Distribution Agreement

Signature:

________________________________ _______________

Jay W. Schwartz Date

Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution

By

Jay W. Schwartz
Doctor of Philosophy
Psychology

Harold Gouzoules, Ph.D.
Advisor

Lynne C. Nygaard, Ph.D.
Committee Member

Frans B. M. de Waal, Ph.D.
Committee Member

Mar M. Sanchez, Ph.D.
Committee Member

Gordon Ramsay, Ph.D.
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

Date

Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution

By

Jay W. Schwartz

M.A., The Ohio State University, 2015

Advisor: Harold Gouzoules, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2020

Abstract

Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution

By Jay W. Schwartz

Research into the vocal expression of emotion has the potential to elucidate both the proximate mechanisms by which vocalizations are produced and perceived, and the ultimate processes by which communication systems evolve. Since Charles Darwin, researchers have hypothesized that emotion is crucial for vocal production in animals, culminating in a recent flourishing of research examining how emotional states influence the acoustic structure of vocalizations. These patterns of vocal emotion expression are broadly consistent across mammalian species, including humans, suggesting they have been conserved over evolutionary time. However, many taxa and vocalization types remain to be studied; in particular, little research exists comparing the effects of emotional states on vocal acoustics between multiple vocalization types within a species. In Study 1, I recorded and analyzed juvenile female rhesus macaque (*Macaca mulatta*) 'coo' and 'scream' vocalizations, occurring in disparate contexts. I compared the acoustic structure of each to the vocalizer's emotional arousal, as inferred from stress hormones for coos, and agonistic intensity for screams. In both coos and screams, several parameters related to fundamental frequency (F0; commonly perceived by humans as pitch) correlated positively with arousal, suggesting that some of arousal's effects on the voice are consistent across vocalization types and contexts. Thus some variation in vocal F0 appears linked to arousal and could, in theory, exhibit a correlated response to selection pressures favoring a broader arousal reaction—and vice versa. In Study 2, participants with and those without job-related experience with rhesus macaques listened to a subset of these vocalizations and indicated their perceptions of the intensity of the vocalizer's emotion. While experience did affect accuracy of perceptions of macaque screams (but not coos), inexperienced and experienced participants alike showed significantly above-chance accuracy with both vocalization types. The F0 of coos and screams—indicated in Study 1 as a cue to emotional arousal—positively influenced participants' perceptions of arousal from these vocalization types, suggesting evolutionary homology in the relationship between vocal F0 and emotional arousal. Collectively this work demonstrates the potential of vocal emotion expression research to elucidate general processes in the evolution of vocalizations, and of communication more broadly.

Vocal emotion expression across contexts, vocalization types, and species: Implications for general processes of vocal evolution

By

Jay W. Schwartz

M.A., The Ohio State University, 2015

Advisor: Harold Gouzoules, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Psychology
2020

Acknowledgments

I am immensely grateful to Dr. Harold Gouzoules for his boundless generosity with feedback and advice, and for always challenging me to make my work the best it could be. This dissertation and my growth over the years would not be possible without his mentorship, nor without the collegiality of Jonathan W. M. Engelberg; our lab has become like a family. On that note, I thank my actual family—Anna, I would never have come this far without your endless support (and insatiable appetite for discussing evolutionary theory and quantitative methods).

I am grateful to my dissertation committee members, Drs. Lynne C. Nygaard, Frans B. M. de Waal, Mar M. Sanchez, and Gordon Ramsay for lending their time and expertise. Special thanks to Dr. Sanchez along with Dr. Mark E. Wilson for their willingness to collaborate and their generosity with resources and data. Thanks also to Manuel Bautista, Patrice Rando, Kelly Bailey, and Natalie Brutto, and to the rest of the research, animal husbandry, and veterinary staff at the Yerkes National Primate Research Center who made this work possible. Finally, I thank my former undergraduate assistants, Jalani Paul and Emma Satty, for their interest in and effort contributed to this research—I hope you learned a skill or two that will continue to be useful no matter which careers you choose.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE – 1343012, and by the National Institutes of Health under award numbers 1R01HD077623 & P51OD011132.

Table of Contents

Figures and Tables

CHAPTER I: GENERAL INTRODUCTION

The study of animal vocal communication holds the potential for far-reaching discoveries, ranging from animal cognition, to general evolutionary processes and principles, to the evolutionary history of human language. From Darwin (1872) onward, there has been a general consensus that emotion plays an important role in vocal production in animals (e.g., Briefer, 2012; Filippi, 2016; Fischer, 2017; Fischer & Price, 2016; Owren, Amoss, & Rendall, 2011; Schamberg, Wittig, & Crockford, 2018; Schwartz, Engelberg, & Gouzoules, in press; Wheeler & Fischer, 2012). Despite this, the specific nature of the relationships between emotion and vocal production has received relatively little direct empirical attention within the field of primate vocal communication. Recently however, prominent authors in this field have called for greater attention to the mechanisms of call production, and how primate vocal communication systems evolve (Fischer, 2017; Fischer & Price, 2016; Owren et al., 2011; Seyfarth & Cheney, 2017; Wheeler & Fischer, 2012, 2015). In this dissertation, I will first present the argument (this chapter) and then demonstrate with empirical studies (Chapters II-III) that an understanding of the role of emotion in vocal production is important for both these goals. I begin by introducing a fundamental concept in biology that will serve to structure this chapter.

1. The Proximate-Ultimate Distinction

In 1961, the evolutionary biologist Ernst Mayr established a distinction between two levels of causal explanation for biological traits (Mayr, 1961; see also: Alcock & Sherman, 1994; Laland, Sterelny, Odling-Smee, Hoppitt, & Uller, 2011; Sherman, 1988). Two years later, founding ethologist Niko Tinbergen (1963) established a similar framework for the study of animal behavior known as his Four Whys (see also Bateson & Laland, 2013); the proximate and ultimate categories can each be viewed as encompassing two of these four questions (Dewsbury, 1999; Nesse, 2013). Hence my focus on the proximate-ultimate distinction here is intended as a celebration of Tinbergen's Four Whys as well.

Proximate-level explanations concern the mechanisms by which a trait manifests within an individual organism. Proximate explanations for behaviors include ontogenetic factors like genetic predispositions, learning, environmental influences, and organizing effects of hormones (Tinbergen's question of "ontogeny"; Tinbergen, 1963), as well as more immediate neural, physiological, and psychological factors such as emotions, motivations, cognition, activating effects of hormones, and neural activity (Tinbergen's question of "causation"). In contrast, *ultimate*-level explanations concern the evolution of traits. The word "ultimate" should be interpreted as meaning "distal," not in the sense of being preeminent or an end-all. Ultimate explanations for behaviors include impacts on fitness (Tinbergen's question of adaptive "function"), as well as the evolutionary forces that have shaped it and the stages through which its evolution has proceeded (Tinbergen's question of "evolution"). A related term is *adaptive*, meaning that a trait (or *adaptation*) functions to increase an individual's reproductive fitness, and evolved through direct natural selection due to the same positive fitness impacts (Williams, 1966).

The critical insight that comes from the proximate-ultimate distinction is that these two levels of explanation are not in competition with one another. Rather, biological phenomena can and should be explained at both levels. A major purpose of the distinction was, and remains, to avoid fruitless debates resulting from conflation of the two levels (Alcock & Sherman, 1994; Dewsbury, 1999; Laland, Odling-Smee, Hoppitt, & Uller, 2013; Laland et al., 2011; Mayr, 1961; Scott-Phillips, Dickins, & West, 2011; Sherman, 1988). Imagine a hypothetical argument between two researchers about why a male howler monkey roars when confronted by a rival male. One asserts howling occurs because it is adaptive—the ancestral howlers that produced such sounds deterred competing males resulting in greater resource acquisition and this behavioral response has been inherited. The other insists that the vocal response occurs because the howler is in a state of high arousal. The insight brought by the proximate-ultimate distinction is that these two researchers' positions can be correct simultaneously: natural selection may have favored a system in which an aroused response to the sight of a competitor motivates male howlers to roar, because that system on average results in the acquisition or retention of resources.

While the proximate-ultimate distinction is critical for understanding causal hypotheses and how they relate to one another, it is more nuanced than a simple dichotomy. Laland et al. (2011, 2013) point out that many explanations do not fit cleanly into one or the other category, and caution against treating the proximate and ultimate levels as totally separate. At the most basic level of evolutionary theory, hypothesizing that a trait is adaptive requires the proximate-level assumption that variation in the trait is heritable (due to genetic variation) to some degree. In some cases the proximate-ultimate distinction has been argued to have had a segregating effect on research, whereby scientists often focus exclusively on either the proximate or the ultimate level while (at least operationally) viewing the other as irrelevant (Laland et al., 2013). This tendency can probably be largely attributed to logistical issues rather than a belief on the part of researchers that one or the other level is truly unimportant. Nonetheless, there might be important insights that only come from synthesizing findings and ideas across the two levels.

This was apparent to Tinbergen, who, when establishing his Four Whys, argued that a full understanding of animal behavior could only come from integrating knowledge from the proximate and ultimate levels (Bateson & Laland, 2013; Nesse, 2013; Tinbergen, 1963). At the most basic level of evolutionary theory, the hypothesis that a trait is adaptive entails the proximate-level assumption that some variation in the trait stems from genetic sources. As Dobzhanski (1973) famously wrote, “nothing in biology makes sense except in the light of evolution”; simultaneously, little in evolution makes complete sense except in consideration of proximate mechanisms. Accordingly, the rest of this chapter will review first the proximate (anatomical, physiological, and emotional) mechanisms of vocal production, and then how selection and other evolutionary processes relating to communication are thought to shape the acoustic properties of vocalizations, ending with a discussion of the insights that may come from integrating these two fields.

2. Proximate (Mechanistic) Explanations for the Acoustic Properties of Vocalizations

2.1. Basics of Bioacoustics and Vocal Production

Sound comprises vibrations through a solid, liquid, or gaseous medium, e.g., the air constituting our atmosphere. We, and most tetrapod animals (cetaceans being a notable exception), sense these vibrations as they are transmitted from the air to a tympanic membrane and subsequently transduced into neural activity within the inner ear. Sound is generated when a vibrating object, termed a *source*, transmits its vibrations to the surrounding medium. In the case of mammalian vocalization, the vocal folds (sometimes called the vocal cords), twin infoldings of mucous membrane within the larynx, typically serve as the source of vibrations (Taylor & Reby, 2010). As the lungs expel air through the trachea, muscles within the larynx may tense and adduct the vocal folds, drawing them together, at which point the airflow causes them to oscillate. This oscillation modulates the expelled air, turning it into sound.

All sounds are characterized by *amplitude* (the degree of change in air pressure over the course of an oscillation, generally perceived as volume), and *frequencies* (oscillations per unit time, represented in Hz or kHz). Pure tones (such as that produced by a tuning fork) consist of a series of frequencies superimposed on one another, including the frequency at which the source is vibrating—called the *fundamental frequency* (*F0*)—and *harmonics*, which are integer multiples of the F0. Such a sound would be described as exhibiting some amount of *energy* at the F0 and each of the harmonic frequencies; for vocalizations, the relative amount of energy at each of these frequencies is determined by the tautness and thickness of the vocal folds, which are determined in turn by tension in various muscles in the larynx (Taylor & Reby, 2010; Titze, 1994).

Because very few natural sounds maintain a constant and pure F0, sounds can be characterized in terms of changes in the energy distribution across different frequencies, or changes in *spectral structure,* over time. This is commonly visualized using spectrograms, which use Fourier analysis to depict the frequency spectrum on the vertical axis, time on the horizontal axis, and energy through contrasting hue or brightness (Figure 1.1). Though spectrograms are derived relative to waveforms, which depict amplitude over time (i.e., the sound in its most basic, physical form), spectrograms are interpreted more intuitively because they map onto the way the mammalian auditory system processes sound: different sound frequencies are segregated within the inner ear and subsequently represented by different neurons, and areas within the primary auditory cortex (tonotopic organization) (Reale & Imig, 1980; Romani, Williamson, & Kaufman, 1982). Consequently, we perceive sounds as changes in energy at different frequencies over time. Humans generally (though not always; McPherson & McDermott, 2018) perceive the F0 of a sound as its pitch, and the relative energy of each of its harmonics as partly determining its timbre. For example, the sounds emitted by a metal and a nylon guitar string playing the same note (same F0) are perceived as having equal pitch, but different timbre, because the structural properties of the strings emphasize different harmonics. Likewise, different profiles of tension in the various laryngeal muscles, and consequent differences in emphasized harmonics of vocal sounds, can lead to perceptual qualities on a spectrum from breathiness to creakiness. Differences in energy at different harmonics are visible in Figure 1.1: other than the F0, the second harmonic shows the greatest emphasis while the third and fourth harmonics are almost invisible.

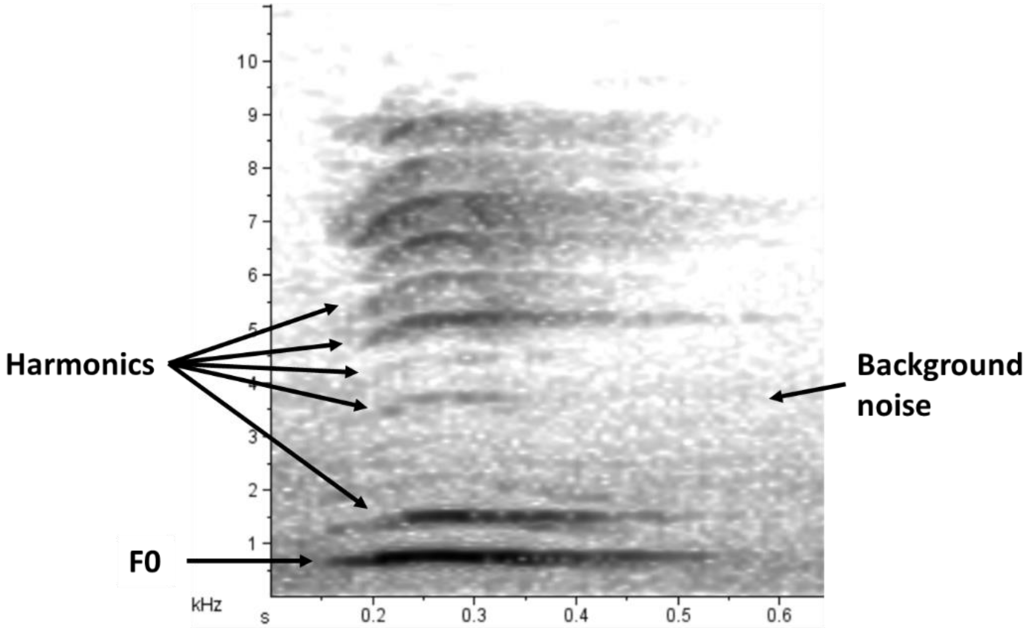

Figure 1.1. A spectrogram depicting a monkey vocalization.

Additional energy at different frequencies in vocal sounds stems from the length and shape of the vocal tract. According to the predominant model of mammalian vocal production, source-filter theory, the spectral structure of vocalizations is the result of sounds first produced by the larynx (the *source*) being *filtered* by the *vocal tract*, which includes the pharynx, the mouth, and the nasal cavity (Taylor & Reby, 2010). The shape of the vocal tract attenuates some source frequencies and amplifies or introduces others. When harmonics from the source match the resonant frequencies of various areas of the vocal tract, these areas resonate, amplifying (increasing the energy present at) those frequencies; these are referred to as *formant frequencies*. Humans control these frequencies by positioning the tongue and lips into particular configurations, resulting in different perceived vowel sounds. Other filter effects are perceived as additional aspects of timbre; for example, the degree of resonance within the nasal cavity determines the “nasaly” perceptual quality of a voice. For some vocal sounds (e.g., whispered speech, unvoiced consonants, whistles, “raspberry,” “kiss-squeak”), the vocal tract rather than the larynx serves as the source, leading in some cases to debate over whether such sounds can be considered vocalizations at all (Owren et al., 2011, pp. 539-540).

Several acoustic properties in addition to F0, harmonics, and formant frequencies are also of interest to mammal bioacousticians. *Duration* (the time elapsed between the onset and offset of a vocalization) is almost ubiquitously included in analyses of vocalizations. Amplitude is another obvious choice, as it reflects the size of the vocal fold oscillations and is determined largely by the rate of airflow, with faster flow resulting in greater amplitude. Measuring amplitude under natural conditions is challenging, however, as it depends in part on the distance from the vocalizer to the recording device, the direction the vocalizer is facing, etc. Consequently it is rarely analyzed in bioacoustic research except under particular controlled situations (Fischer, Noser, & Hammerschmidt, 2013). The spectral structure of a timepoint within a vocalization, or of a vocalization overall, can be characterized as a statistical distribution of energy at different frequencies (referred to as a spectral slice or power spectrum), and described accordingly. This *distribution of the frequency amplitudes* (DFA) is often described using variables that represent the frequency values constituting the upper limits of the first, second, and third quartiles of energy (*DFA25*, *DFA50*, and *DFA75*, respectively). Another commonly measured spectral parameter is *peak frequency* (or *dominant frequency*), the frequency with the greatest energy at a given timepoint, or on average over the course of the call (note that terminology is inconsistent across the literature, with some studies instead using “peak frequency” to refer to the maximum value of the F0; e.g., H. Gouzoules & Gouzoules, 2000). DFA and peak frequency measurements are especially useful for characterizing vocalizations where the F0 is obscured (Fischer et al., 2013). This tends to be the case when calls are particularly *noisy*, reflecting turbulence introduced either by chaotic oscillation of the vocal folds due to factors like high laryngeal muscle tension and/or rapid expulsion of air (Fitch, Neubauer, & Herzel, 2002), or by constriction of some part of the vocal tract. Noisiness itself can be measured by variables such as the *harmonics-to-noise ratio.* Finally, variables like call rate, inter-call interval, call bout duration, and call number are often included, but it should be noted that these better describe *vocal usage*, i.e., calling behavior per se, as opposed to the acoustic structure of vocalizations (for more on this distinction, see Seyfarth & Cheney, 2010).

2.2. Variation Between and Within Vocalization Types

Some acoustic properties of vocalizations are largely consistent within an individual and can reveal information about that individual’s morphology and identity. For example, the F0 of a vocalization can indicate an animal’s body size (e.g., H. Gouzoules & Gouzoules, 1990), as the minimum producible F0 is determined in part by the length of the vocal folds and hence the size of the larynx (Titze, 1994), for the same reason that a longer guitar string produces a lower pitch when plucked. Some filter features reflect the immutable shape of an animal’s vocal tract and can thus carry cues to individual identity, partly underlying many primates’ ability to recognize others from vocalizations alone (Rendall, Owren, & Rodman, 1998). Many acoustic properties of vocalizations, however, exhibit substantial intra-individual variation. This acoustic variation can be categorized as falling between or within vocalization types.

The term *vocalization type* (synonymous with *call type*) refers to a distinct, species-typical category of vocalizations with a largely innate basic acoustic structure, while a species' *vocal repertoire* comprises all of its call types and any within-type variants that might function in communication (e.g., Anikin, Bååth, & Persson, 2018; Dunn & Smaers, 2018; Ficken, Ficken, & Witkin, 1978; Fischer, Wadewitz, & Hammerschmidt, 2016; Hammerschmidt & Fischer, 1998; Rand & Ryan, 1981; Wadewitz et al., 2015). For example, the human vocal repertoire consists of several nonlinguistic vocalization types including screams, moans, laughter, crying, and perhaps others (Anikin et al., 2018). Most mammalian vocalization types (with some exceptions) do not represent totally discrete acoustic categories; they often exhibit within-type variation and grade into one another to some degree, with some vocalizations exhibiting intermediate acoustic structures (reviewed in Fischer et al., 2016). Following the human example, most would agree that laugh vocalizations can sound very different from one another, and can sometimes grade and/or transition into crying ("I don't know if you're laughing or crying!"). Based on this acoustic variation and gradation, contemporary literature around primate vocal communication conceptualizes vocalization types as clusters of vocalizations within a multidimensional acoustic space, separated by fuzzy rather than hard boundaries (Fischer et al., 2016; Tabitha Price et al., 2015; Wadewitz et al., 2015).

Although many species' vocal repertoires are not made up of acoustically discrete and perfectly stereotyped vocalization types, vocalization types are nonetheless biologically real. From a production standpoint, vocalizations within a type are hypothesized to share a pattern of neural activity that distinguishes them from other vocalization types (Jürgens, 2009; Owren et al., 2011). Likewise, receivers of many primate species appear to classify vocalizations into conceptually distinct vocalization type categories with different meanings (reviewed in Seyfarth & Cheney, 2003; Seyfarth et al., 2010). In a classic example, Diana monkeys (*Cercopithecus diana*), once habituated to an aerial predator alarm call, transfer habituation to an eagle's cry but not to a terrestrial predator alarm call, even though the two alarm calls are more acoustically similar than the aerial alarm is to the eagle's cry, suggesting distinct conceptual meanings for the two alarm calls (Zuberbühler, Cheney, & Seyfarth, 1999).

In sum, for most vertebrate species, acoustic variation among an individual's vocalizations fits into a number of vocalization types that are acoustically graded but nonetheless represent, in some cases, natural categories for both senders and receivers. Much research in animal communication over the last 40 years has focused on the usage of different vocalization types and the responses of receivers. This focus is understandable, given the biological significance of vocalization types. However, until recently, there has been a dearth of research examining the significance of acoustic variation within vocalization types. Recent directions have begun to recognize and explore the role of emotion in within-type acoustic variation. This literature is discussed below.

2.3. The Role of Emotion

2.3.1. Defining Emotion

A consideration of the relationships between emotion and vocal production requires first defining "emotion." Many definitions exist in the animal literature (Paul & Mendl, 2018). Here, I use the term to refer to a temporary, multicomponent internal state that motivates or potentiates behavior. “Multicomponent” here refers to the many biological and mental processes that generally (but imperfectly) co-occur in coordination as part of the occurrence of an emotion (see Scarantino, 2012), the most pertinent of which are reviewed below. Note that the motivation-centered definition adopted here, common in the animal literature (e.g., Anderson & Adolphs, 2014; Adolphs & Andler, 2018; de Waal, 2011; Mendl, Burman, & Paul, 2010), differs from higher-order definitions centered on consciously felt experiences (e.g., LeDoux & Brown, 2017) (for commentary, see Adolphs & Andler, 2018; Scarantino, 2018). The relationships between emotions as defined here and conscious experiences are discussed further below.

Emotional states are characterized by where they lie on two dimensions: arousal and valence (Mendl et al., 2010; Russell, 2003). Arousal represents the spectrum from a sleepy, calm, or inactive state to a highly alert or excited state. High arousal (in both positive and negative contexts) involves the endocrine stress response and activation of the sympathetic branch of the autonomic nervous system, resulting in tensing of muscles throughout the body, increased heart rate, increased respiration, and mobilization of energy stores (Charmandari, Tsigos, & Chrousos, 2005; Ralph & Tilbrook, 2016; Romero, 2004) (for more detail, see Section 2.3.2. Measuring Emotion in Animals). Emotional valence is theoretically orthogonal to arousal and ranges from positive to negative. For example, a victim of an attack would typically exhibit a high-arousal, negative-valence emotion, while a low-arousal, positive-valence emotion might be exhibited during a rest just after feeding. While the neural and physiological representation of emotional valence is complex and not completely understood (Lindquist, Satpute, Wager, Weber, & Barrett, 2016), it appears to be at least partly represented in the medial orbitofrontal cortex of the brain (Kragel & LaBar, 2016). The amygdala has historically been viewed as important for the representation of negative emotional valence, but recent evidence instead links amygdalar activity to high-arousal emotions irrespective of valence (Lin et al., 2020).

The above framework of arousal and valence comes from the dimensional or Core Affect view of emotion (e.g., Barrett, 2006; Barrett, Henzi, & Rendall, 2007; Mendl et al., 2010; Russell, 2003), which contrasts with the basic or discrete emotions view (e.g., Ekman, 1992; Izard, 2007; Panksepp, 2007, 2011; Sauter, Eisner, Ekman, & Scott, 2010; Scarantino & Griffiths, 2011; see also Hamann, 2012; Kragel & LaBar, 2016). The dimensional approach has been argued to reflect how emotional states are represented in the brain: separate neural systems underlie arousal and valence, and emotion arises from the interaction of these two systems along with cognitive interpretations and constructions. In contrast, the basic emotions view holds that emotions are discrete "natural kinds" (Izard, 2007)—such as joy, fear, anger, sadness—each underlain by a distinct neural and physiological pattern. For example, a neural circuit involving the amygdala and periaqueductal gray is activated by afferent signals from sensory cortices upon perception of threatening stimuli, and plays a role in defensive behavior, leading to a view of this circuit as central to the emotion of fear (depending on definitions; for contrasting views see LeDoux & Brown, 2017; Panksepp, 2011). The dimensional and discrete views have each received empirical support: brain imaging research suggests that the dimension of valence is represented in the brain, but, consistent with the discrete view, emotions that are experienced categorically are associated with distinct distributed neural systems (Kragel & LaBar, 2016). However, the neural and physiological processes that comprise purportedly basic emotions do not perfectly co-occur, leading proponents of the dimensional view to question whether emotions can be considered "basic" or to be "natural kinds" (Barrett, 2006; Russell, 2003; Scarantino & Griffiths, 2011).

Put differently, if we imagine emotion categories as clusters of points in a multidimensional emotion space, with each dimension representing a suite of neural and physiological processes and each point representing an instance of an emotional state, then the emotion clusters are not perfectly discrete; the dimensional-discrete debate is largely about what this observation means for our understanding of emotions (Scarantino & Griffiths, 2011). In this way, the debate between the dimensional and discrete views of emotion shares issues in common with the ongoing discussion about vocal repertoires and vocalization types reviewed above (Section 2.2. Variation Between and Within Vocalization Types): emotional states can be thought of as analogous to vocalization types, which exhibit intra-type acoustic variation and gradation but might nonetheless constitute natural categories.

Each side of the dimensional-discrete debate has been endorsed in various articles within the vocal emotion expression literature (dimensional view: e.g., Briefer, 2012; discrete view: e.g., Scherer, 2003). This dissertation features much discussion of emotional arousal (and, to a lesser extent, valence) and its role in vocal production and other behavior. This amounts not to a tacit endorsement of the dimensional view of emotion over the basic emotions view, but rather, an assumption as to the existence of the arousal dimension without any assertion as to whether or not discrete emotions lie along that dimension. The existence of the arousal dimension is not in dispute, so the discussion within this dissertation should be acceptable to researchers on all sides of the dimensional-discrete debate.

