, but there was no access to food. Discrimination 85 training was employed to introduce birds to trials in which there was no access to food, but the houselight remained illuminated, in order to prepare birds for transfer trials in which stimuli were neither rewarded nor punished. Discrimination 85 training continued until birds completed two 540-trial blocks with a DR ≥ 0.80.

Transfer Trials. During Transfer testing, the stimuli and reinforcement contingencies from Discrimination 85 were maintained. In addition, 72 stimuli (eight high- and eight low-arousal vocalizations from nine species heard during non-differential training but not discrimination training) were introduced. These new (i.e., transfer) stimuli were each presented once during a 612-trial block (S+ and S- stimuli from Discrimination 85 training were presented five times each; randomly-selected without replacement). Responses to transfer stimuli resulted in a 30-s intertrial interval with the houselight on, but no access to food; we did not differentially reinforce or punish transfer stimuli, and only presented each transfer stimulus once each per bin, so subjects did not learn specific contingencies associated with responding to these transfer stimuli. All birds completed a minimum of three blocks of Transfer trials and these were included for analysis. Following Transfer, birds were returned to their colony rooms.

Response Measures. For each stimulus exemplar, a proportion response was calculated by the following formula: R+/(N-I), where R+ is the number of trials in which the bird went to the feeder, N is the total number of trials, and I is the number of interrupted trials in which the bird left the perch before the entire stimulus played.

For Discrimination and Discrimination 85 training, we calculated a discrimination ratio (DR), by dividing the mean proportion response to all S+ stimuli by the mean proportion response of S+ stimuli plus the mean proportion response of S- stimuli. A DR of 0.50 indicates equal responding to rewarded (S+) and unrewarded (S-) stimuli, whereas a DR of 1.00 indicates perfect discrimination.

In order to analyze responding to each of the 18 stimulus types (nine high arousal, nine low arousal) during Transfer Trials, we calculated the proportion of responding for each stimulus type by averaging the percent response from the birds within each condition.

Statistical Analyses. We conducted independent-samples *t*-tests on the number of trials to criterion and DRs for the True and Pseudo category groups. Analyses of variance (ANOVAs) were conducted on DRs to determine if the birds had similar DRs to all species' vocalizations during Discrimination training. To determine if the True group continued to respond during Transfer to the contingencies they learned during Discrimination training, we ran a repeated measures ANOVA on the proportion of responding, and several paired-samples *t*-tests with Bonferroni corrections ($p < 0.0125$); similar tests were conducted with the Pseudo group. Huynh-Feldt correction was used on all repeated measures tests to correct for any possible violations in sphericity. We also conducted a three-way ANOVA (Condition × Arousal Level × Stimulus Species) on the proportion of responding during Transfer, followed by paired-samples *t*-tests to determine which of the nine species' vocalizations birds were transferring their responding to.

Results

Trials to Criterion

To determine whether birds in the two True category groups differed in their speed of acquisition, we conducted an independent-samples *t*-test on the number of 108-trial blocks to reach criterion for the two True category conditions (S+ High Group, S+ Low Group). There was no significant difference, $t_4 = -0.192$, $p = .857$, $d = .192$, 95% Confidence Interval [CIs] = -10.3290, 8.9957.

To determine whether birds in the two Pseudo category groups differed in their speed of acquisition, we conducted an independent-samples *t*-test on the number of 108-trial blocks to reach criterion for the two Pseudo category conditions (Pseudo Group 1, Pseudo Group 2). There was no significant difference, $t_4 = 0.761$, $p = .489$, $d = .761$, 95% CIs = -7.905, 13.9504. Since there were no differences in the speed of acquisition between the two Pseudo category groups, we combined the two groups in the remaining analyses.

To compare the acquisition performance of the True and Pseudo category groups and to determine if the True group learned to categorize in fewer trials than the Pseudo group, we conducted an independent-samples *t*-test on the number of 108-trial blocks to reach criterion for the True category and Pseudo category groups. There was a significant difference between the groups ($t_{10} = -2.244$, $p = .049$, $d = 1.419$, 95% CIs = -10.9603, -0.0397) in that True birds learned to discriminate significantly faster than Pseudo birds. If we removed the True group bird that was on modified Pretraining criteria, due to low feeding, the difference between group still approached significance, $t_9 = -2.173$, $p = .058$, $d = 1.449$, 95% CIs = -11.9052, 0.2385, meaning that acquisition during Discrimination was slightly slower for this particular bird. If we remove the True group bird that met Pretraining criteria, but was moved on to Discrimination Training before completing the second round, there is still a significant difference between the groups, $t_9 = -2.282$, $p = .048$, $d = 1.521$, 95% CIs = -12.0152, -0.0515.

DR Analysis

To examine if birds learned to discriminate all species' vocalizations equally we conducted a three-way ANOVA for the True group with Condition (S+ High and S+ Low), Arousal Level (High and Low Arousal), and Stimulus Species (Giant Panda, American Alligator, Common Raven, Hourglass Treefrog, Human, African Elephant, Barbary Macaque, Black-capped Chickadee, Domestic Pig) as fixed factors and DR as the dependent variable. There was a significant three-way interaction ($F_{1, 1, 8} = 2.635$, $p = .014$, $\eta^2 = .226$). There were significant main effects of Condition ($F_1 = 6.126$, $p = .016$, $\eta^2 = .078$, 95% CIs = 0.2304, 0.4192) and Arousal Level ($F_1 = 193.788$, $p = .029$, $\eta^2 = .064$, 95% CIs = 0.2356, 0.4242), and a significant interaction of Condition × Arousal Level ($F_{1, 1} = 193.788$, $p < .001$, $\eta^2 = .729$). All other interactions and main effects were non-significant (ps ≥ 0.277).

We conducted a three-way ANOVA for the Pseudo group with Condition (Pseudo 1 and Pseudo 2), Stimulus Type (Rewarded and Unrewarded stimuli), and Stimulus Species (Giant Panda, American Alligator, Common Raven, Hourglass Treefrog, Human, African Elephant, Barbary Macaque, Black-capped Chickadee, Domestic Pig) as fixed factors and DR as the dependent variable. There were significant main effects of Stimulus Species ($F_8 = 4.813$, $p < .001$, $\eta^2 = .348$, 95% CIs = 0.2528, 0.7259), Condition ($F_1 = 4.056$, $p = .048$, $\eta^2 = .053$, 95% CIs = 0.3307, 0.5277), and Stimulus Type ($F_1 = 336.523$, $p < .001$, $\eta^2 = .824$, 95% CIs = 0.5996, 0.7419), as well as a significant interaction of Stimulus Species × Stimulus Type ($F_{8, 1} = 4.548$, $p < .001$, $\eta^2 = .336$). All other interactions and main effects were non-significant (ps ≥ 0.085).

Category Learning

During Transfer trials, birds were intermittently presented with high- and low-arousal stimuli not presented during Discrimination training. A repeated measures ANOVA on the proportion of responding to the four stimulus types [Discrimination (DIS) S+ stimuli, Discrimination S- stimuli, Transfer (TRS) S+ associated stimuli, Transfer S- associated stimuli] by birds in the True Group was significant ($F_{1, 5} = 65.820$, $p = .001$, $\eta^2 = .943$, 95% CIs = 0.2757-0.5069). We then conducted four paired-samples *t*-tests with Bonferroni corrections ($p < .0125$). There were significant differences between responding to rewarded and unrewarded category stimuli during both Discrimination training and Transfer testing (DIS S+ vs. DIS S-, t_5 = 7.264, p = .001, d = 3.3383, 95% CIs = 0.3627, 0.7600; TRS S+ associated stimuli vs. TRS S- associated stimuli, t_5 = 9.085, p < .001, d =1.4662, 95% CIs = 0.1811, 0.32414), demonstrating that during Transfer testing birds continued to respond to the contingencies learned during discrimination training. There were no significant differences between responding to rewarded stimuli during training versus testing (DIS S+ vs. TRS S+ associated stimuli, t_5 = 2.014, p = .100, d = 0.5846, 95% CIs = -0.0335, 0.2758) or to unrewarded category stimuli during training versus testing (DIS S- vs. TRS S- associated stimuli, t_5 = -0.3494, p = .031, d = -1.5329, 95% CIs = -0.3494, -0.0258).

A repeated measures ANOVA on the proportion of responding to the four stimulus types [Discrimination (DIS) S+ stimuli, Discrimination S- stimuli, Transfer (TRS) S+ associated stimuli, Transfer S- stimuli] by birds in the Pseudo Group was significant ($F_{1, 5}$ = 52.258, p = .001, η^2 = .913, 95% CIs = 0.2870, 0.4958). We conducted four paired-samples t-tests with Bonferroni corrections (p < .0125). There were significant differences between responding to rewarded and unrewarded discrimination stimuli during transfer (DIS S+ vs. DIS S-, t_5 = 16.333, p < .001, d = 5.5181, 95% CIs = 0.4572, 0.6280), and transfer stimuli (TRS S+ associated stimuli vs. TRS S- associated stimuli, t_5 = 4.433, p = .007, d = 0.5054, 95% CIs = 0.0422, 0.1589), demonstrating that during transfer testing birds responded to the non-differentially rewarded stimuli according to the contingencies learned during discrimination training. There were no significant differences between responding to rewarded stimuli during training versus testing (DIS S+ vs. TRS S+ associated stimuli, t_5 = 3.777, p = .013, d = 1.4528, 95% CIs = 0.0757, 0.3985) or between responding to unrewarded stimuli during training versus testing (DIS S- vs. TRS S- associated stimuli, t_5 = -3.009, p = .030, d = -1.3637, 95% CIs = -0.3799, -0.0299).

Transfer Trials

To determine if the pattern of learning was the same across all species' vocalizations in transfer, we conducted a Condition (S+ High, S+ Low) × Arousal Level (high, low) × Stimulus Species (Giant Panda, American Alligator, Common Raven, Hourglass Treefrog, Human, African Elephant, Barbary Macaque, Black-capped Chickadee, Domestic Pig) three-way ANOVA on the proportion of responding of the True group during testing. There was a significant three-way interaction of Condition × Arousal Level × Stimulus Species ($F_{1,\,1,\,8}$ = 3.386, p = .002, η^2 = .273). There were also significant interactions of Condition × Stimulus Species ($F_{1,\,8}$ = 3.651, p = .001, η^2 = .289), and Arousal Level × Condition ($F_{1,\,1}$ = 27.836, p < .001, η^2 = .279). All other main effects and interactions were non-significant (ps ≥ 0.164). The main effect of Arousal Level was not significant as the True group was rewarded for either high or low arousal (S+ High, S+ Low), so half of the subjects would be expected to respond more to high arousal stimuli than low arousal stimuli and vice versa. However, the interaction of Condition and Arousal Level was significant, indicating that birds were transferring their learning to appropriately respond to rewarded-contingency high or low arousal novel stimuli, as demonstrated previously (see 'Category Learning'; Figure 2). To further investigate responding across Stimulus Species for each of the true category groups, we conducted paired-samples *t*-tests; see Table 2 for these statistical results. See Figure 4 for S+ High Group and S+ Low Group subjects' responding.

