Domestic Dogs (*Canis lupus familiaris*) are Sensitive to the "Human" Qualities of Vocal Commands

Jennifer M. Gibson[1], Stephanie A. Scavelli[1], Chester J. Udell[2], and Monique A. R. Udell[3*]

[1]Flagler College
[2]University of Oregon
[3]Oregon State University

*Corresponding author (Email: moniqueudell@gmail.com)

Abstract - In recent years, domestic dogs have been recognized for their ability to utilize human communicative gestures in choice tasks, as well as communicate with humans through visual and auditory means. A few dogs have even demonstrated the capacity to learn hundreds to thousands of human words and object labels with extensive training. However less is known about dogs' understanding or perception of human vocalizations in the absence of explicit training. This study was conducted to determine what aspects of human scolding vocalizations dogs would be most responsive to when presented with a choice to consume or avoid available food items. Variables included the gender, authenticity, word clarity and the human quality of the vocal commands. Our results suggest that dogs are generally cautious about novel sounds produced in the proximity of food. However they are most likely to avoid consumption when hearing a vocalization originally produced by a scolding human, suggesting awareness of vocal qualities common to human speech.

Keywords – Dog, *Canis familiaris*, Communication, Vocal commands, Human animal interaction, Cognition

The desire to converse intelligibly with animals has captivated the minds of humans for generations. Talking animals with complex emotional lives are the protagonists of fables dating back thousands of years (e.g., Aesop's Fables) and are equally common in the art, literature, and movie plots of today, often shaping how children and the broader community come to view and interact with the animals around them (Timmerman & Ostertag, 2012). Possibly even more common is the belief that non-human animals – including dogs – can understand human speech (Pongrácz, Miklósi, & Csányi, 2001) as well as accurately interpret and even share our emotions (Morris, Doe, & Godsell, 2008,). While some of these examples may fall into the realm of anthropomorphism, there has also been historical scientific interest in cross-species communication and species continuity in this domain. For example, Charles Darwin (1872) contemplated the possibility of shared traits in the emotional expression of man and animals from an evolutionary perspective. He believed that some characteristics of human and non-human animal vocalizations, gesture, and related emotional content were products of a shared origin, and thus varied only in degree, not in kind.

Indeed, especially in the case of vocalizations, some universal (or at least widespread) acoustic qualities have been identified that are often used to communicate similar contextual or affective information across species. For example, pitch has been correlated with signaler size across multiple species; the lower the pitch of an individual's vocalizations often the larger their size (Hall, Kingma, & Peters, 2013; Taylor, Reby, & McComb, 2010). Individuals of many species have also been found to

lower their pitch in specific contexts to convey threat (Bartholomew & Collias, 1962), superiority- often, but not always, in mating or territorial contexts (Bee, Perrill, & Owen, 2000; Puts, Gaulin, & Verdolini, 2006) or dominance interactions (Puts et al., 2006; Vannoni & McElligott, 2008). Conversely, high pitch vocalizations have been positively correlated with low stress situations (Knutson, Burgdorf, & Panksepp, 2002), non-threatening interactions, such as those between parents and offspring (Bartholomew & Collias, 1962), and submission (Bartholomew & Collias, 1962; Puts et al., 2006). Exceptions to these patterns are often marked by other vocal characteristics (e.g., vocalization length, range of frequencies, rate of production, or harmonic structure) providing additional or alternative contextual information to the receiver (e.g., Jürgens, 1979).

While communication between species and groups with different communication styles or dialects might be limited, shared features may still allow individuals of one group to obtain relevant contextual information from individuals of another group or species to mutual or individual benefit; this may be especially true when it comes to warning vocalizations. For example, some mixed-species monkey troops have been shown to respond appropriately to predator alarm calls made by members of the other species with whom they were cohabitating (Wolters & Zuberbühler, 2003). Birds have also demonstrated responsiveness to distinct calls made by other species- for example in one study at least 10 different species responded in accordance with chickadee mobbing calls but did not respond to a playback of a chickadee song (Hurd, 1996) and at least one species, Nuthatches, appeared to be sensitive to variations within the chickadee mobbing call that correspond with varying degrees of predator risk (Templeton & Greene, 2007). Furthermore, Rainey, Zuberbühler, & Slater (2004), demonstrated that eagle-vulnerable Black-Casqued Hornbills (birds) could distinguish between playbacks of eagle and leopard alarm calls made by Diana and Campbell's monkeys, showing a greater response to the highly relevant eagle calls than to leopard calls, which did not signal immediate risk to this species. This finding suggests that, even across wide taxonomic divides, individuals may come to recognize and utilize the calls of other species as a product of evolutionary adaptation, learning, or both, when doing so results in increased benefit or reduced cost.

In recent years, special attention has been given to the relationship between domestic dogs and humans, especially with regard to their cross-species social and communicative interactions (Bensky, Gosling, & Sinn, 2013; Hare &   , 2007; Serpell, 1995; Udell & Wynne, 2008). Both evolutionary (phylogenetic) and lifetime (ontogenetic) explanations have been proposed to explain the success of dogs in human societies (Udell & Wynne, 2010) and, at least in the Western world, an exceptionally strong cross-species bond with humans (Udell & Wynne, 2008). Whereas the great majority of research in this domain has focused on gestural or visual communication between dogs and humans (see Udell, Dorey, & Wynne, 2010 for a review), several studies have also demonstrated that human subjects can often identify the affective state and recording context of dog vocalizations at above chance levels, based on auditory features alone (Pongrácz, Molnár, Dóka, & Miklósi, 2011; Pongrácz, Molnár, & Miklósi, 2006; Pongrácz, Molnár, Miklósi, & Csányi, 2005). In fact, even human infants as young as six months of age have demonstrated the ability to match aggressive and non-aggressive barks of dogs with the appropriate canine facial expressions (Flom, Whipple, & Hyde, 2009) suggesting that this sensitivity can develop early in life. However, humans are not simply passive consumers of the kind of contextual and affective cues found in canine vocalizations; we are also producers. Like many animal species, humans exhibit sexual dimorphism in the pitch of their vocalizations, including speech, with males having lower pitch voices on average than females; in addition males have been found to lower their voices further when exerting dominance in a competitive situation (Puts et al., 2006). Conversely, high or rising pitched vocalizations are often interpreted as submissive, yielding, and non-threatening across many human cultures (Ohala, 1983).

However, relatively few empirical studies have investigated if and how specific characteristics of human vocalizations influence a dog's behavioral response or predict compliance. One recent study by Scheider, Grassmann, Kaminski, and Tomasello (2011) found that dogs were more likely to follow the point of a human who said "da" ("there" in German) in a high pitched friendly tone of voice, than when the human used a low pitched voice. In the latter case, dogs often sat instead of approaching the container.