Another debate in the emotion literature concerns whether the subjective experience of emotion (hereafter “feelings”; also referred to as “qualia” or “affective consciousness”) exists in nonhuman animals, with some researchers suggesting that feelings emerge from neural processes (and cognitive experiences) unique to humans (Craig, 2002; Dolan, 2002), others arguing that the weight of evidence favors a view of feelings as evolutionarily old and conserved across mammals (de Waal, 2011; Panksepp, 2005, 2011), and others still proposing that we veer away from the topic altogether (Adolphs & Andler, 2018; LeDoux, 2012). The range of processes and experiences that comprise emotional feelings, and the extent to which these are extricable from the subcortical neural circuits associated with emotions, are debated (LeDoux & Brown, 2017; Panksepp, 2011). Overall, the view that nonhuman mammals experience feelings in some way seems most parsimonious. That said, the multicomponent conception of emotion adopted here allows for the possibility of a felt component, but does not assume or rely on the existence thereof in nonhuman species (see Scarantino, 2018). The debate over nonhuman feelings persists precisely because our understanding of nonhuman consciousness (emotional or otherwise) is not sufficiently advanced to generate specific predictions as to how the influence of emotion on behavior (such as vocal production) should differ depending on the presence or absence of feelings per se; hence it is possible, and perhaps preferable, to study emotion while setting the matter of feelings aside (Adolphs & Andler, 2018). This does not amount to denying the existence of feelings or the possibility that feelings might even be integral to mammalian emotions (Panksepp, 2005, 2011), but rather choosing for the time being to address those research questions that can be addressed without *necessarily* invoking feelings.

The behaviors evoked by, or expressing, an emotional state are not considered to be components of the emotional state itself (Anderson & Adolphs, 2014; de Waal, 2011; Mendl et al., 2010; Panksepp, 2011; Paul & Mendl, 2018). Instead, emotions are construed as “intervening variables” between perceived cues to context and behavioral responses (de Waal, 2011). The relationships between context, emotion, and behavior are highly biologically significant; many authors consider the general evolutionary function of emotions to be to motivate behaviors that are on average adaptive in a particular emotionally salient context (Adolphs & Andler, 2018; de Waal, 2011; LeDoux, 2012; Mendl et al., 2010; Nesse, 1990; Panksepp, 2011; Plutchik, 2003), an idea that goes back to Darwin (1872). In many cases, a single emotional state might produce multiple distinct behavioral outputs simultaneously, a characteristic termed emotional *pleiotropy* (after the phenomenon where a single gene contributes to multiple phenotypic traits) (Anderson & Adolphs, 2014). For example, an animal in a state of distress might freeze, flee, vocalize, look around, grimace, or self-scratch, in some combination. On the input side, multiple different contexts might evoke a similar emotional response, a characteristic termed *stimulus generalizability* (Anderson & Adolphs, 2014). A monkey might, for example, enter a physiologically similar state of emotional arousal in response to the presence of a predator as in response to a threat from a high-ranking conspecific. Thus an array of contexts are described as "fanning in" to a particular emotional state, from which multiple behaviors "fan out"; some have argued that this "fan-in, fan-out" architecture is essential to the function of emotions (Adolphs & Andler, 2018; Anderson & Adolphs, 2014).

2.3.2. Measuring Emotion in Animals

Emotions carry importance both intellectually for the study of behavior (reviewed in the below sections) and in practice for maximizing animal welfare (M. S. Dawkins, 2000; Hemsworth, Mellor, Cronin, & Tilbrook, 2015). Measuring emotions in other species is challenging; animals generally cannot tell us how they feel, if they do indeed feel their emotions (although even humans' self-reports of emotions or feelings are flawed indicators of their internal states; Barrett, 2006). However, researchers can assess animals' emotions through any combination of three non-mutually exclusive categories of methods: context-based inference, observation of behavior, and physiological and/or neural measurements and manipulations.

Context-based inference entails observing the stimuli experienced by an animal and making assumptions about the cognitive mechanisms involved in processing those stimuli, specifically the emotions involved. A wide array of contexts have been used in this way. For example, differing arousal levels have been attributed to common ravens (*Corvus corvax*) in agonistic interactions with high-ranking (high arousal) vs. low-ranking (low arousal) opponents (Szipl, Ringler, Spreafico, & Bugnyar, 2017), wild spider monkeys (*Ateles geoffroyi*) separated from their troop to differing degrees (Ordóñez-Gómez, Santillan-Doherty, & Hammerschmidt, 2019), hourglass treefrogs (*Dendropsophus ebraccatus*) exhibiting escalating levels of sexual competition (Filippi et al., 2017), and silver foxes (*Vulpes vulpes*) as they were gradually approached by humans (Gogoleva, Volodina, Volodin, Kharlamova, & Trut, 2010). Differing emotional valence has been attributed to aggressive (negative) vs. playful or affiliative (positive) interactions by dogs (*Canis familiaris*) or pigs (*Sus scrofa*) (Faragó, Pongrácz, Range, Virányi, & Miklósi, 2010; Maigrot, Hillmann, & Briefer, 2018; Taylor, Reby, & McComb, 2009), and in pigs with a partner (positive) vs. alone (negative) (Briefer, Vizier, Gygax, & Hillmann, 2019). One recent study used context to infer both arousal and valence in horses (*Equus caballus*), basing valence on whether groupmates were arriving (positive) or departing (negative), and basing arousal partly on the number of groupmates arriving or departing (Briefer, Maigrot, et al., 2015).

Context-based inference is convenient and noninvasive, but it relies on assumptions about how an animal will perceive and react to a particular situation. Often these assumptions are based in anthropomorphism, which is not invariably problematic but does comprise a bias of which researchers should be cognizant (Arbilly & Lotem, 2017; de Waal, 1999). Basing inferences of emotion on context alone is also risky because individual differences in emotional reactivity to different contexts are likely. Different instances within a certain type of context can vary in emotional salience within an individual as well, depending on details about the incident, e.g., a rhesus monkey's reaction to a conspecific threat depends in part on the degree of rank difference between an individual and an opponent (H. Gouzoules, Gouzoules, & Tomaszycki, 1998). With all of that said, some assumptions are generally safe, e.g., an attack by a predator will elicit a relatively high-arousal, negative emotion; sudden access to food will elicit a relatively positive emotion).

In addition to context, researchers base assessments of emotion on animals' behavior. In some of the above examples, the "context" observed is defined by an animal's behavior (e.g., play, agonism), thus one might ask what distinguishes these two methodological approaches. Returning to the idea of emotion as an "intervening variable" mediating the relationship between contextual inputs and behavioral outputs (de Waal, 2011), context-based inference measures inputs to infer the emotional state, whereas observation of behavior does so by instead measuring outputs. Thus, placing various methods into these categories is a matter of postulating about whether the observed behavior or context is primarily influencing the animal's emotional state, or an expression thereof. This distinction is somewhat fuzzy but conceptually important. Examples of behavioral expressions used as a measure of emotional arousal include call rate in chacma baboons (*Papio ursinus*) (Meise, Keller, Cowlishaw, & Fischer, 2011; Rendall, 2003), and movement intensity in pigs and giant pandas (*Ailuropoda melanoleuca*) (Filippi et al., 2017; Linhart, Ratcliffe, Reby, & Špinka, 2015; Stoeger, Baotic, Li, & Charlton, 2012). Behaviors used as a measure of emotional valence include ear position in horses (Stomp et al., 2018) and latency to investigate a novel object in pigs (with shorter latencies reflecting a more positive or "optimistic" cognitive bias; Friel, Kunc, Griffin, Asher, & Collins, 2019).

Like context-based inferences, using behavior to measure emotion entails certain assumptions, namely that a given behavior is the product of a particular emotional state as opposed (or in addition) to other cognitive processes. Some behaviors are thought to be involuntary products of emotions, such as approach/avoidance behaviors (Mendl et al., 2010). In many cases, however, there may be no perfect one-to-one correspondence between an emotional state and an emotional behavior, including human facial configurations or "expressions" (Barrett, Adolphs, Marsella, Martinez, & Pollak, 2019), which were long thought of as reflexive, and therefore highly valid, readouts of emotions (Darwin, 1872; Ekman, 1992). This fact should not be viewed as incompatible with the idea that these behaviors are emotional, i.e., motivated at least in part by emotion: emotions can, in addition to or rather than directly evoking behavior, prepare or dispose an individual to behave a certain way, at which point other internal or external factors might amplify or suppress that urge. Many behaviors likely reflect a combination of emotion and other cognitive processes such as mentally represented goals and executive functions; vocal usage appears to generally fall into this category (Schwartz et al., in press**)**. As measures of emotion, such behaviors are indirect and limited but still potentially useful.

Context-based inference and observation of behavior partially measure the inputs and outputs of emotions, respectively, and thus measure emotion indirectly. Emotions are transient states that involve physiological and neural activity; measuring or stimulating this activity provides more direct access to animals' emotional states. Distinct central neural systems are hypothesized to underlie different emotional states (Kragel & LaBar, 2016; Panksepp, 2011), though as reviewed above, the associations between emotions and neural activity are generally imperfect and, in any case, not fully known. That said, a few emotional neural systems are well established. The dopaminergic mesolimbic “wanting” and “liking” systems are involved in emotions surrounding food motivation and reward (Berridge, 1996). The amygdala, hypothalamus, and dorsal periaqueductal gray are involved in emotions surrounding unconditioned threat responses (e.g., LeDoux, 2012; Panksepp, 2011). Measuring activity in these neural circuits, or causing it through brain stimulation, can be highly informative about the relationships between emotion and behavior. For example, brain stimulation allowed for a direct investigation of the acoustic correlates of aversion in squirrel monkey (*Saimiri sciureus*) vocalizations (Fichtel, Hammerschmidt, & Jürgens, 2001). Although informative, neural measurements and manipulations are highly invasive and resource-intensive. They are usually salient and distressing to subjects, and might influence behavior, undermining the goal of understanding natural emotional experiences or responses. Furthermore, if a goal of monitoring an animal’s emotions is to improve that individual’s psychological welfare, administering distressing invasive procedures might be counterproductive (although scientific research utilizing these techniques can potentially be broadly applied to improve the lives of many animals).

Measuring the peripheral physiological components of emotions allows researchers to directly access components of emotional states in a manner that, although invasive, is generally less so than neural measurements. In particular, heightened arousal (a component of various emotions) comprises several measurable physiological changes in the body stemming from the activation of various neural and neuroendocrine systems. One of these is the sympathetic branch of the autonomic nervous system, general activation of which increases respiration, cardiac effort (e.g., heart rate), and tension in various muscles throughout the body, functioning in part to potentiate adaptive behavioral responses to salient stimuli, including social behavior and communication (e.g., Porges, 2003). In general, upon perception of arousing stimuli and other stressors (any real or perceived disruptor of homeostasis), these peripheral neural and physiological changes occur in coordination with activation of two neuroendocrine systems: the hypothalamic-pituitary-adrenal (HPA) axis and the sympathoadrenal system (reviewed in Charmandari et al., 2005; Ralph & Tilbrook, 2016). The sympathoadrenal system causes release of catecholamines including epinephrine (also known as adrenaline) and norepinephrine. Catecholamines carry distributed effects throughout the body including increased cardiac output and respiration rate, and redistributed blood flow to organs involved in survival behaviors. This usually occurs in tandem with activation of the HPA axis: the hypothalamus secretes corticotropin-releasing hormone, stimulating secretion of adrenocorticotropin hormone by the anterior pituitary. Adrenocorticotropin hormone travels through the bloodstream to the adrenal glands, where it stimulates synthesis and release of glucocorticoids, a class of steroid hormones of which the principle one is corticosterone in rodents and birds, and cortisol in all mammals other than rodents. Glucocorticoids act to mobilize energy stores, functioning to restore homeostasis as an animal responds behaviorally to a stressor. Activation of these systems, and resulting physiological changes, in response to acute stressors presumably occur as components of a broader emotional response, and measuring those physiological changes thus allows researchers a degree of access into the emotional state of an animal.

Researchers studying the role of emotion in vocal behavior have measured aspects of physiological arousal including heart rate in horses (Briefer, Maigrot, et al., 2015), goats (*Capra hircus*) (Baciadonna, Briefer, Favaro, & McElligott, 2019; Briefer, Tettamanti, & McElligott, 2015) and common marmosets (*Callithrix jacchus*) (Liao, Zhang, Cai, & Ghazanfar, 2018), and cortisol concentrations in rhesus macaques (*Macaca mulatta*) (Bayart, Hayashi, Faull, Barchas,

& Levine, 1990), yellow-bellied marmots (*Marmota flaviventris*) (Blumstein & Chi, 2012), pigs (Schrader & Todt, 1998), and sheep (*Ovis aries*) (Sèbe et al., 2012). These techniques allow subjects to behave freely during a period of observation: heart rate can be measured continuously using a noninvasive monitor attached to an animal's chest, while cortisol remains in circulation for minutes to hours after exposure to a stressor and can be assessed by sampling an animal's blood, saliva, or feces after the observation period ends. There is a several-minute time lag between exposure to a stressor and measurable increases in plasma glucocorticoids (Schlotz et al., 2008), meaning that such a measurement will not be affected by the procedure used to obtain it. In these ways, physiological measurements generally allow for greater ecological validity than neural techniques.

It is important to reiterate that emotions are multicomponent states consisting of imperfectly co-occurring processes (Scarantino, 2012; Scarantino & Griffiths, 2011). Thus, no one neural or physiological process can be equated to an emotion. As for physiological techniques, while these can provide solid information about arousal, they are poor indicators of emotional valence, largely because high-arousal, negative states and high-arousal, positive states entail similar physiological processes. Rewarding experiences, like aversive ones, evoke synthesis and release of glucocorticoids and catecholamines (reviewed in Ralph & Tilbrook, 2016). Additional problems with using glucocorticoids as a measure of emotion stem from other, non-stress-related fluctuations. HPA activation occurs not only in response to emotionally salient stimuli, but also in response to caloric intake, temperature changes, immune challenge, and exercise, and additionally exhibits circadian and ultradian rhythms (Charmandari et al., 2005; Devenport, Knehans, Sundstrom, & Thomas, 1989; Fulkerson & Tang, 1979; Ralph & Tilbrook, 2016; Turner, Rivalland, Clarke, & Tilbrook, 2010). Therefore a glucocorticoid concentration measurement alone is virtually meaningless with respect to emotions (Ralph & Tilbrook, 2016). A common solution is to obtain a baseline plasma glucocorticoid measurement, keeping time of day consistent to control for circadian rhythm, then expose the animal to a specific and controlled stressor, and finally obtain a peak measurement; under these conditions, the absolute increase in glucocorticoid concentrations from baseline to peak reflects the degree of HPA activation in response to the stressor (reviewed in Romero, 2004). Again, HPA activation does not equal the emotion of stress—rather the former is a component of the latter—but under such controlled conditions, glucocorticoid increases can be reasonably interpreted as an indicator of an animal's emotional arousal reaction.

The above discussion highlights a broader principle, namely that the best assessments of animal emotional states come from integration of results from multiple measures (de Vere & Kuczaj, 2016), and interpretation of these measures should always be framed by an animal's behavior and the context and stimuli to which an animal is reacting (Romero, 2004). Use of these three methods (context-based inference, observation of behavior, and physiological and/or neural measurements) in conjunction will undoubtedly yield the most accurate interpretations regarding animals' emotional states. Recent work by Briefer and colleagues is exemplary in this regard (Baciadonna et al., 2019; Briefer, Maigrot, et al., 2015; Briefer, Tettamanti, et al., 2015; Briefer, Vizier, et al., 2019).

2.3.3. Vocal Emotion Expression: Effects of Emotion on Vocal Production and Acoustics

I will hereafter use the term *vocal emotion expression* to collectively refer to the ways in which acoustic characteristics of vocalizations change depending on the emotional state of the vocalizer; this term is intended as an analogue to "emotional prosody" of speech (e.g., Filippi, 2016), but without the implication of a linguistic aspect to the vocalization. The dimensional view of emotion has predominated the literature around vocal emotion expression in nonhuman animals, with authors typically purporting to have identified acoustic correlates of emotional arousal and/or valence in a species (e.g., Briefer, 2012; Briefer, Maigrot, et al., 2015; Briefer, Vizier, et al., 2019; Friel et al., 2019; Linhart et al., 2015; Maigrot et al., 2018; Szipl et al., 2017). Many additional relevant studies simply report the correlation between a particular context and an acoustic change, but the role of the animal's emotional reaction to the context can be inferred (reveiwed in Briefer, 2012). Altogether, this literature has identified general consistencies in vocal emotion expression across mammalian (and some bird) species, likely owing to similarities in the physiology of emotion and the mechanisms of vocal production (Briefer, 2012; Zimmermann, Leliveld, & Schehka, 2013). I review these trends below.

Some insights into vocal emotion expression come from comparisons among different vocalization types. For example, a recent study demonstrated that pigs emit more "low-frequency grunts" in a low-arousal (as indicated by heart rate) isolation condition and more "high-frequency grunts" in during high-arousal interactions with familiar humans (Leliveld, Düpjan, Tuchscherer, & Puppe, 2017), suggesting a positive correlation between arousal and call frequency. Such comparisons between vocalization types are undermined by a potentially confounding third variable, namely, the usage of different call types appears to involve not only emotion but also an array of other cognitive processes, including executive functioning (Schamberg et al., 2018; Schwartz et al., in press). Evidence for this includes "audience effects" wherein the presence of particular receivers influences the odds of calling (Fichtel & Manser, 2010); for example, chimpanzees have been shown to modify their recruitment screams based partly on the rank composition of the audience, exaggerating the screams when there was a listener present who outranked the aggressor (Slocombe & Zuberbühler, 2007). Chimpanzees also modify usage of alert vocalizations based on whether the receiver appears to have registered the presence of the threat (Crockford, Wittig, Mundry, & Zuberbühler, 2012; Schel, Townsend, Machanda, Zuberbühler, & Slocombe, 2013). In principle, differences in a sender's emotional state in the presence versus the absence of a particular audience, or when a receiver is ignorant versus knowledgeable, could account for audience effects (Owren et al., 2011), but as one considers greater and greater context-dependency in call usage, the idea that it is all underlain by subtle differences in sender affect becomes more burdened (Cheney & Seyfarth, 2018).

In contrast, the precise acoustic structure of an individual utterance, i.e., acoustic variation within vocalization types, appears to be determined more directly by the sender's emotional arousal or valence (Schamberg et al., 2018; Schwartz et al., in press). In support of this, a recent study with common marmosets showed that within-type acoustic structure was associated with sender heart rate (and thus arousal), while call type was associated with the presence/absence and distance of a conspecific but not heart rate (Liao et al., 2018). Many recent studies on the vocal communication of emotion have at least implicitly, and sometimes explicitly, recognized the necessity of assessing the significance of variation within vocalization types (e.g., Briefer, 2018; Briefer, Linhart, et al., 2019; Briefer, Maigrot, et al., 2015; Briefer, Vizier, et al., 2019; Faragó, Takács, Miklósi, & Pongrácz, 2017; Friel et al., 2019; Linhart et al., 2015; Maigrot et al., 2018; Schwartz & Gouzoules, 2019). Accordingly, most of the evidence reviewed below regarding patterns of vocal emotion expression comes from studies examining within-type acoustic variation.

The vocal expression of arousal has been studied in a wide variety of species. One of the best-established acoustic correlates of arousal is F0 (defined, along with other acoustic parameters, in Section 2.1. Basics of Bioacoustics and Vocal Production), with increases in arousal generally associated with increases in mean F0 (Briefer, Maigrot, et al., 2015; Briefer, Tettamanti, et al., 2015; Collins, McGreevy, Wheatley, & Harcourt, 2011; Filippi et al., 2017; Meise et al., 2011; Rendall, 2003; Scherer, 2003; Szipl et al., 2017; Yeon et al., 2011 for a possible exception, see Ordóñez-Gómez et al., 2019), as well as F0 variability (Bayart et al., 1990; Lingle, Wyman, Kotrba, Teichroeb, & Romanow, 2012; Norcross & Newman, 1993; Rendall, 2003; Sugiura, 2007; Yamaguchi, Izumi, & Nakamura, 2010). These effects of arousal on F0 are attributed to heightened activity in sympathetic neurons innervating the diaphragm and/or laryngeal muscles, resulting in increased muscle tension there and, therefore, faster vibration of the vocal folds and greater changes in vibration frequency (Briefer, 2012; Porges, 2003; Riede, 2010; Scherer, 1986; Titze, 1994). The general relationship between arousal and mean F0 of vocalizations appears to hold true for humans (Scherer, 2003), although it might be attenuated in speakers of tonal languages where F0 is linguistically meaningful (Anolli, Wang, Mantovani, & De Toni, 2008). F0 appears to influence human listeners' estimations of the arousal level of vocalizations by humans (Juslin & Laukka, 2003; Sauter, Eisner, Calder, & Scott, 2010; Schwartz & Gouzoules, 2019) as well as other animals (Faragó et al., 2014; Filippi et al., 2017; Kelly et al., 2017). In one study, the higher F0 of infant bonobo (*Pan paniscus*) vocalizations relative to those of infant chimpanzees (*P. troglodytes*) led human listeners to erroneously perceive the former as reflecting greater distress than the latter despite the vocalizations having been recorded in similar contexts (Kelly et al., 2017).

Emotional arousal shows associations with several other acoustic properties of vocalizations in addition to F0. Increased tension and action in the diaphragm resulting from arousal can increase air pressure and thus vocal amplitude (Briefer, 2012; Scherer, 1986), as evidenced in species where arousal and vocal amplitude are positively correlated (Linhart et al., 2015; Patel & Owren, 2007; Szipl et al., 2017). Greater arousal is also associated with increases in the frequencies of formants and, more generally, upward shifts in the distribution of energy across the frequency spectrum (Briefer, Maigrot, et al., 2015; Gogoleva et al., 2010; Yeon et al., 2011), perhaps as a result of arousal constricting the pharynx, and/or attenuating salivation thereby drying out and changing the resonant properties of the vocal tract (Briefer, 2012; Scherer, 1986). Finally, arousal can lead to noisiness, through some combination of increased air pressure, pharyngeal constriction, and/or irregular ("chaotic") vocal fold vibration due to laryngeal tension (Fitch et al., 2002; Scherer, 1986). Indeed, noisiness has shown positive correlations with arousal in some studies (Liao et al., 2018; Meise et al., 2011; Puppe, Schön, Tuchscherer, & Manteuffel, 2005; Siebert, Langbein, Schön, Tuchscherer, & Puppe, 2011; Stoeger, Charlton, Kratochvil, & Fitch, 2011), though others have shown no, or conflicting, results (Blumstein & Chi, 2012; Linhart et al., 2015; Szipl et al., 2017). Experiments with human listeners have also shown mixed results with respect to perceptions of arousal from noisiness (Faragó et al., 2014; Filippi et al., 2017). These mixed findings come from studies ranging across contexts, vocalization types, and species, and might reflect variation in patterns of vocal emotion expression across these.

In addition to arousal, differences in emotional valence can also result in changes to the voice, though studies examining this were rare until recently. Vocalization duration appears to map onto valence in some species, with shorter vocalizations generally associated with more positive emotions and contexts (Briefer, 2012; Briefer, Maigrot, et al., 2015; Briefer, Vizier, et al., 2019; Fichtel et al., 2001; Friel et al., 2019; Taylor et al., 2009). Consistent with these findings, human listeners have tended to perceive longer vocalizations by dogs (*Canis familiaris*) and humans as more negative (Faragó et al., 2014, 2017), though listeners also perceive longer vocalizations as more emotionally “intense” (i.e., aroused) (Faragó et al., 2014; Schwartz & Gouzoules, 2019). In addition to duration, several studies have suggested that filter features reflect valence in addition to or rather than arousal, with negative valence (like greater arousal) associated with a greater amount of energy at higher frequencies, and higher formants (Briefer, 2012; Faragó et al., 2010; Goudbeek & Scherer, 2010; Maigrot et al., 2018; Yeon et al., 2011), though others have failed to support this relationship (Briefer, Maigrot, et al., 2015; Briefer, Tettamanti, et al., 2015; Friel et al., 2019; Taylor et al., 2009). Again, this might reflect differences in vocal emotion expression across contexts, vocalization types, and/or species. Whereas arousal carries clear physiological effects on the body including the vocal apparatus, the mechanisms by which emotional valence might influence acoustic properties of vocalizations are less clear. It has been suggested that valence-dependent changes in energy distribution and formants might come about as a result of a vocalizer extending the face (and thus the vocal tract) forward in an approach response to positive stimuli, vs. recoiling and restricting the oral and nasal orifices in an avoidance response to negative stimuli (Scherer, 1986). Hypotheses regarding the causal mechanisms underlying the correlation between emotional valence and vocalization duration are lacking, despite the findings of several studies reporting this correlation.

In sum, vocal emotion expression exemplifies the “pleiotropic” nature of emotions (Anderson & Adolphs, 2014 see also Section 2.3.1. Defining Emotion), that is, the multiplicity of behavioral outputs of a single emotional state. It is clear from the behavioral observations reviewed above, as well as what is known about the physiological mechanisms by which emotion affects vocal production, that any given emotional state can yield many different effects on vocal acoustics. Even so simple a change as increased tension/action in the diaphragm can produce changes in at least four emotionally prominent acoustic parameters: vocal amplitude, duration, F0, and noisiness. But the effects of emotional arousal are not limited to a single muscle or muscle group, rather, they are global, affecting most of the muscles involved in vocalization, as well as the resonance properties of the vocal tract, leading to numerous coordinated acoustic changes. Many emotional states consist not only of changes in arousal, but also other physiological and/or neural phenomena, which may lead to additional acoustical effects. Thus, any given within-vocalization-type acoustic change observed to occur in a given context is best understood at the proximate level as part of a broader set of acoustic, and other, changes associated with an emotional reaction to that context.

Another noteworthy observation from the literature reviewed above is that the studies showing a relationship between an acoustic parameter and an emotional dimension often represent a wide array of emotional contexts. For example, increased arousal has correlated with increases in vocal F0 in horses as their stablemates were removed or returned (Briefer, Maigrot, et al., 2015), goats during feeding (Briefer, Tettamanti, et al., 2015), chacma baboons when handling infants or grooming, and common ravens in agonistic interactions (Szipl et al., 2017), to name a few. The relationship between emotional valence and vocalization duration is consistent across pigs exhibiting valenced cognitive biases (Friel et al., 2019), squirrel monkeys experiencing brain stimulation (Fichtel et al., 2001), and dog growls in playful or agonistic interactions (Faragó et al., 2010; Taylor et al., 2009). Thus, in addition to pleiotropy, vocal emotion expression exemplifies the “stimulus-generalizability” characteristic of emotions (Anderson & Adolphs, 2014), that is, different contexts can evoke similar emotions (and thus similar changes to the acoustic structure of vocalizations). The pleiotropy and stimulus generalizability represented in vocal emotion expression carry potential evolutionary significance (explored below). However, the role of emotion in vocal production in many species, contexts, and vocalization types has yet to be studied; assessing the evolutionary and other significance of vocal emotion expression requires further research in this area.

3. Animal Communication and Ultimate (Evolutionary) Explanations for Acoustic Properties of Vocalizations

Discussing the significance of emotion for vocal evolution requires first reviewing traditional views on the evolution of animal communication and, specifically, how selection is thought to shape the acoustic properties of vocalizations. The evolution of communication is exceptionally dynamic because it is inherently coevolutionary, involving ongoing interplay between senders and receivers. In many communication systems, the same individuals comprise both senders and receivers, undergoing simultaneous evolutionary change in both the mechanisms of production and perception (an exception being some instances of sexual signaling in which the primary function of signals emitted by one sex is to attract the other sex). Bradbury & Vehrencamp (1998) outlined two broad models describing these evolutionary dynamics: the *sender-precursor model* and the *receiver-precursor model*. I briefly review these two models below; although the models apply to communication across modalities, I limit the below discussion to examples around acoustic properties of vocalizations.