Discussion: Experiment 1

We demonstrated that black-capped chickadees are capable of discriminating vocalizations based on the arousal context, even though the vocalizations were produced by multiple species that our subjects likely had no prior experience with (subjects had experience with conspecific, human, and possibly raven vocalizations). The two True-category groups did not differ in speed of acquisition, nor did the two Pseudo-category groups; however, the True and Pseudo groups differed in speed of acquisition in that the True group learned to discriminate the high arousal stimuli from the low arousal stimuli in fewer trials compared to the Pseudo group. Due to the difference in speed of acquisition (i.e., the True group learned in fewer trials than the Pseudo group in training), it appears that chickadees might be using arousal-based category information over rote memorization. True group birds then transferred their training to novel stimuli based on the ‘rules’ they learned from discrimination. This is supported by the fact that there were no significant differences between responding by the True group to rewarded category stimuli during training versus testing (i.e., subjects continued to respond to the transfer stimuli based on the contingency learned during discrimination training); however, there was a significant difference between responding by the True group to unrewarded category stimuli during training versus testing.

Overall, we demonstrated that chickadees are capable of discriminating between vocalizations and transferring prior training to several species’ vocalizations based on vocal components related to arousal. Specifically, chickadees demonstrated transfer of training to giant panda, human, and black-capped chickadee vocalizations, one group trended toward transfer of training to domestic pig vocalizations (See ‘General Discussion’ for further explanation of S+ Low group’s transfer of training to low arousal domestic pig vocalizations), but neither group (S+ High or S+ Low) transferred their training to American alligator, common raven, hourglass treefrog, African elephant, or Barbary macaque vocalizations. After demonstrating successful categorical training based on arousal in a non-human animal species, we wanted to test human learning and transfer of training on a comparative task to provide a direct comparison with our chickadee results.

Experiment 2: Humans

In previous work, Filippi et al. (2017) demonstrated that humans can discriminate between high and low arousal vocalizations produced by multiple species. However, it is unclear how they would perform on a task that is directly analogous to that used to test chickadees. Thus, we designed a go/no-go discrimination task for humans to train then test their ability to discriminate based on arousal without explicit instructions on the categorical information, replicating our chickadee experiment with humans.

Methods

Participants

University of Alberta undergraduate students participated in this study between July 13, 2015 to August 10, 2015 in exchange for partial research credit towards their introductory psychology class. Students were required to have normal hearing and were not allowed to have previously participated in similar acoustic experiments.

Twenty-one participants were excluded from the study; three had incomplete surveys, 16 exhibited high responding during training (i.e., >95%), one exhibited low responding during training (i.e., <10%), and one participant was removed due to an incorrect file being used in testing. Thus, there was a total of 16 S+ High Arousal participants (seven males, nine females), 14 S+ Low Arousal (five males, nine females), 12 Pseudo 1 (four males, eight females), and 13 Pseudo 2 (six males, seven females). We processed the data of the remaining 55 participants (M_{age} = 21 years; SD_{age} = 3.121 years; 22 males and 33 females). Eighteen subjects were native

English speakers, whereas 37 were non-native English speakers. Sixteen of our participants had one or more pets (dog, cat, hamster, and/or fish), and the average reported experience with animals was a rated three out of seven ($M_{experience}$ = 2.936).

Ethical Note

All human studies were conducted under Songbird Neuroethology Laboratory's Auditory Perception and Cognition protocols with approval from the University of Alberta's Research Ethics & Management Online (REMO; "Auditory perception and cognition", Pro00016997). Participants were recruited via the University of Alberta, Department of Psychology's Research Pool in exchange for partial course credit toward the introductory psychology course in which they were actively enrolled.

Apparatus

Each participant was seated in front of an LG FLATRON W2442PA computer, equipped with an Intel Core i7 CPU and Windows 7 Professional. A pair of SENNHEISER HD 280 Pro headphones with nominal impedance of 64 Ohms was connected to each computer. The volume on each computer was set to a standardized level (i.e., headphone jack set to 20% of maximum volume), but participants were allowed to adjust the volume to the level they felt was most comfortable. A program from a previous experiment conducted in our laboratory (Hoeschele, Weisman, & Sturdy, 2012), utilizing a go/no-go task, was installed and placed on the desktop of each computer. Participants heard auditory stimuli through the headphones and used the computer mouse to make responses.

Stimuli

Experiment 2 used the same 180 stimuli from Experiment 1. One hundred and eight stimuli were used (six high- and six low-arousal vocalizations from each of the nine species) in the discrimination training phase, and 180 stimuli (108 discrimination stimuli plus an additional four high- and four low-arousal vocalizations from each of the nine species) during the transfer testing phase.

Procedure

Up to four participants completed the experiment at a time, randomly assigned to one of four conditions (High S+, Low S+, Pseudo 1, or Pseudo 2 group; Figure 2); they first read the information letter and signed a consent form. Then they filled out a survey in regards to their age, sex, native language, duration of musical training, experiences in sharing common area with pets at home, and level of animal experience (Appendix B).

Discrimination Practice Phase. A short Practice phase was completed before the Training phase to allow participants to understand the basics of using the program. The Practice phase included six stimuli, presented in a fixed order. Human infant cries were used as stimuli during this phase, but were not used as stimuli during the remainder of the experiment. The procedure for this phase was the similar to the other experimental phases, but utilized only these six stimuli, three high- and three low-arousal vocalizations.

Discrimination Training Phase. One hundred and eight stimuli were presented in the Training phase, consisting of six high-arousal and six low-arousal vocalizations from each of the nine species. During this training phase, participants were to categorize sounds that they heard into an “S+” category or a second, unnamed category. They received feedback from the program following their categorization, informing them of whether their response was “CORRECT” or “INCORRECT”. The goal of this stage was for the participants to successfully discriminate the sounds into the arbitrary category that had been chosen by the experimenters (i.e., “S+”). Experimental groups, and respective stimuli, were the same as for the birds (see Experiment 1:

Discrimination Training). Specifically, if an individual had been randomly assigned to S+ High Arousal Group, they would view a rewarded phrase of “CORRECT” if they clicked the S+ button when a high-arousal sound was heard (i.e., a correct “go” response; Appendix A, Supplemental Figure 3). Contrary, they would view an unrewarded response of “INCORRECT” if they clicked the S+ button when a low-arousal sound was heard (i.e., an incorrect “go” response; Appendix A, Supplemental Figure 4). The individual would not receive any feedback if they choose to not press the S+ button whenever they heard a stimulus. This was viewed as a correct response when clicking S+ would have been incorrect (i.e., a correct “no-go” response). This concept was also applied when low arousal vocalizations were rewarded. For both of the Pseudo subgroups, responding (i.e., pressing the S+ button) to half of the high- and half of the low-arousal stimuli resulted in “CORRECT”. All participants moved onto Transfer Testing after completing 108 trials.

Transfer Testing Phase. A set of 180 stimuli were utilized in the Transfer Testing stage in which four additional high- and four additional low-arousal vocalizations of each of the nine species were added, for a total of 10 high- and 10 low-arousal stimuli for each of the nine species. These new stimuli provided the opportunity for us to test participants’ ability to generalize their knowledge of the categories based on what they learned from the training phase. Feedback was not provided in this stage (the feedback window was no longer depicted on the computer screen; Appendix A, Supplemental Figure 4). We expected that participants in the Pseudo group would respond non-differentially to the high and low arousal stimuli during this stage.

Statistical Analyses. All analyses conducted in Experiment 2 were the same as Experiment 1 (see Experiment 1’s ‘Statistical Analyses’), except that we could not conduct independent-samples *t*-tests on the number of trials to criterion as every participant moved on to testing following a certain number of trials.

Results

Discrimination Responding

To examine if humans in the True group learned to discriminate all species' vocalizations equally well we conducted a three-way ANOVA with Condition (S+ High and S+ Low Groups), Arousal Level (High and Low Arousal), and Stimulus Species (Giant Panda, American Alligator, Common Raven, Hourglass Treefrog, Human, African Elephant, Barbary Macaque, Black-capped Chickadee, Domestic Pig) as fixed factors and DR as the dependent variable. There was a significant three-way interaction ($F_{1, 1, 8} = 5.252$, $p < .001$, $\eta^2 = .077$). There were significant interactions of Stimulus Species × Condition ($F_{8, 1} = 10.009$, $p < .001$, $\eta^2 = .137$) and Arousal Level × Condition ($F_{1, 1} = 45.826$, $p < .001$, $\eta^2 = .083$). All other interactions and main effects were non-significant (*p*s ≥ 0.188).

We also conducted a three-way ANOVA for the Pseudo Group with Condition (Pseudo 1 and Pseudo 2), Stimulus Type (Rewarded and Unrewarded stimuli), and Stimulus Species (Giant Panda, American Alligator, Common Raven, Hourglass Treefrog, Human, African Elephant, Barbary Macaque, Black-capped Chickadee, Domestic Pig) as fixed factors and DR as the dependent variable. There was no significant three-way interaction, $F_{1, 1, 8} = 0.120$, $p = .998$, $\eta^2 = .002$. There was a significant main effect of Condition ($F_1 = 17.200$, $p < .001$, $\eta^2 = .040$, 95% CIs = 0.6328, 0.7048), but all other main effects and interactions were non-significant (*p*s ≥ 0.486).

Category Learning

To determine if the True group continued to respond to the contingencies learned in training, we conducted a repeated measures ANOVA on the proportion of responding to the four stimulus types (Discrimination (DIS) S+ stimuli, Discrimination S- stimuli, Transfer (TRS) S+ associated stimuli, Transfer S- associated stimuli) for the True group, $F_{1, 29} = 283.007$, $p < .001$, $\eta^2 = .907$, 95% CIs = 0.4126, 0.4821. We then conducted four independent-samples *t*-tests with Bonferroni corrections ($p > .0125$). There were significant differences between responding to rewarded and unrewarded stimuli during both Discrimination training and Transfer testing (DIS S+ vs. DIS S-, $t_{29} = 6.814$, $p < .001$, $d = 1.2562$, 95% CIs = 0.1464, 0.2720; TRS S+ associated stimuli vs. TRS S- associated stimuli, $t_{29} = 5.688$, $p < .001$, $d = 0.9877$, 95% CIs = 0.1081, 0.2296), with participants responding significantly more to rewarded category stimuli than unrewarded. There were also significant differences in responding to rewarded category stimuli during training versus testing (DIS S+ vs. TRS rewarded-contingency stimuli, $t_{29} = 3.660$, $p = .001$, $d = 0.2289$, 95% CIs = 0.0147, 0.0520), but no significant difference between responding to unrewarded category stimuli during training versus testing (DIS S- vs. TRS S-, $t_{29} = -0.6577$, $p = .516$, $d = -0.0375$, 95% CIs = -0.0288, 0.0148), as humans responded more to rewarded category stimuli from training than from testing but showed no difference in responding to unrewarded training and testing stimuli.

A repeated measures ANOVA on the proportion of responding to the four stimulus types (Discrimination (DIS) S+ stimuli, Discrimination S- stimuli, Transfer (TRS) S+ associated stimuli, Transfer S- associated stimuli) by humans in the Pseudo group was not significant ($F_{1, 24} = 188.993$, $p < .001$, $\eta^2 = 0.887$, 95% CI = 0.4961, 0.5752). There was a significant difference between unrewarded training versus testing (DIS S- vs. TRS S-, $t_{24} = 4.279$, $p < .001$, $d = 0.4862$, 95% CIs = 0.0256, 0.0733). All other paired-samples *t*-tests were non-significant, *p*s > .057.