In some cases dogs have shown sensitivity to specific phonemes within tape-recorded human commands. For example, Fukuzawa, Mills, and Cooper (2005) systematically altered phonemes within vocal recordings of the commands "sit" and "come". When "sit" was played back to the dog subjects as **CH**it, s**A**t, and si**K**, in each case altering one of three possible phonemes while holding the others constant, dogs who had previously been trained and tested for their compliance with the command "sit" generally failed to sit when one of the altered commands was given. Instead, on average, they engaged in nonspecific responses such as lifting their paw or orienting towards the trainer. Dogs also appear sensitive to the context of human cues, at least when the contextual information is visual. For example, Fukuzawa et al. (2004) found that a human's posture, distance from the dog, and the presence of sunglasses over the humans eyes all influenced the likelihood that dogs would comply with a verbal sit and come command. Virányi, Topál, Gácsi, Miklósi, and Csányi, (2004) also found that a dog was most likely to comply when a recorded "sit" command was played if the instructor was looking at them instead of looking at another human, away from the dog, or when the instructor could not be seen. Likewise several studies have demonstrated that dogs are more likely to comply with a verbal instruction to 'leave' forbidden food untouched when the human remains in the room after the initial scolding, maintaining visual contact with the dog and food (Bräuer, Call, & Tomasello, 2004; Call, Bräuer, Kaminski, & Tomasello, 2003) and when the room is well lit - disobeying more often when the room or the location of the food is dark (Kaminski, Pitsch, & Tomasello, 2012). It is important to note, that many studies looking at dog responsiveness to human vocal commands also include (or focus on the importance of) accompanying visual cues or gestures; only occasionally have auditory qualities of human speech served as the sole independent variable under test (e.g., Fukuzawa et al., 2005). As a result there is still much to learn about the range of vocal features that may influence dog behavior in a variety of training and interaction contexts with humans, including the degree to which dogs are responding to common evolutionary characteristics of human vocalizations (tone/pitch cues), learned associations (human words/referents) or a combination of these features.

Dogs are especially interesting subjects in research on responsiveness to human vocal cues for a number of reasons. As the first domesticated species, they have lived in close proximity to humans for over 14,000 years (Nobis, 1979) and this has shaped their evolution in significant ways (Udell et al., 2010). For example, domestication is responsible for a shift in the critical period for socialization in dogs compared to their wild counterparts (Lord, 2013; Trut, 1999) allowing dogs to form bonds with other species, including humans, later into their lives when encounters with individuals outside of their family group are more likely. Domesticated animals, including dogs, have also been found to maintain a wider range of vocalizations into adulthood (Price, 1984), which could in turn make them more interested in or sensitive to the wide range of auditory stimuli produced by humans. Of course, domestication, by definition, has also resulted in animals that maintain closer proximity to humans throughout their lives (Coppinger & Coppinger, 2001). This is especially true of pet and working dogs. This proximity may result in ample opportunity for dogs to learn about human vocalizations and even word meaning (Udell & Wynne, 2008). Indeed, learning has been found to play a critical role in dogs' responsiveness to human body language and gestures (Bentosela, Barrera, Jakovcevic, Elgier, & Mustaca, 2008; Elgier, Jakovcevic, Mustaca, & Bentosela, 2009; Udell, Dorey, & Wynne, 2011; Udell, Hall, Morrison, Dorey, & Wynne, 2013). Given the large body of evidence demonstrating the importance of both phylogeny (evolution) and ontogeny (maturation and learning) in the development of dogs' response to human gestures (Reid, 2009; Udell et al., 2010; Udell & Wynne, 2010), it is reasonable to predict that both phylogeny and ontogeny should influence dogs' response to human vocalizations as well.

In parts of the world where pet dogs are common -for example the United States where there are over 83 million dogs living as pets (APPA, 2013)- many humans willingly bring dogs into their homes adopting them as companions, partners, or even children. In fact, one study found that the way human owners speak to their dogs is quite similar to how they talk to human infants in some respects: high pitched voice, short statements, proper grammar, and repetitiveness in both contexts (Mitchell, 2001). However, some differences existed as well. For example, owners were less likely to ask their dogs questions or treat them as an active conversant. Nonetheless, studies such as this suggest that pet dogs

have ample opportunity to learn something about words, tones, or other qualities of speech and possibly their relationship to specific objects, contexts, or outcomes within the human home. Such regular, albeit informal, learning opportunities may allow for a more sophisticated responsiveness to human speech than would be expected from a common evolutionary response to shared qualities of mammalian vocalizations alone. In other words, dogs may be acting out a naturalistic experiment, mirroring those attempted with non-human primates from decades past (Hayes & Hayes, 1952), providing an opportunity to study what a non-human animal might learn about human speech as a consequence of child-like immersion into the human home.

Recently several important scientific studies have shown that some pet dogs, after many years of dedicated training, can learn to respond accurately to as many as 1,000 human words or labels (Kaminski, Call, & Fischer, 2004; Pilley & Reid, 2011) and even some elements of human syntax (Pilley, 2013). However there is still much to learn about the kind and amount of information dogs, especially those without explicit language training, glean from human vocalizations in the absence of gestural information, and under what conditions they are most likely to respond in accordance with those vocalizations. For example, how important are words, formants, tone/pitch, or even 'humanness' to a dog's behavioral response in common human-dog communication scenarios, such as basic training, praise, and scolding events?

In the current study we presented domestic dogs with a choice to consume or avoid available, but forbidden, food items in the presence of a pre-recorded scolding. Some recordings were of an authentic human scolding; others were manipulated or manufactured to alter or distort specific qualities of human speech. We specifically chose to focus on auditory scolding in line with prior scientific investigation on visual or visual + auditory cues associated with scolding in the presence of forbidden food (e.g., Call et al., 2003; Horowitz, 2009). At the time of this study, dogs' response to different qualities of vocal cues associated with human scolding had not yet been investigated. We asked:

(1) Do dogs differentiate between human scolding vocalizations and dehumanized scolding vocalizations (containing some, but not all, features associated with human speech)?

(2) Does human gender or correspondence between visual and auditory cues of gender influence dog compliance in a scolding scenario?

(3) Are dogs sensitive to auditory cues associated with authentic, versus inauthentic, human scolding?

(4) Do dogs respond to words associated with scolding in the absence of other auditory features associated with human speech (e.g., pitch, tone, spectral/harmonic content)?

Method

Subjects

A total of 56 dogs (male, $N = 27$; female, $N = 29$), of mixed breeds residing at the Flagler County Humane Society were included in this study. Dogs ranged in age from four to 132 months of age ($M = 29.8$ months). To eliminate the possibility of order effects or habituation, each dog participated in only one of the seven possible conditions, resulting in eight experimentally naive dogs per condition. Dogs were randomly assigned to their testing condition at the time of enrollment. The only selection criteria for a dog's inclusion were: (1) a willingness to eat food out of the experimenter's hand without signs of fear or aggression and (2) good health (as certified by a veterinarian). Time at the shelter (number of days since intake) was recorded for each dog at the time of testing. Testing dogs in this environment allowed us to look at duration of time in a shelter environment as a possible variable. It also provided a consistent testing and living environment for all dogs during the time of test, where the amount and kind of recent interaction with humans, diet, and feeding schedule was equivalent across subjects.