3.1. The Sender-Precursor Model

The sender-precursor model begins with a *cue*, defined as a morphological or behavioral trait that is intrinsically linked with some other characteristic of the sender. For example, as reviewed above, variation in F0 and other acoustic parameters within vocalization types often correlates with, and thus comprise potential cues to, the sender’s emotional state. If it is beneficial for receivers to perceive a cue against background noise and to vary their behavior accordingly, selection can be expected to favor the evolution of increased sensitivity to the cue and/or refined behavioral reactions, i.e., increased capacity to perceive and respond appropriately to the information available from the cue. Put differently, selection may favor receivers who can access the *information* available within a cue (sensu Seyfarth et al., 2010). Critics have advocated against the term “information” for animal communication, arguing that it evokes a misleading comparison to human language (R. Dawkins & Krebs, 1978; Rendall, Owren, & Ryan, 2009). On the contrary, to say that a trait conveys information merely means that receivers have the capacity to predict something based on the trait; the term entails no assumptions about the cognitive processes at play in receivers, which can range in complexity from simple stimulus-response reflexes to (in humans) understanding of symbolic meaning in a linguistic sense (Seyfarth et al., 2010). For example, it could benefit a receiver’s fitness to be able to use emotion-related acoustic variation to predict a sender’s subsequent behavior, or to react to a stimulus that is not directly perceivable by a receiver but to which a sender’s vocal response might serve as cue.

Evolutionary change in receiver behavior creates novel selection pressures on senders that depend on how receivers’ evolved responses affect senders’ fitness. If receiver behavior carries positive fitness consequences for senders, one might expect selection to favor evolutionary change in the cue to make it a stronger elicitor of receiver behavior; this evolutionary change is referred to as *ritualization* and comprises the transition from a cue to a *signal*, a morphological or behavioral trait whose structure can be explained partly as a product of direct natural selection stemming from communicative function. For example, it is possible that many species-typical aggressive and submissive vocalization types began as cues to the sender's motivational state (i.e., reflecting that the sender is disposed to attack or to submit, respectively). The motivation-structural rules model (Morton, 1977) posits that a capacity for receivers to respond adaptively to these cues might conceivably reduce incidences of physical aggression, benefiting both senders and receivers; consequent selection favoring more effective communication might partly account for the exaggerated acoustic differences between dominant and submissive vocalization types in many taxa (e.g., canine barks vs. whimpers) (see also August & Anderson, 2006).

The screams of some catarrhine primates serve as another example of signal evolution where senders' and receivers' fitness interests overlap. Victims of aggression emit screams in response, serving to inform third-party receivers (including kin and allies) about the event; senders benefit by potentially soliciting aid from receivers, while receivers, by intervening, might enhance their inclusive fitness and/or protect their own place in the social hierarchy (de Waal & Van Hooff, 1981; H. Gouzoules, 2005; H. Gouzoules & Gouzoules, 1995; H. Gouzoules, Gouzoules, & Marler, 1985; H. Gouzoules et al., 1998; S. Gouzoules, Gouzoules, & Marler, 1984). Thus one might expect selection to favor screams exhibiting an acoustic structure that more effectively solicits intervention. This could conceivably account for the high amplitude and generally chaotic structure of many primate screams, which could make them more attention-grabbing to receivers (Owren & Rendall, 2001).

The above examples illustrate how selection can influence the acoustic parameters that characterize a given vocalization type—what might be called the vocalization type's baseline or centroid acoustic structure. It is also possible to imagine that the particular acoustic structure of a given vocalization *within a given context* (i.e., within-type acoustic variation) might be influenced by selection. For example, in many mammals, contact call variants emitted when the sender is isolated (relative to merely separated, e.g., by a glass partition) show greater F0 modulation (Lingle et al., 2012); it has been suggested that this vocal change evolved because it makes vocalizations more detectable and localizable, increasing the chances of reunion with groupmates when senders are lost (Sugiura, 2007; Yamaguchi et al., 2010). In addition, acoustic variation within vocalization types mapping onto senders' emotional states (reviewed in Section 2.3.3. Vocal Emotion Expression: Effects of Emotion on Vocal Production and Acoustics) could allow receivers to predict the senders' subsequent behavior and/or external factors of emotional salience to senders. For example, in meerkats (*Suricata suricatta*), alarm call type conveys information about predator class, whereas arousal-related acoustic variation conveys information about the urgency of the threat (Manser, 2001; Manser, Bell, & Fletcher, 2001; Manser, Seyfarth, & Cheney, 2002).

When responses that are adaptive for receivers are harmful to senders' fitness, selection is expected to favor *decoupling* of the cue from the characteristic about which it initially provided information. For example, it is generally in predators' interests to use anti-predator alarm calls as a cue to the sender's location, whereas this carries obvious negative consequences for senders; this might account for the generally short duration of alarm calls, functioning to decouple the call from information about location (e.g., Caro, 2005). In cases where receivers' evolved responses to different variants of a cue or signal carry differential fitness consequences for senders, selection might favor senders who emit signals that are exaggerated or otherwise "deceptive" (in an evolutionary, not cognitive, sense). For example, in many taxa, male-male competition for mates and/or resources is decided partly by vocal contests in which the F0 and/or spacing of the formant frequencies is used as a cue to gauge competitor body size or strength (reviewed in Fitch & Hauser, 2006). Howler monkeys (*Alouatta* spp.) have evolved a highly derived vocal apparatus, in which an enlarged hyoid bulla acts as a resonating chamber amplifying low frequencies and an elongated vocal tract results in wider formant spacing, presumably due to selection favoring the ability to convey exaggerated information about body size (Kitchen et al., 2015; for a similar example in frogs, see Bee et al., 2002). Cases in which senders' and receivers' fitness interests conflict yet cues remain stable over evolutionary time are thought to involve coevolution of an increased ability on the part of receivers to discriminate honest from dishonest signals (Maynard Smith & Harper, 2003), and/or evolutionary constraints preventing decoupling (Fitch & Hauser, 2006). Constraints are discussed further in Section 4.1. Phylogenetic Inertia and Evolutionary Homology.

3.2. The Receiver-Precursor Model

The receiver-precursor model begins with some kind of predisposition or *sensory bias* in receivers, creating a selection pressure favoring senders who evolve signals that exploit these biases and "manipulate" receivers' behavior (again, in an evolutionary, not cognitive, sense). For example, male túngara frogs (*Physalaemus pustulosus*) have evolved a call sequence for which females show a phonotactic preference, composed of a "whine" call followed by a number of "chucks" (Ryan, 1990; Ryan & Rand, 1993). The female preference for these calls seems to be a byproduct of the structure of the species' auditory system, which contains two organs that each happen to be most sensitive to the frequencies emphasized in the whine and the chuck respectively; the calls are thought to have evolved to emphasize these frequencies in order to more strongly elicit female responses (Ryan & Rand, 1993, 2003). As another example, many animals are thought to hold a tendency to abate aggression toward infants, presumably due to selection pressures against infanticide; the motivation-structural rules model, introduced above, posits that tonal, upward-sweeping vocalizations associated with submission might have evolved partly through selection pressures to pacify aggressive receivers by mimicking infant vocalizations (Morton, 1977). In turn, Lingle et al. (2012) suggested that some infant calls might have evolved to bear acoustic similarities to the courtship calls of males, in order to encourage female attraction toward those sounds. The motivation-structural rules model also posits that lower-frequency, noisy vocalizations, signaling aggression, might have evolved partly through pressures stemming from receivers' tendency to balk in contests with individuals of large body size, of which low frequency presumably once served as an honest cue, as discussed above.

If receivers' reactions to such signals are beneficial to their fitness, selection is expected to favor ritualization, as in the sender-precursor model, as well as *tuning* or enhancement of receivers' sensitivity to the signal, whereas if the fitness consequences are negative, selection is expected to favor receivers who can resist the signal or discriminate among signal variants (Maynard Smith & Harper, 2003). This can result in an evolutionary arms race between senders and receivers (who, it bears repeating, often comprise the same individuals) in which receivers evolve ever greater resistance while senders evolve ever more exaggerated signals (Arak & Enquist, 1995). The extremely loud, noisy, and low-frequency vocalizations exhibited by howlers, discussed above, might have evolved through such a process (Kitchen et al., 2015).

While this chapter is primarily focused on the significance of emotion for vocal production, it is important to note here that emotion can be an important mechanism underlying receiver behavior. The affect induction model posits that vocalizations generally function by evoking an emotional response in receivers (Owren & Rendall, 1997, 2001). For example, it is possible that infant distress vocalizations evolved their noisy quality, rapid amplitude increase at onset, and dramatic frequency modulation in order to directly induce arousal in receivers through features of the auditory nervous system, thereby manipulating a mother into providing for her infant even beyond her own fitness interests (Owren & Rendall, 2001). Emotional contagion, the process by which one individual’s emotional state spreads to others through automatic perception-action processes (Preston & de Waal, 2002), is one potentially widespread mechanism of affect induction in which the emotional state of vocalizers, conveyed by their vocalizations, directly induces a similar emotional state in receivers (Briefer, 2018). This could potentially play a role in many instances in which receivers respond more urgently to acoustic variants of calls given in more urgent situations—a phenomenon seen, as noted above, in meerkats’ responses to alarm calls (Manser et al., 2002).

3.3. Environmental Pressures

When signaling is beneficial for senders, selection favors acoustic characteristics that defy or counteract deleterious effects of distance and the environment (Morton, 1975). Effects of the environment on the acoustics of signals can include overall attenuation and frequency-dependent filtering due to absorption of energy by the atmosphere, ground, and obstacles (e.g., grass, trees), introduction of amplitude modulation by air turbulence and reverberation, and interference by the sounds of other con- and heterospecifics (Bee & Micheyl, 2008; Forrest, 1994; Marten & Marler, 1977; Waser & Waser, 1977; Wiley & Richards, 1978). The simplest way to counteract attenuation over distance is to make signals louder, and another is to lower their emphasized frequencies, since low-frequency sound travels farther (Wiley & Richards, 1978). Selection is especially likely to favor louder and lower-frequency signals when senders and receivers are separated by distances and obstacles; this may partly explain why howler monkeys, who call to communicate with faraway groups in dense tropical forests, have evolved such loud, low-frequency calls (da Cunha, de Oliveira, Holzmann, & Kitchen, 2015). To account for frequency-dependent filtering by the environment, and noise from other species, some bird and primate species appear to have evolved loud calls that emphasize the frequencies least attenuated within their habitats (the "sound window") (Marten & Marler, 1977; Morton, 1975; Waser & Waser, 1977). Although vocal flexibility is generally limited in primates, common marmosets have demonstrated a capacity to adaptively and preemptively modify the acoustic structure of their vocalizations after experiencing interfering environmental sounds (Zhao, Rad, & Wang, 2019). The impacts of environmental factors can exert selection pressures not only on the acoustic characteristics of signals, but on the entire communication systems of species. For example, still water has a much lesser attenuating effect than ground, perhaps partly explaining why acoustic communication is such an important aspect of life for frogs and toads, who commonly mate in and around ponds and lakes (Forrest, 1994).

4. Integrating Proximate and Ultimate Explanations: The Significance of Emotion for Vocal Evolution

The sender- and receiver-precursor models of animal communication clearly have enormous value for explaining the evolution of the acoustic properties of vocalizations. At the same time, these models heavily emphasize the action of direct selection pressures. Understanding how other evolutionary processes—such as phylogenetic inertia, evolutionary constraints on vocal production, and correlated response to selection—might influence the acoustic structures of vocalizations requires considering the proximate mechanisms of vocal production, specifically the role of emotion. This argument is developed below, with the central thesis being that understanding the proximate-level role of emotion in vocal production is not incidental, but rather critical, to ultimate-level understanding of vocal communication.

4.1. Phylogenetic Inertia and Evolutionary Homology

In *The Expression of the Emotions in Man and Animals*, Charles Darwin conceived of emotional facial expressions as nonadaptive vestiges of behaviors that served an ancestral function (Darwin, 1872). According to this reasoning, we snarl when provoked not because the snarl is itself adaptive, but because our ancestors may have bitten at the neck of their opponent; the snarl is a remnant of that behavior, which evolution has since removed from our behavioral repertoire. Put differently, Darwin viewed facial expressions as the product of *phylogenetic inertia*, the tendency of traits to remain stable over evolutionary time in the absence of direct selection pressures or even despite negative pressures (for a review of definitions, see Shanahan, 2011). Phylogenetic inertia can result in retention of traits long after they have lost their utility, or even become detrimental, to an organism. For example, a valley population Darwin's leaf-eared mouse (*Phyllotis darwini*) still exhibits adaptations for high altitude, 450,000 years after descending from its ancestral habitat in the Andean plateau (Bacigalupe, Nespolo, Opazo, & Bozinovic, 2015).

One can especially expect to observe phylogenetic inertia when a trait is under neutral or stabilizing selection, but even when pressures favor evolutionary change, *evolutionary constraints* may cause traits to remain stable over evolutionary time. For example, as discussed above, receivers of many species use acoustic cues to attain information about senders' body size or strength; it is possible to imagine that selection might favor a decoupling of these cues, allowing senders to "exaggerate," as is thought to have occurred in, e.g., frogs and howler monkeys (Bee et al., 2002; Kitchen et al., 2015). One possible factor constraining the evolution of exaggerated signals in some taxa has to do with the anatomical mechanisms of vocal production: selection on the mammalian larynx is constrained by the organ's dual role as the vocal source and as the valve protecting the respiratory tract from food and other matter (Fitch & Hauser, 2006). Any evolutionary change in laryngeal morphology or placement due to selective pressure relating to vocal production will be constrained by the larynx's role as gatekeeper to the respiratory tract, interference with which may result in elevated risk of choking to death. According to this account, vocal evolution in many mammalian taxa has been kept in check, perhaps partly explaining the generally limited vocal flexibility of mammals relative to birds, whose vocal production organ (the syrinx) is divorced from the "gatekeeper" function (Fitch & Hauser ibid.). Strong pressures favoring increased vocal flexibility in humans are thought to be responsible for the descent of the larynx in our species' evolutionary history, despite increased choking risks suffered by humans relative to other mammals, making us the exception that proves this rule (Fitch, 2000). Although constraints on selection are but one mechanism of phylogenetic inertia, this discussion elegantly illustrates the necessity of considering proximate mechanisms when attempting to generate ultimate explanations: without understanding the mechanisms of vocal production, we would have no idea of the competing evolutionary forces in play over the course of vocal evolution.

From a macroevolutionary perspective, phylogentic inertia operating in multiple related lineages can result in *evolutionary homology*, the phenomenon where multiple extant species exhibit shared traits (or *homologies*) inherited from a common ancestor. The concept of evolutionary homology is traceable to *On the Origin of Species*, where Darwin coined the law of "Unity of Type" (Darwin, 1859, p. 206), to capture the observation that many species exhibit some similarities in form despite different ecologies (see Shanahan, 2011). A classic example is seen in the forelimbs of vertebrates: the human arm, the foreleg of a horse, a bat’s wing, a dolphin’s fin, and a penguin’s flipper all share a basic skeletal structure. Darwin argued, correctly, that such homologies are the result of common descent. Each species inherits its form from an ancestral species and subsequently undergoes evolution by natural selection, but retains the legacy of its ancestry through phylogenetic inertia.

Evolutionary homology is visible in the ways in which emotions manifest in the acoustic properties of vocalizations in different species (Briefer, 2012; Filippi et al., 2017; Zimmermann et al., 2013). Evidence supporting this view includes cross-species similarities in vocal emotion expression as well as humans’ ability to accurately judge the emotional content of the vocalizations of other species (reviewed in Section 2.3.3. Vocal Emotion Expression: Effects of Emotion on Vocal Production and Acoustics). Homologies in vocal emotion expression are likely due in large part to conserved proximate mechanisms around physiology and vocal anatomy. For example, emotional arousal tenses the vocal folds, resulting in increased vocal F0, in different species alike. This is at least generally true for mammalian taxa; some non-mammal vertebrates (e.g., ravens) show similar relationships between arousal and vocal acoustics to those of mammals (Filippi et al., 2017; Szipl et al., 2017), and this could be the result of homology or convergent evolution. Homologous similarities in vocal emotion expression across species suggest that the acoustic properties of some vocalizations, especially strongly emotional vocalizations, are better explained by phylogenetic inertia than by selective pressures relating to the focal species’ particular socioecology. That is not to say that selection is incapable of tweaking vocal emotion expression to create exceptions to mammal-typical patterns, but simply to say that phylogenetic inertia appears to be a significant factor in ultimate explanations for acoustic properties of vocalizations, particularly when emotion plays a role.

The available evidence appears to support greater cross-species consistency, and thus a greater role of phylogenetic inertia, in the acoustic correlates of emotional arousal and valence within vocalization types than between vocalization types. The diversity of acoustic structures of different vocalization types across taxa suggests that selection is free to alter the baseline or centroid acoustic structure of a given vocalization type, sometimes in ways that contradict predictions based on emotion. For example, although anti-predator alarm calls are associated with relatively high urgency and presumably negative emotional valence—generally associated with longer vocal duration (reviewed in Section 2.3.3. Vocal Emotion Expression: Effects of Emotion on Vocal Production and Acoustics)—they tend to be short in duration relative to other vocalization types (Caro, 2005), perhaps because selection has favored alarm calls that minimize the information conveyed to predators about the location of the sender (as discussed above). Correlations between emotion and within-type acoustic structure are not universal—for example, “whinneys” of spider monkeys and the “rumbles” of Asian elephants (*Elephas maximus*) appear to show a negative rather than positive correlation between arousal and F0 (Ordóñez-Gómez et al., 2019; Sharma, Prakash S, Kohshima, & Sukumar, 2020)—but such exceptions to cross-species trends vis-à-vis vocal emotion expression within vocalization types are fewer and farther between than exceptions in between-vocalization-type differences. One possible explanation is that within-type acoustic variation generally functions to convey information about senders’ emotional states, and has therefore remained under stabilizing selection. Another possible factor is that, as discussed earlier (Section 2.3.3. Vocal Emotion Expression: Effects of Emotion on Vocal Production and Acoustics), the relationships between emotion and within-type acoustic variation are more straightforwardly causal at the proximate level than those between emotion and between-type variation. Specifically, whereas within-type acoustic variation appears to be determined more exclusively by emotional states, call usage appears to involve an array of different cognitive processes including executive functioning (Schamberg et al., 2018; Schwartz et al., in press).

4.2 Correlated Response to Selection

In addition to direct selection and phylogenetic inertia, *correlated response to selection* also plays an important role in the evolutionary trajectory of some traits (Trevor Price & Langen, 1992). This can occur when two traits are intrinsically linked or integrated in some way, for example, genetic pleiotropy (the protein produced by a single gene influences both traits), or the development or production of one trait directly impacts that of the other trait. In such cases, selective pressures on only one of the two traits can cause evolutionary change in both. This process exemplifies the interdependence of proximate- and ultimate-level processes: it is impossible to understand correlated response to selection without understanding how (in a proximate sense) two or more traits are linked together (Park, 2007; Pigliucci & Kaplan, 2000). Behavioral examples hypothesized in the literature include multiple mating by females evolving through correlated response to selection favoring multiple mating by males (Halliday & Arnold, 1987), alloparenting as a product of correlated response to parenting (arising with the evolution of communal breeding; Jamieson, 1989), and rejection of one’s own eggs evolving through pressures favoring rejection of brood parasites’ eggs (Davies & Brooke, 1988). The critical insight that comes from consideration of correlated response to selection is that traits need not necessarily be functional or beneficial in order to evolve through selection (Gould & Lewontin, 1979). However, not all traits that evolve through correlated response to selection are evolutionarily neutral (Trevor Price & Langen, 1992). If a one trait negatively impacts fitness, negative selection might constrain the evolution of a linked beneficial trait. If a trait that initially arose through correlated response is potentially useful to an organism, it can be *exapted* (co-opted by selection to serve a function) (Gould & Vrba, 1982); for example, it has been argued that alloparenting, although initially perhaps arising through correlated response to selection, was likely subsequently favored (Emlen et al., 1991). That said, evolutionary biologists recognize the importance of distinguishing between traits that evolved primarily through direct natural selection and traits that initially arose through other processes (such as correlated response to selection) were subsequently exapted (Gould, 1997). Applying this distinction to animal communication theory (Bradbury & Vehrencamp, 1998), the former category would apply to a signal that initially arose due to positive selection favoring exploitation of receiver biases (i.e., the receiver-precursor model), whereas a signal that initially arose as a cue and then underwent ritualization (i.e., the sender-precursor model) could be viewed as having been exapted.

The argument put forth by Fitch & Hauser (2006) about the role of constraints in the evolution of vocal signaling, reviewed earlier (Section 4.1. Phylogenetic Inertia and Evolutionary Homology), can be interpreted through the lens of correlated response to selection. Fitch & Hauser (ibid.) argued that selection on vocal production is constrained by the dual function of the larynx as vocal source and gatekeeper to the respiratory tract; another way of putting this is that there is an intrinsic link between vocal production and respiration. Thus a hypothetical mutation altering laryngeal morphology resulting in beneficial effects on vocal production (e.g., increased attractiveness of male vocalizations to mates) might also carry deleterious impacts on respiration (e.g., increased risk of choking, or even inability to breathe). In that case, selection might remove the mutant allele despite any fitness advantages conferred by the change in vocal production (again, humans serve as a case in which such a mutation(s) was favored because the benefits outweighed the costs associated with choking risk; Fitch, 2000). This example features a link between vocal production and a separate trait, but different acoustic features of vocalizations can also be viewed as linked with one another, as becomes clear upon consideration of the mechanisms of vocal production. For example, an increase in lung size would allow an animal to expel greater volumes of air, and would therefore be expected to allow an animal to produce longer calls *as well as* louder calls. Thus a hypothetical selection pressure favoring louder calls might lead to an evolved increase in lung size, and as a result, we might observe a correlated response to selection in call duration resulting in longer calls.

4.3 The Theoretical Significance of Emotion for Correlated Response to Selection

For heuristic purposes, above, I have attempted an accessible, and therefore probably oversimplified, explanation of correlated response to selection in behavior. In reality, the picture is complicated by intra-individual variation in behavior, the relationship of this variation with context, and variation in the context-behavior relationship itself, all of which make thinking about how behaviors might evolve through correlated response to selection potentially onerous, though no less necessary for that (Buss, Haselton, Shackelford, Bleske, & Wakefield, 1998; Park, 2007). However, invoking emotion as an “intervening variable” between context and behavior (de Waal, 2011) can clarify these relationships and help to generate specific hypotheses. Specifically, I hypothesize that the proximate links between emotions and emotional behaviors (including vocal emotion expression) make such behaviors particularly susceptible to evolution by correlated response to selection. In the paragraphs below, I will substantiate this hypothesis by developing speculative accounts regarding how certain acoustic properties of vocalizations in certain contexts might have evolved through correlated response to selection on other emotional behaviors. It is not my intention to advocate adopting these alternative explanations over more traditional ones at this time. Rather, these alternative explanations are intended to illustrate the potential epistemological value of the emotion concept for the study of the evolution of vocal production, acoustics, and communication, and for evolutionary science more broadly.

An important feature of emotions is that two instances of a similar emotional state can yield multiple coordinated behavioral outputs (de Waal, 2011); this principle has been referred to as “emotional pleiotropy” (Anderson & Adolphs, 2014; see also Section 2.3.1. Defining Emotion). The pleiotropic nature of emotion is thought to generally facilitate adaptive behavioral responses (Adolphs & Andler, 2018; Anderson & Adolphs, 2014; de Waal, 2011), but also implies that correlated response to selection might represent an important process in the evolution of emotional behavior. Genetic pleiotropy (a single gene contributing to multiple phenotypic traits) is a well-established mechanism of correlated response to selection: selection on one of those traits might result in evolutionary change in the frequency of that gene, which would then be expected to produce an incidental change in the other trait (Trevor Price & Langen, 1992). Extending this reasoning to emotional pleiotropy, selection favoring one behavioral outcome of an emotional state, in a particular context, might be expected to yield a correlated response in other behavioral outcomes of the emotional state, without those other behaviors needing to be adaptive in said context.

For example, Owren & Rendall (2001) suggested that the noisy quality of primate alarm calls might function to grab the attention of conspecific receivers. Let us assume for the sake of argument that this noisiness is partly attributable at the proximate level to a high-arousal emotional state, with arousal tensing the diaphragm and laryngeal muscles and constricting the pharynx generating air turbulence. Physiological arousal per se might comprise an adaptive response to the presence of a predator, because it increases blood-flow to the muscles, mobilizes energy stores, etc., preparing an animal to escape (Masterson & Crawford, 1982). It is plausible that selection might have favored an arousal response to predators on this basis, resulting in the evolution of heightened activation of the sympathetic nervous system and/or the HPA and sympathoadrenal axes. In that case, these evolved changes could be expected to yield additional distributed effects on physiology, cognition, and behavior in the context of predator threat, including call noisiness and other acoustic properties of vocalizations. The noisiness of alarm calls could then be viewed as partly a product of correlated response to selection favoring a broader arousal response, in addition to any selection pressures acting directly on call noisiness itself.

In the above example, correlated response to selection is invoked to explain the baseline or centroid acoustic structure of a vocalization type, or variation among vocalization types, but this reasoning can be extended to explain the evolution of context-dependent acoustic variation *within* vocalization types as well. For example, the contact calls of many mammals exhibit greater F0 modulation when the sender is isolated—a change that is hypothesized to be due at the proximate level to increased arousal in isolation (Lingle et al., 2012). This proximate-level account is credible in that tensing of the laryngeal muscles should be expected to result in more F0-variable calls, a trend that is supported by evidence across various contexts in several mammalian taxa (Briefer, 2012). This acoustic change is postulated to have evolved to make these calls more localizable in situations where localizability is more important to call function, namely, reunion with groupmates (Sugiura, 2007; Yamaguchi et al., 2010). This account is plausible, but other selection pressures on other outputs of emotional arousal in the context of isolation might also have played a role. Selection could be expected to favor enhanced readiness to fight or flee from predators in contexts of isolation, since an isolated animal is more vulnerable; insofar as arousal causes changes in the F0 modulation of calls, this selection pressure favoring a broader physiological arousal response could be expected to produce, through a correlated response, the increased F0 modulation observed in contact calls emitted in isolation relative to other contexts. Indeed, in principle, such a selection pressure favoring a broader arousal response could potentially suffice to explain the change in F0 modulation, even in the absence of any direct fitness impact of the acoustic change itself, though correlated response and direct selection are not incompatible and could both have played a role simultaneously. Moreover, it is plausible that selection favoring changes in other acoustic parameters in the context of isolation might also have yielded a correlated response in F0 modulation. For example, selection could have favored increased call amplitude functioning to increase the distance traveled by the signal, and this pressure could have resulted in an evolved increase in sender arousal in that context; such an evolutionary process might theoretically be expected to yield a correlated response in F0 modulation.