Transfer Trials

We conducted a Condition (S+ High, S+ Low) × Arousal Level (high, low) × Stimulus Species (Giant Panda, American Alligator, Common Raven, Hourglass Treefrog, Human, African Elephant, Barbary Macaque, Black-capped Chickadee, Domestic Pig) three-way ANOVA on the proportion of responding during testing to determine if the pattern of learning was the same across all species' vocalizations. There was a significant three-way interaction of Condition × Arousal Level × Stimulus Species ($F_{1,1,8} = 4.955$, $p < .001$, $\eta^2 = .073$). There was a main effect of Condition ($F_1 = 4.439$, $p = .036$, $\eta^2 = .009$, 95% CIs = 0.3789, 0.4533). There was also a significant interaction of Arousal Level × Condition ($F_{8,1} = 44.070$, $p < .001$, $\eta^2 = .080$), and Stimulus Species × Condition ($F_{8,1} = 8.326$, $p < .001$, $\eta^2 = .117$). All other interactions and main effects were non-significant (*p*s ≥ 0.118). As in Experiment 1, Arousal Level was not significant as a main effect, but was significant as an interaction of Condition by Arousal Level; this indicates that the True group were responding appropriately by transferring their training to high and low arousal novel stimuli (See Figure 5 for group responding to categories by human participants during transfer testing). To further investigate the effect of Stimulus Species, we conducted paired-samples *t*-tests; see Table 3 for these statistical results. See Figure 6 for S+ High Group and S+ Low Group human participants' responding.

We also conducted a Pseudo Group (Pseudo 1, Pseudo 2) × Arousal Level (high, low) × Stimulus Species (nine species' vocalizations) three-way ANOVA on the proportion of responding during testing to determine if the pattern of responding was the same across all species' vocalizations. There were no main effects or interactions (*p*s ≥ .075). Thus, responding did not change with Arousal Level according to Condition as the Pseudo group did not respond to novel stimuli based on arousal, as expected for a Pseudo (control) group.

Human Experience

In order to determine if ‘experience’, a term used here to describe participants’ current interaction level with animals, affected the ability to transfer discrimination abilities between high and low arousal stimuli, we ran four one-way ANOVAs. Participants’ experience (1-7; Appendix B) did not influence their responding (S+ High: $F_{143} = 0.633$, $p = .728$, $\eta^2 = .031$; S+ Low: $F_{125} = 0.984$, $p = .512$, $\eta^2 = .270$; Pseudo 1: $F_{107} = 0.686$, $p = .894$, $\eta^2 = .200$; Pseudo 2: $F_{116} = 1.131$, $p = .317$, $\eta^2 = .191$; CIs for listed experiences reported in Table 4).

Discussion: Experiment 2

In keeping with the results of Filippi et al. (2017), we demonstrated that humans are capable of discriminating vocalizations produced by multiple species based on the arousal context. Additionally, we predicted that human participants with more animal experience would be able to categorize at a higher level than those without such experience. However, we found that participants’ experience, according to our survey, did not affect their responding. The fact that experience did not affect performance could be interpreted as support for the existence of perceivable acoustic differences in vocalizations that differ in levels of arousal, regardless of familiarity with animals; however, there were limitations on our survey as the wording used on the survey did not directly ask about prior experience with animals such as owning a pet when young.

Filippi and colleagues (2017) showed that humans are capable of discriminating across all species’ vocalizations when a high and low arousal stimulus were paired; it may have been difficult for the human participants in the current study to learn the categorization task since they were presented with one stimulus at a time and were not given the specific instructions to categorize based on arousal. In contrast, we found that humans could discriminate giant panda, human, and black-capped chickadee vocalizations when presented consecutively (using the same stimuli as the previous experiment utilizing concurrent presentation; Filippi et al., 2017). The difference in the results from the current study and Filippi et al. (2017) is likely methodological, with the go/no-go design used here being more conceptually difficult than a choice task. In addition, vocalizations produced by some species were quite short, potentially increasing the difficulty for birds and humans to obtain arousal-based information in this design.

For both rewarded (S+) and unrewarded (S-) stimuli, True group participants continued to respond during Transfer testing based on the contingencies from Discrimination training, however, there was a significant difference in responding to rewarded stimuli during training versus testing such that they responded less to testing stimuli than to training stimuli. There was a significant difference in responding by the True group to high and low arousal transfer stimuli, demonstrating transfer of training to untrained stimuli. This indicates that humans can not only perceive and learn categories of arousal, but also transfer that learned categorization to some species' vocalizations. The human participants in the True group successfully transferred their training to the same three species that the chickadees transferred to (i.e., giant panda, human, and black-capped chickadee vocalizations; S+ Low Arousal group approached significance for responding to black-capped chickadee vocalizations, $p = .053$). In addition, participants in the S+ High arousal group transferred their training to vocalizations produced by African elephants. This indicates that acoustic features may vary significantly between vocalizations of high and low arousal level, as most humans do not have extensive experience with the majority of species whose vocalizations they were trained with, especially the giant panda vocalizations to which they demonstrated transfer of training. This also provides direct comparative findings of arousal perception as chickadees and humans demonstrated similar responding on analogous go/no-go discrimination tasks without instructions about the categories.

Bioacoustic Analysis

To examine the acoustic variation that black-capped chickadees and humans may have attended to when discriminating vocalizations that differed in arousal, we conducted bioacoustic analyses and discriminant function analyses on the discrimination and transfer stimuli presented during the go/no-go discrimination task.

Methods

For the bioacoustic analysis, we measured four acoustic features that had been previously measured to access variation in acoustic arousal (Filippi et al., 2017; Maruščáková et al., 2015): (1) duration of the vocalization, (2) initial fundamental frequency (F0), (3) spectral centre of gravity (SCG; a measure of the average frequency height), and (4) harmonic-to-noise ratio (HNR; harshness of a sound, a measure of relative pure and noisy signal components). We then conducted two categories of stepwise DFA using these acoustic features from the high and low arousal stimuli from each of the nine species. The first was a single DFA where vocalizations from all species and all arousal levels were categorized concurrently (i.e., 18 ‘vocal categories’ total) to determine which stimulus groups were most likely to be misclassified. The second was a series of nine DFAs in which the vocalizations for each species was categorized as high or low arousal separate from the other species. The purpose of these two types of DFAs was to differentiate between overall classification and within species classification. By comparing the misclassifications in the DFA to the black-capped chickadee and human performance errors, we hoped to determine if these acoustic features could explain the observed results. The four acoustic measures (duration, F0, SCG, HNR) were included as independent variables. The Discrimination training stimuli were used to create the discriminant function, a model to predict which vocal category a stimulus belongs to based on its four acoustic measures. In addition, based on the prediction errors from the DFA, we evaluated into which vocal categories the stimuli were incorrectly classified. Next, we used the discriminant function to classify the Transfer testing stimuli; again, we evaluated which stimuli were incorrectly classified by examining the prediction errors from the DFA. Some acoustic measures were not completed as the structure of certain vocalizations did not allow for the measurement of some features (i.e., ‘noisy’ vocalizations; one common raven, two African elephant, two Barbary macaque, and one domestic pig vocalization were not measured on F0 and HNR); the missing values were replaced with the mean of the remaining vocalizations for the species within the same arousal level for the statistical analysis. For each of the nine stepwise DFA, our criterion for a variable to be entered in the analysis was a minimum partial F = 3.84, and to be removed was a maximum partial F = 2.71, the same criterion as the first DFA.

While the behavioural task was to categorize high- versus low-arousal vocalizations, since stimuli included vocalizations produced by nine distinct species, it is possible that subjects were treating the task as nine separate categorization tasks. In order to determine which acoustic measures were driving the discrimination between high and low arousal vocalizations for each species, we conducted nine stepwise DFAs that classified the high- and low-arousal stimuli of each species separately (e.g., high arousal giant panda vs. low arousal giant panda vocalizations). Again, the Discrimination training stimuli were used to create a discriminant function and, based on the prediction errors from the DFA, we evaluated the percentage of stimuli that were incorrectly classified and which species’ vocalizations were most likely to be incorrectly classified. We then used the discriminant function to classify the Transfer testing stimuli.

Results

For the discriminant function we constructed to classify the stimuli based on the category of vocalization (high or low arousal level for each species), the overall Wilks' λ was significant [Wilks' $\lambda = 0.003$, $\mathcal{X}^2$ (68, $N = 108$) = 521.076, $p < 0.001$]. In addition, each residual Wilks' λ was significant [test of functions 2 through 4: Wilks' $\lambda = 0.043$, $\mathcal{X}^2 = 288.493$, $p < 0.001$; test of functions 3 through 4: Wilks' $\lambda = 0.250$, $\mathcal{X}^2 = 127.518$, $p < 0.001$; and test of function 4: Wilks' $\lambda = 0.610$, $\mathcal{X}^2 = 45.516$, $p < 0.001$].

The four discriminant functions assigned 88.0% of the original grouped cases to correct category of vocalizations (chance = 1/18 = 5.6%). After creating the discriminant functions with Discrimination training stimuli measures, we used the discriminant functions to predict the category membership for the Transfer training stimuli. The four discriminant functions assigned 52.8% of Transfer training stimuli grouped cases to correct category of vocalizations (chance = 1/18 = 5.6%; see Table 5).

For each of our nine stimulus species, we conducted a stepwise DFA to determine whether our four acoustic measures (duration, F0, SCG, HNR) could predict arousal level (high or low); see Table 6 for the statistical output. For the Discrimination stimuli, the DFA was able to correctly classify 100% of the giant panda vocalizations as high vs. low arousal, 91.7% of the American alligator vocalizations as high vs. low arousal, 83.3% of the common raven vocalizations as high vs. low arousal, 100% of the hourglass treefrog vocalizations as high vs. low arousal, 100% of the human vocalizations as high vs. low arousal, 100% of the African elephant vocalizations as high vs. low arousal, 91.7% of the black-capped chickadee vocalizations as high vs. low arousal, and 83.3% of the domestic pig vocalizations as high vs. low arousal (chance for each = 50%).

Next, we assessed how well the discriminant functions created with the Discrimination stimuli would classify the Transfer stimuli. We found that 87.5% of the giant panda vocalizations, 100% of the American alligator vocalizations, 62.5% of the common raven vocalizations, 100% of the hourglass treefrog vocalizations, 87.5% of the human vocalizations, 87.5% of the African elephant vocalizations, 62.5% of the black-capped chickadee vocalizations, and 75% of the domestic pig vocalizations were correctly classified according to arousal level. Although organisms may not consistently pay attention to all available acoustic features, these findings suggest that there was sufficient acoustic variation for discrimination between the high and low arousal transfer stimuli for each of these species.