We considered that a dog's age or the time that had elapsed since coming to the shelter might influence responsiveness to human scolding vocalizations (with younger dogs or dogs with a longer shelter stay possibly showing a less pronounced response to scolding/ greater food consumption). For this

analysis we used the four unaltered human vocalizations following the methods described below. The mean age of dogs across these four conditions was roughly equal (mean age between groups varied by less than 1 year, full age range: 4 months-11 years). We did not find a strong correlation between a dog's age and food consumption (Pearson correlation, $r = -0.01$, $p = 0.61$) nor between the duration of time a dog had been living at the shelter and food consumption (Pearson correlation, $r = 0.01$, $p = 0.68$). This result suggests that responsiveness to human scolding vocalizations, as measured in this study, had likely already developed or been learned by four months of age. Once established, dog responsiveness to scolding also appears to be at least moderately resistant to later environmental change, i.e., relocation from a home environment to a shelter environment.

Vocal Recordings/Experimental Stimuli

Recording of human vocalizations. To capture human scolding vocalizations one adult male and one adult female, naive to the purpose of this study, were recruited to record an unscripted and a scripted version of a forbidden food scolding.

Unscripted/authentic scolding (male & female). The goal of this session was to create recordings of authentic scolding/verbal deterrents that could be played back through a speaker at a later point in the experiment.

The male scolder was asked to stand in an empty hallway selected for the recordings. A trained sound engineer and audio expert (C.U.) stood next to him, holding an H4N Zoom Flash Audio Recorder (Samson Technologies, Hauppauge, NY) and RØDE ® NTG2 microphone (RØDE Microphones LLC, Long Beach, CA). Prior to bringing a dog into the recording area, the scolder was given a bowl of dry dog food and told that he could say whatever he needed to keep the dog from eating food from the bowl for 30 s. However, he could not touch the dog, say the dog's name, use gestures, or physically block the food from the dog.

An assistant then led a dog into the hallway. The scolder was then told to place the food bowl on the floor in front of him and the audio recording began, capturing the scolder attempting to keep the dog away from the food using verbal instruction alone. After 30 s of scolding, the audio expert cued the scolder indicating that the recording session had ended. The dog was then allowed to consume the food.

The same recording procedure was later followed for the female scolder.

Although the dog readily ate food from the bowl both before and after each scolding (once permission had been granted), both scolders successfully deterred the dog from eating the food for the full 30 s of their recording.

Scripted/inauthentic scolding (male & female). After both male and female volunteers had recorded their authentic unscripted scolding, the experimenter replayed the recordings in private and wrote down what each individual had said, thereby, creating a written script of each scolding. The male and female were then given their script individually in the absence of the dog. Each scolder was recorded in the same hallway once again, this time reading their script aloud in a neutral context. Afterwards, the scripted version of the scolding was modified in *Amadeus Pro™* (HairerSoft, United Kingdom), an audio editing and analysis software, so that the timing of each word spoken in the scripted version of the recording matched that of the unscripted scolding along the 30 s timeline. This was done so the only salient feature that differed between the two scolding types was the contextual information: authentic or inauthentic scolding tone.

After all four recordings were captured, background noise was removed using *PVC 1.0™* (Paul Koonce, high-precision spectral filtering algorithms) and all recordings were normalized to the same volume (100% RMS, Root Mean Square) using *Amadeus Pro* by C.U.

Creation of dehumanized vocalizations. For both dehumanized conditions, the authentic female scolding was modified in the manner described below.

Distorted vocalization (female). The objective was to create a distorted speech stream derived from the original female recording so that the same human speech patterns and tone were still discernible, but the words were no longer identifiable, not unlike the "wawa" vocalizations of the teacher from the

Peanuts™ cartoons. A spectral FFT based "Formants Mapper" technique was employed by C.U. using *PVC*, where the formants of the analyzed speech sequence were shifted to the nearest formants of a target spectrum. The target spectrum in this case was an isolated vowel from the original speech stream: The "o" in "No." This resulted in the speech stream's formant structure being constrained to a single static spectrum, degrading the word intelligibility without destroying the other "human-specific" features of the vocalization. In other words, this vocalization had the timbral qualities, pitch, rhythmic pacing, and spectral articulations of human speech, and the same 30 s duration as the original recording, only the word content was no longer identifiable.

   Robotic vocalization (female). In this case the words spoken in the human scolding, were typed into a computer generated speech program (*Verbalize* ™ Corporation Eight), played back and recorded. This preserved the word content of the scolding, while removing the human idiosyncrasies and tone. In other words, this condition was the antithesis of the distorted vocalization. As in the case of the scripted scolding, this vocalization was modified so that the timing of the words matched that of the unscripted recording over the 30 s timeline.

   Control/silence. Thirty seconds of silence was recorded. This file was used during control trials (as opposed to simply not playing a file) to additionally control for any extraneous variables that might have been present in the experimental conditions (movements involved with playing the sound, electrical signals associated with playing an audio file, etc.).

   All audio files were saved using the Audio Interchange File Format (AIFF) at a sample rate of 44100 Hz and a bit depth of 16 (CD quality) and placed on an iPod nano, capable of transmitting the full frequency range of human speech (Frequency range of iPod nano: 20 Hz – 20,000 Hz).

Testing Layout

   The testing layout was similar to that used in Faragó, Pongrácz, Range, Virányi, and Miklósi (2010), see Figure 1. A 4'X4' square was marked out on the floor with blue painter's tape at one end of the room. A medium sized dog crate was placed in the center of the 4'X4' taped square. A YAMAHA® MSP5 Studio speaker was placed on top the crate. The crate and speaker were concealed with a cloth crate cover. The iPod containing the vocalizations was connected to the speaker and sat on top, uncovered. The experimenter stood neutrally on a marked spot to the right of the speaker. An assistant video recorded the dog's behavior through a window opposite the experimenter, from outside the room. A tripod holding a second camera was set up to the right of the cage to capture the gaze of the subject from the experimenter's perspective. The same testing layout and procedure was used for all experimental and control conditions, so that the only manipulated variable was the sound/vocalization coming from the speaker.

Forbidden Food

   Twenty pieces of dry dog food and a single medium sized milk bone served as the forbidden food items in this study. Both the bone and dry food pieces were deposited in a metal dog bowl. The experimenter placed the baited food bowl on a marked position on the floor, in front of the speaker, before the start of each session.