Another important characteristic of emotion is the multiplicity of different contexts that might evoke a similar set of emotional states and, therefore, behavioral responses—a feature that has been referred to as “stimulus generalizability” (Anderson & Adolphs, 2014). In many cases, this feature of emotion serves as an evolutionary shortcut that allows animals to respond adaptively to different contexts without needing to evolve separate stimulus-response circuits for every conceivable situation (Adolphs & Andler, 2018; Anderson & Adolphs, 2014; de Waal, 2011). In other cases, the stimulus generalizability feature of emotion can potentially result in nonadaptive behavior. Darwin (1872) acknowledged this, using the example of a cat kneading a blanket (see also Anderson & Adolphs, 2014). Kneading when in contact with a nursing mother's abdomen is adaptive for a kitten because it stimulates milk flow. This appears to have resulted in selection favoring kneading as an expression of some low-arousal, positive-valence emotion (comfort or contentedness). This emotion occurs, and yields kneading, not only when a cat is in contact with a nursing mother's abdomen but also in other contexts, including when in contact with a blanket. Blanket-kneading can thus be thought of as having evolved at least in part through a correlated response to selection favoring belly-kneading (this does not preclude the possibility that the kneading of surfaces other than the mother's belly might have been exapted to serve some communicative function, i.e., the sender-precursor model).

As discussed earlier, vocal emotion expression exemplifies the stimulus generalizability characteristic of emotion: similar acoustic changes occur in association with changes in emotional arousal or valence across disparate contexts (and vocalization types). It follows that certain acoustic properties of vocalizations in certain contexts might be analogous to the behavior of kneading a blanket, that is, products of correlated response to selection favoring similar acoustic properties in other distinct, but emotionally similar, contexts. To illustrate this hypothesis, let us return to the example of increased F0 modulation in contact calls in the context of isolation—an acoustic change that is probably due at the proximate level to emotional arousal, and has been hypothesized to increase the odds of successful reunion with groupmates by making the calls more localizable (Sugiura, 2007; Yamaguchi et al., 2010). Contact calls of many primates occur not only in isolation but also in a wide variety of other contexts including in response to high-quality food and preceding the initiation of friendly contact. Above I discussed how, to the extent that arousal is pleiotropic, the degree of F0 modulation of contact calls might respond not only to direct selection pressure but also to selection acting on other behavioral and acoustic outputs of arousal. Similarly, insofar as arousal exhibits stimulus generalizability, the acoustics of contact calls in isolation might respond not only to selection stemming from fitness consequences in the context of isolation, but also to pressures stemming from the communicative functions of contact calls in other contexts. Extending this reasoning to the extreme, for the sake of argument, it is even conceivable that selection pressures on the acoustic properties of one vocalization type in one context might influence the effects of arousal and other components of emotions on the anatomical mechanisms of vocal production, yielding a correlated evolutionary response in different acoustic properties of a different vocalization type in a different context.

In summary, changes in vocal acoustics associated with emotional arousal or valence can be viewed as potentially integrated with each other and with other outputs of emotional states, in similar as well as disparate (but emotionally similar) contexts; consequently, selection favoring any one of those outputs might exert a pull on them all (Figure 2). That is not to negate the potential impact of selection acting directly on vocal acoustics through the kinds of pressures described in Section 3. Animal Communication and Evolutionary Explanations for the Acoustic Properties of Vocalizations—in particular, insofar as emotional acoustic variation comprises a cue to senders’ emotional states, one may expect it to be subjected to the various pressures described within the sender-precursor model. Rather, emotion-related acoustic changes and other outputs of emotional states can, to the degree that they are integrated with each other, be expected to *collectively* exhibit evolutionary change in a direction consistent with the *sum* of all selection pressures acting on all integrated outputs.

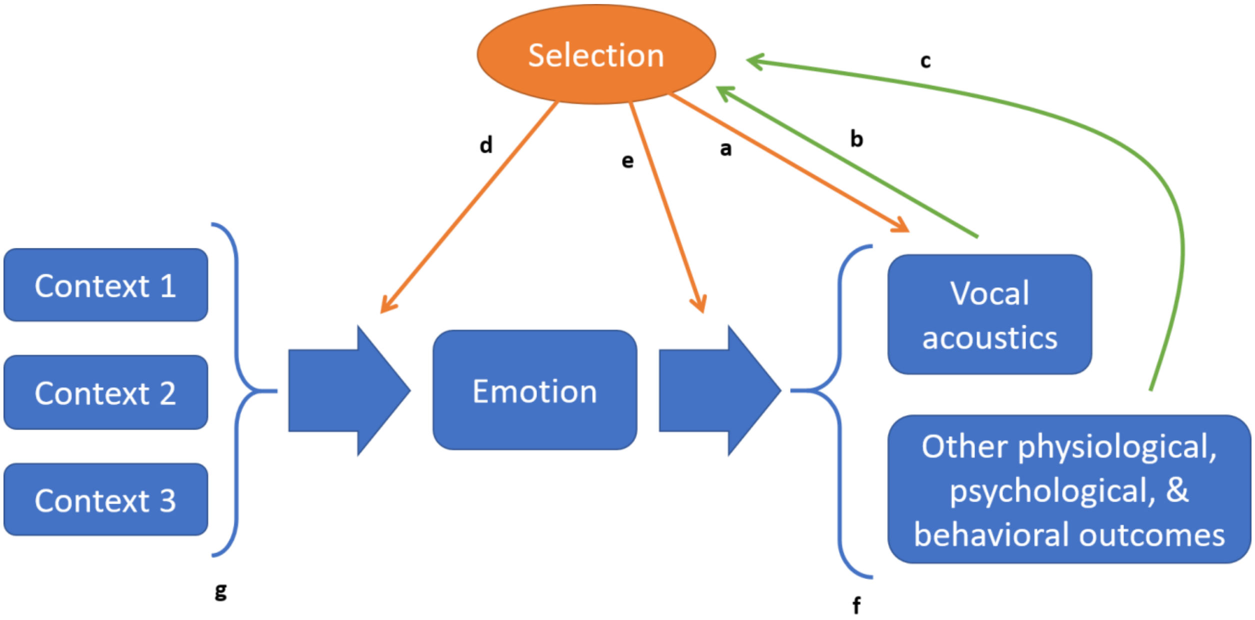

Figure 1.2. A theoretical model summarizing how the proximate-level relationships between contexts, emotion, and vocal and other behavior might facilitate various evolutionary processes. Selection may act directly on the acoustic properties of a vocalization type (**a**), based on the fitness consequences of these acoustic properties (**b**), i.e. the sender- and receiver-precursor models. Direct selection probably plays a central role in determining the typical acoustic structures of different vocalization types in a given species. Selection pressures affecting emotional responses to a given context might stem from the fitness consequences of within-vocalization-type acoustic variation in that context (**b**) and/or the fitness consequences of any of the other outcomes of an emotional state (**c**). These pressures might result in evolutionary changes to the ways in which the context evokes an emotional reaction (**d**) and/or the ways in which an emotion manifests in behavior etc. (**e**). To the extent that the multiple outputs of an emotional state are linked to that emotional state and to each other (**f**), selection pressures favoring a particular change to physiology, psychology, and/or behavior in a given context (**e**) might yield a *correlated response* in vocal acoustics in that context, and vice versa. To the extent that multiple different contexts fan into a set of similar emotional states (**g**), selection pressures favoring a particular response to a particular context (**d**) might result in the evolution of a similar response in other, emotionally similar contexts.

It bears repeating that my intention has been to illustrate the potential epistemological significance of emotion for the study of the evolution of communication, rather than to seriously evaluate competing accounts for the evolution of specific acoustic properties of vocalizations. A scientific theory is only valuable to the extent that it can generate testable hypotheses. My hope is that I have clarified some of the hypotheses regarding the evolution of vocal communication that become apparent upon consideration of the proximate role of emotion in vocal production. In the next section, I will discuss how empirical research into vocal emotion expression might help resolve these hypotheses, and introduce the specific contributions of the research included in this dissertation.

5. The Evolutionary Potential of Vocal Emotion Expression Research

Above, I have delineated several evolutionary processes that might be involved in the evolution of vocal communication, especially the evolution of vocal production and how the acoustic properties of vocalizations vary depending on context. One of these processes is direct selection, in which vocal acoustics and their relationship to context undergo evolutionary change due to effects on communicative function and sender fitness. Another is phylogenetic inertia, in which vocal acoustics and their relationship to context remain stable over evolutionary time—potentially resulting in homologous similarities among related taxa—due to neutral selection, stabilizing selection, and/or evolutionary constraints. Finally, I discussed the potential for a process in which vocal acoustics and their relationship to context undergo evolutionary change due to the action of selection on other phenotypic characteristics that are linked to vocal acoustics at the proximate level, e.g., through the physiological components of emotion.

The question, then, is how do the predictions of these different evolutionary hypotheses differ, and how can they be tested? The hypothesis that vocal emotion expression has been conserved over evolutionary time predicts that changes in emotional arousal and valence should each be associated with similar changes in vocal acoustics across contexts, vocalization types, and species—a prediction which has received considerable empirical support (reviewed in Briefer, 2012; Zimmermann et al., 2013) but which remains untested in many taxa.

Discriminating empirically between the roles of direct selection and correlated response to selection in vocal evolution is more challenging. In theory, the hypothesis that any given acoustic change occurring in a given context might exhibit a correlated response to selection pressures acting on other acoustic changes, and/or other integrated outputs of an emotional state, is only plausible to the extent that acoustic parameters of vocalizations are linked to emotional states and integrated with each other. One approach to gaining insight into the evolutionary dynamics at play in vocal evolution is, therefore, to test the degree of this integration. If the ways in which emotional arousal and valence manifest in the acoustic properties of vocalizations are relatively consistent across contexts, vocalization types, and species, this observation would be consistent with the idea that vocal acoustics are strongly linked to emotion, theoretically increasing the odds that the model of correlated response to selection outlined above might has been a significant factor in vocal evolution broadly. Such evidence would support a view of the acoustic and other outputs of emotional states as comprising an integrated whole that should be expected to respond collectively to the sum of selection pressures acting on each output. In contrast, if the ways in which emotional arousal and valence manifest in vocal acoustics vary substantially across contexts, vocalization types, and species, this would be inconsistent with the above model of integration and correlated response to selection. The aim of this dissertation is to contribute to testing these predictions and resolving these broader hypotheses.

In Chapter II, I report a study examining the correlations between emotional arousal and vocal acoustics within two different rhesus macaque vocalization types—coos and screams—in two different contexts—separation from conspecifics during a behavioral test, and naturally occurring agonistic interactions, respectively. If the acoustic correlates of emotional arousal within coos are different from those within screams, this would suggest arousal-related acoustic changes occurring in these two vocalization types in these two contexts are independent. If, however, the acoustic correlates of arousal are similar within coos and within screams, despite the disparity between the two contexts in which these two types of vocalizations were recorded, this would suggest that the vocal changes observed with increasing arousal are integrated across contexts and vocalization types, supporting an evolutionary model in which the acoustic properties of one vocalization type observed in one context might be expected to respond to selection pressures acting on the acoustic and other outputs of emotional arousal in other vocalization types and other contexts. Based on the abundance of literature indicating homologous similarities in vocal emotion expression, I hypothesize that the acoustic correlates of arousal will be generally similar between rhesus macaque coos and screams.

In Chapter III, I report a study examining human perceptions of the emotions represented in rhesus macaque vocalizations. The hypothesis that vocal emotion expression shows homologous similarities across species suggests that listeners of one species might be sensitive to the emotional significance of vocalizations of another species, even if they lack familiarity with that species (Filippi et al., 2017). I therefore tested the accuracy of humans, some experienced with rhesus macaques and some not, on a forced choice task in which they indicated which of two rhesus macaque vocalizations represented greater emotional arousal. If even inexperienced listeners exhibit above-chance accuracy on this task, this would suggest that vocal emotion expression is similar between humans and rhesus macaques, likely due to evolutionary homology. This would support a significant role of phylogenetic inertia in the relationships between arousal and vocal acoustics across catarrhine primates.

To summarize, this research promises a better understanding of vocal communication at both the proximate and ultimate level. At the proximate level, it provides an improved understanding of the mechanistic role of emotion in vocal production. It also carries the potential to help researchers and staff at primate research centers, zoos, and sanctuaries to better monitor and improve the emotional welfare of rhesus macaques living in human care. At the ultimate level, this research helps resolve competing hypotheses about the evolutionary history of the acoustic properties of the vocalizations under study, as well as about how vocal communication evolves more generally. The significance of the results reported in Chapters II and III for the evolutionary issues outlined in the present chapter are discussed in Chapter IV.

References

Adolphs, R., & Andler, D. (2018). Investigating emotions as functional states distinct from feelings. *Emotion Review*, *10*(3), 191–201. https://doi.org/10.1177/1754073918765662

Alcock, J., & Sherman, P. (1994). The utility of the proximate-ultimate dichotomy in ethology. *Ethology*, *96*(1), 58–62. https://doi.org/10.1111/j.1439-0310.1994.tb00881.x

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

Anikin, A., Bååth, R., & Persson, T. (2018). Human non-linguistic vocal repertoire: Call types and their meaning. *Journal of Nonverbal Behavior*, *49*(2), 53–80. https://doi.org/10.1007/s10919-017-0267-y

Anolli, L., Wang, L., Mantovani, F., & De Toni, A. (2008). The voice of emotion in Chinese and Italian young adults. *Journal of Cross-Cultural Psychology*, *39*(5), 565–598. https://doi.org/10.1177/0022022108321178

Arak, A., & Enquist, M. (1995). Conflict, receiver bias and the evolution of signal form. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *349*(1330), 337–344. https://doi.org/10.1098/rstb.1995.0122

Arbilly, M., & Lotem, A. (2017). Constructive anthropomorphism: a functional evolutionary approach to the study of human-like cognitive mechanisms in animals. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1865), 20171616. https://doi.org/10.1098/rspb.2017.1616

August, P. V., & Anderson, J. G. T. (2006). Mammal sounds and motivation-structural rules: A test of the hypothesis. *Journal of Mammalogy*, *68*(1), 1–9. https://doi.org/10.2307/1381039

Baciadonna, L., Briefer, E. F., Favaro, L., & McElligott, A. G. (2019). Goats distinguish between positive and negative emotion-linked vocalisations. *Frontiers in Zoology*, *16*(1), 25. https://doi.org/10.1186/s12983-019-0323-z

Bacigalupe, L. D., Nespolo, R. F., Opazo, J. C., & Bozinovic, F. (2015). Phenotypic flexibility in a novel thermal environment: Phylogenetic inertia in thermogenic capacity and evolutionary adaptation in organ size. *Physiological and Biochemical Zoology*, *77*(5), 805–815. https://doi.org/10.1086/422229

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58. https://doi.org/10.1111/j.1745-6916.2006.00003.x

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Barrett, L. F., Henzi, P., & Rendall, D. (2007). Social brains, simple minds: does social complexity really require cognitive complexity? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 561–575. https://doi.org/10.1098/rstb.2006.1995

Bateson, P., & Laland, K. N. (2013). Tinbergen’s four questions: An appreciation and an update. *Trends in Ecology and Evolution*, *28*(12), 712–718. https://doi.org/10.1016/j.tree.2013.09.013

Bayart, F., Hayashi, K. T., Faull, K. F., Barchas, J. D., & Levine, S. (1990). Influence of maternal proximity on behavioral and physiological responses to separation in infant rhesus monkeys (*Macaca mulatta*). *Behavioral Neuroscience*, *104*(1), 98–107. https://doi.org/10.1037/0735-7044.104.1.98

Bee, M. A., & Micheyl, C. (2008). The cocktail party problem: What is it? How can it be solved?

And why should animal behaviorists study it? *Journal of Comparative Psychology*, *122*(3), 235–251. https://doi.org/10.1037/0735-7036.122.3.235

Bee, M. A., Perrill, S. A., & Owen, P. C. (2002). Male green frogs lower the pitch of acoustic signals in defense of territories: A possible dishonest signal of size? *Behavioral Ecology, 11*(2), 169–177. https://doi.org/10.1093/beheco/11.2.169

Berridge, K. C. (1996). Food reward: Brain substrates of wanting and liking. *Neuroscience & Biobehavioral Reviews*, *20*(1), 1–25. https://doi.org/10.1016/0149-7634(95)00033-B

Blumstein, D. T., & Chi, Y. Y. (2012). Scared and less noisy: Glucocorticoids are associated with alarm call entropy. *Biology Letters*, *8*(2), 189–192. https://doi.org/10.1098/rsbl.2011.0832

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*. Sunderland, MA: Sinauer.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *7*(1), 1–20. https://doi.org/10.1098/rsbl.2011.0832

Briefer, E. F. (2018). Vocal contagion of emotions in non-human animals. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873), 20172783. https://doi.org/10.1098/rspb.2017.2783

Briefer, E. F., Linhart, P., Policht, R., Špinka, M., Leliveld, L. M. C., Düpjan, S., … Tallet, C. (2019). Vocal expression of emotional valence in pigs across multiple call types and contexts. *PeerJ Preprints*. https://doi.org/10.7287/peerj.preprints.27934v1

Briefer, E. F., Maigrot, A. L., Mandel, R., Freymond, S. B., Bachmann, I., & Hillmann, E. (2015). Segregation of information about emotional arousal and valence in horse whinnies. *Scientific Reports, 4,* 1–11. https://doi.org/10.1038/srep09989

Briefer, E. F., Tettamanti, F., & McElligott, A. G. (2015). Emotions in goats: Mapping physiological, behavioural and vocal profiles. *Animal Behaviour*, *99*, 131–143. https://doi.org/10.1016/J.ANBEHAV.2014.11.002

Briefer, E. F., Vizier, E., Gygax, L., & Hillmann, E. (2019). Expression of emotional valence in pig closed-mouth grunts: Involvement of both source- and filter-related parameters. *The Journal of the Acoustical Society of America*, *145*(5), 2895–2908. https://doi.org/10.1121/1.5100612

Buss, D. M., Haselton, M. G., Shackelford, T. K., Bleske, A. L., & Wakefield, J. C. (1998). Adaptations, exaptations, and spandrels. *American Psychologist*, *53*(5), 533–548. https://doi.org/10.1037/0003-066X.53.5.533

Caro, T. M. (2005). *Antipredator defenses in birds and mammals*. Chicago: University of Chicago Press.

Charmandari, E., Tsigos, C., & Chrousos, G. (2005). Endocrinology of the stress response. *Annual Review of Physiology*, *67*(1), 259–284. https://doi.org/10.1146/annurev.physiol.67.040403.120816

Cheney, D. L., & Seyfarth, R. M. (2018). Flexible usage and social function in primate vocalizations. *Proceedings of the National Academy of Sciences*, *115*(9), 1974–1979. https://doi.org/10.1073/pnas.1717572115

Collins, K. T., McGreevy, P. D., Wheatley, K. E., & Harcourt, R. G. (2011). The influence of behavioural context on Weddell seal (*Leptonychotes weddellii*) airborne mother-pup vocalisation. *Behavioural Processes, 87*(3), 286–290. https://doi.org/10.1016/j.beproc.2011.06.005

Craig, A. D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*(8), 655–666. https://doi.org/10.1038/nrn894

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology*, *22*(2), 142–146. https://doi.org/10.1016/J.CUB.2011.11.053

da Cunha, R. G. T., de Oliveira, D. A. G., Holzmann, I., & Kitchen, D. M. (2015). Production of loud and quiet calls in howler monkeys. In M. M. Kowalewski, P. A. Garber, L. Cortés-Ortiz, B. Urbani, & D. Youlatos (Eds.), *Howler monkeys: Adaptive radiation, systematics, and morphology* (pp. 337–368). New York: Springer.

Darwin, C. (1859). *On the origin of species by natural selection*. London: Murray. https://doi.org/10.1016/S0262-4079(09)60380-8

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Marry. https://doi.org/10.1037/h0076058

Davies, N. B., & Brooke, M. de L. (1988). Cuckoos versus reed warblers: Adaptations and counteradaptations. *Animal Behaviour*, *36*(1), 262–284. https://doi.org/10.1016/S0003-3472(88)80269-0

Dawkins, M. S. (2000). Animal minds and animal emotions. *American Zoologist*, *40*(6), 883–888. https://doi.org/10.1093/icb/40.6.883

Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation. In J. R. Krebs & N. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 282–309). Oxford: Blackwell.

de Vere, A. J., & Kuczaj, S. A. (2016). Where are we in the study of animal emotions? *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(5), 354–362. https://doi.org/10.1002/wcs.1399 de Waal, F. B. M. (1999). Anthropomorphism and anthropodenial: Consistency in our thinking about humans and other animals. *Philosophical Topics*, *27*(1), 255–280. https://doi.org/10.5840/philtopics199927122

de Waal, F. B. M. (2011). What is an animal emotion? *Annals of the New York Academy of Sciences*, *1224*(1), 191–206. https://doi.org/10.1111/j.1749-6632.2010.05912.x

de Waal, F. B. M., & Van Hooff, J. A. R. A. M. (1981). Side-directed communication and agonistic interactions in chimpanzees. *Behaviour*, *77*(3), 164–198. https://doi.org/10.1163/156853981X00211

Devenport, L., Knehans, A., Sundstrom, A., & Thomas, T. (1989). Corticosterone's dual metabolic actions. *Life Sciences*, *45*(15), 1389–1396. https://doi.org/10.1016/0024-3205(89)90026-X

Dewsbury, D. A. (1999). The proximate and the ultimate: Past, present, and future. *Behavioural Processes*, *46*(3), 189–199. https://doi.org/10.1016/S0376-6357(99)00035-2

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher*, *75*, 87–91. Retrieved from http://papodeprimata.com.br/wp-content/uploads/2016/02/Dobzhansky.pdf

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194. https://doi.org/10.1126/science.1076358

Dunn, J. C., & Smaers, J. B. (2018). Neural correlates of vocal repertoire in primates. *Frontiers in Neuroscience*, *12*, 534. https://doi.org/10.3389/fnins.2018.00534

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, *6*(3–4), 169–200. https://doi.org/10.1080/02699939208411068

Emlen, S. T., Reeve, H. K., Sherman, P. W., Wrege, P. H., Shellman-Reeve, J., & Shellman-

Reeve, J. (1991). Adaptive versus nonadaptive explanations of behavior: The case of alloparental helping. *The American Naturalist, 138*(1), 259–270. https://doi.org/10.1086/285216

Faragó, T., Andics, A., Devecseri, V., Kis, A., Gacsi, M., & Miklosi, A. (2014). Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biology Letters*, *10*(1), 20130926. https://doi.org/10.1098/rsbl.2013.0926

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*(4), 917–925. https://doi.org/10.1016/J.ANBEHAV.2010.01.005

Faragó, T., Takács, N., Miklósi, Á., & Pongrácz, P. (2017). Dog growls express various contextual and affective content for human listeners. *Royal Society Open Science, 4*(5), 170134. https://doi.org/10.1098/rsos.170134

Fichtel, C., Hammerschmidt, K., & Jürgens, U. (2001). On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behavior, 138*, 97–116. https://doi.org/doi:10.1163/15685390151067094

Fichtel, C., & Manser, M. (2010). Vocal communication in social groups. In P. Kappeler (Ed.), *Animal Behaviour: Evolution and Mechanisms* (pp. 29–54). Heidelberg, Heidelberg: Springer Berlin. https://doi.org/10.1007/978-3-642-02624-9

Ficken, M. S., Ficken, R. W., & Witkin, S. R. (1978). Vocal repertoire of the Black-capped Chickadee. *Auk*, *95*, 34–48. https://doi.org/10.2307/4085493

Filippi, P. (2016). Emotional and interactional prosody across animal communication systems: A comparative approach to the emergence of language. *Frontiers in Psychology*, *7*, 1393. https://doi.org/10.3389/fpsyg.2016.01393

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, 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: Biological Sciences, 284*(1859), 20170990. https://doi.org/10.1098/rspb.2017.0990

Fischer, J. (2017). Primate vocal production and the riddle of language evolution. *Psychonomic Bulletin and Review*, *24*(1), 72–78. https://doi.org/10.3758/s13423-016-1076-8

Fischer, J., Noser, R., & Hammerschmidt, K. (2013). Bioacoustic field research: A primer to acoustic analyses and playback experiments with primates. *American Journal of Primatology*, *75*(7), 643–663. https://doi.org/10.1002/ajp.22153

Fischer, J., & Price, T. (2016). Meaning, intention, and inference in primate vocal communication. *Neuroscience and Biobehavioral Reviews*, *82*, 22–31. https://doi.org/10.1016/j.neubiorev.2016.10.014

Fischer, J., Wadewitz, P., & Hammerschmidt, K. (2016). Structural variability and communicative complexity in acoustic communication. *Animal Behaviour*, *134*, 229–237. https://doi.org/10.1016/j.anbehav.2016.06.012

Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267. https://doi.org/10.1016/S1364-6613(00)01494-7

Fitch, W. T., & Hauser, M. D. (2006). Unpacking “honesty”: Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic communication* (pp. 65–137). New York: Springer. https://doi.org/10.1007/0-387-22762-8_3

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, *63*(3), 407–418. https://doi.org/10.1006/anbe.2001.1912

Forrest, T. G. (1994). From sender to receiver: Propagation and environmental effects on acoustic signals. *American Zoologist*, *34*(6), 644–654. https://doi.org/10.1093/icb/34.6.644

Friel, M., Kunc, H. P., Griffin, K., Asher, L., & Collins, L. M. (2019). Positive and negative contexts predict duration of pig vocalisations. *Scientific Reports*, *9*(1), 2062. https://doi.org/10.1038/s41598-019-38514-w

Fulkerson, W. J., & Tang, B. Y. (1979). Ultradian and circadian rhythms in the plasma concentration of cortisol in sheep. *Journal of Endocrinology*, *81*(1), 135–141. https://doi.org/10.1677/joe.0.0810135

Gogoleva, S. S., Volodina, E. V., Volodin, I. A., Kharlamova, A. V., & Trut, L. N. (2010). The gradual vocal responses to human-provoked discomfort in farmed silver foxes. *Acta Ethologica*, *13*(2), 75–85. https://doi.org/10.1007/s10211-010-0076-3

Goudbeek, M., & Scherer, K. R. (2010). Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. *The Journal of the Acoustical Society of America*, *128*(3), 1322–1336. https://doi.org/10.1121/1.3466853

Gould, S. J. (1997). The exaptive excellence of spandrels as a term and prototype. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(20), 10750–10755. https://doi.org/10.1073/pnas.94.20.10750

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society B: Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Gould, S. J., & Vrba, E. S. (1982). Exaptation—A missing term in the science of form.