General Discussion

Here we have demonstrated that black-capped chickadees and humans, can learn to discriminate between high and low arousal vocalizations. Across Experiment 1 and 2, the majority of True groups showed transfer of training to giant panda, human, and black-capped chickadee vocalizations. Human and giant panda vocalizations correspond to the species Filippi et al. (2017) found human participants were best capable of identifying when asked which was the high arousal vocalization in paired evaluations, with participants identifying black-capped chickadee vocalizations at 85% correct. Black-capped chickadee subjects' responding to black-capped chickadee and human vocalizations could be explained by familiarity; the subjects in this experiment were wild-caught black-capped chickadees housed in captivity with daily exposure to humans (i.e., lab staff and researchers). Familiarity has been demonstrated to be a factor in identifying information contained within vocalizations (e.g., content, context, etc.). For example, cat vocalizations (i.e., meow sounds) that had been recorded from different cats in multiple behavioural contexts were best classified by human participants with more exposure to cats (Nicastro & Owren, 2003). In the current study, human participants' significant transfer of training to human and black-capped chickadee vocalizations may have been due to familiarity, although it is possible that some participants did not have experience with black-capped chickadees. In addition, it is not likely that black-capped chickadee subjects or human participants had much (if any) exposure to giant panda vocalizations, which they significantly transferred their responding to during the task. Scheumann and colleagues (2014) found that human participants recognized the emotional content of human vocalizations, but had mixed results for animal vocalizations depending on individual experience with each species; they concluded that human recognition of the emotional content of animal vocalizations is determined by familiarity rather than acoustic universal-like coding. We found that our participants' overall experience with animals did not improve their responding when discriminating between vocalizations of high and low arousal; however, we did not collect information on the specific species that participants had experience with, so it is possible that all of our human participants had limited (if any) experience with all of the species whose vocalizations we used. Therefore, while familiarity may play a role in differentiating arousal level in vocalizations, this alone does not explain the responding by black-capped chickadees and humans during this task, especially to the giant panda stimuli.

We conducted a DFA to better understand which acoustic features could be contributing to the discrimination of high and low arousal vocalizations across all of the species whose vocalizations we used as stimuli. The DFA correctly classified more than half of the giant panda and black-capped chickadee vocalizations (chance = 5.6%), which is consistent with the responding by black-capped chickadees and humans, as both chickadees and humans significantly transferred their responding to vocalizations produced by these two species. However, there was a clear discrepancy between the classification performance of the DFA and that of the chickadees and humans; the DFA misclassified high arousal human vocalizations entirely (0% accuracy, with 75% accuracy in classifying low arousal human vocalizations), while the chickadee and human subjects significantly transferred their responding to human vocalizations. From the additional DFAs, we determined which acoustic features may be utilized to correctly classify between high and low arousal vocalizations for each species. Both birds and humans transferred to giant panda and black-capped chickadee vocalizations, stimuli in which the DFAs utilized the initial fundamental frequency (F0). Black-capped chickadees also transferred to domestic pig vocalizations. Maruščáková and colleagues (2015) found that the acoustic properties of piglet vocalizations were effective in human participants' judging of emotional content; specifically, that simple acoustic parameters (e.g., F0) were more effective than complex acoustic properties (e.g., harmonic-to-noise ratio: HNR). Consistent with this, the results from our DFA suggest that F0 is a useful acoustic feature to differentiate between high and low arousal domestic pig vocalizations. Taken together, these results suggest that F0 may be one fundamental acoustic feature that is useful when differentiating arousal level in various vocalizations. Although research (for instance Bowling et al., 2017) has found that F0 correlates negatively with signaler's body size within species, the link between the expression of emotional arousal and body size through modulation of F0 remains open to investigation. For black-capped chickadee vocalizations, in addition to using F0, the DFA also used duration to classify based on arousal. The only other DFA that used duration to classify based on arousal was the DFA classifying African elephant vocalizations. In our behavioural experiments, in addition to significant transfer of training to black-capped chickadee vocalizations, there was some transfer of training to African elephant vocalizations (i.e., in Experiment 2). These results suggest that duration may also be a useful feature when discriminating between vocalizations of arousal. Consistent with this, previous studies have demonstrated that animals discriminate among acoustic stimuli using duration. For example, Briefer's (2012) review concluded that physiological arousal (e.g., a change in respiration) is typically reflected in vocalization changes, including longer durations and higher F0.

When classifying American alligator, common raven, and human vocalizations, the DFA used the spectral centre of gravity (SCG). SCG has previously been utilized as a reliable acoustic feature as Sauter et al. (2010) and Faragó et al. (2014) found that higher intensity ratings consistently corresponded with higher SCG for human and dog vocalizations, respectively. Filippi et al. (2017) also demonstrated that F0 and SCG predicted accuracy in humans' ability to identify the high-arousal vocalizations. However, due to a lack of transfer to alligator and raven vocalizations, SCG is unlikely to be a useful acoustic feature for this type of discrimination. Belin and colleagues (2008) demonstrated that humans could discriminate between human vocalizations of positive and negative valence (affectivity; different from the calming/excitement of arousal), but not vocalizations produced by other species (i.e., cats and rhesus monkeys); however, fMRI imaging indicated that there was appropriate discrimination at a neurobiological, albeit unconscious, level based on cerebral blood flow. Thus, while behavioural responses failed to demonstrate discrimination of heterospecific vocalizations, accurate discrimination was revealed from the fMRI result. It is possible that chickadees and humans in the current experiment were capable of transfer to all nine species' vocalizations (negatively valenced), but did not demonstrate the discrimination at a behavioural level, similar to the results of Belin et al. (2008).

In conclusion, we demonstrated that birds and humans (without being given instruction about the nature of the task) that were previously rewarded for responding to vocalizations based on arousal level responded significantly more to untrained vocalizations associated with the same arousal level that was previously rewarded. Both species transferred their training to novel stimuli produced by a third of the species that we included in our task based on their previous training; however, consistent transfer of training was not demonstrated to the vocalizations of the remaining species by either birds or humans. Similar in that it is adaptive for red-breasted nuthatches to eavesdrop on black-capped chickadee mobbing calls (Templeton & Greene, 2007) or non-vocal iguanas to understand the emotional content of mockingbirds’ alarm calls regarding a nearby shared predator (the Galapagos hawk; Vitousek et al., 2007), it would be advantageous to have the ability to perceive arousal in vocalizations produced by a variety of species, including unfamiliar species. We propose that future studies utilize non-vocal learning study species to further investigate the perception of arousal, and potentially incorporate fMRI techniques to provide a more thorough investigation, especially of unconscious perception. Our findings demonstrate that a species of songbird and humans, both vocal learners, perceive variations of arousal in vocalizations produced by multiple species (Mammalia and Aves), suggesting that acoustic features may exist that vary sufficiently between high- and low-arousal vocalizations to allow these two exemplars, and perhaps other species, to discriminate based on level of arousal.

Acknowledgements

We thank Lou Omerzu for his technical assistance in the chickadee operant chambers. In addition, we would like to acknowledge our Research Assistants who supervised our chickadee operant chambers at the time, Josh J. H. Yong and Erica Cheung. We would also like to thank the many undergraduate students who made the human experiment possible: Natalie Kuo-Hsuan, Kevin Lui, Avnit Gewal, and Parisa Asghari-Voloshkolae. AHH, PF, KAC are co-second authors.

This research was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant and Discovery Accelerator Supplement, an Alberta Ingenuity Fund (AIF) New Faculty Grant, a Canada Foundation for Innovation (CFI) New Opportunities Fund (NOF) and Infrastructure Operating Fund (IOF) grants along with start-up funding and CFI partner funding from the University of Alberta (UofA) to CBS. JVC was supported by both an Alexander Graham Bell Canada Graduate Scholarship-Master’s (NSERC CGS M) and -Doctoral (NSERC CGS D).

References

Aubin, T. & Jouventin, P. (2002). Localisation of an acoustic signal in a noisy environment: the display call of the king penguin *Aptenodytes patagonicus*. *The Journal of Experimental Biology*, *205*, 3793-3798.

Avey, M. T., Hoeschele, M., Moscicki, M. K., Bloomfield, L. L., & Sturdy, C. B. (2011). Neural correlates of threat perception: Neural equivalence of conspecific and heterospecific mobbing calls is learned. *PLoS ONE*, *6*, 1-7.

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2007). Human cerebral response to animal affective vocalizations. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 473-481.

Bowling, D. L., Gingras, B., Han, S., Sundararajan, J., & Opitz, E. C. L. (2013). Tone of voice in emotional expression: Relevance for the affective character of musical mode. *Journal of Interdisciplinary Music Studies*, *7*, 29-44.

Briefer, E. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*, 1-20.

Britton, A. R. C. (2001). Review and classification of call types of juvenile crocodilians and factors affecting distress calls. In *Crocodilian biology and evolution* (eds G.C. Grigg, F. Seebacher, C.E. Franklin), pp. 364-377. Chipping Norton, Australia: Surrey, Beatty, & Sons.

Caro, T. (2005). Antipredator Defenses in Birds and Mammals, The University of Chicago Press.

Christie, P. J., Mennill, D. J., & Ratcliffe, L. M. (2004). Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. *Behavioral Ecology and Sociobiology*, *55*, 341-348.

Darwin C. (1871). The descent of man and selection in relation to sex. London, UK: John Murray.

Faragó, T., Andics, A., Devecseri, V., Kis, A., Gácsi, M., & Miklósi, Á. (2014). Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biology Letters*, *10*, 1-5.

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., Hoeschele, M., Ocklenburg, S., de Boer, B., Sturdy, C. B., Newen, A., & Güntürkün, O. (2017). Humans recognize emotional arousal in vocalizations across all classes of terrestrial vertebrates: Evidence for acoustic universals. *Proceedings of the Royal Society B*, *284*, 20170990.

Fischer, J., Hammerschmidt, K., & Todt, D. (1995). Factors affecting acoustic variation in Barbary-macaque (*Macaca sylvanus*) disturbance calls. *Ethology*, *101*, 51-66.

Gerhardt, H. C., Humfeld, S. C., & Marshall, V. T. (2007). Temporal order and the evolution of complex acoustic signals. *Proceedings of the Royal Society B: Biological Sciences*, *274,* 1789-1794.

Green, S. & Marler, P. (1979). The analysis of animal communication. In P. Marler & J. G. Vandenbergh (Eds.), *Handbook of behavioral neurobiology: III. Social behavior and communication* (pp. 73-158). New York: Plenum Press.

Griffiths, R., Double, M. C., Orr, K., & Dawson, R. J. (1998). A DNA test to sex most birds. *Molecular Ecology*, *7,* 1071-1075.

Guillette, L. M., Hahn, A. H., Hoeschele, M., Przyslupski, A. M., & Sturdy, C. B. (2015). Individual differences in learning speed, performance accuracy and exploratory behaviour in black-capped chickadees. *Animal Cognition*, *18*, 165-178.

Hahn, A. H., Guillette, L. M., Hoeschele, M., Mennill, D. J., Otter, K. A., Grava, T., Ratcliffe, L. M., & Sturdy, C. B. (2013). Dominance and geographic information contained within black-capped chickadee (*Poecile atricapillus*) song. *Behaviour*, *150*, 1601-1622.

Hahn, A. H., Hoang, J., McMillan, N., Campbell, K., Congdon, J., & Sturdy, C. B. (2015). Biological salience influences performance and acoustic mechanisms for the discrimination of male and female songs. *Animal Behaviour*, *104*, 213-228.