   While prior studies have used contact with a bone as the primary measure of a dog's compliance or discomfort in the presence of an audio recording (e.g., Faragó et al., 2010), we predicted that a less portable food item, like small pieces of dry dog food, might allow for a more sensitive indicator of a dog's level of compliance and physiological state. Therefore both food types were included in the current study. The predictive value of each food item with respect to overall approach behavior and consumption across conditions was compared during analysis.
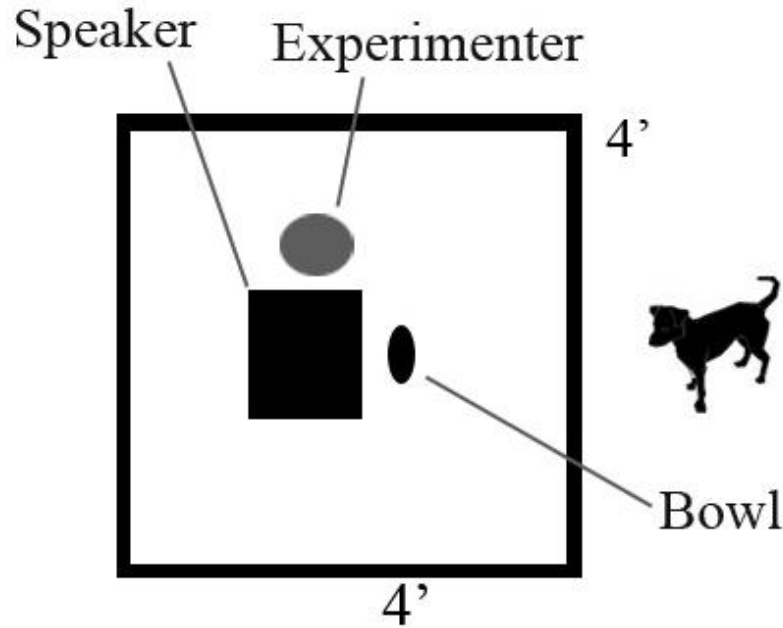
Figure1. Experimental layout.

Experimental Testing

Each dog participated in only one of the experimental or control conditions to ensure that prior experimental exposure (order effects) could not account for differences found between conditions. Before the start of a session the dog was brought into the testing room and allowed 5 min to habituate to its surroundings, no food or bowl was present during this time. A female experimenter stood on a marked spot to the right of the speaker looking at the dog with a neutral expression.

The female experimenter was present during all testing and control conditions from the time the dog entered the room until the end of the trial, to give the impression that any vocalization could be coming from a human who could enforce it. Prior studies have found that dogs often ignore vocal commands given by humans (or recordings of humans) if no human is physically present (Call et al., 2003; Fukuzawa et al., 2004). Therefore the constant presence of a human was intended to increase the likelihood that the dog would attend to the vocal recordings played from the speaker. The same human was also present during control conditions where the recording contained no sound. This controlled for any influence that the presence of a human might have on subject behavior independent of the auditory variables under test.

After the 5 min habituation period, the assistant who was outside of the room during testing tapped on the wall opposite the experimenter to temporarily draw the dogs attention away from the testing area; this marked the start of the trail. At this cue, the experimenter placed the bowl containing the edible dog bone and 20 pieces of dry food on the floor and simultaneously hit play on the iPod to begin the 30 s condition-specific audio track (male unscripted, male scripted, female unscripted, female scripted, distorted, robot or control/silence). After the 30 s audio recording finished playing (indicated by an audible "ping" at the end of the track), the experimenter picked up the bowl and recorded the number of pieces of food eaten by the dog (20 minus the number of pieces still in the bowl) during the trial. The experimenter also recorded if the bone had been removed from the bowl by the dog, and if either the bone or dry food had been visibly consumed during the test. Each dog was tested in one of the seven conditions in this manner until eight dogs had been tested per condition.

The remaining behavioral observations were recorded from video by a coder (J. G.) at a later date. A second video coder, naive to the purpose of the study, was randomly assigned to code 38% of the session videos (3 dogs per condition). Before coding began, a third party (M.U.) renamed all video files so that the condition being coded was unknown to the coders. Inter-rater reliability scores were high (IRR: Average time in contact with bowl 90%; Bone taken out of the innermost square 100%; Time looking at speaker 85%; Time looking at human 100%).

Data Analysis

Non-parametric statistics were used in all analyses due to the moderate sample size ($n = 8$) for each condition. All statistical tests were two-tailed, with an alpha of 0.05 unless otherwise noted. Specific statistical tests employed are listed with the results.

Results

Sensitivity of Measurement Based on Food Type

We predicted that of the two food items used in the current study, a consumable dog bone and 20 small pieces of dry food, dry food consumption would serve as a more sensitive measure of a dog's response to an auditory stimulus/scolding vocalization.

Our data supported this prediction. Similar numbers of dogs interacted with the bone across conditions independent of the auditory stimulus presented: Four dogs in Female Scripted (FS); Five dogs in Female Unscripted (FU), Male Unscripted (MS) and Distorted (D); Six dogs in Male Unscripted (MU), Robot (R) & Control(C). Furthermore, half of the subjects took the bone out of the bowl and out of the inner 4' square even during the control condition -when no scolding stimulus was being played over the speaker- therefore bone removal could not strictly be interpreted as avoidance behavior even when it occurred.

Because exact consumption measurements were difficult with the bone (it was often crushed and scattered without actual consumption) we instead attempted to measure the percent of time each subject visibly held part or all of the bone in their mouth. Again, no significant difference between conditions was identified (Kruskal-Wallis Test, $H(6) = 1.52$, $p = 0.96$). Furthermore, while inter-rater reliability on this measure was not exceptionally low (75% agreement), it was the lowest of any other measure used in our analysis, making it a less than ideal measure of subject compliance or affective state.

Conversely, dry food consumption could be measured precisely. The number of pieces consumed also positively correlated with the percent of time a dog spent in contact with the food bowl (Spearman's rho correlation $= 0.74$, $p = 0.01$), and consequently in proximity to the speaker. Finally, in the control condition –where no sound was played- every subject consumed at least some dry food (whereas two of the eight dogs did not interact with the bone at all) suggesting that the dry food was universally palatable. Therefore dry food consumption proved to be the more sensitive measure of food consumption or avoidance behavior on the part of the dog, and was therefore used for the remainder of the analysis.