Paleobiology, *8*(1), 4–15. https://doi.org/10.1017/S0094837300004310

Gouzoules, H. (2005). Monkeying around with symbolism: Are vocalizations simple symbols...or more like cymbals? In L. Namy (Ed.), *Symbol use and symbolic representation: Developmental and comparative perspectives.* (pp. 245–265). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Gouzoules, H., & Gouzoules, S. (1990). Body size effects on the acoustic structure of pigtail macaque (Macaca nemestrina) screams. *Ethology*, *85*(4), 324–334. https://doi.org/10.1111/j.1439-0310.1990.tb00411.x

Gouzoules, H., & Gouzoules, S. (1995). Recruitment screams of pigtail monkeys (*Macaca nemestrina*): Ontogenetic perspectives. *Behaviour*, *132*(5), 431–450. https://doi.org/10.1163/156853995X00658

Gouzoules, H., & Gouzoules, S. (2000). Agonistic screams differ among four species of macaques: The significance of motivation-structural rules. *Animal Behaviour*, *59*(3), 501–512. https://doi.org/10.1006/anbe.1999.1318

Gouzoules, H., Gouzoules, S., & Marler, P. (1985). External reference and affective signaling in mammalian vocal communication. In G. Zivin (Ed.), *The development of expressive behavior: Biology environment interactions* (pp. 77–101). New York: Academic Press.

Gouzoules, H., Gouzoules, S., & Tomaszycki, M. (1998). Agonistic screams and the classification of dominance relationships: Are monkeys fuzzy logicians? *Animal Behaviour*, *55*(1), 51–60. https://doi.org/10.1006/anbe.1997.0583

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour, 32*(1), 182–193. https://doi.org/10.1016/S0003-3472(84)80336-X

Halliday, T., & Arnold, S. J. (1987). Multiple mating by females: a perspective from quantitative genetics. *Animal Behaviour*, *35*(3), 939–941. https://doi.org/10.1016/S0003-3472(87)80138-0

Hamann, S. (2012). Mapping discrete and dimensional emotions onto the brain: Controversies and consensus. *Trends in Cognitive Sciences*, *16*(9), 458–466. https://doi.org/10.1016/j.tics.2012.07.006

Hammerschmidt, K., & Fischer, J. (1998). The vocal repertoire of Barbary macaques: A quantitative analysis of a graded signal system. *Ethology*, *104*, 203–216. https://doi.org/10.1111/j.1439-0310.1998.tb00063.x

Hemsworth, P., Mellor, D., Cronin, G., & Tilbrook, A. (2015). Scientific assessment of animal welfare. *New Zealand Veterinary Journal*, *63*(1), 24–30. https://doi.org/10.1080/00480169.2014.966167

Izard, C. E. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, *2*(3), 260–280. https://doi.org/10.1111/j.1745-6916.2007.00044.x

Jamieson, I. G. (1989). Behavioral heterochrony and the evolution of birds' helping at the nest: An unselected consequences of communal breeding? *The American Naturalist*, *133*(3), 394–406. https://doi.org/10.1086/284925

Jürgens, U. (2009). The neural control of vocalization in mammals: A review. *Journal of Voice*, *23*(1), 1–10. https://doi.org/10.1016/j.jvoice.2007.07.005

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kelly, T., Reby, D., Levréro, F., Keenan, S., Gustafsson, E., Koutseff, A., & Mathevon, N. (2017). Adult human perception of distress in the cries of bonobo, chimpanzee, and human infants. *Biological Journal of the Linnean Society*, *120*(4), 919–930. https://doi.org/10.1093/biolinnean/blw016

Kitchen, D. M., da Cunha, R. G. T., Holzmann, I., & de Oliveira, D. A. G. (2015). Function of loud calls in howler monkeys. In M. M. Kowalewski, P. A. Garber, L. Cortés-Ortiz, B. Urbani, & D. Youlatos (Eds.), *Howler monkeys: Adaptive radiation, systematics, and morphology* (pp. 369–399). New York: Springer. https://doi.org/10.1007/978-1-4939-1957-4_14

Kragel, P. A., & LaBar, K. S. (2016). Decoding the nature of emotion in the brain. *Trends in Cognitive Sciences*, *20*(6), 444–455. https://doi.org/10.1016/j.tics.2016.03.011

Laland, K. N., Odling-Smee, J., Hoppitt, W., & Uller, T. (2013). More on how and why: Cause and effect in biology revisited. *Biology & Philosophy*, *28*(5), 719–745. https://doi.org/10.1007/s10539-012-9335-1

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is Mayr’s proximate-ultimate dichotomy still useful? *Science*, *334*(6062), 1512–1516. https://doi.org/10.1126/science.1210879

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

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016-E2025. https://doi.org/10.1073/pnas.1619316114

Leliveld, L. M. C., Düpjan, S., Tuchscherer, A., & Puppe, B. (2017). Vocal correlates of emotional reactivity within and across contexts in domestic pigs (*Sus scrofa*). *Physiology and Behavior*, *181*, 117–126. https://doi.org/10.1016/j.physbeh.2017.09.010

Liao, D. A., Zhang, Y. S., Cai, L. X., & Ghazanfar, A. A. (2018). Internal states and extrinsic factors both determine monkey vocal production. *Proceedings of the National Academy of Sciences*, *115*(15), 3978–3983. https://doi.org/10.1073/pnas.1722426115

Lin, H., Müller-Bardorff, M., Gathmann, B., Brieke, J., Mothes-Lasch, M., Bruchmann, M., ... & Straube, T. (2020). Stimulus arousal drives amygdalar responses to emotional expressions across sensory modalities. *Scientific Reports*, *10*(1898), 1-12. https://doi.org/10.1038/s41598-020-58839-1

Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2016). The brain basis of positive and negative affect: evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex*, *26*(5), 1910-1922. https://doi.org/10.1093/cercor/bhv001

Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology*, *58*(5), 698–726. https://doi.org/10.1093/czoolo/58.5.698

Linhart, P., Ratcliffe, V. F., Reby, D., & Špinka, M. (2015). Expression of emotional arousal in two different piglet call types. *PLoS ONE*, *10*(8), 1–13. https://doi.org/10.1371/journal.pone.0135414

Maigrot, A.-L., Hillmann, E., & Briefer, E. F. (2018). Encoding of emotional valence in wild boar (*Sus scrofa*) calls. *Animals*, *8*(6), 85. https://doi.org/10.3390/ani8060085

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1483), 2315–2324. https://doi.org/10.1098/rspb.2001.1773

Manser, M. B., Bell, M. B., & Fletcher, L. B. (2001). The information that receivers extract from alarm calls in suricates. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1484), 2485–2491. https://doi.org/10.1098/rspb.2001.1772

Manser, M. B., Seyfarth, R. M., & Cheney, D. L. (2002). Suricate alarm calls signal predator class and urgency. *Trends in Cognitive Sciences*, *6*(2), 55–57. https://doi.org/10.1016/S1364-6613(00)01840-4

Marten, K., & Marler, P. (1977). Sound transmission and its significance for animal vocalization: I. Temperate habitats. *Behavioral Ecology and Sociobiology*, *2*(3), 271–290. https://doi.org/10.1007/BF00299740

Masterson, F. A., & Crawford, M. (1982). The defense motivation system: A theory of avoidance behavior. *Behavioral and Brain Sciences*, *5*(4), 661–675. https://doi.org/10.1017/S0140525X00014114

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford: Oxford University Press.

Mayr, E. (1961). Cause and effect in biology: Kinds of causes, predictability, and teleology are viewed by a practicing biologist. *Science*, *134*(3489), 1501–1506. https://doi.org/10.1126/science.134.3489.1501

McPherson, M. J., & McDermott, J. H. (2018). Diversity in pitch perception revealed by task dependence. *Nature Human Behaviour*, *2*(1), 52–66. https://doi.org/10.1038/s41562-017-0261-8

Meise, K., Keller, C., Cowlishaw, G., & Fischer, J. (2011). Sources of acoustic variation: Implications for production specificity and call categorization in chacma baboon (*Papio ursinus*) grunts. *The Journal of the Acoustical Society of America*, *129*(3), 1631–1641. https://doi.org/10.1121/1.3531944

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

Morton, E. S. (1975). Ecological sources of selection on avian sounds. *The American Naturalist*, *109*(965), 17–34. https://doi.org/10.1086/282971

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855–869. https://doi.org/10.1086/283219

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, *1*(3), 261–289. https://doi.org/10.1007/BF02733986

Nesse, R. M. (2013). Tinbergen's four questions, organized: A response to Bateson and Laland. *Trends in Ecology & Evolution*, *28*, 681–682. https://doi.org/10.1016/j.tree.2013.10.008

Norcross, J. L., & Newman, J. D. (1993). Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls. *American Journal of Primatology*, *30*(1), 37–54. https://doi.org/10.1002/ajp.1350300104

Ordóñez-Gómez, J. D., Santillan-Doherty, A. M., & Hammerschmidt, K. (2019). Acoustic variation of spider monkey (*Ateles geoffroyi*) contact calls is related to caller isolation and affects listeners' responses. *PLOS ONE*, *14*(4), e0213914. https://doi.org/10.1371/journal.pone.0213914

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, *73*(6), 530–544. https://doi.org/10.1002/ajp.20913

Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocal signaling. In D. H. Owings, M. D. Beecher, & N. S. Thompson (Eds.), *Perspectives in ethology* (Vol. 12). New York: Penum Press.

Owren, M. J., & Rendall, D. (2001). Sound on the rebound: Bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology*, *10*(2), 58–71. https://doi.org/10.1002/evan.1014

Panksepp, J. (2005). Affective consciousness: Core emotional feelings in animals and humans. *Consciousness and Cognition*, *14*(1), 30–80. https://doi.org/10.1016/j.concog.2004.10.004

Panksepp, J. (2007). Neurologizing the psychology of affects: How appraisal-based constructivism and basic emotion theory can coexist. *Perspectives on Psychological Science*, *2*(3), 281–296. https://doi.org/10.1111/j.1745-6916.2007.00045.x

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: Do animals have affective lives? *Neuroscience and Biobehavioral Reviews*, *35*(9), 1791–1804. https://doi.org/10.1016/j.neubiorev.2011.08.003

Park, J. (2007). Distinguishing byproducts from non-adaptive effects of algorithmic adaptations. *Evolutionary Psychology*, *5*(1), 47–51.

Patel, E. R., & Owren, M. J. (2007). Acoustics and behavioral contexts of "gecker" vocalizations in young rhesus macaques (Macaca mulatta). *The Journal of the Acoustical Society of America*, *121*(1), 575–585. https://doi.org/10.1121/1.2390662

Paul, E. S., & Mendl, M. T. (2018). Animal emotion: Descriptive and prescriptive definitions and their implications for a comparative perspective. *Applied Animal Behaviour Science*, *205*, 202–209. https://doi.org/10.1016/j.applanim.2018.01.008

Pigliucci, M., & Kaplan, J. (2000). The fall and rise of Dr Pangloss: Adaptationism and the *Spandrels* paper 20 years later. *Trends in Ecology and Evolution*.

https://doi.org/10.1016/S0169-5347(99)01762-0

Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*. Washington, DC: American Psychological Association. Retrieved from https://psycnet.apa.org/record/2003-04005-000

Porges, S. W. (2003). The polyvagal theory: Phylogenetic contributions to social behavior. *Physiology & Behavior, 79*(3), 503-513. https://doi.org/10.1016/S0031-9384(03)00156-2

Preston, S. D., & de Waal, F. B. M. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences*, *25*(1), 1–20. https://doi.org/10.1017/S0140525X02000018

Price, Tabitha, Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, *5*, 1–11. https://doi.org/10.1038/srep13220

Price, Trevor, & Langen, T. (1992). Evolution of correlated characters. *Trends in Ecology & Evolution*, *7*(9), 307–310. https://doi.org/10.1016/0169-5347(92)90229-5

Puppe, B., Schön, P. C., Tuchscherer, A., & Manteuffel, G. (2005). Castration-induced vocalisation in domestic piglets, *Sus scrofa*: Complex and specific alterations of the vocal quality. *Applied Animal Behaviour Science*, *95*(1–2), 67–78. https://doi.org/10.1016/j.applanim.2005.05.001

Ralph, C. R., & Tilbrook, A. J. (2016). Invited Review: The usefulness of measuring glucocorticoids for assessing animal welfare. *Journal of Animal Science*. https://doi.org/10.2527/jas2015-9645

Rand, A. S., & Ryan, M. J. (1981). The adaptive significance of a complex vocal repertoire in a neotropical frog. *Zeitschrift Für Tierpsychologie*, *57*(3–4), 209–214.

https://doi.org/10.1111/j.1439-0310.1981.tb01923.x

Reale, R. A., & Imig, T. J. (1980). Tonotopic organization in auditory cortex of the cat. *The Journal of Comparative Neurology*, *192*(2), 265–291. https://doi.org/10.1002/cne.901920207

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America*, *113*(6), 3390–3402. https://doi.org/10.1121/1.1568942

Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *The Journal of the Acoustical Society of America*, *103*(1), 602–614. https://doi.org/10.1121/1.421104

Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour*, *78*(2), 233–240. https://doi.org/10.1016/j.anbehav.2009.06.007

Riede, T. (2010). Elasticity and stress relaxation of rhesus monkey (*Macaca mulatta*) vocal folds. *Journal of Experimental Biology*, *213*, 2924–2932. https://doi.org/10.1242/jeb.044404

Romani, G. L., Williamson, S. J., & Kaufman, L. (1982). Tonotopic organization of the human auditory cortex. *Science*, *216*(4552), 1339–1340. https://doi.org/10.1126/science.7079770

Romero, L. M. (2004). Physiological stress in ecology: Lessons from biomedical research. *Trends in Ecology & Evolution*, *19*(5), 249–255. https://doi.org/10.1016/J.TREE.2004.03.008

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

Ryan, M. J. (1990). Sexual selection, sensory systems and sensory exploitation. *Oxford Surveys* *in Evolutionary Biology*, *7*, 157–195.

Ryan, M. J., & Rand, A. S. (1993). Sexual selection and signal evolution: The ghost of biases past. *Philosophical Transactions - Royal Society of London, B*, *340*(1292), 187–195. https://doi.org/10.1098/rstb.1993.0057

Ryan, M. J., & Rand, A. S. (2003). Mate recognition in túngara frogs: A review of some studies of brain, behavior, and evolution. *Acta Zoological Sinica*, *49*(6), 713–726.

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Sauter, D. A., Eisner, F., Ekman, P., & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408–2412. https://doi.org/10.1073/pnas.0908239106

Scarantino, A. (2012). How to define emotions scientifically. *Emotion Review*, *4*(4), 358–368. https://doi.org/10.1177/1754073912445810

Scarantino, A. (2018). Are LeDoux's survival circuits basic emotions under a different name? *Current Opinion in Behavioral Sciences*, *24*, 75-82. https://doi.org/10.1016/j.cobeha.2018.06.001

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review*, *3*(4), 444–454. https://doi.org/10.1177/1754073911410745

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: A new take on ultimate and proximate influences in vocal production. *Biological Reviews*, *93*(4), 2071–2082. https://doi.org/10.1111/brv.12437

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013).

Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE*, *8*(10), e76674. https://doi.org/10.1371/journal.pone.0076674

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Schlotz, W., Kumsta, R., Layes, I., Entringer, S., Jones, A., & Wüst, S. (2008). Covariance between psychological and endocrine responses to pharmacological challenge and psychosocial stress: A question of timing. *Psychosomatic Medicine*, *70*(7), 787–796. https://doi.org/10.1097/PSY.0b013e3181810658

Schrader, L., & Todt, D. (1998). Vocal quality is correlated with levels of stress hormones in domestic pigs. *Ethology*, *104*(10), 859–876. https://doi.org/10.1111/j.1439-0310.1998.tb00036.x

Schwartz, J. W., & Gouzoules, H. (2019). Decoding human screams: Perception of emotional arousal from pitch and duration. *Behaviour*, *156*(13–14), 1283–1307. https://doi.org/10.1163/1568539X-00003566

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (in press). Evolving views on cognition in animal communication: Contributions from scream research. *Animal Behavior and Cognition.*

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science*, *6*(1), 38–47. https://doi.org/10.1177/1745691610393528

Sèbe, F., Poindron, P., Andanson, S., Chandeze, H., Delval, E., Despres, G., … Boissy, A.

(2012). Bio-acoustic analyses to assess emotion in animals: Acoustic patterns are linked to behavioural, cardiac and hormonal responses of ewes to the separation from their lambs. *Bioacoustics*, *21*(1), 54.

Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, *54*(1), 145–173. https://doi.org/10.1146/annurev.psych.54.101601.145121

Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language*, *115*(1), 92–100. https://doi.org/10.1016/j.bandl.2009.10.003

Seyfarth, R. M., & Cheney, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour*, *124*, 339–346. https://doi.org/10.1016/j.anbehav.2016.05.020

Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K., & Hammerschmidt, K. (2010). The central importance of information in studies of animal communication. *Animal Behaviour*, *80*(1), 3–8. https://doi.org/10.1016/j.anbehav.2010.04.012

Shanahan, T. (2011). Phylogenetic inertia and Darwin’s higher law. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *42*(1), 60–68. https://doi.org/10.1016/J.SHPSC.2010.11.013

Sharma, N., Prakash S, V., Kohshima, S., & Sukumar, R. (2020). Asian elephants modulate their vocalizations when disturbed. *Animal Behaviour*, *160*, 99–111. https://doi.org/10.1016/j.anbehav.2019.12.004

Sherman, P. W. (1988). The levels of analysis. *Animal Behaviour*, *36*(2), 616–619. https://doi.org/10.1016/S0003-3472(88)80039-3

Siebert, K., Langbein, J., Schön, P. C., Tuchscherer, A., & Puppe, B. (2011). Degree of social isolation affects behavioural and vocal response patterns in dwarf goats (*Capra hircus*). *Applied Animal Behaviour Science*, *131*(1–2), 53–62. https://doi.org/10.1016/j.applanim.2011.01.003

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104*(43), 17228–17233. https://doi.org/10.1073/pnas.0706741104

Stoeger, A. S., Baotic, A., Li, D., & Charlton, B. D. (2012). Acoustic features indicate arousal in infant giant panda vocalisations. *Ethology*, *118*(9), 896–905. https://doi.org/10.1111/j.1439-0310.2012.02080.x

Stoeger, A. S., Charlton, B. D., Kratochvil, H., & Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America*, *130*(3), 1700–1710. https://doi.org/10.1121/1.3605538

Stomp, M., Leroux, M., Cellier, M., Henry, S., Hausberger, M., & Lemasson, A. (2018). Snort acoustic structure codes for positive emotions in horses. *The Science of Nature*, *105*(9–10), 57. https://doi.org/10.1007/s00114-018-1582-9

Sugiura, H. (2007). Effects of proximity and behavioral context on acoustic variation in the coo calls of Japanese macaques. *American Journal of Primatology*, *69*, 1412–1424. https://doi.org/10.1002/ajp.20447

Szipl, G., Ringler, E., Spreafico, M., & Bugnyar, T. (2017). Calls during agonistic interactions vary with arousal and raise audience attention in ravens. *Frontiers in Zoology*, *14*(1), 1–13. https://doi.org/10.1186/s12983-017-0244-7

Taylor, A. M., & Reby, D. (2010). The contribution of source-filter theory to mammal vocal communication research. *Journal of Zoology*, *280*(3), 221–236.

https://doi.org/10.1111/j.1469-7998.2009.00661.x

Taylor, A. M., Reby, D., & McComb, K. (2009). Context-related variation in the vocal growling behaviour of the domestic dog (*Canis familiaris*). *Ethology*, *115*(10), 905–915. https://doi.org/10.1111/j.1439-0310.2009.01681.x

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie*, *20*, 410–433.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Turner, A. I., Rivalland, E. T. A., Clarke, I. J., & Tilbrook, A. J. (2010). Stressor specificity of sex differences in hypothalamo-pituitary-adrenal axis activity: cortisol responses to exercise, endotoxin, wetting, and isolation/retraint stress in gonadectomized male and female sheep. *Endocrinology*, *151*(9), 4324–4331. https://doi.org/10.1210/en.2010-0234

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires - Hard vs. soft classification approaches. *PLoS ONE*, *10*(4), e0125785. https://doi.org/10.1371/journal.pone.0125785

Waser, P. M., & Waser, M. S. (1977). Experimental studies of primate vocalization: Specializations for long-distance propagation. *Zeitschrift Für Tierpsychologie*, *43*(3), 239–263. https://doi.org/10.1111/j.1439-0310.1977.tb00073.x

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology*, *21*(5), 195–205. https://doi.org/10.1002/evan.21319

Wheeler, B. C., & Fischer, J. (2015). The blurred boundaries of functional reference: A response to Scarantino & Clay. *Animal Behaviour*, *100*, 9–13. https://doi.org/10.1016/j.anbehav.2014.11.007

Wiley, R. H., & Richards, D. G. (1978). Physical constraints on acoustic communication in the atmosphere: Implications for the evolution of animal vocalizations. *Behavioral Ecology and Sociobiology*, *3*(1), 69–94. https://doi.org/10.1007/BF00300047

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press. https://doi.org/10.1086/332352

Yamaguchi, C., Izumi, A., & Nakamura, K. (2010). Time course of vocal modulation during isolation in common marmosets (*Callithrix jacchus*). *American Journal of Primatology*, *72*(8), 681–688. https://doi.org/10.1002/ajp.20824

Yeon, S. C., Kim, Y. K., Park, S. J., Lee, S. S., Lee, S. Y., Suh, E. H., … Lee, H. J. (2011). Differences between vocalization evoked by social stimuli in feral cats and house cats. *Behavioural Processes*, *87*(2), 183–189. https://doi.org/10.1016/j.beproc.2011.03.003

Zhao, L., Rad, B. B., & Wang, X. (2019). Long-lasting vocal plasticity in adult marmoset monkeys. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1905), 20190817. https://doi.org/10.1098/rspb.2019.0817

Zimmermann, E., Leliveld, L. M. C., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: A comparative approach to mammalian voices. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of emotional communication: From sounds in nonhuman mammals to speech and music in men* (pp. 117–132). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199583560.001.0001

Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, *113*(1), 33–42. https://doi.org/10.1037/0735-7036.113.1.33

CHAPTER II: EXPRESSION OF EMOTIONAL AROUSAL ACROSS TWO VOCALIZATIONS TYPES IN RHESUS MACAQUES (*MACACA MULATTA*): INTEGRATING PROXIMATE AND ULTIMATE PERSPECTIVES

Abstract

The study of emotional communication in animals goes back to Charles Darwin and has the potential to improve scientific understanding of the mechanisms of signal production as well as how signals evolve. We examined the relationships between emotional arousal and selected acoustic characteristics of coo and scream vocalizations by juvenile female rhesus macaques (*Macaca mulatta*). Arousal was assessed, in the analysis of coos, through measures of increased concentrations of stress hormones exhibited in response to a stress test (the Human Intruder Test), and, in the analysis of screams, from the intensity of agonism experienced by the vocalizer during natural social interactions. Both vocalization types showed a positive relationship between emotional arousal and overall fundamental frequency (F0) of the voice. Among coos, greater arousal was additionally associated with wider F0 modulation and shorter duration. Coos showed decreases in certain F0 parameters and noisiness, and increases in jitter and duration, from 6 months of age, to 16 months, to menarche (21.3-31.3 months); additionally, the correlations between stress hormone increases and some acoustic properties of coos were dampened as monkeys aged. Our results regarding the acoustic correlates of emotional arousal are broadly consistent with findings in other species, supporting Darwin’s hypothesis of evolutionary continuity in emotion expression. They also illustrate how understanding the proximate-level link between emotional arousal and vocal acoustics can elucidate general processes in the evolution of vocal communication.

Introduction

The study of animal vocal communication holds the potential for far-reaching discoveries, ranging from cognition, to general evolutionary processes and principles, to the evolutionary history of human language. After decades of research and much debate in this field, authors agree that emotion generally plays an important role in vocal production (Darwin, 1872; Fischer, 2017; Fischer & Price, 2016; Morton, 1977; Owren, Amoss, & Rendall, 2011; Scarantino & Clay, 2015; Schamberg, Wittig, & Crockford, 2018; Seyfarth & Cheney, 2003a, 2003b; Wheeler & Fischer, 2015, 2012). Yet the precise manner in which emotion influences the voice has received relatively little empirical attention in primate research (though see Fichtel, Hammerschmidt, & Jürgens, 2001; Lemasson, Remeuf, Rossard, & Zimmermann, 2012; Liao, Zhang, Cai, & Ghazanfar, 2018; Rendall, 2003; Schehka & Zimmermann, 2009); instead, many of the advances in this area have taken place in research with domesticated animals (Baciadonna, Briefer, Favaro, & McElligott, 2019; Briefer, 2012; Briefer, Linhart, et al., 2019; Briefer, Maigrot, et al., 2015; Briefer, Tettamanti, & McElligott, 2015; Briefer, Vizier, Gygax, & Hillmann, 2019; Faragó, Pongrácz, Range, Virányi, & Miklósi, 2010; Faragó, Takács, Miklósi, & Pongrácz, 2017; Leliveld, Düpjan, Tuchscherer, & Puppe, 2017; Maigrot, Hillmann, & Briefer, 2018). Recently, prominent authors in the field of primate vocal communication have called for more research investigating the mechanisms of call production, and general processes in the evolution of primate vocal communication systems (Fischer, 2017; Fischer & Price, 2016; Owren et al., 2011; Seyfarth & Cheney, 2017; Wheeler & Fischer, 2012, 2015). Research into the role of emotion in vocal production is key to both these goals.

Discussing the relationships between emotion and behavior requires first defining “emotion” (Paul & Mendl, 2018). Here, we use the term to refer to a temporary, internal state defined by its valence (from negative, e.g., fear, to positive, e.g., joy), and its level of arousal, a framework that comes from the Dimensional (or Core Affect) view of emotions (e.g., Mendl, Burman, & Paul, 2010; Russell, 2003). This approach differs from the view of emotions as discrete states (but is not incompatible with this view; see Scarantino & Griffiths, 2011), and from higher-order definitions centered on consciously felt experiences (e.g., LeDoux & Brown, 2017). Arousal represents a spectrum from a sleepy, calm, or tranquil state to a highly alert or excited state. Emotional arousal comprises a suite of physiological processes that generally (but imperfectly) co-occur (see Scarantino & Griffiths, 2011) as part of a broader emotional reaction to arousing stimuli in both positive and negative contexts. These processes include heightened activation of the sympathetic nervous system resulting in increased heart rate, respiration, and tensing of muscles throughout the body, and activation of the hypothalamic-pituitary-adrenal (HPA) axis resulting the production and release of glucocorticoid hormones (Ralph & Tilbrook, 2016; Romero, 2004). Thus HPA activation is an aspect of, not equal to, emotional arousal, but the former is used as an indicator of the latter (e.g., Bayart, Hayashi, Faull, Barchas, & Levine, 1990; Blumstein & Chi, 2012; Schrader & Todt, 1998; Sèbe et al., 2012).