Hahn, A. H., Hoeschele, M., Guillette, L. M., Hoang, J., McMillan, N., Congdon, J. V., Campbell, K. A., Mennill, D. J., Otter, K. A., Grava, T., Ratcliffe, L. M., & Sturdy, C. B. (2016). Black-capped chickadees categorize songs based on features that vary geographically. *Animal Behaviour*, *112*, 93-104.

Hoang, J. (2015). Pitch perception is not unitary: Evidence for the perception of pitch chroma in black-capped chickadees. Unpublished Psychology M.Sc. thesis (University of Alberta).

Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, & Psychophysics*, *74*, 1742-1760.

Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithological Science*, *5*, 5-14.

Klecka, W. R. (1980). Discriminant analysis. SAGE Publications, Newburg.

Linhart, P., Ratcliffe, V. F., Reby, D. & Spinka, M. (2015). Expression of Emotional Arousal in two different piglet call types. *PLoS ONE, 10*, 1-13.

Marler, P. (1976). Social organization, communication and graded signals: The chimpanzee and the gorilla. In P. P. G. Bateson & R. A. Hinde (Eds), *Growing points in ethology*. Cambridge: Cambridge University Press.

Marler, P. (2004). Bird calls: their potential for behavioral neurobiology. *Annals of the New York Academy of Sciences*, *1016*, 31-44.

Maruščáková, I. L., Linhart, P., Ratcliffe, V. F., Tallet, C., Reby, D., & Špinka, M. (2015). Humans (*Homo sapiens*) judge the emotional content of piglet (*Sus scrofa domestica*) calls based on simple acoustic parameters, not personality, empathy, nor attitude toward animals. *Journal of Comparative Psychology*, *129*, 1-32.

Massen, J. J. M., Pašukonis, A., Schmidt, J., & Bugnyar, T. (2014). Ravens notice dominance reversals among conspecifics within and outside their social group. *Nature Communications*, *5*, 1-7.

Maynard-Smith, J. & Harper, D. (2004). *Animal Signals*. Oxford University Press.

McComb, K., Reby, D., Baker, L., Moss, C., & Sayialel, S. (2003). Long distance communication of acoustic cues to social identity in African elephants. *Animal Behavior*, *65*, 317-329.

McComb, K., Reby, D., Baker, L., Moss, C., & Sayialel, S. (2003). Long-distance communication of acoustic cues to social identity in African elephants. *Animal Behaviour*, *65*, 317-329.

McMillan, N., Hahn, A. H., Congdon, J. V., Campbell, K. A., Hoang, J., Scully, E. N., Spetch, M. L., & Sturdy, C. B. (2017). Chickadees discriminate contingency reversals presented consistently, but not frequently. *Animal Cognition*, *20*, 655-663.

Morton, E. S. (1977). On the occurrence and significance of motivational-structural rules in some bird and mammal sounds. *American Naturalist*, *111*, 855-869.

Mukhin, A., Chernetsov, N., & Kishkinev, D. (2008). Acoustic information as a distant cue for habitat recognition by nocturnally migrating passerines during landfall. *Behavioural Ecology*, *19*, 716-723.

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, *1*, 261-289.

Nicastro, N. & Owren, M. J. (2003). Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology*, *117*, 44-52.

Njegovan, M., Hilhorst, B., Ferguson, S., & Weisman, R. (1994). A motor-driven feeder for operant training in song birds. *Behavior Research Methods, Instruments, & Computers*, *26*, 26-27.

Owren, M. J. & Rendall, D. (2001). Sound on the rebound: Returning form and function to the forefront of understanding nonhuman primate signaling. *Evolutionary Anthropology*, *10*, 58-71.

Pyle (1997). Identification guide to North American birds. Slate Creek Press, Bolinas.

Palya, W. L., & Walter, D. E. (2001). *Document set for the high-performance experiment controller*. Retrieved 25 October 2014 from http://www.jsu.edu/depart/psychology/sebac/Exp-Ctl.html.

Reber, S. A., Nishimura, T., Janisch, J., Robertson, M., & Fitch, W. T. (2015). A Chinese alligator in heliox: Formant frequencies in a crocodilian. *Journal of Experimental Biology*, *218*, 2442-2447.

Reichert, M. S. (2011). Aggressive calls improve leading callers' attractiveness in the treefrog *Dendropsophus ebraccatus*. *Behavioral Ecology*, *22*, 951-959.

Reichert, M. S. (2013). Patterns of variability are consistent across signal types in the treefrog, *Dendropsophus ebraccatus*. *Biology J. Linn. Soc.*, *109*, 131-145.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161-1178.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *The Quarterly Journal of Experimental Psychology*, *63*, 2251-2272.

Scheumann, M., Hasting, A. S., Kotz, S. A., & Zimmermann, E. (2014). The voice of emotion across species: How do human listeners recognize animals' affective states? *PLoS One*, *9*, 1-10.

Schmidt, K.A. & Ostefeld, R.S. (2008). Eavesdropping squirrels reduce their future value of food under the perceived presence of cache robbers. *The American Naturalist*, *171*, 388-393.

Seddon, N., Merrill, R. M., & Tobias, J. A. (2008). Sexually selected traits predict patterns of species richness in a diverse clade of suboscine birds. *The American Naturalist*, *171*, 620-631.

Smith, S. M. (1991). *The black-capped chickadee: Behavioral ecology and the natural history*. Ithaca, NY: Cornell University Press.

Stocker, M., Munteanu, A., Stöwe, M., Schwab, C., Palme, R., & Bugnyar, T. (2016). Loner or socializer? Ravens' adrenocortical response to individual separation depends on social integration. *Hormones and Behavior*, *78*, 194-199.

Stoeger, Charlton, Kratochvil, & Fitch (2011). Vocal cues indicate level of arousal in infant African elephant roars. *Journal of Acoustical Society of America*, *130*, 1700-1710.

Stoeger, A. S., Baotic, A., Li, D., & Charlton, B. D. (2012). Acoustic features indicate arousal in infant giant panda vocalisations. *Ethology, 118*, 896-905.

Sturdy, C. B., & Weisman, R. G. (2006). Rationale and methodology for testing auditory cognition in songbirds. *Behavioural Processes*, *72*, 265-272.

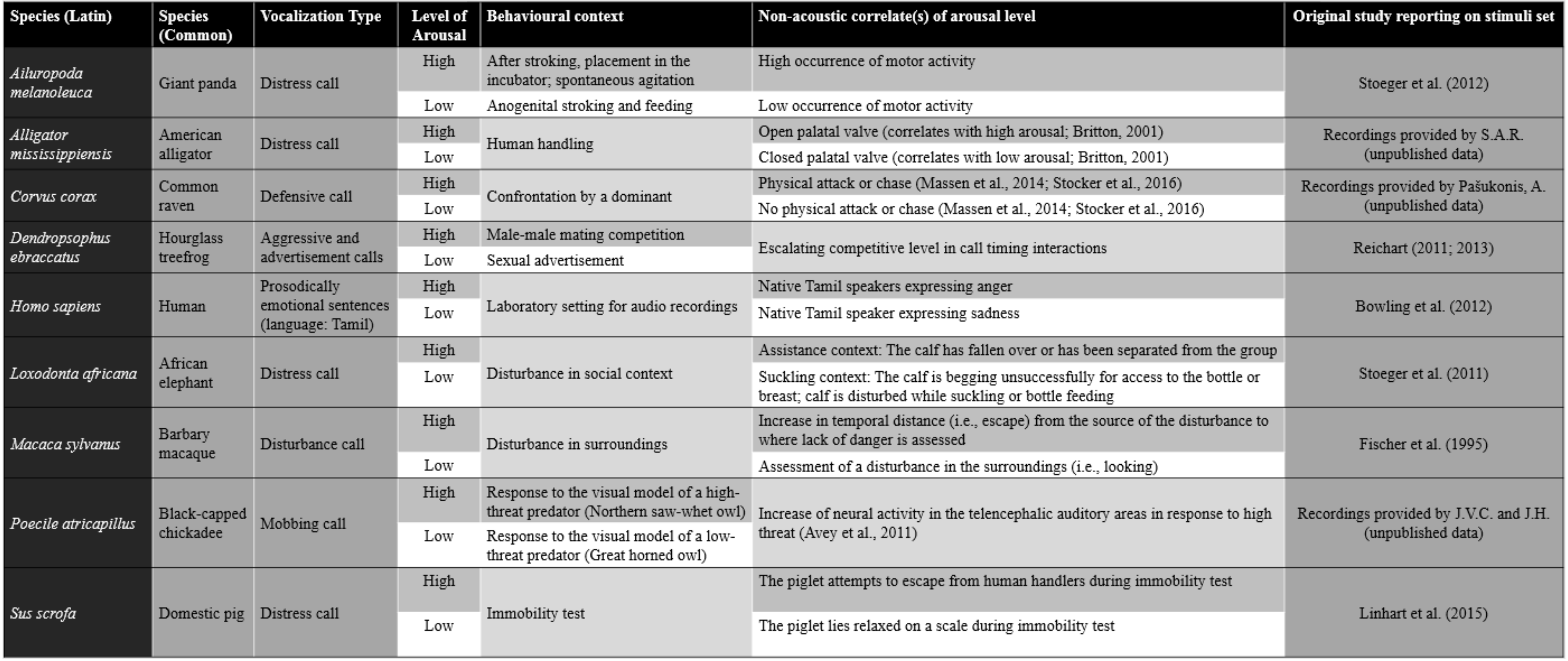
Templeton, C. N. & Greene, E. (2007). Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. *Proceedings of the National Academy of Sciences*, *104,* 5479-5482.

Templeton, C. N., Greene, E., & Davis, K. (2005). Allometry of alarm calls: Black-capped chickadees encode information about predator size. *Science*, *308*, 1934-1937.

Vitousek, M. N., Adelman, J. S., Gregory, N. C., & St Clair, J. J. (2007). Heterospecific alarm call recognition in a non-vocal reptile. *Biology Letters*, *3*, 632-634.

Table 1

The nature (i.e., vocalizations type, behavioural context under which vocalizations were recorded, and non-acoustic correlate(s) of arousal level) of the high- and low-arousal vocalizations collected from nine species that were used as stimuli for all three experiments explained with references (modified from Filippi et al., 2017's Table S1).