Response to Scolding Vocalizations

Significant differences were found in the number of pieces of food consumed across the three broad experimental categories: Control/silence, dehumanized scolding and human scolding (Kruskal-Wallis Test, $H(6) = 18.77$ $p = 0.005$). The greatest number of food pieces was consumed in the control condition, where no scolding occurred, followed by the dehumanized conditions, and the least amount of food was consumed during the human scolding conditions (Figure 2).
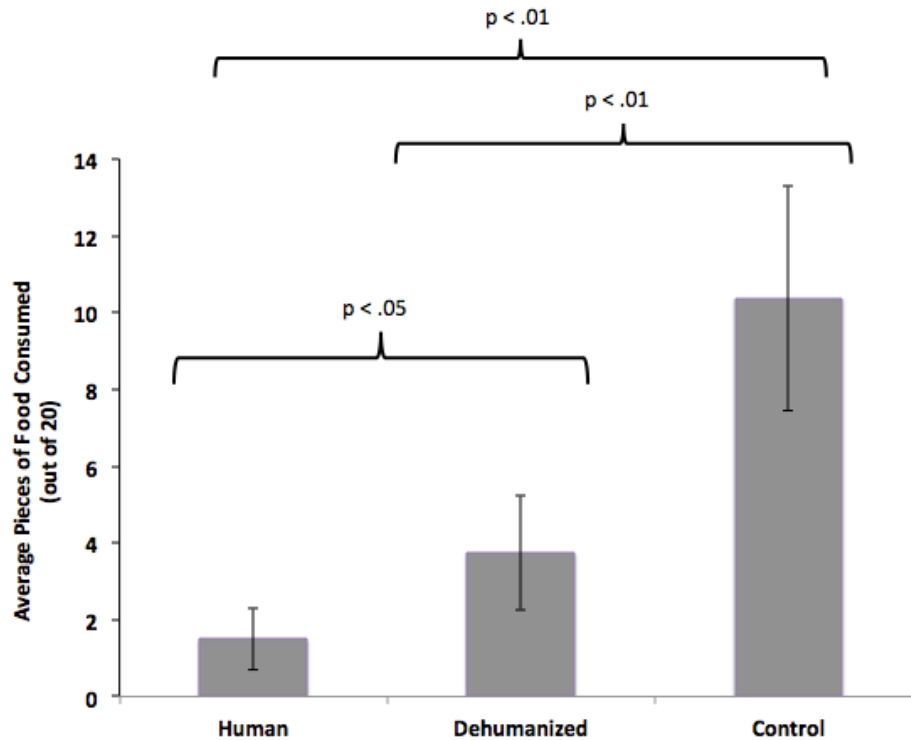
Figure 2. Average amount of dry food consumed across experimental categories. Error Bars = ± SEM

Does human scolding affect food consumption? If recorded human scolding vocalizations deter dogs from consuming forbidden food, then we would expect to see a reduced level of food consumption in each human scolding condition when compared to the control condition (in the absence of sound). Indeed, in each case (FS, FU, MS, MU), dogs consumed less food on average in the experimental (human scolding) conditions than in the control condition (quantity of food consumed during Control Trials, $M = 10.38$ pieces; Female Unscripted, $M = 0.88$, Mann–Whitney $U = 3$, $p = 0.001$; Female Scripted, $M = 1.62$, $U = 7.5$, $p = 0.007$); Male Unscripted, $M = 3.13$, $U = 10.5$, $p = 0.02$); Male Scripted, $M = 0.38$, $U = 2$, $p = 0.001$; Figure 3). A significant difference was also found between the average amount of food eaten across all human scolding conditions combined and the average amount of food consumed by dogs in the control condition (control, $M = 10.38$; human scolding, $M = 2.25$, Mann–Whitney $U = 47$, $p < 0.001$).

Do human gender cues matter? Two elements that could have influenced a dogs' response to the human scolding vocalizations were absolute deepness of voice – an evolutionary cue correlated with gender in humans (Puts et al., 2006) – or a learned visual or olfactory correspondence between the gender of the experimenter (female) and the gender cues present in scolding recording (female or male).

For example, dogs might be expected to consume less and spend less time in the proximity of the bowl and speaker when the scolding recording featured a male voice, as lower tone has been associated with avoidance of a target containing food in previous studies (Scheider et al., 2011) and is associated with dominance in humans (Puts et al., 2006). Alternatively, dogs might be expected to consume less and spend less time in the proximity of the bowl and speaker when the recording was of a female voice – in the presence of visual and auditory correspondence – because the only person present (capable of enforce the scolding) was always female. Similarly, if dogs recognized that a male voice would likely not come from the female experimenter they should consume more under these circumstances, much like dogs have been shown to do when the human issuing a "leave it" command is not in the room (e.g., Call et al., 2003).

Interestingly, no significant difference was found in dogs' responsiveness to male and female scolding recordings (see Figure 3). On average dogs consumed roughly the same amount of dry food independent of whether the voice was male ($M = 1.75$ pieces) or female ($M = 1.25$) (Mann–Whitney $U = 127$, $p = 0.98$); average food consumption was low for both groups suggesting compliance with the scolding in both cases. Likewise no significant difference was found in the percent of time spent in physical contact with the bowl independent of whether the voice was male ($M = 13\%$) or female ($M = 16\%$) (Mann–Whitney $U = 109$, $p = 0.49$). The percent of time spent looking at the speaker (male $M = 12\%$, female $M = 7\%$, Mann–Whitney $U = 117.5$, $p = 0.70$) and human experimenter (male $M = 2\%$, female $M = 2\%$, Mann–Whitney $U = 110.5$, $p = 0.52$) also did not differ based on the gender cues associated with the recordings. Therefore, on average, dogs appeared to respond similarly to the scolding vocalizations made by males and females, independent of gender specific auditory cues or auditory-visual correspondence. However, further investigation of gender correspondence might be warranted before strong conclusions are drawn. For example, it is still possible that dogs would respond differently (consuming significantly more or less food) in the presence of male visual + auditory cues, a combination not tested in the current study. Nonetheless, the current study suggests that auditory gender cues alone were not sufficient to predict differential responding to scolding vocalization in dogs.
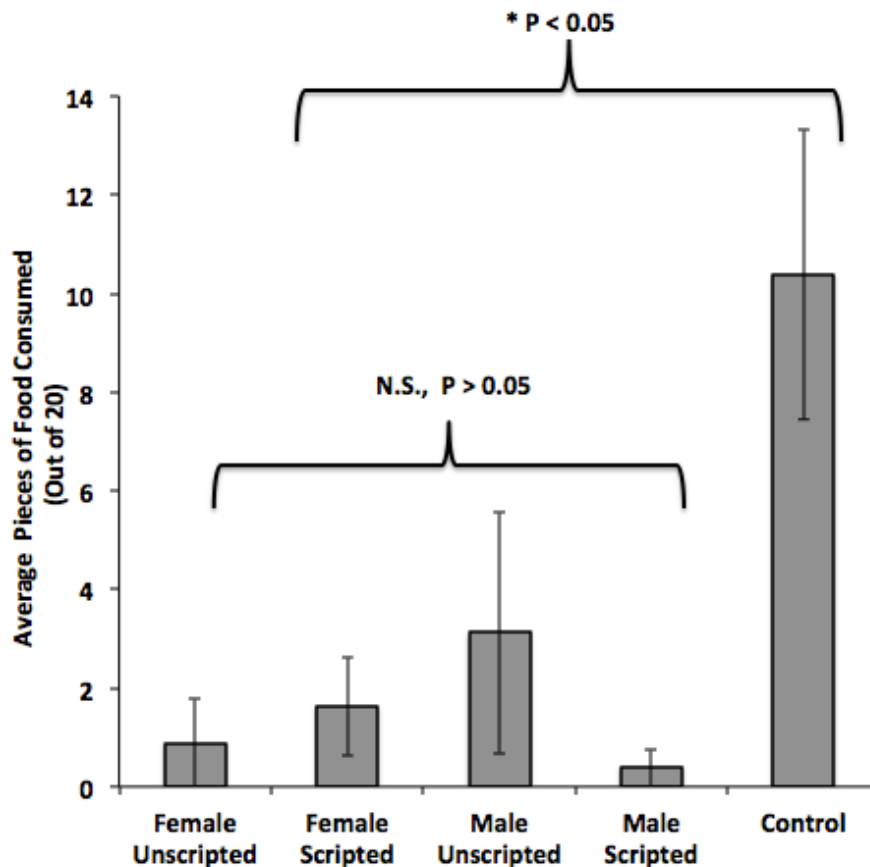