From a proximate perspective, the effects of emotion on vocal production are thought to largely result from physiological processes, such as arousal, acting on the vocal apparatus (Briefer, 2012; Scherer, 1986; Zei Pollermann & Archinard, 2002). One of the best-established acoustic correlates of arousal is vocal fundamental frequency (F0; determined by the rate of oscillation of the vocal folds and commonly perceived as pitch by humans) with increases in arousal generally associated with increases in mean F0 (Briefer, Maigrot, et al., 2015; Briefer, Tettamanti, et al., 2015; Collins, McGreevy, Wheatley, & Harcourt, 2011; Filippi et al., 2017; Meise, Keller, Cowlishaw, & Fischer, 2011; Rendall, 2003; Scherer, 2003; Szipl, Ringler, Spreafico, & Bugnyar, 2017; Yeon et al., 2011; for a possible exception, see Ordóñez-Gómez, Santillan-Doherty, & Hammerschmidt, 2019), as well as F0 modulation, e.g., the F0 range or slope (Lingle, Wyman, Kotrba, Teichroeb, & Romanow, 2012; Norcross & Newman, 1993; Rendall, 2003; Sugiura, 2007; Yamaguchi, Izumi, & Nakamura, 2010). Arousal might achieve these effects on F0 by increasing tension and action in the diaphragm and/or laryngeal muscles, resulting in faster vibration of the vocal folds and greater changes in vibration frequency (Briefer, 2012; Riede, 2010; Scherer, 1986; Titze, 1994). Effects of arousal on the acoustic structure of vocalizations also include changes in vocal duration (Filippi, Hoeschele, Spierings, & Bowling, 2019; Scherer, 1986; though see Briefer, 2012; Briefer, Maigrot, et al., 2015; Briefer, Vizier, et al., 2019; Friel, Kunc, Griffin, Asher, & Collins, 2019), upward shifts in the distribution of energy across the frequency spectrum (Briefer, 2012; Briefer, Maigrot, et al., 2015; Gogoleva, Volodina, Volodin, Kharlamova, & Trut, 2010; Yeon et al., 2011), as well as increased noisiness (Liao et al., 2018; Meise et al., 2011; Puppe, Schön, Tuchscherer, & Manteuffel, 2005; Siebert, Langbein, Schön, Tuchscherer, & Puppe, 2011; Stoeger, Charlton, Kratochvil, & Fitch, 2011; though see Blumstein & Chi, 2012; Linhart, Ratcliffe, Reby, & Špinka, 2015; Szipl et al., 2017).

Although this literature includes both studies comparing different call types and those looking at acoustic variation within a single call type, arousal appears to relate more strongly to acoustic variation within a call type than to call usage (i.e., whether a call, and which call type, occurs) (Schamberg et al., 2018; Schwartz, Engelberg, & Gouzoules, in press). Senders use calls in a goal-directed fashion (Schamberg et al., 2018; Schel, Townsend, Machanda, Zuberbühler, & Slocombe, 2013) and in some cases base calling on the composition or knowledge state of the audience (Crockford, Wittig, Mundry, & Zuberbühler, 2012; Fichtel & Manser, 2010), suggesting call usage involves not only emotion but also other cognitive processes, e.g., executive function. In contrast, the precise acoustic structure of a vocalization (i.e., within-type variation) might generally be more strongly influenced by the effects of physiological components of emotions (especially arousal) on the mechanisms of vocal production (Briefer, 2012; Schamberg et al., 2018; Schwartz et al., in press)

From an ultimate perspective, research into the role of emotion in vocal production can potentially elucidate general processes in the evolution of communication. Emotions are generally thought to have evolved to potentiate coordinated behavioral responses in contexts in which they are adaptive (Adolphs & Andler, 2018; Anderson & Adolphs, 2014; Darwin, 1872; de Waal, 2011; Mendl et al., 2010; Nesse, 1990; Panksepp, 2011). Facilitating this function, broadly similar emotional states can be evoked by a range of emotionally salient contexts or stimuli, and can potentiate, motivate, or manifest in a range of behavioral and physiological outputs (Anderson & Adolphs, 2014; de Waal, 2011); thus emotion is described as exhibiting a “fan-in, fan-out architecture” (Adolphs & Andler, 2018; Anderson & Adolphs, 2014). While this architecture is critical to the hypothesized functionality of emotion, it also might constrain or otherwise complicate the evolution of vocal and other behaviors.

To illustrate this principle, consider the effects of emotional arousal: because arousal potentiates (or “fans out” to) multiple correlated vocal, physiological, and behavioral outputs (e.g., increased mean F0 of vocalizations, mobilization of energy stores, increased vigilance), one might hypothesize that a selection pressure favoring just one of these outputs within a particular context might result in the evolution of a stronger emotional arousal response to relevant stimuli, and this might incidentally cause the other outputs to simultaneously be expressed. For example, a juvenile social mammal separated from the group might benefit from the increased vigilance that comes with greater arousal; consequent evolutionary change in the arousal reaction to separation might, then, result in changes in the acoustic properties of vocalizations (Lingle et al., 2012). If these acoustic changes enhanced communicative function, they might be favored by selection, resulting in signal evolution. If, however, these acoustic changes were evolutionarily neutral, they might persist over evolutionary time, even if hypothetically they carried no evolutionary benefit to communicative function. It is even possible to imagine potentially conflicting selection pressures acting on arousal's effects on behavior, physiology, and vocal acoustics. These kinds of evolutionary dynamics are only plausible to the extent that the acoustic and behavioral outputs of emotions are linked to emotional arousal and to one another. One approach to gaining insight into the evolution of vocal acoustics is, therefore, to test the degree to which acoustic correlates of arousal are consistent not only across species but also across vocalization types and contexts within a species.

Overall, although scientific understanding of the vocal expression of emotion is growing, many taxa remain unstudied, and few studies have compared multiple vocalization types within a species (Bastian & Schmidt, 2008; Briefer, Linhart, et al., 2019; Gogoleva et al., 2010; Linhart et al., 2015; Yeon et al., 2011), resulting in limited knowledge of the degree of consistency in vocal emotion expression. The aim of the present study was to compare the acoustic correlates of emotional arousal across two vocalization types in juvenile female rhesus macaques (*Macaca mulatta*)—coos and screams—in order to contribute toward ongoing efforts to test the hypothesis that the effects of emotional arousal on vocal production are generally consistent across vocalization types, contexts, and species (Briefer, 2012; Filippi et al., 2017). Coos, a tonal call with a soft chevron-shaped F0 contour ranging from 400-3000 Hz, are emitted by rhesus macaques as well as several other macaque species in a variety of contexts including anticipation of food, affiliative social interaction, and mother-infant separation, and are thought to function to stimulate friendly contact and/or convey the location of the sender (Bayart et al., 1990; de Waal & Bauers, 1991; Green, 1975; Hansen, 1976; Hauser, 1991; Hauser & Marler, 1993; Simons & Bielert, 1973). Rhesus macaque screams are high-F0 (3-10 kHz) vocalizations; like in other as other catarrhine primates, macaque screams are generally uttered during agonistic interactions and function to recruit aid, usually from kin (de Waal & Van Hooff, 1981; H. Gouzoules, 2005; H. Gouzoules & Gouzoules, 1995, 2000; H. Gouzoules, Gouzoules, & Marler, 1985; H. Gouzoules, Gouzoules, & Tomaszycki, 1998; S. Gouzoules, Gouzoules, & Marler, 1984).

The vocal behavior of rhesus macaques has been well studied (Bayart et al., 1990; Fitch, de Boer, Mathur, & Ghazanfar, 2016; H. Gouzoules & Gouzoules, 2000; S. Gouzoules et al., 1984; Hammerschmidt, Newman, & Champoux, 2000; Hauser, 1991; Hauser, Evans, & Marler, 1993; Hauser & Marler, 1993; Jovanovic & Gouzoules, 2001; Jovanovic & Maestripieri, 2010; Le Prell, Hauser, & Moody, 2002; Maestripieri, Jovanovic, & Gouzoules, 2000; Patel & Owren, 2007; Pfefferle et al., 2016; Rendall, Owren, & Rodman, 1998; Riede, 2010; Rowell & Hinde, 1962; Tomaszycki, Davis, Gouzoules, & Wallen, 2001). Classic studies of vocal communication in macaques hypothesized that call usage and acoustic variation might be attributable to the emotional states of vocalizers (e.g., Green, 1975; Rowell & Hinde, 1962), yet few studies have directly investigated the effects of emotion on vocal production and acoustics. Bayart et al. (1990) collected vocal recordings and hormone measurements from infant rhesus macaques kept in total or partial isolation from their mothers, and analyzed the infants' coo calls. Coos emitted in total isolation were qualitatively described as showing increased F0 modulation and duration; isolation was also associated greater plasma concentrations of cortisol, suggesting that these acoustic changes might have resulted from greater arousal. Japanese macaques (*Macaca fuscata*) have shown a similar correlation between social isolation and F0 modulation of coos (Sugiura, 2007). Increased vocal effort associated with arousal has also been invoked to explain acoustic changes to infant rhesus macaques' gecker vocalizations in more intense contexts, including increased amplitude and number of sound pulses (Patel & Owren, 2007).

Research into rhesus macaque scream usage provides additional insight on of the possible role of emotion in vocal production. S. Gouzoules et al. (1984) showed that screams of this species cluster into acoustically distinct scream classes correlating with the rank of the opponent and the degree of intensity of aggression received (i.e. whether contact was involved) (see also H. Gouzoules, 2005; H. Gouzoules & Gouzoules, 1995). The idea that these scream classes are meaningfully different from one another has been challenged in part on the grounds that they exhibit acoustic gradation (Le Prell et al., 2002), but acoustic gradation is not inconsistent with the idea of functionally distinct vocalization classes (Fischer, Wadewitz, & Hammerschmidt, 2016; Price et al., 2015; Wadewitz et al., 2015). To the extent that contextual factors such as the nature of an agonistic interaction has some effect on the emotional state of a monkey, acoustic differences between scream classes emitted in different agonistic situations could potentially be attributable to emotion. For example, noisy screams occur more commonly with physical aggression than tonal screams (S. Gouzoules et al., 1984); it might be that physical aggression evokes greater arousal in the screamer, leading to increased muscle tension in the vocal apparatus, and hence more turbulent airflow and decreased harmonicity in noisy screams relative to tonal screams (see Fitch, Neubauer, & Herzel, 2002). In line with this idea, Jovanovic & Gouzoules (2001) and Maestripieri et al. (2000) showed that infant rhesus macaques emit more noisy vocalizations in response to abusive and/or prolonged maternal restraint, and Bercovitch, Hauser, & Jones (1995) demonstrated that stressed macaques emit more noisy alarm calls.

These studies all focused on patterns of vocal usage (see Seyfarth & Cheney, 2010), i.e., which scream classes occur most frequently in different contexts. As noted earlier, acoustic variation within vocalization classes probably represents a purer indicator of emotional arousal than call usage—decisions about whether to vocalize and which vocalization class to use can involve not only emotion but also other cognitive processes (Schamberg et al., 2018; Schwartz et al., in press)—yet studies of the contextual and psychological factors influencing acoustic variation within scream classes are lacking. Research into how arousal affects acoustics within scream classes has the potential to determine whether changes in arousal might be responsible for pushing scream acoustics along the gradient from one scream class into another. In the present study, we examined the relationship between arousal and the acoustics of coos and screams in juvenile female rhesus macaques. Ideally, comparing the effects of arousal on vocal acoustics across vocalization types and contexts would entail examining multiple vocalization types *each* within multiple different contexts, but doing so is challenging given that the usage of different vocalization types tends to correlate roughly with context. We instead recorded coos and screams occurring in two separate contexts.

Coos were recorded during Human Intruder Tests (HIT) (Kalin & Shelton, 1989) preceded and followed by cortisol measurements; the increase in cortisol from the first to the second measurement serves as a proxy for the level of HPA activation evoked by the test, and thus the emotional arousal response to it. The HIT entailed three stages characterized by escalating interactions with an unfamiliar human. We treated the escalating intensity of the stages as an additional source of variation in monkeys’ arousal levels. Screams were recorded during natural social interactions; the intensity of agonism experienced by a screamer served as a proxy for emotional arousal. Relationships between coo acoustics and cortisol concentrations have been reported in the past, though in a qualitative rather than quantitative way (Bayart et al., 1990). Differences in scream acoustics associated with differing degrees of aggression have likewise been reported previously (S. Gouzoules et al., 1984), but not while controlling for acoustic variation between scream classes. In accordance with the hypothesis that the acoustic correlates of arousal are shared across taxa (Briefer, 2012; Filippi et al., 2017; Zimmermann, Leliveld, & Schehka, 2013), we predicted that these vocalization types would each exhibit arousal-related acoustic changes represented in the literature including increases in F0, F0 modulation, noisiness, and peak frequency, and upward shifts in energy distribution.

In addition to arousal, we examined the effect of juveniles' age on the acoustics of both coos and screams. Previous research has shown decreases in coo F0 and F0 modulation, and increases in coo duration, in this species over the first several weeks of development (Hammerschmidt et al., 2000). We predicted that we would observe similar trends over the first few years of subjects' lives, especially decreases in F0 due to growth of the body and larynx (e.g., Fitch & Hauser, 2006). In the closely related pigtail macaque (*Macaca nemestrina*), screams become more acoustically stereotyped to agonistic context over development (H. Gouzoules & Gouzoules, 1989). We explored how developmental changes would compare or contrast between coos and screams.

Methods

Subjects and Housing

Forty-three female rhesus macaques served as the sample for this study. Subjects were housed in three outdoor enclosures, each measuring three-quarter acres with an attached indoor area, at the Yerkes National Primate Research Center (YNPRC) Field Station in Lawrenceville, GA. Each enclosure housed a social group composed of 1-2 adult males, 40-60 adult females with a linear matrilineal dominance hierarchy, and their immature offspring. Subjects represented a wide range of matrilineal ranks. Subjects had *ad libitum* access to water and low-fat, high-fiber pellets (Purina Mills Int., Lab Diets, St. Louis, MO, USA), and were provisioned with fresh fruit and vegetables daily. Some subjects had additional *ad libitum* access to high-calorie pellets (D14051502B, Research Diets, Inc., New Brunswick, NJ, USA) as part of concurrent research into diet and social behavior.

Ethical Statement

The Emory University Institutional Animal Care and Use Committee approved all procedures in accordance with the Animal Welfare Act and the U.S. Department of Health and Human Services "Guide for Care and Use of Laboratory Animals." Human Intruder tests and blood draws took place as part of the Stress, Obesity, and Diabetes research project (Principle Investigators: M. M. Sanchez & M. E. Wilson) at the YNPRC, independently of this investigation of vocal emotion expression.

Coo Recording, Preparation, and Selection

Data collection for coos took place during a total of 111 HIT conducted from October, 2014 through December, 2017. The HIT elicits HPA activation resulting in increases in plasma cortisol, as well as cooing and other behaviors which vary consistently between individuals depending in part on temperament and other characteristics (Gottlieb & Capitanio, 2013; Hamel et al., 2017; Kalin, Larson, Shelton, & Davidson, 1998; Kalin, Shelton, Rickman, & Davidson, 1998). Subjects underwent up to three tests, at ages 6 months, 16 months, and menarche; 4 of the 43 subjects were removed from the project prior to 16 months of age, and an additional 10 were removed prior to menarche, and therefore only underwent the first one or two tests respectively, leaving 29 subjects who completed all three tests. To determine when subjects reached menarche, pubertal sexual swellings were monitored and, once mature, menstrual status was checked twice per week using vaginal swabs. The menarche test occurred at ages ranging from 21.3-31.3 months. At each age point, at sunrise (to control for circadian variation in cortisol), and following training, the subject was prompted to enter the indoor area, using procedures to which subjects were habituated (described in McCormack, Newman, Higley, Maestripieri, & Sanchez, 2009). The subject was then transferred into a cage, and 1 mL of blood was drawn from the saphenous vein (unanesthetized) within 10 minutes from initial group disturbance (to minimize effects of these procedures on baseline cortisol hormone levels), collected in pre-chilled tubes containing EDTA. This sample served as the baseline for analysis of stress-induced cortisol increase at the end of the HIT (see below).

The subject was then transported to a cage in a 10x10-ft testing room. The cage was large enough for the subject to move about freely, and had one plexiglass wall facing a Sony DCR-SR85 video camera (Tokyo, Japan). The room also contained a HoMedics SS-2000 noise generator (Commerce Township, MI, USA) set to the “waterfall” setting (brown noise) at high volume, to create a controlled and constant auditory background for testing. The transfer of the subject into the cage marked the beginning of the HIT; the video camera and microphone began recording the subject’s behaviors and vocalizations, all researchers promptly left the room, and a stopwatch began keeping track of elapsed time.

The HIT (Kalin & Shelton, 1989) involved three conditions each lasting 10 minutes: Alone, Profile, and Stare. When exactly 10 min had passed with the subject alone in the testing room (Alone stage), a researcher wearing a latex mask of a human faces (one for each of the three age points; serving to avoid any effects of individual recognition, since subjects interacted with researchers regularly) entered the room and sat 3 meters from the subject and presenting the face profile, moving as little as possible for the next 10 min (Profile stage). After 10 min, the human shifted to face the subject, maintaining a direct gaze at the subject’s eyes, but did not move in any other way for the next 10 min (Stare stage). The end of the Stare stage marked the end of the 30-min test, at which point another 1-mL blood sample was collected and placed in a pre-chilled EDTA tube on ice. The increase in cortisol levels present in this blood sample from baseline levels served as a relative measure of physiological stress reactivity in response to the HIT.

Blood samples were immediately centrifuged, and plasma was stored at -80°C until assayed. Cortisol assays were performed by the Biomarker Core Laboratory at the YNRPC. After thawing to 4° C, 20 µl of plasma were pipetted into 100 µl of 50 ng/ml internal standard (d4-cortisol) in acetonitrile in a 96-well block, and the mixture was spun and extracted. Cortisol and d4-cortisol were eluted using an Acquity UPLC BEH C18 column (Waters, Milford, MA, USA), and measured using ultra-performance liquid chromatography electrospray ionization tandem mass spectrometry (SCIEX Triple Quad 6500, Framingham, MA, USA).

Cortisol concentrations for each sample were calculated using linear regression analysis of a standard curve.

Video and audio recordings of HITs were obtained with the video camera (Sony DCR-SR85, Tokyo, Japan). Raw audio was extracted from the video and converted to a mono, 44.1-kHz, 24-bit .WAV format using Adobe Audition 3.0 (Adobe Systems, San Jose, CA, USA). To mitigate effects of the noise machine present in the testing room, prior to coo extraction, audio was prepared using the "Noise Reduction (process)" function in Audition: a noise profile was generated from a 5-s sample of uninterupted noise occurring immediately after a coo (the closest coo to the middle of the HIT), and then the whole sound file underwent reduction of noise fitting that profile. The function was set to reduce noise by 50% and by 20 dB, resulting in substantial reduction but not complete elimination of background noise; more extreme noise reduction interfered with the spectral structure of the coos. Although noise reduction algorithms are best avoided prior to acoustic analyses where possible, the high intensity level and acoustic homogeneity of background noise generated by the noise machine made noise reduction both necessary and effective.

A spectrogram of a coo is visible in Figure 1.1. Thirty-nine out of 43 subjects emitted coos, during 84 of the 111 HITs. The number of coos emitted in each stage of the HIT (Alone, Profile, Stare) was measured. Up to five coos were randomly selected from each of the three stages of each HIT. If fewer than five coos were emitted during a single stage, then all coos from that stage were isolated. In no case were two coos emitted within an inter-call interval 1 s of one another both selected/isolated. Eight hundred fifty-five coos were isolated in total.

Scream Recording, Preparation, and Selection

Data collection for screams took place over 111 hours from September, 2016 through December, 2017, at an enclosure containing a subset of 18 of the 43 subjects, and was limited to these 18 subjects. Ages associated with recorded screams ranged from 5.4-40.0 months. Standing 3 m away from the fence marking the edge of the enclosure, JWS or a trained assistant recorded all sounds emanating from the enclosure and surrounding area, using a Sennheiser ME66 directional microphone with MZW66 windshield (Wennebostel, Germany) and a Marantz PMD671 solid state recorder (Tokyo, Japan), manually adjusting the microphone sensitivity as needed to maximize the clarity of monkey vocalizations and minimize extraneous noise. Recordings were digitized at 44.1 kHz as 16- or 24-bit .WAV files. Screams and accompanying behavioral observations were collected using an all-occurrences sampling procedure (Altmann, 1974): each time one or more screams were emitted by one of the 18 subjects, the recording was accompanied by a verbal note that included the identity of the screamer and the nature of the eliciting behavior (Table 2.1). Occasionally a bout of screaming instigated by a threat was followed minutes later by a spontaneous (i.e., not elicited by further agonistic behavior) scream bout by the same subject. In these cases, the subsequent scream bout was not sampled, since the associated arousal level was especially uncertain (the subject might or might not still be highly aroused by the previous agonistic interaction). The assistant achieved reliability with JWS prior to collecting data (32/33 correct identifying screamer, 17/20 correct identifying behavior). In total, we recorded 224 scream bouts with positive IDs from among the 18 subjects; 181 of these were accompanied by successful behavioral observations. Subsequent analyses were limited to these 181 bouts.

Table 2.1. Categories of agonism accompanying screams, along with numbers of incidents accompanied by each behavior type analyzed vs. recorded during the data collection period.

Intensity	**Behavior**	**Description**	**N** analyzed (observed)
1	**No threat**	Subject received no visible agonistic behavior	24 (33)
2	**Non-contact threat**	Subject received a visible threat (e.g., forward lunge, dominant facial expression) by another individual, no physical contact	64 (81)
3	**Threat with contact**	Subject received a visible threat including physical contact, often grabbing	36 (44)

4	**Chase**	Subject ran from and was pursued by another individual for at least 2 s	13 (15)
	Attack	Subject was physically pinned or restrained by another individual for at least 2 s, often involving biting	7 (8)

To count the numbers of screams in each scream bout, we examined recordings aurally and visually, inspecting spectrograms, using Adobe Audition. We classified each scream into a subclass (following S. Gouzoules et al., 1984). Tonal screams and pulsed screams occurred much more often than other subclasses; analyses were therefore limited to these two subclasses. Tonal screams have a clearly identifiable F0 contour, and a duration lasting 0.25-1 s; pulsed screams are distinguished by their short duration (< 0.25 s), and range in tonality from noisy and cough-like to tonal and chirp-like (Figure 2.1). One hundred forty-four scream bouts contained at least 1 tonal and/or 1 pulsed scream that was suitable for acoustic analyses (no clipping, uninterrupted by other vocalizations or loud sounds, e.g., a monkey striking an enrichment structure during a display). Up to 2 screams of each subclass (tonal and pulsed) were selected per bout—on the basis of recording quality or, if more than 2 screams of a subclass were of comparable recording quality, then randomly—resulting in 362 screams (178 tonal screams and 184 pulsed screams) sampled. The maximum time elapsed between the onset of screaming (usually coincident with the eliciting threat) and a sampled scream was 26 s.

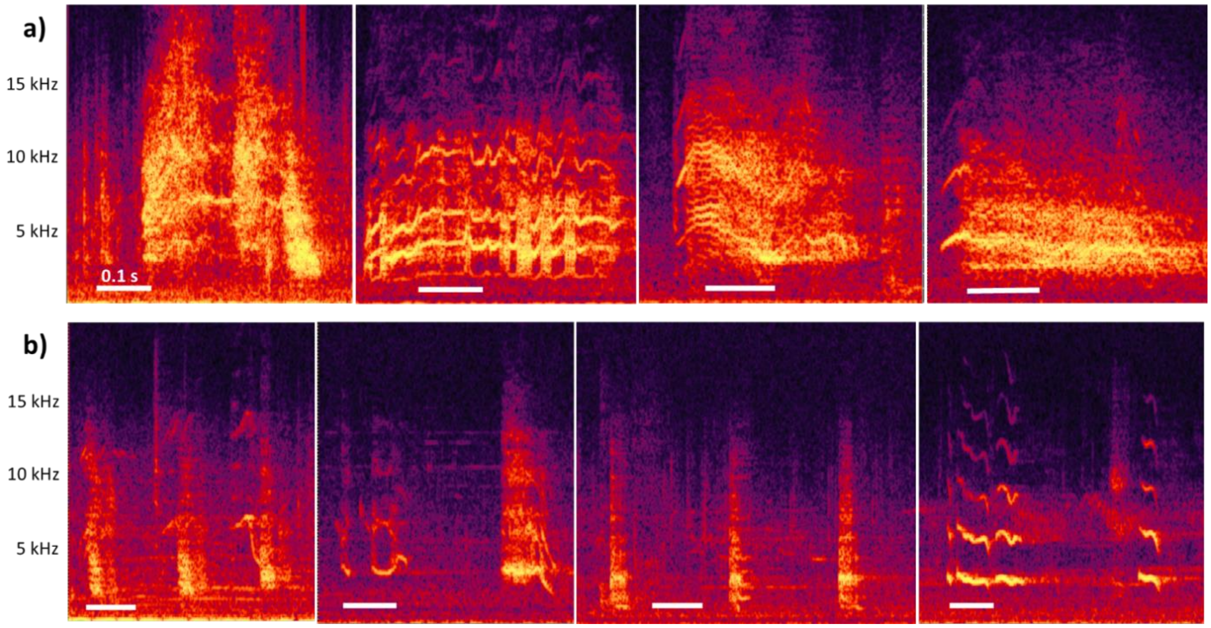

Figure 2.1. Representative spectrograms of (a) tonal and (b) pulsed screams

Acoustic Analyses

The selected vocalizations were isolated from the larger recording and saved as individual .WAV files, using Adobe Audition. The background noise, noise reduction procedure, and microphone quality (built-in microphone of digital video camera) of the HIT recordings precluded valid measurements of amplitude or fine spectral features of coos (e.g., peak frequency, energy distribution). However, the coo recordings were well suited to analysis of duration and F0 contour. In accordance with the recommended practice of adopting a hypothesis-testing (rather than an exploratory) approach in acoustic analyses (Fischer, Noser, & Hammerschmidt, 2013), we analyzed 11 parameters that have been indicated in the literature as a correlate of emotional arousal or valence in other species (Table 2.2).