Species (Latin)	Species (Common)	Vocalization Type	Level of Arousal	Behavioural context	Non-acoustic correlate(s) of arousal level	Original study reporting on stimuli set
Ailuropoda melanoleuca	Giant panda	Distress call	High	After stroking, placement in the incubator; spontaneous agitation	High occurrence of motor activity	Stoeger et al. (2012)
			Low	Anogenital stroking and feeding	Low occurrence of motor activity	
Alligator mississippiensis	American alligator	Distress call	High	Human handling	Open palatal valve (correlates with high arousal; Britton, 2001)	Recordings provided by S.A.R. (unpublished data)
			Low		Closed palatal valve (correlates with low arousal; Britton, 2001)	
Corvus corax	Common raven	Defensive call	High	Confrontation by a dominant	Physical attack or chase (Massen et al., 2014; Stocker et al., 2016)	Recordings provided by Pašukonis, A. (unpublished data)
			Low		No physical attack or chase (Massen et al., 2014; Stocker et al., 2016)	
Dendropsophus ebraccatus	Hourglass treefrog	Aggressive and advertisement calls	High	Male-male mating competition	Escalating competitive level in call timing interactions	Reichart (2011; 2013)
			Low	Sexual advertisement		
Homo sapiens	Human	Prosodically emotional sentences (language: Tamil)	High	Laboratory setting for audio recordings	Native Tamil speakers expressing anger	Bowling et al. (2012)
			Low		Native Tamil speaker expressing sadness	
Loxodonta africana	African elephant	Distress call	High	Disturbance in social context	Assistance context: The calf has fallen over or has been separated from the group	Stoeger et al. (2011)
			Low		Suckling context: The calf is begging unsuccessfully for access to the bottle or breast; calf is disturbed while suckling or bottle feeding	
Macaca sylvanus	Barbary macaque	Disturbance call	High	Disturbance in surroundings	Increase in temporal distance (i.e., escape) from the source of the disturbance to where lack of danger is assessed	Fischer et al. (1995)
			Low		Assessment of a disturbance in the surroundings (i.e., looking)	
Poecile atricapillus	Black-capped chickadee	Mobbing call	High	Response to the visual model of a high-threat predator (Northern saw-whet owl)	Increase of neural activity in the telencephalic auditory areas in response to high threat (Avey et al., 2011)	Recordings provided by J.V.C. and J.H. (unpublished data)
			Low	Response to the visual model of a low-threat predator (Great horned owl)		
Sus scrofa	Domestic pig	Distress call	High	Immobility test	The piglet attempts to escape from human handlers during immobility test	Linhart et al. (2015)
			Low		The piglet lies relaxed on a scale during immobility test	

Table 2

Results of the paired-samples *t*-tests conducted on black-capped chickadee subjects' responding between high vs. low arousal vocalizations for each species during Transfer of Training.

Species	***t-test***	***p-value***	***Cohen's d***	***Confidence Intervals***
S+ High Group (*df* = 2)				
Giant panda	**4.491**	**0.046**	**6.351**	**.0252, 1.1769**
American alligator	0.948	0.443	1.341	-.2483, .3887
Common raven	-1.540	0.263	-2.178	-.2203, .1042
Hourglass treefrog	3.273	0.082	4.629	-.1310, .9644
Human	**6.062**	**0.026**	**8.573**	**.1693, .9974**
African elephant	0.378	0.742	0.535	-.1049, .1251
Barbary macaque	1.801	0.213	2.547	-.2139, .5220
Black-capped chickadee	**5.450**	**0.032**	**7.707**	**.1054, .8957**
Domestic pig	3.024	0.097	4.277	-.0944, .5409
S+ Low Group (*df* = 2)				
Giant panda	**-25.000**	**0.002**	**-35.355**	**-.8140, -.5749**
American alligator	-1.696	0.232	-2.398	-.7042, .3060
Common raven	1.214	0.349	1.719	-.4831, .8627
Hourglass treefrog	-1.211	0.349	-1.713	-1.4866, .8335
Human	**-10.771**	**0.009**	**-15.233**	**-.5728, -.2458**
African elephant	3.308	0.081	4.678	-.0533, .4079
Barbary macaque	-1.053	0.403	-1.489	-.4932, .2993
Black-capped chickadee	**-10.583**	**0.009**	**-14.967**	**-.3978, -.1678**
Domestic pig	**-6.804**	**0.021**	**-9.622**	**-.6596, -.1485**

Bold font indicates significance.

Table 3

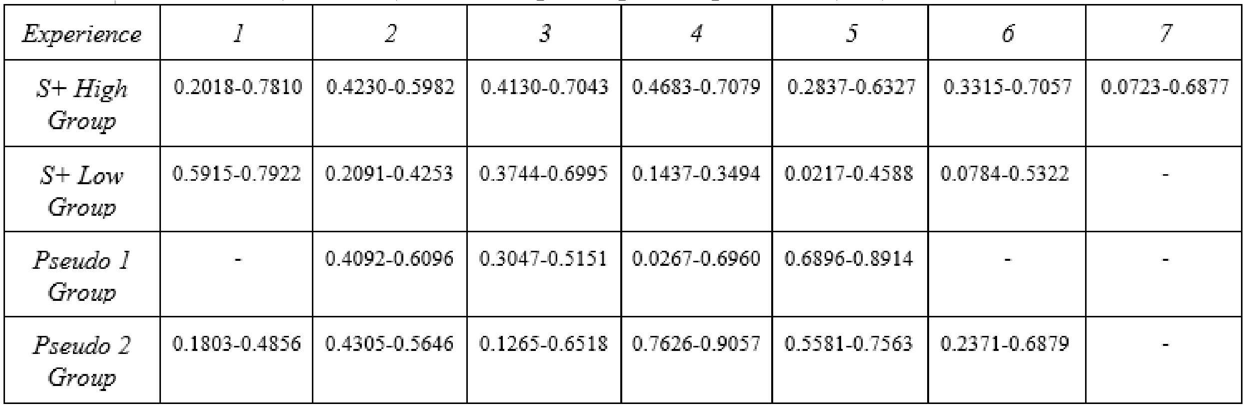
Results of the paired-samples *t*-test results of human participants' responding to each species' vocalizations during Transfer of Training.

Species	*t-test*	*p-value*	*Cohen's d*	*Confidence Intervals*
S+ High Group (*df* = 15)				
Giant panda	**4.901**	**<0.001**	**2.531**	**.2150, .5460**
American alligator	-0.490	0.631	-0.253	-.0836, .0523
Common raven	-0.382	0.708	-0.197	-.1389, .0967
Hourglass treefrog	0.544	0.595	0.281	-.1059, .1784
Human	**6.079**	**<0.001**	**3.139**	**.3790, .7882**
African elephant	**4.276**	**0.001**	**2.208**	**.1099, .3285**
Barbary macaque	0.746	0.467	0.385	-.0673, .1398
Black-capped chickadee	**6.199**	**<0.001**	**3.201**	**.3051, .6249**
Domestic pig	2.005	0.063	1.035	-.0082, .2701
S+ Low Group (*df* = 14)				
Giant panda	**-2.160**	**0.049**	**-1.155**	**-.5998, -.0022**
American alligator	-1.179	0.258	-0.630	-.1720, .0500
Common raven	-1.175	0.260	-0.628	-.1738, .0501
Hourglass treefrog	-2.062	0.058	-1.102	-1.935, .0038
Human	**-2.328**	**0.035**	**-1.244**	**-.07054, -.0289**
African elephant	-1.141	0.273	-0.610	-.2088, .0638
Barbary macaque	-0.322	0.753	-0.172	-.1278, .0945
Black-capped chickadee	-2.113	0.053	-1.128	-.4598, .0035
Domestic pig	0.127	0.901	0.068	-.0845, .0952

Bold font indicates significance, and underlined font indicates levels approaching significance.

Table 4

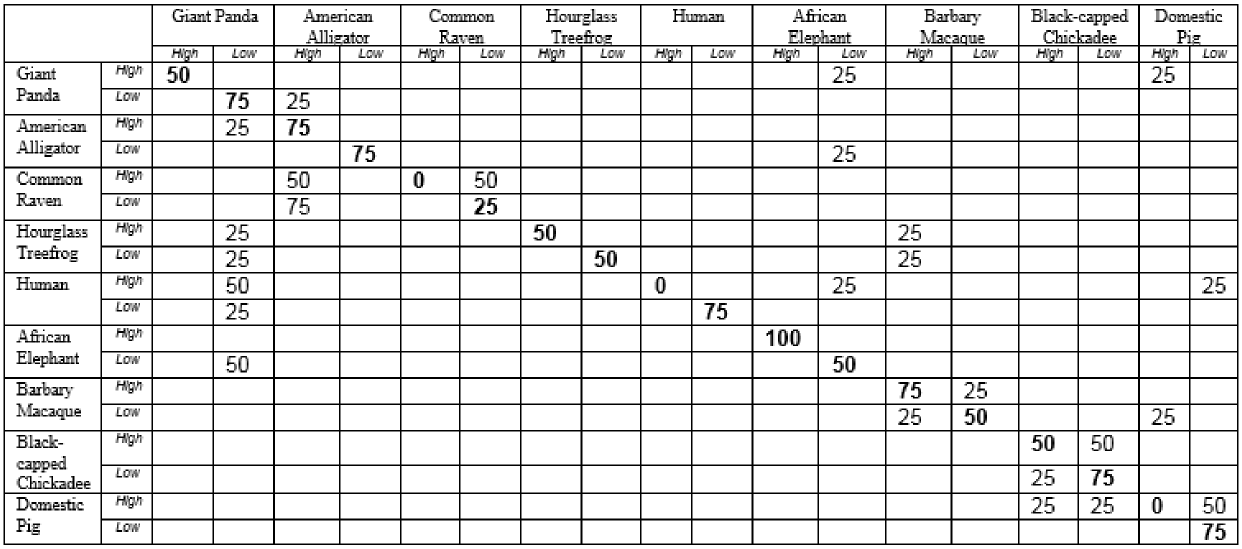
Confidence intervals (95% CIs) for listed participant experience (1-7).



Experience	*1*	*2*	*3*	*4*	*5*	*6*	*7*
S+ High Group	0.2018-0.7810	0.4230-0.5982	0.4130-0.7043	0.4683-0.7079	0.2837-0.6327	0.3315-0.7057	0.0723-0.6877
S+ Low Group	0.5915-0.7922	0.2091-0.4253	0.3744-0.6995	0.1437-0.3494	0.0217-0.4588	0.0784-0.5322	-
Pseudo 1 Group	-	0.4092-0.6096	0.3047-0.5151	0.0267-0.6960	0.6896-0.8914	-	-
Pseudo 2 Group	0.1803-0.4856	0.4305-0.5646	0.1265-0.6518	0.7626-0.9057	0.5581-0.7563	0.2371-0.6879	-

Table 5

Matrix of classification of Transfer testing stimuli by vocalization type using cross-validation in a Discriminant Function Analysis (DFA).



		Giant Panda		American Alligator		Common Raven		Hourglass Treefrog		Human		African Elephant		Barbary Macaque		Black-capped Chickadee		Domestic Pig	
		High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Giant Panda	High	**50**											25					25	
	Low		**75**	25															
American Alligator	High		25	**75**															
	Low				**75**								25						
Common Raven	High			50		**0**	50												
	Low			75			**25**												
Hourglass Treefrog	High		25					**50**						25					
	Low		25						**50**					25					
Human	High		50							**0**			25						25
	Low		25								**75**								
African Elephant	High											**100**							
	Low		50										**50**						
Barbary Macaque	High													**75**	25				
	Low													25	**50**			25	
Black-capped Chickadee	High															**50**	50		
	Low															25	**75**		
Domestic Pig	High															25	25	**0**	50
	Low																		**75**

The percentage of correct classifications is provided in bold along the diagonal. The percentage of misclassifications is given in corresponding rows and columns. Overall, 52.8% of the cross-validated cases were correctly classified. Empty cells indicate a percentage of zero.