Figure 3. Average amount of dry food consumed across human scolding conditions. Error Bars = ± SEM.

Does authenticity/tone matter? We also asked whether dogs would show a differential response to unscripted/authentic versus scripted/inauthentic human scolding recordings. If dogs were sensitive to the context-specific auditory cues associated with human scolding, we would predict a greater hesitation to consume food in the presence of authentic versus scripted recordings. Instead, dogs consumed roughly the same amount of dry food on average in each case (Unscripted/authentic, $M = 2$ pieces; Scripted/inauthentic, $M = 1$ piece, Mann–Whitney $U = 126$, $p = 0.96$). There was also not a significant difference in the average percent of time dogs in each group spent in contact with the bowl (scripted, $M = 15\%$; unscripted, $M = 14\%$; Mann–Whitney $U = 113$, $p = 0.59$) or in the average time spent looking at the speaker (scripted, $M = 12\%$; unscripted, $M = 7\%$, $U = 118$, $p = 0.72$) or human experimenter (scripted, $M = 3\%$, unscripted, $M = 1\%$, $U = 107$, $p = 0.45$). This finding suggests that either the dogs in this study were not sensitive to context-specific auditory cues associated with authentic/inauthentic human scolding, or that these cues – if detected – did not translate to different behavioral responses on the part of the dog. Instead, on average, dogs complied with the human scolding independent of authenticity, avoiding the food.

Dehumanized scolding: Do human words or formants matter? Dehumanized scolding vocalizations also reduced dogs' average food consumption when compared with the control condition (control, $M = 10.38$ pieces; dehumanized vocalizations, $M = 3.75$ pieces; Mann–Whitney $U = 24$, $p = 0.01$).

However, dogs still ate more food on average in the presence of the dehumanized scolding recordings, than in the human scolding conditions (dehumanized scolding, $M = 3.75$ pieces; human scolding, $M = 2.25$; Mann–Whitney $U = 180.5$, $p < 0.05$).

There was no significant difference in food consumption between the Robot and Distorted conditions (robot, $M = 4.25$ pieces; distorted, $M = 3.25$, Mann–Whitney $U = 29$, $p = 0.80$) suggesting that the preservation of human words/formants alone or the preservation of human tone, pitch, rhythmic pacing, and spectral articulations (with the loss of the other features) led to roughly equivalent outcomes, with moderate levels of effectiveness in deterring food consumption when compared to control and human scolding conditions.

Discussion

The current study examined dogs' response to human and dehumanized scolding vocalizations in the presence of forbidden food. Recordings of a human scolding were the most effective in reducing dogs' forbidden food consumption. All categories of human scolding tested – unscripted and scripted, male and female – were equally effective deterrents, significantly reducing the amount of food consumed when compared to controls. Interestingly, a dog's age and time living in the shelter did not correlate with compliance. However, such questions might benefit from further study, where age or duration of time in a shelter environment is treated as the primary independent variable under test. Dehumanized scolding also resulted in decreased food consumption compared to the control condition (silence); however they did not deter consumption as effectively as the recordings of human scolding.

We originally predicted that dogs might be more responsive to some characteristics of human scolding than to others. As a result we asked if either the Robotic recording – a reproduction of human words stripped of human tonal, pitch and spectral qualities – taken alone, would be more influential in deterring forbidden food consumption than the Distorted recording- a recording that preserved human tonal, pitch and spectral information but rendered the word content unintelligible. Ultimately, no significant difference in consumption was found between these conditions. However the presence of both of these qualities combined (human words + human tone, pitch, and spectral information) was significantly more effective in deterring food consumption that either of these features alone. Therefore the combination of words and human vocal qualities is likely important to dogs' recognition of human speech and compliance with scolding vocalizations. Future studies should consider other elements of human speech that might be salient to dogs either in isolation or combination. For example, it has been suggested that dogs treat acoustic stimuli, including strings of words in the form of sentences, as one

signal (McBride, 1995), thus a 30-s long vocalization might make the content of individual words less salient then conditions where only one word (or a few words) are uttered together in a consistent order (e.g., Pilley & Reid, 2011). Other factors worth exploring further include the frequency and timing of individual utterances, the harmonic qualities of human vocalizations produced in different contexts, and recognition of auditory qualities or words typically associated with different categories of human vocalization (e.g., scolding versus praise).

The current data suggests that there is a quality to human scolding vocalizations (*humanness*) that inspires greater compliance in dogs (less forbidden food consumption) when compared with other sounds or the presence of a human in the absence of sound. This responsiveness is likely the product of interacting evolutionary and learning based mechanisms, and may be akin to the finding that pet dogs are more responsive to points made by a human than by a non-human object (Udell, Giglio, & Wynne, 2008). Dogs' ability to distinguish between human speech and other similar auditory stimuli may hold special value both within the lifetime of an individual and on an evolutionary scale. Dogs living in homes, shelters, or in working environments are exposed to a plethora of sounds, from music and traffic to the rumbling of a washing machine. Yet despite this noise, humans typically maintain and enforce the expectation that their dogs should listen to and comply with their commands. In other words even outside of a laboratory setting, dogs may experience daily discrimination training or learning opportunities that could result in the kind of differential response to human vocalizations identified in this study.

Evolutionary pressures experienced early in the domestication process, when dogs primarily lived as scavengers at the fringes of human society (much like many of the world's dogs still do today) may have also contributed to dogs' sensitivity to human vocalizations, especially scolding. Because free-living dogs are often treated as unwanted pests, human vocalizations may warn these dogs of an impending threat, allowing them to safely flee from approaching humans (Ortolani, Vernooij, & Coppinger, 2009). Under such circumstances, dogs that are quick to accurately identify and avoid the sound of human voices may be the most effective scavengers – avoiding potential threats from human inhabitants while reducing the costs associated with false alarms (abandoning a food source in the presence of non-human sounds). Therefore additional research is needed to fully understand the range of relevant mechanisms that may contribute to the development of dogs' response to human scolding vocalizations and to human vocalizations in general.