Table 2.2. Acoustic parameters obtained from coos, and selected references indicating a relationship with emotion

Parameter	Description	References
Duration (s)	Time elapsed between onset and end of a single vocalization	Bayart et al., 1990 (rhesus macaques); Briefer, Maigrot, et al., 2015 (horses); Briefer, Vizier, et al., 2019 (domestic pigs); Collins et al., 2011 (Weddell seals); Fichtel et al., 2001 (squirell monkeys); Gogoleva et al., 2010 (silver foxes); Lemasson et al., 2012 (cercopithecine species); Linhart et al., 2015 (domestig pigs); Maestripieri et al., 2000 (rhesus macaques); Maigrot et al., 2018 (wild boars); Meise, Keller, Cowlishaw, & Fischer, 2011 (chacma baboons); Schehka & Zimmermann, 2009 (tree shrews); Stoeger et al., 2011 (African elephants); Sugiura, 2007 (Japanese macaques)
Mean F0 (Hz)	Mean value of the fundamental frequency (F0) over the vocalization	Briefer, Maigrot, et al., 2015 (horses); Briefer, Tettamanti, & McElligott, 2015 (goats); Briefer et al., 2019 (domestic pigs); Collins et al., 2010 (Weddell seals); Meise et al., 2011 (chacma baboons); Rendall, 2003 (chacma baboons); Scherer, 2003 (humans); Szipl et al., 2017 (ravens); Yeon et al., 2011 (domestic cats)
Max F0 & Min F0 (Hz)	Maximum and minimum values of the F0 over the vocalization; **F0 range** is calculated as the difference	Bayart et al., 1990 (rhesus macaques); Briefer, Maigrot, et al., 2015 (horses); Briefer, Tettamanti, et al., 2015 (goats); Briefer et al., 2019 (domestic pigs); Lemasson et al., 2012 (red capped mangabeys); Ordóñez-Gómez et al., 2019 (spider monkeys) Scherer, 2003 (humans); Sugiura, 2007 (Japanese macaques); Yamaguchi et al., 2010 (common marmosets)
Onset F0 & End F0 (Hz)	F0 values at onset and offset of the vocalization	Briefer, Maigrot, et al., 2015 (horses); Briefer, Tettamanti, et al., 2015 (goats); Lemasson et al., 2012 (red capped mangabeys); Norcross & Newman, 1993 (common marmosets); Ordóñez-Gómez et al., 2019 (spider monkeys); Rendall, 2003 (chacma baboons); Schehka & Zimmermann, 2009 (tree shrews); Sugiura, 2007 (Japanese macaques); Yamaguchi et al., 2010 (common marmosets)
Relative time of max F0 (proportion)	The timing of the maximum value of the F0 as a proportion of total duration	Meise et al., 2011 (chacma baboons); Ordóñez-Gómez et al., 2019 (spider monkeys); Rendall, 2003 (chacma baboons); Yamaguchi et al., 2010 (common marmosets)
Onset F0 slope (Hz/s)	The difference between max and onset F0 divided by the time distance between the two points	Norcross & Newman, 1993 (common marmosets); Rendall, 2003 (chacma baboons); Sugiura, 2007 (Japanese macaques)

Mean harmonics-to-noise ratio (dB)	A measure of the relative amounts of tonal and noisy (chaotic) energy in a vocalization	Briefer et al. 2019 (domestic pigs); Gogoleva et al., 2010 (silver foxes); Linhart et al., 2015 (domestic pigs); Maigrot et al. 2018 (wild boars); Meise et al., 2010 (chacma baboons); Puppe et al., 2005 (domestic pigs); Siebert et al., 2011 (dwarf goats); Stoeger et al., 2011 (African elephants)
Jitter (proportion)	Cycle-to-cycle variation in period length; a measure of F0 instability	Bachorowski & Owren, 1995 (humans); Lehoczki, Szamosvölgyi, Miklósi, & Faragó, 2019 (dogs); Rendall, 2003 (chacma baboons); Szipl et al., 2017 (common ravens)

To analyze these parameters, spectrograms were generated from the coos using fast-Fourier transform in Praat 6.0.29 (Boersma & Weenink, 2013). The coo was highlighted manually, visually inspecting the spectrogram and waveform and listening to the vocalization, to obtain its duration while excluding any reverberation. F0 measurements were then obtained from the highlighted portion using the Quantify Source command in the GSU Praat Tools package, Version 1.9 (Owren, 2008). The default settings for this command were used, with the exception that the pitch ceiling was set to 3000 Hz to account for the F0 range of juvenile rhesus macaque coos. The command uses Praat's To Pitch autocorrelation function to estimate a F0 contour, which the user then inspects and manually corrects if necessary (e.g., octave correction, removal of any unvoiced segments) (Owren, 2008). Automatic mean F0, min F0, and max F0 measurements were obtained in this way. The Quantify Source command also produces a mean HNR measurement based on Praat's To Harmonicity autocorrelation function, and a jitter measurement based on Praat's PointProcess: Get jitter function. Jitter represents the cycle-to-cycle variation in F0; we used the relative average perturbation (RAP) jitter measurement. Onset F0 and end F0 were measured by placing the cursor at the onset or offset of the vocalization and recording the F0 estimate produced by the To Pitch autocorrelation function at that time point. Finally, the time of the maximum F0 was measured by highlighting the segment of the coo beginning at onset and ending at the F0 maximum, and recording the duration of this segment. Onset F0 slope was calculated by dividing the frequency difference between start and max F0 values by the time distance between these two points; coos where this time distance comprised less than 5% of the total call duration were excluded from this measurement. Although coos are often chevron-shaped, many of the coos analyzed here ended at the F0 maximum, therefore we did not analyze offset F0 slope.

The scream recordings exhibited variation in the level of low-frequency (< 1,000 Hz, well below the F0 of recorded screams) energy emanating from nearby electric feeders, which could be expected to influence energy distribution measurements. To account for this, we used the marquee selection tool in Audition to reduce this clearly extraneous noise by 5-10 dB depending on the distance from the feeders to the microphone, in order to achieve a similar, low level of background noise for all scream vocalizations. The same parameters measured for the coos (Table 2.2) were obtained for the screams, excluding the onset F0 slope variable (since screams do not exhibit the chevron-shaped F0 contour characteristic of coos), using the Quantify Source command within the GSU Tools package in Praat. 37 out of the 178 tonal screams, and 20 out of the 184 pulsed screams, lacked Praat measurements because either the automatically estimated F0 contour did not match the F0 contour visible in the spectrogram and could not be manually altered to match, or else the F0 was not visible in the spectrogram (for screams exhibiting subharmonics, acoustic noise, or other nonlinear phenomena). In cases where the estimated F0 contour could not be made to match but the contour was visible in the spectrogram, we manually measured onset F0, end F0, min F0, max F0, and time to max F0 by inspecting the spectrogram in Praat. In addition, we measured 6 spectral parameters associated with emotion across mammalian species (Table 2.3), by applying a 1,024-pt fast-Fourier transform and obtaining automated measurements available in Raven Pro sound analysis software (Center for Conservation Bioacoustics, 2014).

Table 2.3. Additional acoustic parameters obtained from screams, and selected references indicating a relationship with emotion

Parameter	Description	References
Mean peak frequency (Hz)	The frequency with the most energy on average over the vocalization	Fichtel et al., 2001 (squirrel monkeys); Gogoleva et al., 2010 (silver foxes); Puppe, Schön, Tuchscherer, & Manteuffel, 2005 (domestic pigs); Schehka & Zimmerman, 2009 (tree shrews); Yeon et al., 2011 (domestic cats)
Dfa25, dfa50, dfa75 (Hz)	Distribution of frequency amplitudes: frequency values of the upper limits of the first 3 quartiles of energy, on average over the vocalization	Briefer et al., 2015 (horses); Fichtel et al., 2001 (squirrel monkeys); Gogoleva et al., 2010 (silver foxes); Linhart et al., 2015 (domestic pigs); Puppe et al., 2005 (domestic pigs); Scherer, 2003 (humans); Siebert, Langbein, Schön, Tuchscherer, & Puppe 2011 (dwarf goats); Yeon et al., 2011 (domestic cats)
Interquartile range (Hz)	Difference between dfa75 and dfa25; width of energy distribution across the frequency spectrum	Meise et al., 2011 (chacma baboons); Puppe et al., 2005 (domestic pigs); Schrader & Todt, 1998 (domestic pigs); Siebert et al., 2011 (dwarf goats)
Average entropy (bits)	A measure of the relative amounts of tonal and noisy (chaotic) energy in a vocalization	See references for harmonic-noise ratio in Table 2.2.

Statistical Analyses

For the coo analysis, the increase in plasma cortisol from baseline (pre-HIT) to after the HIT served as a proxy for a monkey’s stress and emotional arousal reaction to the test. To examine the relationship between cortisol increase and coo acoustics, we fit several linear mixed models (LMM), each with an acoustic parameter as the outcome variable. In addition to cortisol increase, fixed effects included age in months and stage of the HIT (Alone, Profile, Stare). To control for potential differences in cortisol reactivity depending on baseline cortisol levels (e.g., high baseline cortisol blunting cortisol reactivity), baseline cortisol and its interaction with cortisol increase were both included as fixed effects. HIT and subject were included as nested random effects, to account for non-independence of observations from within a HIT and a subject. Upon inspecting the data, the effects of cortisol increase on some acoustic parameters of coos appeared to change as subjects aged over the course of the study; to test this, we refit the LMM including as an additional fixed effect the interaction between cortisol increase and age.

The analysis of screams was also carried out using LMM with acoustic parameters as outcome variables. The eliciting behavior (Table 2.1) served as a proxy for emotional arousal, and was included as a categorical fixed effect. Attack and chase were lumped together into a single category, due to low frequency counts for these two types of instances and the assumption that attacks and chases both evoke in victims greater arousal than other documented eliciting behaviors. Other fixed effects included age, scream class (tonal or pulsed), and the interaction between eliciting behavior and scream class, to determine whether and how the effects of arousal on scream acoustics might differ between tonal and pulsed screams. Scream bout and subject were included as random effects.

LMM were fit by restricted maximum likelihood, with a gaussian distribution and identity link, using the lmer function (lme4 package; Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2018). Degrees of freedom and p values were estimated using Satterthwaite’s method via the summary and anova functions in the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017). Nonsignificant interaction effects were removed from the models because they can interfere with estimation of main effects (Engqvist, 2005). Posthoc pairwise comparisons between levels of categorical fixed effects were conducted in R using the Kenward-Roger method, with Tukey HSD p adjustment, via the lsmeans package (Lenth, 2016). Residuals of each LMM were checked visually for homoscedasticity and normality. Based on these checks, the following were tested with log-link, penalized quasi-likelihood generalized linear mixed models (GLMM), using the glmmPQL function in the MASS package (Venables & Ripley, 2013), instead of LMM: for coos, jitter; for screams, duration, min F0, onset F0, end F0, jitter, DFA25. For these, overall p values of categorical fixed effects were estimated using Wald tests, via the wald.test function in the aod package (Lesnoff & Lacelot, 2012). Counts of numbers of vocalizations were tested with zero-inflated poisson GLMM, using the glmmTMB package (Brooks et al., 2017), instead of LMM. All significance testing used an alpha of 0.05.

Results

Descriptive statistics for acoustic measurements are presented in Table 2.4.

Table 2.4. Mean ± SD values of acoustic parameters.

Acoustic parameter	**Coos**	**Tonal screams**	**Pulsed screams**
Call duration (s)	0.600 ± 0.230	0.440 ± 0.164	0.123 ± 0.054
Mean F0 (Hz)	1125.4 ± 296.0	4394.4 ± 1440.9	3877.7 ± 1348.9
Min F0 (Hz)	855.7 ± 222.9	2969.6 ± 973.4	3228.8 ± 1186.3
Max F0 (Hz)	1273.7 ± 365.7	5883.8 ± 2196.6	4508.8 ± 1749.1
F0 Range (Hz)	418.0 ± 244.1	2907.7 ± 2111.9	1253.3 ± 1415.0
Onset F0 (Hz)	866.6 ± 217.2	3910.2 ± 1781.0	3910.8 ± 1532.1
End F0 (Hz)	1168.8 ± 325.2	3945.2 ± 1489.0	3541.5 ± 1412.8
Time of Max F0 (%)	52.02 ± 17.47	47.40 ± 27.34	35.17 ± 30.58
Onset F0 Slope (Hz/s)	1514.6 ± 1039.5	----------	----------
Mean HNR (dB)	11.10 ± 5.02	1.52 ± 2.90	3.04 ± 3.68
Jitter (RAP)	$7.09\text{x}10^{-5}$ ± $9.26\text{x}10^{-5}$	$2.58\text{x}10^{-5}$ ± $6.91\text{x}10^{-5}$	$2.15\text{x}10^{-5}$ ± $2.94\text{x}10^{-5}$
Mean peak freq. (Hz)	----------	4769.4 ± 1329.9	4248.8 ± 1453.1
DFA25 (log Hz)	----------	3760.3 ± 977.3	3642.2 ± 1237.5
DFA50 (Hz)	----------	4692.5 ± 1351.8	4340.8 ± 1664.7
DFA75 (Hz)	----------	6000.5 ± 1705.6	5314.7 ± 2090.9
IQR (Hz)	----------	2240.2 ± 1405.5	1672.6 ± 1368.4

Average entropy (bits)	----------	5.68 ± 0.61	5.32 ± 0.76

Coos

Almost every tested acoustic parameter exhibited a significant relationship with arousal as measured by the cortisol increase from before until after the HIT (Table 2.5). HIT trials associated with greater cortisol increases yielded higher numbers of coos, with shorter duration, higher absolute and mean F0 values, greater F0 modulation (range, onset slope, and jitter). Only one parameter, coo duration, showed significant main effect of baseline cortisol; the effect of cortisol increase on coo duration was significantly dampened as baseline cortisol increased (i.e., significant interaction effect). Most parameters also changed significantly over the course of the HIT, with the Alone stage producing coos with longer duration, higher F0, greater F0 modulation (range, onset slope), and more noise than the other stages, though generally these effect sizes were miniscule (Figure 2.2). As monkeys aged from 6 months to 16 months to menarche, their coos exhibited increased duration, decreases in some F0 parameters, decreased noisiness, and increased jitter (Table 2.5). Upon visual inspection of data plots, a possible interaction between age and cortisol increase on acoustics became apparent; we therefore added a cortisol-increase-by-age interaction effect to each model. This was significant for several F0 parameters, with the positive effect of cortisol increase on F0 being dampened as the monkeys aged (Table 2.5, Figure 2.3); the overall main effect of cortisol increase remained significant for all of these parameters. One menarche-aged subject exhibited an extremely high cortisol increase and coos showing relatively low F0 parameters; results were not altered significantly by filtering out this trial.

Table 2.5. Coo LMM results. Nonsignificant interactions were removed, then individually re-added to estimate their effect size and p value. p < 0.05 presented in bold.

Acoustic variable	Cortisol Increase (µg/mL)		Baseline Cortisol (µg/mL)		Increase*Baseline		HIT Stage	Age (months)		Cortisol Incr*Age	
	Coeff ± SE	p	Coeff ± SE	p	Coeff ± SE	p	p	Coeff ± SE	p	Coeff ± SE	p
Number of calls (log) [a]	**0.017 ± 0.006**	**<0.001**	0.006 ± 0.006	0.492	Not estimable	Not estim.	Not estim.	Not estimable	Not estim.	Not estimable	Not estim.
Call duration (s)	**-0.0127 ± 0.0049**	**0.011**	**-0.0175 ± 0.0065**	**0.009**	**5.51x10^{-4} ± 2.55x10^{-2}**	**0.035**	**<0.001**	**0.0124 ± 0.0017**	**<0.001**	2.53x10^{-4} ± 2.04x10^{-4}	0.218
Mean F0 (Hz)	**17.03 ± 4.19**	**<0.001**	1.44 ± 3.32	0.666	-0.18 ± 0.33	0.577	**<0.001**	-6.44 ± 4.52	0.159	**-0.63 ± 0.21**	**0.003**
Min F0 (Hz)	**9.54 ± 2.84**	**0.001**	-1.01 ± 2.25	0.656	-0.31 ± 0.17	0.165	**<0.001**	**-8.61 ± 3.06**	**0.006**	**-0.37 ± 0.14**	**0.011**
Max F0 (Hz)	**21.26 ± 5.36**	**<0.001**	2.37 ± 4.32	0.584	-0.11 ± 0.42	0.797	**<0.001**	-6.01 ± 5.84	0.307	**-0.76 ± 0.27**	**0.006**
F0 Range (Hz)	**12.50 ± 3.90**	**0.002**	3.56 ± 3.14	0.260	0.23 ± 0.31	0.460	**<0.001**	3.28 ± 4.26	0.444	**-0.40 ± 0.20**	**0.046**
Onset F0 (Hz)	**3.74 ± 1.73**	**0.034**	-0.91 ±2.41	0.707	-0.12 ± 0.21	0.568	**<0.001**	**-15.47 ±1.45**	**<0.001**	-0.23 ± 0.15	0.121
End F0 (Hz)	**19.68 ± 4.74**	**<0.001**	4.25 ± 3.75	0.261	-0.14 ± 0.37	0.703	**<0.001**	-4.97 ± 5.11	0.334	**-0.73 ± 0.23**	**0.003**
Time of Max F0 (%)	0.16 ± 0.15	0.307	0.34 ± 0.20	0.105	1.3x10^{-4} ± 1.8x10^{-4}	0.455	0.222	**-0.39 ± 0.12**	**0.002**	-2.1x10^{-4} ± 1.3x10^{-4}	0.105
Onset F0 Slope (Hz/s)	**57.85 ± 15.70**	**<0.001**	16.39 ± 12.77	0.203	0.75 ± 1.24	0.547	**<0.001**	7.22 ± 17.31	0.678	**-1.85 ± 0.81**	**0.025**
Mean HNR (dB)	-0.044 ± 0.054	0.422	-0.063 ± 0.076	0.405	-0.0052 ± 0.0066	0.436	**0.002**	**-0.165 ± 0.045**	**<0.001**	0.0064 ± 0.0047	0.175
Jitter (log RAP) [b]	**0.016 ± 0.004**	**<0.001**	0.006 ± 0.008	0.437	4.8x10^{-4} ± 5.7x10^{-4}	0.403	0.81	**0.065 ± 0.005**	**<0.001**	7.9x10^{-4} ± 5.3x10^{-4}	0.147

[a] Fit by GLMM with poisson distribution, log link, and zero-inflation parameter (glmmTMB function, glmmTMB package), in place of LMM. Estimates presented on log scale. Due to failure to converge, effects were estimated one-at-a-time in separate models, and some were not estimable.

[b] After visual inspection of LMM residuals, GLMM with penalized quasi-likelihood, Gaussian distribution, and log link (glmmPQL function, MASS package) used in place of LMM. p values for categorical fixed effects calculated using Wald Test (wald.test function). Estimates presented on log scale.

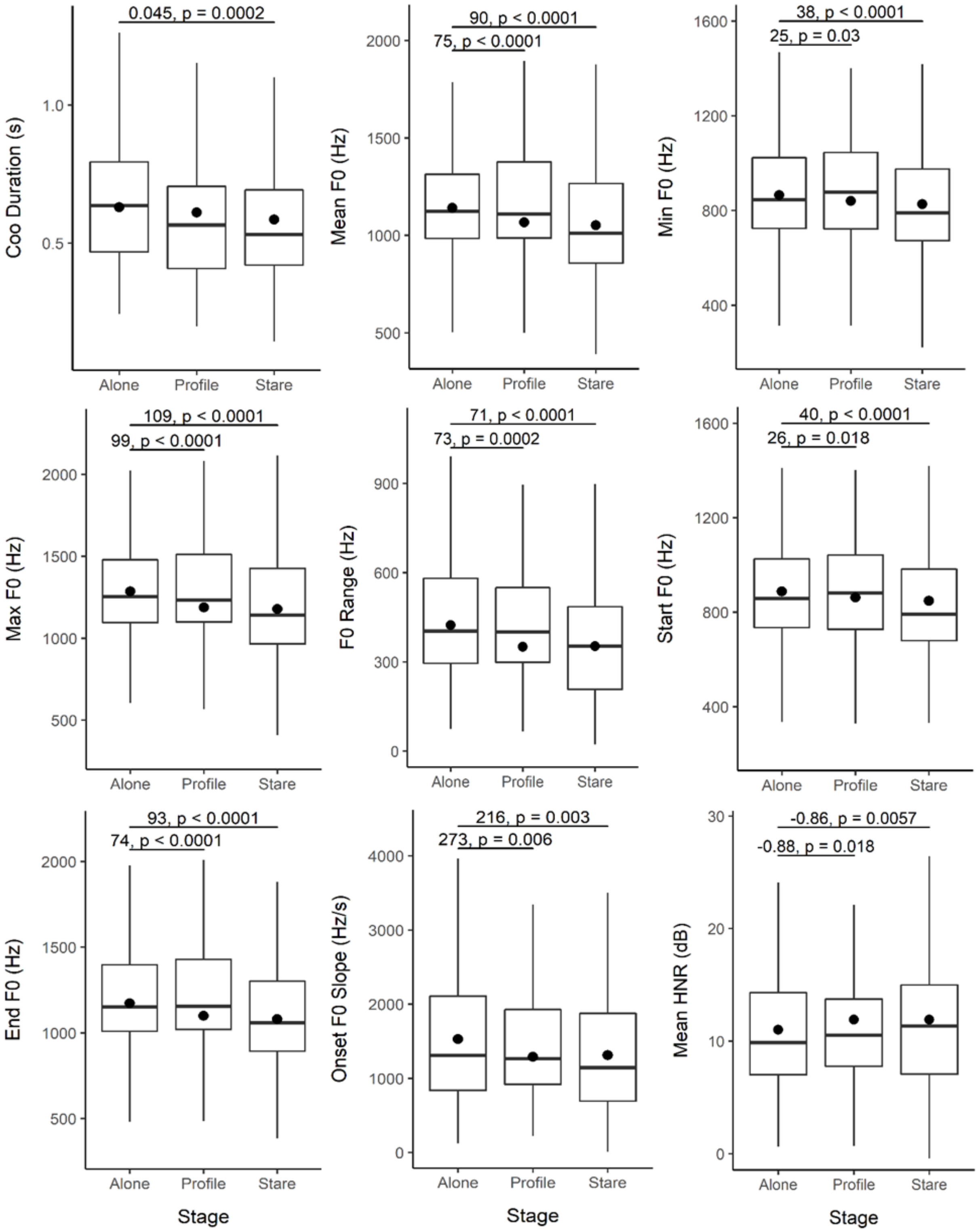

Figure 2.2. Pairwise comparisons of coo acoustics by stage of the HIT test. Boxes show quartiles and median, whiskers show 5- and 95-percentiles, points show least-squares means estimated through (G)LMM. Outliers omitted for clarity. Significant differences ($p < 0.05$) indicated with horizontal lines; first number indicates least-squares mean difference.

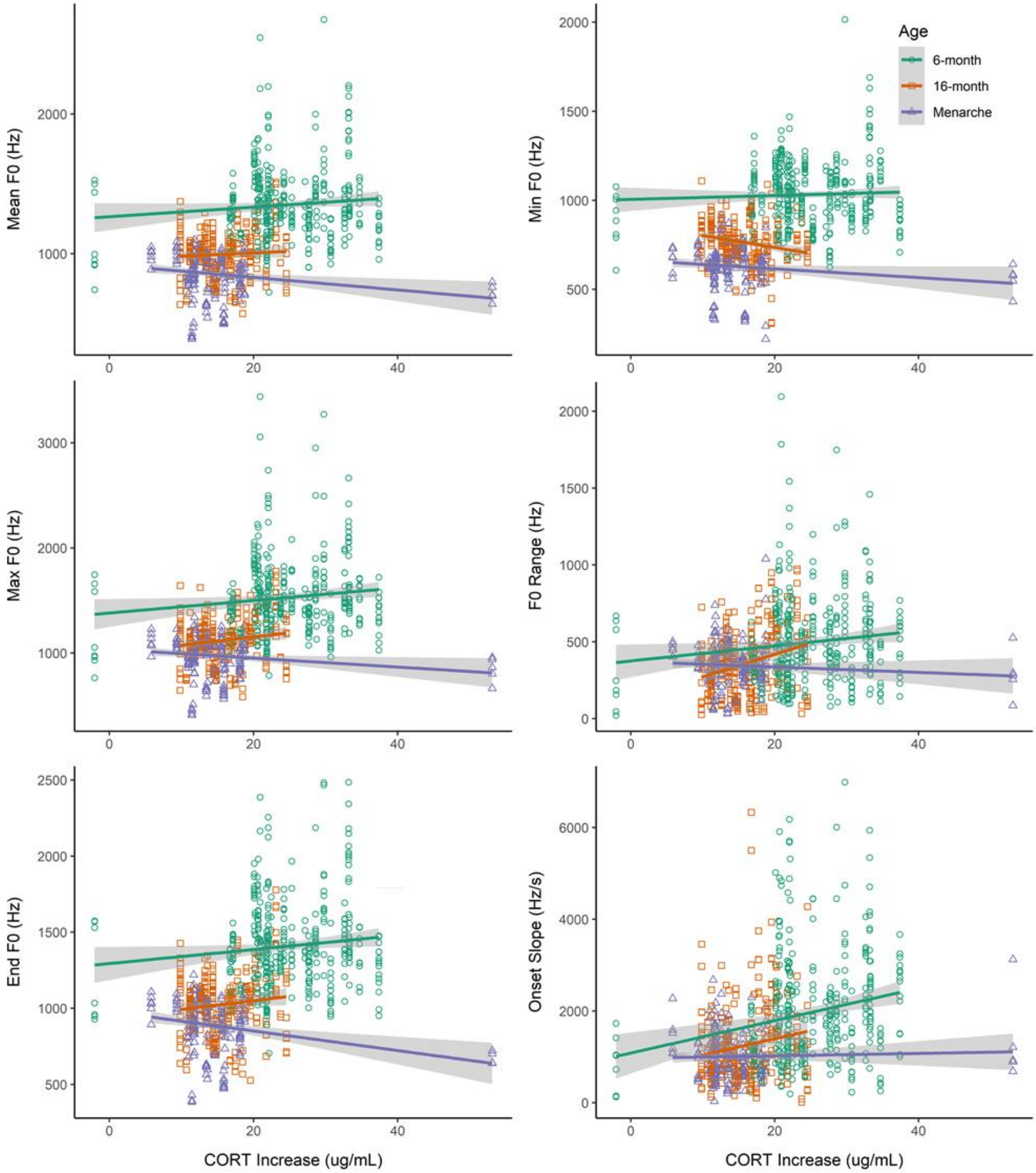

Figure 2.3. Interaction effect of cortisol increase by age on acoustic parameters of coos. Points represent individual coos. Age represented categorically for visualization purposes; statistical analyses used age in months (continuous). Menarche trendline slopes were not altered signficiantly by excluding the outlying trial far to the right.

Screams

Tonal versus pulsed screams differed across several temporal and spectral acoustic parameters (Table 2.6). Scream usage varied significantly depending on context, with attack and chase contexts yielding higher numbers of tonal screams, and less intense contexts yielding more pulsed screams (Figure 2.4). In addition, several acoustic parameters exhibited significant context-dependent variation (Table 2.6). Specifically, high-arousal attack and chase contexts were associated with higher mean F0 and onset F0 than lower-arousal contexts (Figure 2.5). Pulsed screams elicited by attack and chase contexts exhibited higher min F0 and DFA25 than those emitted in lower-arousal contexts, but tonal screams did not exhibit significant context-dependent variation in these parameters (i.e., signficant context-by-class interaction) (Figure 2.5). Although LMM indicated a significant effect of context on max F0, posthoc comparisons indicated no significant pairwise differences. Of all parameters, only the number of tonal screams showed a signficant effect of age, with monkeys emitting fewer tonal screams as they aged.

Table 2.6. Scream LMM results. Nonsignificant interactions were removed, then individually re-added to estimate their effect size and p value. p < 0.05 presented in bold.