Table 6

The acoustic measures used by the stepwise DFA to determine which four acoustic measures (duration, F0 SCG, HNR) could predict arousal level (high or low), and respective statistical output. Using the inclusion/exclusion criterion, only one or two acoustic measures were included in the discriminant function classifying each species:

Species	*Acoustic measure(s)*	*Wilks' λ*	$\mathcal{X}^2$	*p-value*
Giant panda	F0	0.138	18.826	<0.001
American alligator	SCG	0.198	15.389	<0.001
Common raven	SCG	0.494	6.698	0.010
Hourglass treefrog	F0/HNR	0.016	37.004	<0.001
Human	SCG	0.156	17.665	<0.001
African elephant	HNR/Duration	0.123	14.663	0.001
*Barbary macaque**	-	-	-	-
Black-capped chickadee	Duration/F0	0.307	10.617	0.005
Domestic pig	F0	0.386	9.035	0.003

*Note: Barbary macaque vocalization measures did not produce a discriminant function as none of the features met our criterion for inclusion in the model.

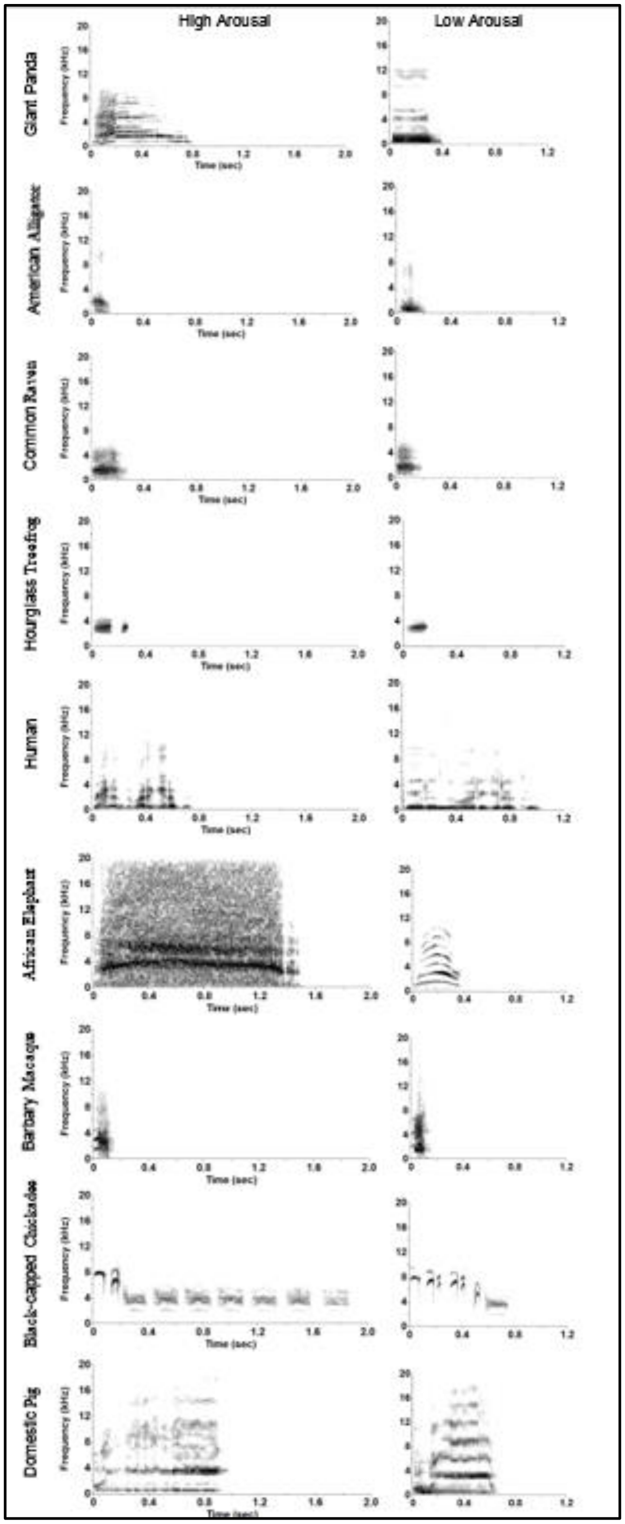


Figure 1. Spectrograms of high and low arousal vocalizations produced by each of the nine species (Giant Panda, American Alligator, Common Raven, Hourglass Treefrog, Human, African Elephant, Barbary Macaque, Black-capped Chickadee, and Domestic Pig), with Time on the x-axis (sec) and Frequency (kHz) on the y-axis.

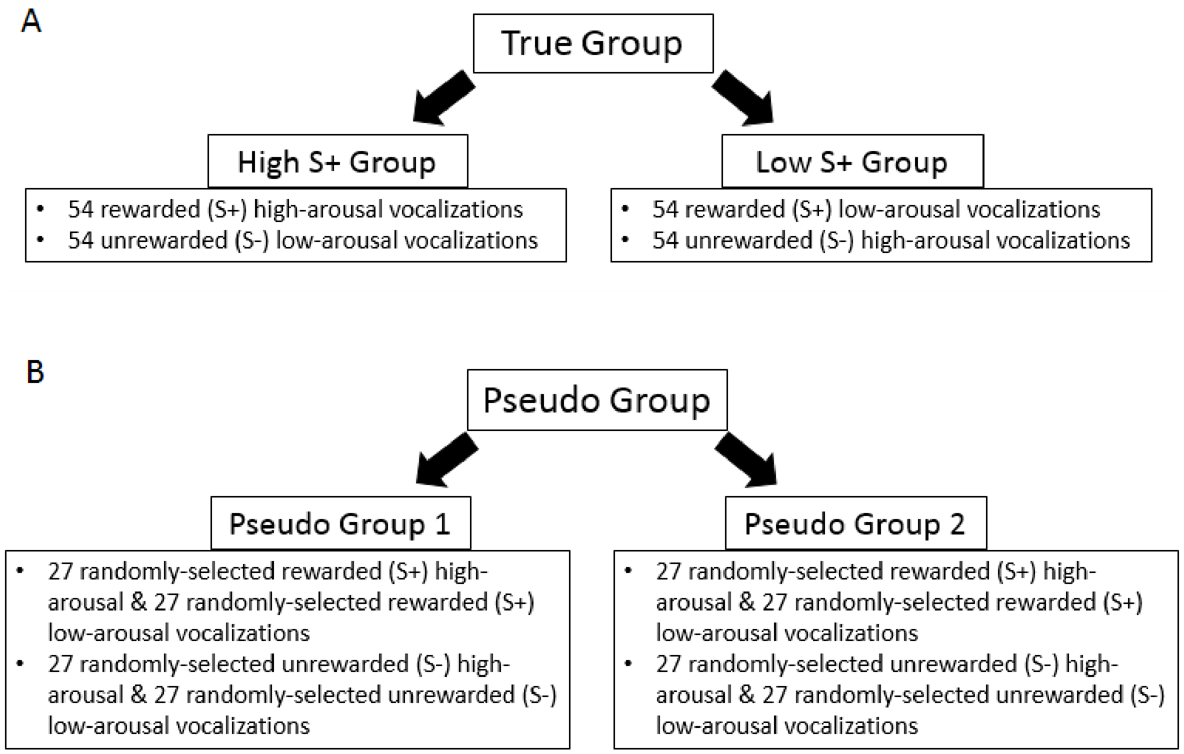


Figure 2. Black-capped chickadee subjects (Experiment 1) and human participants (Experiment 2) were randomly assigned to the (**A**) True or (**B**) Pseudo group and rewarded (S+) and punished (S-) for responding to different acoustic stimuli.

Experiment 1

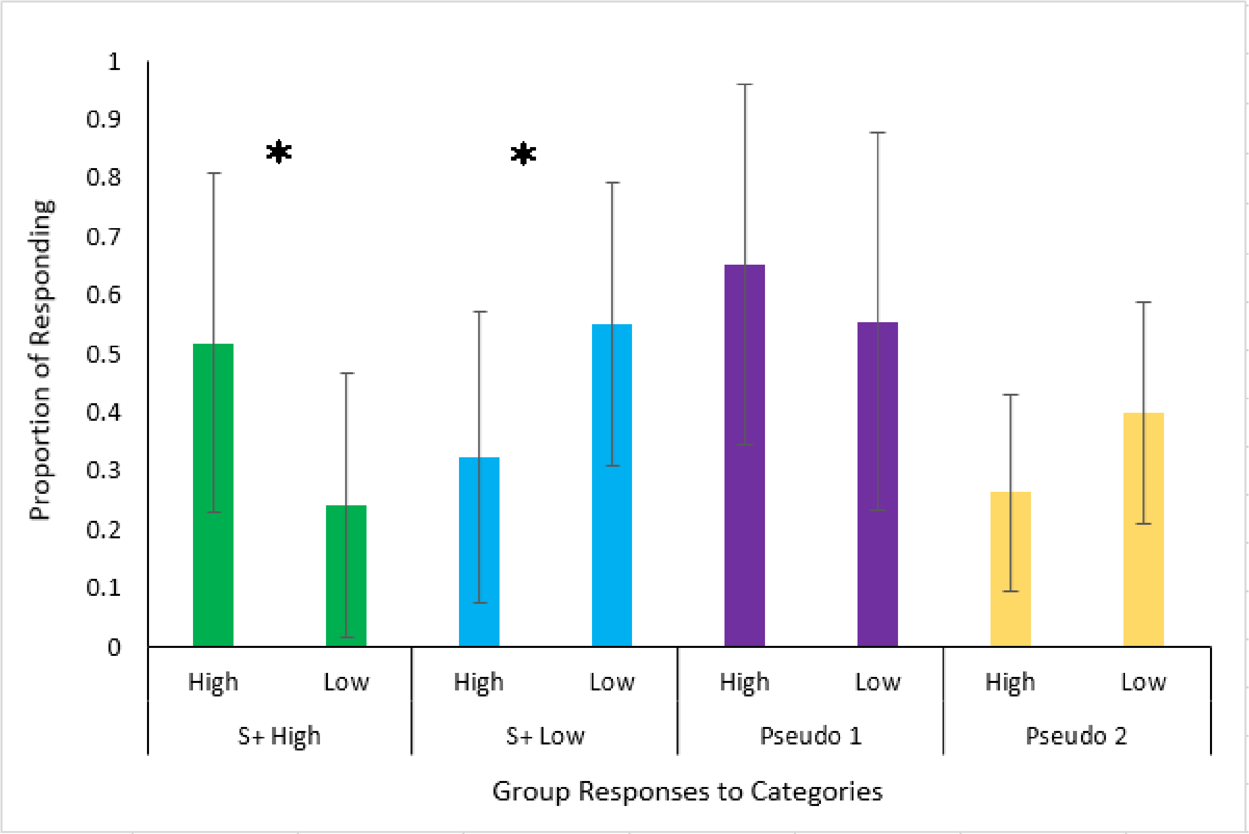


Figure 3. Proportion of responding to high vs. low arousal stimuli for each condition by black-capped chickadees (*n* = 12) during the transfer testing phase in Experiment 1 ± CI.
* indicates a significant difference between two adjacent bars.