Interestingly, auditory cues associated with 'authentic' scolding scenarios (as opposed to scripted) appear to be less important to dogs than the 'human' quality of the vocalization. It is possible that in the absence of corresponding gestural cues, all human scolding vocalizations become ambiguous, causing dogs to err on the side of caution. There is evidence that dogs placed in an ambiguous scolding situation – the owner believes their dog has stolen forbidden food, but in reality they have not – will nonetheless engage in appeasement behavior (Horowitz, 2009). This response appears to be beneficial to the dog, whether they engaged in a forbidden act, as owners are less likely to punish a dog that "looks guilty" than one who does not (Hecht, Miklósi, & Gácsi, 2012). Research has also suggested that dogs sometimes continue to comply with human gestural cues, for example following a point to a container, even when physical cues (smell or other visual stimuli) suggest that they would be more likely to find food by ignoring the point (e.g., Szetei, Miklósi, Topál, & Csányi, 2003). Therefore, dogs' tendency to comply with all forms of human scolding presented in this study (even scripted/inauthentic ones) may stem from similar motivations. In fact, it is possible that dogs may respond to all human vocalizations as if they were deterrents in the current food based task, not just scolding vocalizations specifically. In other words, it is possible that in the absence of explicit word/language training dogs may not be as sensitive to subtle nuances of human speech, including tone and contextual information, as previously thought; even though they can be quite responsive to human vocalizations in general. Future studies will be needed to more precisely determine the range and cause of such overgeneralizations. It also seems likely that different populations of dog will show different levels of sensitivity to human vocalizations (or even specific categories of human vocalizations) based on their individual experiences; therefore comparative studies investigating the responsiveness of dogs living in different environments (homes, shelters, feral, etc.) are of future interest.

Acknowledgements

References

APPA. (2013). National pet owners survey: Industry statistics & trands. *American Pet Products Association*. Retrieved from http://www.americanpetproducts.org/press_industrytrends.asp

Bartholomew, G. A., & Collias, N. E. (1962). The role of vocalization in the social behaviour of the northern elephant seal. *Animal Behaviour*, *10*, 7–14. doi:10.1016/0003-3472(62)90124-0

Bee, M. A., Perrill, S. A., & Owen, P. C. (2000). Male green frogs lower the pitch of acoustic signals in defense of territories: A possible dishonest signal of size? *Behavioral Ecology*, *11*, 169–177. doi:10.1093/beheco/11.2.169

Bensky, M. K., Gosling, S. D., & Sinn, D. L. (2013). Chapter five - The world from a dog's point of view: A review and synthesis of dog cognition research. In T. J. R. H. Jane Brockmann (Ed.), *Advances in the study of behavior,* Vol. 45 (pp. 209–406). Oxford, UK: Academic Press.

Bentosela, M., Barrera, G., Jakovcevic, A., Elgier, A. M., & Mustaca, A. E. (2008). Effect of reinforcement, reinforcer omission and extinction on a communicative response in domestic dogs (*Canis familiaris*). *Behavioural Processes*, *78*, 464–469. doi:10.1016/j.beproc.2008.03.004

Bräuer, J., Call, J., & Tomasello, M. (2004). Visual perspective taking in dogs (*Canis familiaris*) in the presence of barriers. *Applied Animal Behaviour Science*, *88*, 299–317. doi:10.1016/j.applanim.2004.03.004

Call, J., Bräuer, J., Kaminski, J., & Tomasello, M. (2003). Domestic dogs (*Canis familiaris*) are sensitive to the attentional state of humans. *Journal of Comparative Psychology*, *117*, 257–263. doi:10.1037/0735-7036.117.3.257

Coppinger, R., & Coppinger, L. (2001). *Dogs: A startling new understanding of canine origin, behavior & evolution* (1st ed.). New York: Scribner.

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

Elgier, A. M., Jakovcevic, A., Mustaca, A. E., & Bentosela, M. (2009). Learning and owner-stranger effects on interspecific communication in domestic dogs (*Canis familiaris*). *Behavioural Processes*, *81*, 44–49. doi:10.1016/j.beproc.2008.12.023

Faragó, T., Pongrácz, P., Range, F., Virányi, Z., & Miklósi, Á. (2010). "The bone is mine": Affective and referential aspects of dog growls. *Animal Behaviour*, *79*, 917–925. doi:10.1016/j.anbehav.2010.01.005

Flom, R., Whipple, H., & Hyde, D. (2009). Infants' intermodal perception of canine (*Canis familairis*) facial expressions and vocalizations. *Developmental Psychology*, *45*, 1143–1151. doi:10.1037/a0015367

Fukuzawa, M., Mills, D. S., Cooper, J. J. (2004). More than just a word: Non-semantic command variables affect obedience in the domestic dog (*Canis familiaris*). *Applied Animal Behaviour Science*, *91*, 129–141. doi: 10.1016/j.applanim.2004.08.025

Fukuzawa, M., Mills, D. S., & Cooper, J. J. (2005). The effect of human command phonetic characteristics on auditory cognition in dogs (*Canis familiaris*). *Journal of Comparative Psychology*, *119*, 117–120. doi:10.1037/0735-7036.119.1.117

Hall, M. L., Kingma, S. A., & Peters, A. (2013). Male songbird indicates body size with low-pitched advertising songs. *PLoS ONE*, *8*, e56717. doi:10.1371/journal.pone.0056717

Hare, B., & Tomasello, M. (2005). Human-like social skills in dogs? *Trends in Cognitive Sciences*, *9*, 439–444. doi:10.1016/j.tics.2005.07.003

Hayes, K. J., & Hayes, C. (1952). Imitation in a home-raised chimpanzee. *Journal of Comparative and Physiological Psychology*, *45*, 450–459. doi: 10.1037/h0053609

Hecht, J., Miklósi, Á., & Gácsi, M. (2012). Behavioral assessment and owner perceptions of behaviors associated with guilt in dogs. *Applied Animal Behaviour Science*, *139*(1-2), 134–142. doi:10.1016/j.applanim.2012.02.015

Horowitz, A. (2009). Disambiguating the "guilty look": Salient prompts to a familiar dog behaviour. *Behavioural Processes*, *81*, 447–452. doi:10.1016/j.beproc.2009.03.014

Hurd, C. R. (1996). Interspecific attraction to the mobbing calls of black-capped chickadees (*Parus atricapillus*). *Behavioral Ecology and Sociobiology*, *38*, 287–292. doi:10.1007/s002650050244

Jürgens, U. (1979). Vocalization as an emotional indicator: A neuroethological study in the squirrel monkey. *Behaviour*, *69*, 88–117. doi:10.1163/156853979X00412

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog: evidence for "fast mapping." *Science*, *304*, 1682–1683. doi:10.1126/science.1097859

Kaminski, J., Pitsch, A., & Tomasello, M. (2012). Dogs steal in the dark. *Animal Cognition*, *16*, 385–394. doi:10.1007/s10071-012-0579-6

Knutson, B., Burgdorf, J., & Panksepp, J. (2002). Ultrasonic vocalizations as indices of affective states in rats. *Psychological Bulletin*, *128*, 961–977. doi:10.1037/0033-2909.128.6.961

Lord, K. (2013). A comparison of the sensory development of wolves (*Canis lupus lupus*) and dogs (*Canis lupus familiaris*). *Ethology*, *119*, 110–120. doi:10.1111/eth.12044

McBride, A. (1995). The human–dog relationship. In Robinson, I. (ed). The waltham book of human–animal interaction: benefits and responsibilities of pet ownership. Oxford: Pergamon.