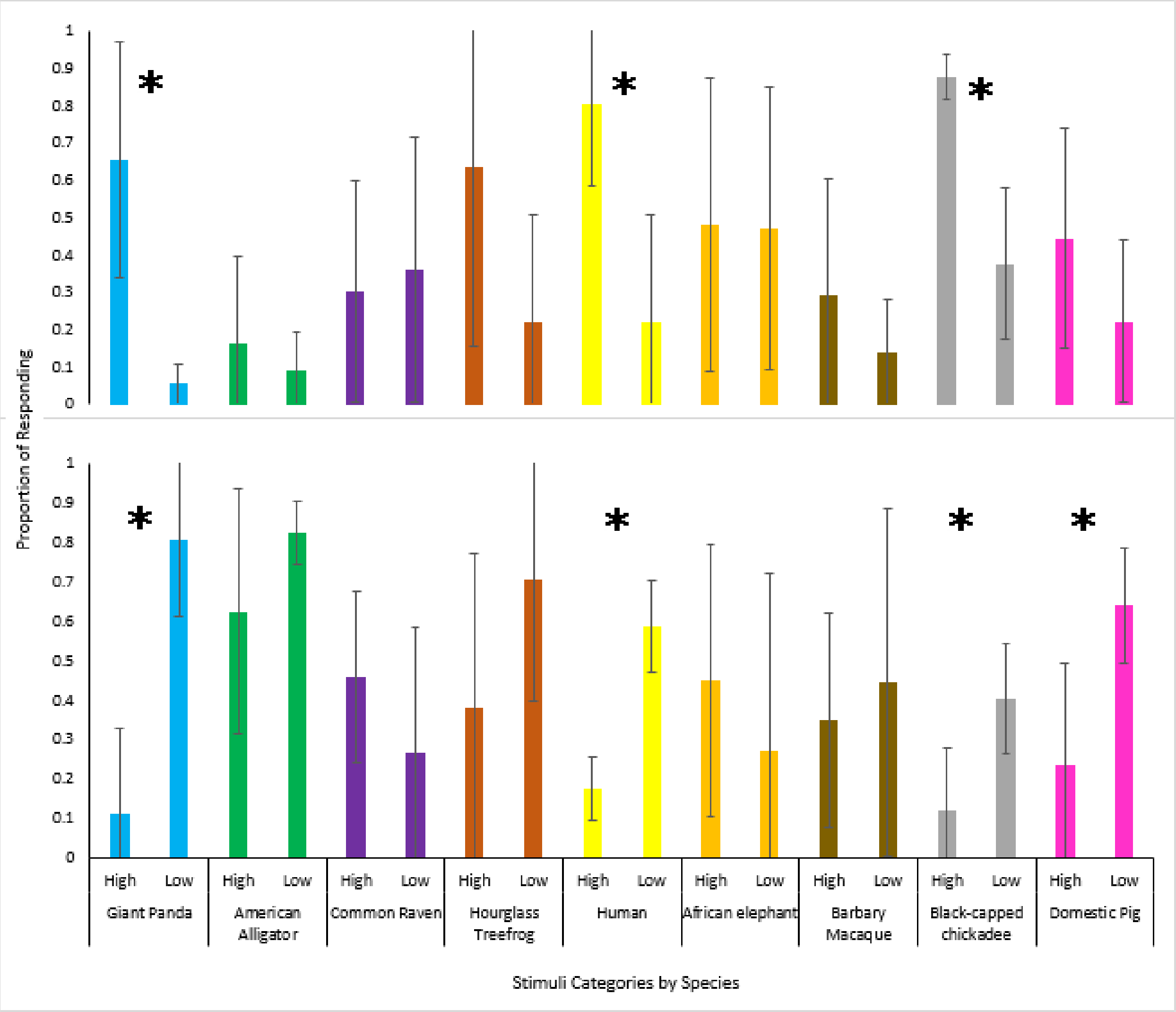


Figure 4. Proportion of responding to high arousal (previously rewarded) vs. low arousal (previously unrewarded) stimuli to each stimulus category (i.e., species) by black-capped chickadees in the S+ High group during the Transfer testing phase in Experiment 1 ± CI [top], and proportion of responding to low arousal (previously rewarded) vs. high arousal (previously unrewarded) stimuli to each stimulus category (i.e., species) by black-capped chickadees in the S+ Low group during the transfer testing phase in Experiment 1 ± CI [bottom].
* indicates a significant difference between two adjacent bars.

Experiment 2

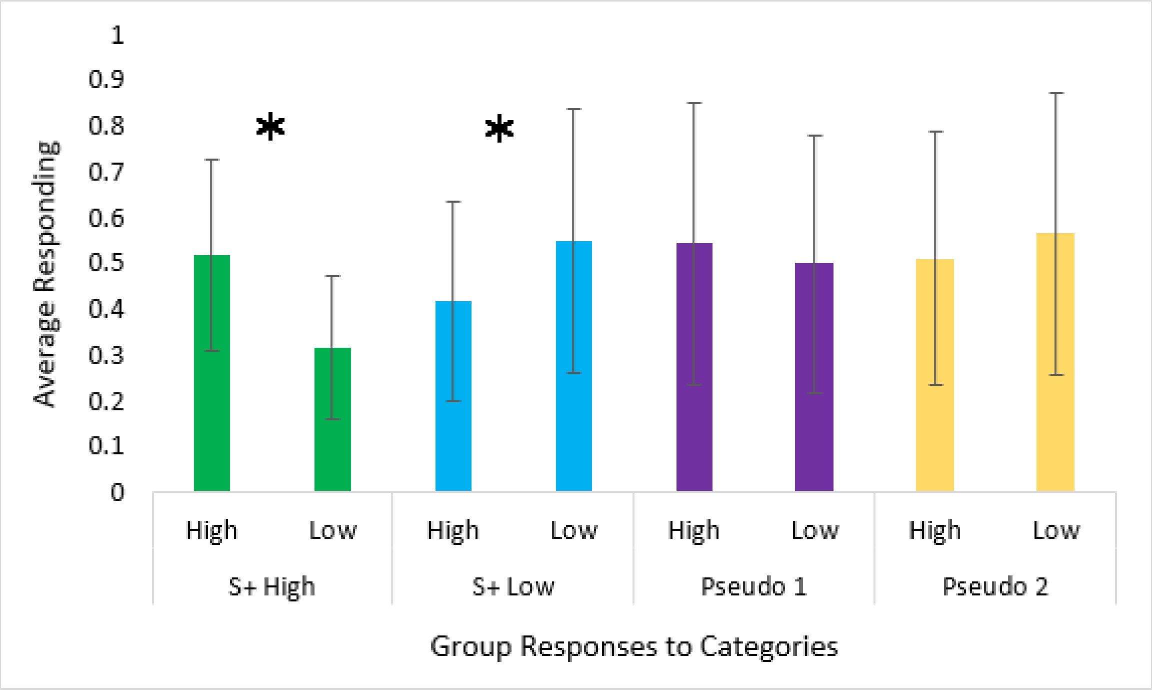


Figure 5. Average responding to high vs. low arousal stimuli for each condition by human participants (*n* = 76) during the transfer testing phase in Experiment 2 ± CI.
* indicates a significant difference between two adjacent bars.

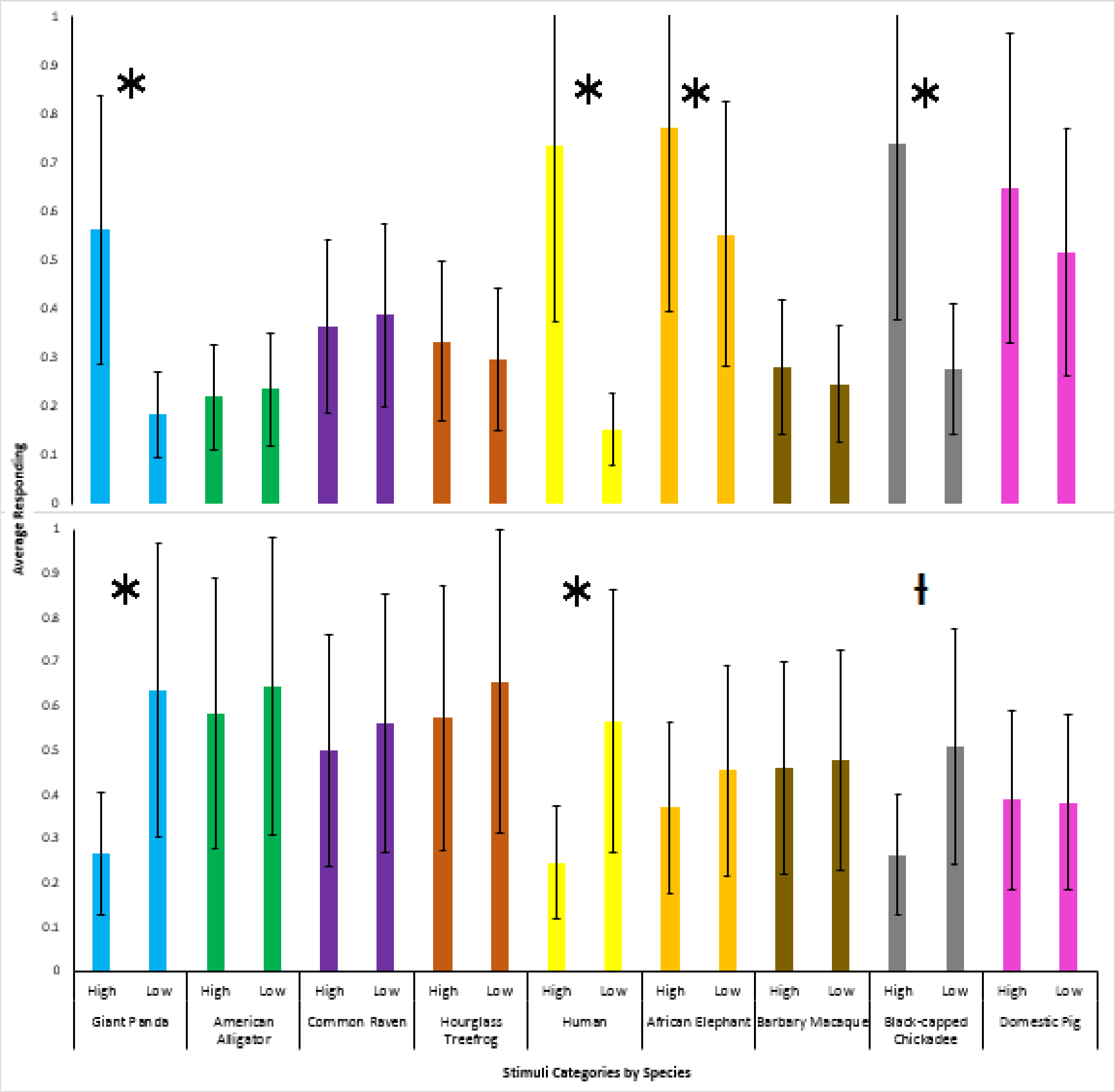


Figure 6. Average responding to high arousal (previously rewarded) vs. low arousal (previously unrewarded) stimuli to each stimulus category (i.e., species) by human participants in the S+ High group during the transfer testing phase in Experiment 2 ± CI [top], and average responding to low arousal (previously rewarded) vs. high arousal (previously unrewarded) stimuli to each stimulus category (i.e., species) by participants in the S+ Low group during the transfer testing phase in Experiment 2 ± CI [bottom].

* indicates a significant difference between two adjacent bars.

ǂ indicates a difference approaching significance between two adjacent bars.