, A. (2007). *Dog behaviour, evolution, and cognition*. New York: Oxford University Press.

Mitchell, R. W. (2001). Americans' talk to dogs: Similarities and differences with talk to infants. *Research on Language & Social Interaction*, *34*, 183–210. doi:10.1207/S15327973RLSI34-2_2

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

Nobis, G. (1979). Der älteste Haushunde lebte vor 14000 Jahren. [The oldest domestic dog lived 14,000 years ago]. *Umschau*, *79*, 610.

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*, 1–18. doi:10.1159/000261678

Ortolani, A., Vernooij, H., & Coppinger, R. (2009). Ethiopian village dogs: Behavioural responses to a stranger's approach. *Applied Animal Behaviour Science*, *119*, 210–218. doi:10.1016/j.applanim.2009.03.011

Pilley, J. W. (2013). Border collie comprehends sentences containing a prepositional object, verb, and direct object. *Learning and Motivation*, *44*, 229–240. doi:10.1016/j.lmot.2013.02.003

Pilley, J. W., & Reid, A. K. (2011). Border collie comprehends object names as verbal referents. *Behavioural Processes*, *86*, 184–195. doi:10.1016/j.beproc.2010.11.007

Pongrácz, P., Miklósi, Á., & Csányi, V. (2001). Owner's beliefs on the ability of their pet dogs to understand human verbal communication: A case of social understanding. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition*, *20*, 87–107.

Pongrácz, P., Molnár, C., Dóka, A., & Miklósi, Á. (2011). Do children understand man's best friend? Classification of dog barks by pre-adolescents and adults. *Applied Animal Behaviour Science*, *135*, 95–102. doi:10.1016/j.applanim.2011.09.005

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

Pongrácz, P., Molnár, C., Miklósi, Á., & Csányi, V. (2005). Human listeners are able to classify dog (*Canis familiaris*) barks recorded in different situations. *Journal of Comparative Psychology*, *119*, 136–144. doi:10.1037/0735-7036.119.2.136

Price, E. O. (1984). Behavioral aspects of animal domestication. *The Quarterly Review of Biology*, *59*, 1–32.

Puts, D. A., Gaulin, S. J. C., & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*, 283–296. doi:10.1016/j.evolhumbehav.2005.11.003

Rainey, H. J., Zuberbühler, K., & Slater, P. J. B. (2004). The responses of black-casqued hornbills to predator vocalisations and primate alarm calls. *Behaviour*, *141*, 1263–1277.

Reid, P. J. (2009). Adapting to the human world: Dogs' responsiveness to our social cues. *Behavioural Processes*, *80*, 325–333. doi:10.1016/j.beproc.2008.11.002

Serpell, J. (1995). *The domestic dog: Its evolution, behaviour, and interactions with people*. New York: Cambridge University Press.

Szetei, V., Miklósi, Á., Topál, J., & Csányi, V. (2003). When dogs seem to lose their nose: An investigation on the use of visual and olfactory cues in communicative context between dog and owner. *Applied Animal Behaviour Science*, *83*, 141–152. doi:10.1016/S0168-1591(03)00114-X

Taylor, A. M., Reby, D., & McComb, K. (2010). Size communication in domestic dog, *Canis familiaris*, growls. *Animal Behaviour*, *79*, 205–210. doi:10.1016/j.anbehav.2009.10.030

Templeton, C. N., & Greene, E. (2007). Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 5479–5482. doi:10.1073/pnas.0605183104

Timmerman, N., & Ostertag, J. (2012). Too many monkeys jumping in their heads: Animal lessons within young children's media. *Canadian Journal of Environmental Education (CJEE)*, *16*, 59–75.

Trut, L. (1999). Early canid domestication: The farm-fox experiment. *American Scientist*, *87*, 160. doi:10.1511/1999.2.160

Udell, M. A. R., Dorey, N. R., & Wynne, C. D. L. (2010). What did domestication do to dogs? A new account of dogs' sensitivity to human actions. *Biological Reviews*, *85*, 327–345. doi:10.1111/j.1469-185X.2009.00104.x

Udell, M. A. R., Dorey, N. R., & Wynne, C. D. L. (2011). Can your dog read your mind? Understanding the causes of canine perspective taking. *Learning & Behavior*, *39*, 289–302. doi:10.3758/s13420-011-0034-6

Udell, M. A. R., Giglio, R. F., & Wynne, C. D. L. (2008). Domestic dogs (*Canis familiaris*) use human gestures but not nonhuman tokens to find hidden food. *Journal of Comparative Psychology (Washington, D.C.: 1983)*, *122*, 84–93. doi:10.1037/0735-7036.122.1.84

Udell, M. A. R., & Wynne, C. D. L. (2008). A review of domestic dogs' (*Canis Familiaris*) human-like behaviors: Or why behavior analysts should stop worrying and love their dogs. *Journal of the Experimental Analysis of Behavior*, *89*, 247–261. doi:10.1901/jeab.2008.89-247

Udell, M. A. R., & Wynne, C. D. L. (2010). Ontogeny and phylogeny: Both are essential to human-sensitive behaviour in the genus *Canis*. *Animal Behaviour*, *79*, e9–e14.

Udell, M., Hall, N. J., Morrison, J., Dorey, N. R., & Wynne, C. D. L. (2013). Point topography and within-session learning are important predictors of pet dogs' (*Canis lupus familiaris*) performance on human guided tasks. *Revista Argentina de Ciencias del Comportamiento*, *5*, 3–20.

Vannoni, E., & McElligott, A. G. (2008). Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males. *PloS One*, *3*, e3113. doi:10.1371/journal.pone.0003113

Virányi, Z., Topál, J., Gácsi, M., Miklósi, Á., & Csányi, V. (2004). Dogs respond appropriately to cues of humans' attentional focus. *Behavioural Processes*, *66*, 161–172. doi:10.1016/j.beproc.2004.01.012

Wolters, S., & Zuberbühler, K. (2003). Mixed-species associations of Diana and Campbell's monkeys: The costs and benefits of a forest phenomenon. *Behaviour*, *140*, 371–385.