Acoustic variable	Scream Class (Pulsed -> Tonal)		Context	Age (days)		Context *Class
	Coeff ± SE	p	p	Coeff ± SE	p	p
Number of Tonal (log) [a]	-------------	------	**<0.001**	**$-1.14\text{x}10^{-3}$ ± $3.7\text{x}10^{-4}$**	**0.002**	----------
Number of Pulsed (log) [a]	-------------	------	**<0.001**	$-3.04\text{x}10^{-4}$ ± $3.25\text{x}10^{-4}$	0.348	----------
Scream duration (log s) [b]	**1.27 ± 0.06**	**<0.001**	0.37	$4.8\text{x}10^{-5}$ ± $1.42\text{x}10^{-4}$	0.73	0.63
Mean F0 (Hz)	**532.23 ±121.80**	**<0.001**	**0.033**	1.07 ±0.74	0.160	0.523
Min F0 (log Hz) [b]	**-0.327 ± 0.086**	**<0.001**	**<0.001**	$-2.5\text{x}10^{-5}$ ± $1.03\text{x}10^{-4}$	0.807	**0.037**

Max F0 (Hz)	**1345.1 ± 184.3**	**<0.001**	**0.043**	1.17 ± 0.98	0.235	0.249
F0 Range (Hz)	**1664.7 ± 184.4**	**<0.001**	0.282	0.99 ± 0.80	0.221	0.173
Onset F0 (log Hz) [b]	-0.030 ± 0.045	0.511	**<0.001**	$4.2\text{x}10^{-5} \pm 1.24\text{x}10^{-4}$	0.733	0.34
End F0 (log Hz) [b]	**0.097 ± 0.043**	**0.024**	0.22	$1.9\text{x}10^{-5} \pm 1.16\text{x}10^{-4}$	0.872	0.083
Time of Max F0 (%)	**12.80 ± 3.08**	**<0.001**	0.167	$3.14\text{x}10^{-3} \pm 1.05\text{x}10^{-2}$	0.767	0.057
Mean HNR (dB)	**-1.47 ± 0.36**	**<0.001**	0.75	$-6.98\text{x}10^{-4} \pm 1.4\text{x}10^{-3}$	0.628	0.298
Jitter (log RAP) [b]	0.263 ± 0.266	0.325	0.59	$1.42\text{x}10^{-4} \pm 7.57\text{x}10^{-4}$	0.852	0.94
Mean peak freq. (Hz)	**516.12 ± 129.90**	**<0.001**	0.345	0.575 ± 0.604	0.347	0.683
DFA25 (log Hz) [b]	**-0.174 ± 0.728**	**0.018**	**<0.001**	$-1.12\text{x}10^{-4} \pm 8.3\text{x}10^{-5}$	0.17	**0.013**
DFA50 (Hz)	295.10 ± 144.45	**0.042**	0.177	0.20 ± 0.63	0.750	0.234
DFA75 (Hz)	**559.00 ± 177.78**	**0.002**	0.231	0.88 ± 0.85	0.309	0.526
IQR (Hz)	**468.24 ± 131.67**	**<0.001**	0.448	1.02 ± 0.61	0.103	0.126
Average entropy (bits)	**0.316 ± 0.065**	**<0.001**	0.663	$1.79\text{x}10^{-4} \pm 2.69\text{x}10^{-4}$	0.512	0.887

[a] Fit by GLMM with poisson distribution, log link, and zero-inflation parameter (glmmTMB function, glmmTMB package). Estimates presented on log scale.

[b] After visual inspection of LMM residuals, GLMM with penalized quasi-likelihood, gaussian distribution, and log link (glmmPQL function, MASS package) used in place of LMM. p values for categorical fixed effects calculated using Wald Test (wald.test function). Estimates presented on log scale.

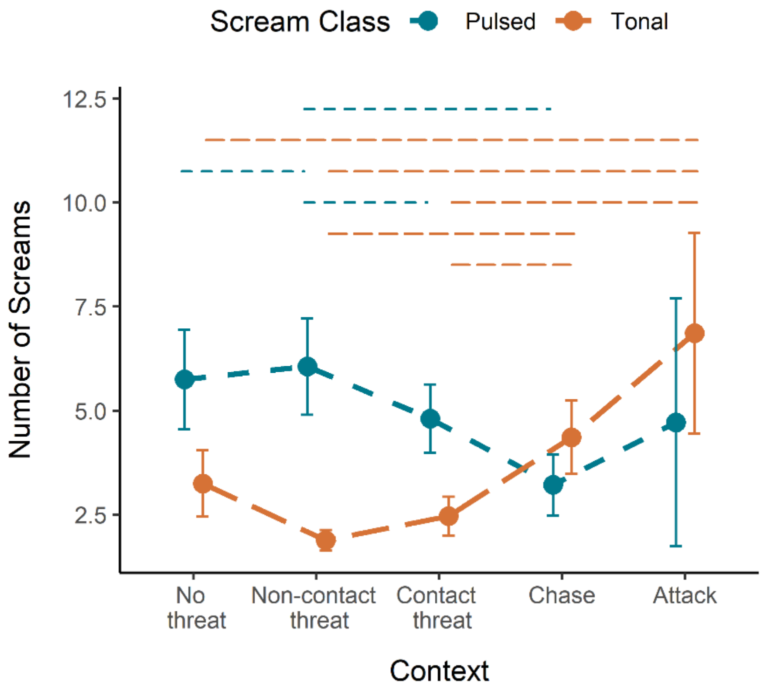

Figure 2.4. Mean and SE numbers of pulsed and tonal screams uttered in different contexts. Horizontal lines denote overall significant pairwise differences between contexts within a scream class, indicated by zero-inflated poisson GLMM ($p < 0.05$).

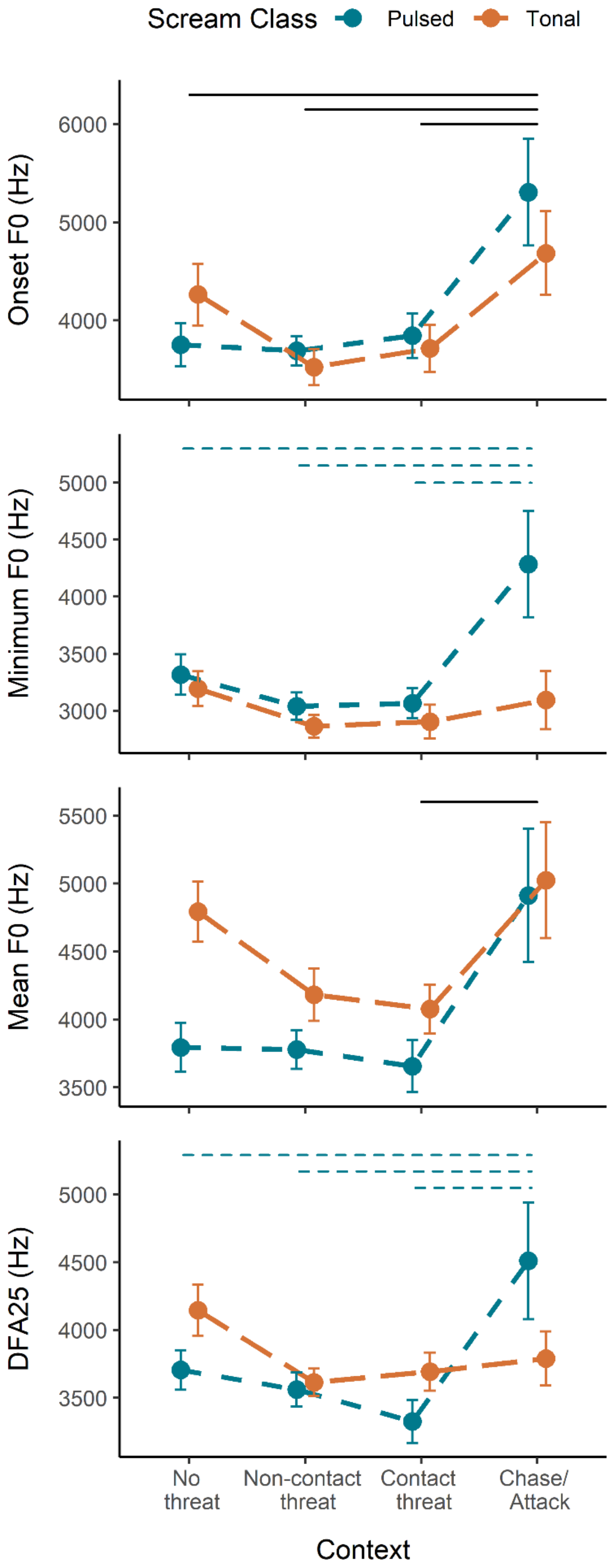

Figure 2.5. Mean and SE acoustic parameters of screams by context and class. Horizontal lines denote overall significant pairwise differences indicated by LMM ($p < 0.05$); dotted indicates class-specific pairwise difference (i.e., significant context-by-class interaction).

Discussion

The present study identified similarities in the acoustic correlates of emotional arousal between two rhesus macaque vocalization types, coos and screams: within each vocalization type, greater arousal was associated with increases in mean F0 and other F0 measures. These similar patterns are striking in light of the highly disparate contexts in which the coos and screams were recorded, and the different methods used to measure arousal for each, namely, cortisol measurement in a controlled behavioral test for coos, and intensity of aggression received in naturally occurring social interactions for screams. These results along with those of similar studies (noted below) suggest a degree of evolutionary homology in the cognitive, physiological, and anatomical mechanisms of vocal production, both between these vocalization types and between rhesus macaques and other species, and carry broader implications regarding general processes in vocal evolution. Coos showed additional arousal-dependent variation in duration and F0 modulation parameters, whereas we did not observe similar trends among screams. Coo acoustics and their relationships with arousal also exhibited significant changes over the course of juvenile development, that were not seen among screams. These contrasts might be due to evolved differences in the physiology of emotional arousal and mechanisms of vocal production between these two vocalization types and the contexts in which they occur, or to the different methods used to study them, or a combination of these.

Comparing the Acoustic Correlates of Emotional Arousal in Coos and Screams

Arousal entails increased action and tension in muscles throughout the body, including within the vocal apparatus, with predictable effects on the acoustic properties of the voice. Numerous studies have supported a positive relationship between emotional arousal and vocal

F0, likely due to increased tension and action in laryngeal muscles (e.g., Briefer, 2012; Green, 1975; Scherer, 2003; Titze, 1994). However, very few studies prior to the present one have explicitly examined the effects of emotional arousal on acoustic variation within vocalization types of rhesus macaques (Bayart et al., 1990; Patel & Owren, 2007), and to our knowledge, no study has previously done so across multiple vocalization types. In line with our hypotheses, results showed a positive relationship between arousal and overall F0 in both coos (mean, min, max, start, and end F0) and screams (mean, min, and onset F0)—suggesting homology with other previously studied mammalian taxa (Briefer, 2012) including other primates (e.g., Rendall, 2003).

In addition to overall F0, the literature also indicates F0 modulation as a significant indicator of arousal (e.g., Lehoczki et al., 2019; Lingle et al., 2012). Consistent with this trend, and in support of our hypothesis, the coos of subjects showing stronger arousal reactions to the HIT (i.e., greater cortisol increase) exhibited greater F0 modulation (including F0 range, slope of the upward segment of the F0 contour, and jitter). Lingle et al. (2012) pointed out that a higher max F0 (and therefore greater overall F0 range) is one of the major features distinguishing separation/isolation calls from contact calls, and suggested that, in general, the acoustic differences between these vocalization types might be attributable to arousal differences. In rhesus macaques, coos are uttered in a wide range of contexts varying in arousal level (e.g., Bayart et al., 1990; Hauser & Marler, 1993); the correlation observed in the present study between F0 modulation and cortisol increases appears to show the graded acoustic variation hypothesized by Lingle et al. (ibid.) at a small scale, i.e., *within* the class of coos that might be interpreted based on context as “isolation calls”. That said, it is important to reiterate here that coos occur not only in isolation but also preceding friendly contact and in anticipation of food

(Bayart et al., 1990; de Waal & Bauers, 1991; Green, 1975; Hansen, 1976; Hauser, 1991; Hauser & Marler, 1993; Simons & Bielert, 1973), each of which is probably associated with a different emotional state; whether F0 modulation correlates with arousal in coos within these other contexts remains open to investigation.

The results of this study suggest that socioecological selection pressures favoring changes in emotional arousal would likely affect acoustic properties of vocalizations, and vice versa. To illustrate this principle, consider work by Sugiura (2007) on the coos of the closely related Japanese macaque, demonstrating greater F0 modulation when callers were farther from group members. This result was interpreted as evidence that selection had favored acoustic changes making coos more localizable when callers were separated from the group (Sugiura, 2007; for a similar result and interpretation, see Yamaguchi et al., 2010). This explanation is plausible, but consideration of the proximate role of emotional arousal illuminates some additional evolutionary possibilities. Extrapolating our results with rhesus macaques to Japanese macaques, at the proximate level, the wider F0 modulation seen in Japanese macaque coos during separation from group mates might be due to greater arousal (see Lingle et al., 2012). If that is the case, then any selection pressure favoring an arousal response to separation from group mates would be expected to incidentally result in wider F0 modulation, through a correlated response to selection. It seems to us likely that any number of selection pressures might favor an aroused emotional reaction to separation (Lingle et al., 2012), e.g., increased vigilance for predators when individuals are most vulnerable, and mobilization of energy stores that would aid in escape. Thus selection pressures favoring increased vigilance and ability to escape danger could have worked in tandem with other pressures favoring increased localizability of coos to produce this observed relationship between isolation and coo acoustics. On the other hand, that we failed to observe a relationship in screams between F0 modulation measures and arousal (as measured by the intensity of agonism received) suggests the possibility that F0 modulation might vary according to arousal in contact/separation/isolation calls (e.g., macaque coos) across taxa, but not in some other vocalization types and contexts. This could potentially reflect selection favoring an especial strengthening of the link between arousal and F0 modulation in coos in the context of maternal/group separation, due to the benefits of encoding information about location within this particular vocalization type in this particular context.

Similarly, while coos were shorter in association with greater cortisol increases, scream duration did not vary significantly depending on the intensity of agonism received by the screamer, at least while accounting for scream class (the two classes unsurprisingly showed a large difference in duration, discussed below). The literature contains examples of support for call duration as an indicator of arousal (Filippi et al., 2019; Lemasson et al., 2012), but also suggestions that duration correlates more strongly with emotional valence (Briefer, Maigrot, et al., 2015; Briefer, Vizier, et al., 2019; Friel et al., 2019). Our mixed findings contribute further ambiguity to the question of the relationship between call duration and emotion. More broadly, the contrasts between coos and screams raise the possibility that selection might be capable of tinkering with the acoustic correlates of emotion in some vocalization types—this possibility is discussed further below.

Acoustic Variation Within and Between Scream Classes

Rhesus macaque screams have received considerable empirical attention, largely focused on the possibility of acoustically and functionally distinct scream classes and the factors determining their usage (H. Gouzoules, 2005; S. Gouzoules et al., 1984; Le Prell et al., 2002;

Tomaszycki et al., 2001) and inter-individual acoustic variation (Fugate, Gouzoules, & Nygaard, 2008; Rendall, Rodman, & Emond, 1996). Scream classes are hypothesized to correlate with the intensity of agonism received and the rank difference between the screamer and the opponent, functioning to inform kin (in the sense of reduction of uncertainty; Seyfarth et al., 2010) and influence decisions by allies to intervene (H. Gouzoules, 2005; S. Gouzoules et al., 1984). This account is not incompatible with the notion that acoustic variation within scream classes might be influenced by emotion, or that changes in arousal might "push" scream acoustics along the gradient from one scream class into another. For example, noisy screams are uttered in the most intense and threatening agonistic contexts (S. Gouzoules et al., 1984); in principle, it could be that high arousal associated with these contexts results in pharyngeal constriction and increased air pressure, and thus acoustic noise, transitioning a tonal scream into a noisy scream. Prior to the present study, little attention has been paid to the relationships between emotion and acoustic variation within rhesus macaque scream classes. In order to investigate this, we analyzed two classes of scream, tonal screams and pulsed screams (Figure 2.1), and accounted for between-class acoustic variation by including scream class as a factor in our analyses.

We found a significant main effect of the intensity of the preceding agonism (our operational measure of arousal for screams) on mean F0 and onset F0, irrespective of scream class. Additionally, DFA25 and min F0 correlated with agonism intensity among pulsed screams, showing greater values in high-arousal attack and chase contexts, though not among tonal screams. Counter to the notion that arousal underlies acoustic gradations between scream classes, mean, min, and start F0 and DFA25 are not seemingly among the parameters by which scream classes are distinguished (S. Gouzoules et al., 1984); temporal characteristics and those related to spectral noise would better fit that criterion, yet these did not exhibit significant arousal-dependent variation within scream classes. This is consistent with the hypothesis that, while arousal can influence acoustic variation within vocalization classes, decisions about which vocalization class to use likely involve other cognitive processes (Schamberg et al., 2018; Schwartz et al., in press).

While our primary goal in including scream class as a factor in our analyses was to account for between-class variation while examining relationships with arousal, analyses revealed several differences between scream classes (Table 2.5) that warrant discussion. Pulsed screams were significantly shorter in duration than tonal screams, which is to be expected since duration is one of the qualities originally used to distinguish these two scream classes. Tonal screams exhibited substantially higher F0 and peak frequency, and wider F0 range and energy distribution, than pulsed screams. Pulsed screams included relatively low-F0 exemplars whereas tonal screams are more consistently high-F0, and the longer duration of tonal screams likely led to greater modulation of the F0 and spectral energy. Surprisingly, tonal screams contained more spectral noise than pulsed screams (lower HNR, higher entropy). To our knowledge, most of these between-class acoustic differences have not previously been formally described. However, because our primary goal was to control for between-class acoustic variation, not to describe it per se, our statistical methods were not ideal for this task—soft cluster analysis would be more appropriate (Wadewitz et al., 2015)—so these results should be interpreted cautiously.

Tonal and pulsed screams are just two of the scream classes exhibited by rhesus macaques, and might not be representative of screams as a whole in this species, especially since they elicit weaker responses than noisy and arched screams (corresponding to severe aggression and rank challenge, respectively). Bearing that caveat in mind, the function of distinct scream classes provides an interesting lens through which to speculate about the possible evolutionary history of the differences in acoustic correlates of arousal between coos and screams. While arousal showed relationships with overall F0, F0 modulation, and duration among coos, we did not find F0 modulation and duration to correlate significantly with arousal among screams. Negative results must be interpreted cautiously, but it is worth noting that F0 modulation and duration are both apparently significant for distinguishing scream classes from one another: F0 modulation distinguishes tonal, arched, and undulated screams, while duration distinguished pulsed and other types of screams (in the present study as well as previously) (S. Gouzoules et al., 1984). Indeed, perceptual experimentation with laboratory-housed rhesus macaques suggested that the max F0 (and thus, F0 range) of screams might be a critical factor for distinguishing at least among tonal and arched screams (Le Prell et al., 2002). It is conceivable that, in the ancestral state, greater arousal might have been associated with screams of shorter duration and wider F0 modulation in addition to higher overall F0 (as is the case in coos), but selection might have favored a decoupling of arousal and these acoustic parameters in screams, in order to prevent variation in caller arousal from interfering with the communicative function of scream classes. This idea is highly speculative, but could serve as a jumping-off point for future research into primate scream production, structure, and function. Alternatively, the differences in our results between coos and screams could stem from the different methods used to measure arousal used for these two call types.

Methods for Measuring Emotional Arousal

Methods used to gauge the emotional states of animals include inferences based on contextual factors such as the intensity of agonism (Filippi et al., 2017; Szipl et al., 2017), and direct measurement of physiological components of emotions, such as glucocorticoids released as a result of activation of the HPA axis in response to stress (Bayart et al., 1990; Blumstein & Chi, 2012; Ralph & Tilbrook, 2016; Schrader & Todt, 1998; Sèbe et al., 2012). The present study utilized both of these types of methods—the former for screams and the latter for coos. Neither method is perfect. Context-based inferences about emotional states are convenient and noninvasive, but rely on assumptions about how an animal perceives and reacts cognitively to a particular situation. In particular, monkeys of different ages might exhibit different emotional and behavioral reactions to similar social situations (e.g., Cheney, 1977) suggesting caution in extrapolating results from the present study to adult rhesus macaques. However, we believe it is reasonable to assume that an attack or chase elicits greater arousal in victims of all ages than, e.g., a mere lunge forward (non-contact threat). Glucocorticoids such as cortisol are also limited as indicators of emotion: HPA activation occurs not only in response to arousing stimuli, but also in response to caloric intake, temperature changes, immune challenge, and exercise, and additionally exhibits circadian and ultradian rhythms that are regulated differently from the responses to the above stressors (Charmandari, Tsigos, & Chrousos, 2005; Devenport, Knehans, Sundstrom, & Thomas, 1989; Fulkerson & Tang, 1979; Ralph & Tilbrook, 2016; Turner, Rivalland, Clarke, & Tilbrook, 2010). We partially controlled for this by measuring the absolute increase in cortisol concentrations from baseline (pre-HIT) to after the HIT, which occurred at the same time of day across trials; this method effectively gauges the degree of HPA activation in response to a stressor (Romero, 2004). With that said, HPA activation does not equal emotional arousal—the former does not exclusively occur in emotional contexts, and in the context of an emotional response, the latter is a broader suite of imperfectly co-occurring physiological changes that generally (but not always) includes the former—but under such controlled conditions, stress-related glucocorticoid increases are reasonably interpreted as an indicator of an animal's emotional arousal reaction.

The coos and screams were investigated using different methods, so drawing comparisons and contrasts between vocalization types requires caution. Specifically, it could be that the differences in F0 modulation and duration as acoustic correlates of arousal between coos and screams reflect differences in method rather than true differences in the acoustic correlates of arousal across these two call types. The potential effects of method on conclusions are especially apparent upon consideration of the differences between the acoustic correlates of cortisol increases versus HIT stage among coos: most of the acoustic parameters that increased with greater cortisol increases also decreased from the alone stage to the other stages of the HIT test, albeit with very small effect sizes. This is the opposite of what we had predicted based on the premise that the presence of an unfamiliar human intruder would evoke greater arousal than isolation, illustrating the potential for different measures of emotion to yield different results. To speculate, it is possible that this might have to do with adaptive responses to the different conditions within the HIT: the acoustic changes associated with greater arousal might make vocalizations more attention-grabbing to receivers (Briefer, 2018), and it might benefit a vulnerable juvenile monkey separated from her group to emit less attention-grabbing vocalizations in the presence of an unknown and potentially dangerous individual (the human intruder). More broadly, despite differences in method, we found similarity between coos and screams regarding the relationship between arousal and overall F0, a pattern that rhesus macaques share with other species (Briefer, 2012).

Effects of Age

We predicted that we would observe decreases in overall F0 and F0 modulation, and increases in duration, of coos over the first few years of development, in line with past research (Hammerschmidt et al., 2000). In general, these predictions were supported; coos also exhibited decreased noisiness as monkeys aged. By contrast, the only parameter that showed a significant effect of age in screams was the number of tonal screams, which decreased over development. The reasons for this contrast between coos and screams are ambiguous; one possibility is, insofar as rhesus macaque screams fall into distinct classes, they might be somewhat acoustically stereotyped from birth, as are the predator-specific alarm calls of many primate species (Egnor & Hauser, 2004; Seyfarth & Cheney, 2010). However, the screams of pigtail macaques do appear to undergo some acoustic stereotyping over the course of development (H. Gouzoules & Gouzoules, 1989).

Although an interaction between age and arousal was not represented in our hypotheses, we observed that the effects of cortisol increase on coos' mean, min, max, and end F0, F0 range, and onset F0 slope became dampened over the course of development (although the main effect of cortisol increase remained significant for all of these parameters). This may be related to developmental changes in how monkeys perceive and react to the HIT, including separation from the mother. Infant primates' regulation of emotional and physiological responses to stressors depends in part on the presence of the mother, i.e., "maternal buffering," with youngsters gaining the capacity to self-regulate the stress response as they develop (Sanchez, McCormack, & Howell, 2015). Thus, it is possible that development and experience might alter the relationships between arousal, vocal production, and glucocorticoid concentrations. Alternatively, it is possible that the relationship between arousal and vocal acoustics is easier to detect statistically in younger monkeys simply because they generally showed stronger HPA activation in response to the HIT than older monkeys (Figure 2.3) (presumably due to the greater importance of maternal buffering at younger ages).

From an ultimate perspective, it is worth considering the possibility that the functions of coos might change over development, and that selection might therefore have favored ontogenetic changes in the relationships between arousal and the acoustic structure of coos. For example, since separation presumably represents a more significant threat to fitness in vulnerable infants than adults, it is plausible that a positive link between arousal and F0 modulation within coos could be more beneficial for infants, as it might make the calls more localizable, increasing the odds of reunion with the mother or other group mates (Lingle et al., 2012; Sugiura, 2007); this function might become less important as monkeys age, and relaxed selection on coo acoustics in older monkeys might have resulted in an ontogenetic dampening of the effects of arousal on coo acoustics. On the other hand, if anything, the results of the present study point to the complex relationships between arousal and vocal acoustics, and therefore caution against overly simplistic adaptationist thinking—or at least highlight the wide range of possible evolutionary processes that should be considered to explain the acoustic structures of vocalizations and the development thereof.

Conclusions

The results of the present study support one clear conclusion: rhesus macaque coos and screams alike show evidence of a positive association between arousal and F0, despite differences between the contexts in which these vocalizations were recorded. This similarity likely represents homology with other mammalian taxa. It also suggests that any selection pressures acting on emotional arousal in rhesus macaques, or indeed in other mammals, should be expected to produce correlated evolutionary responses in the F0 of vocalizations, and similarly, selection pressures acting on the F0 of vocalizations might yield or be constrained by correlated responses in other emotional behaviors or processes. Thus, vocal F0 should be viewed as integrated at the proximate level with the other outputs, and physiological components, of emotions, and potentially constrained in their ability to evolve independently from one another. On the other hand, contrasts between coos and screams raise interesting, if currently speculative, possibilities regarding the interplay between emotion, natural selection, and vocal production. Namely, perhaps the different functions of rhesus macaque screams and coos, within the disparate contexts in which they occur, have resulted in the evolution of adaptive differences in the effects of arousal and/or age on these two vocalization types in their respective contexts. These hypotheses require further study before any definitive adaptive conclusions can be drawn, but they illustrate how future research and discussion on the evolution of the acoustic structure of animal vocalizations stand to benefit from careful consideration of proximate mechanisms, especially the role of emotional states.

References

Adolphs, R., & Andler, D. (2018). Investigating emotions as functional states distinct from feelings. *Emotion Review*, *10*(3), 191–201. https://doi.org/10.1177/1754073918765662

Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, *49*(3), 227–267. https://doi.org/10.1080/14794802.2011.585831

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

Bachorowski, J.-A., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, *6*(4), 219–224. https://doi.org/10.1111/j.1467-9280.1995.tb00596.x

Baciadonna, L., Briefer, E. F., Favaro, L., & McElligott, A. G. (2019). Goats distinguish between positive and negative emotion-linked vocalisations. *Frontiers in Zoology*, *16*(1), 25. https://doi.org/10.1186/s12983-019-0323-z

Bastian, A., & Schmidt, S. (2008). Affect cues in vocalizations of the bat, *Megaderma lyra* , during agonistic interactions. *The Journal of the Acoustical Society of America*, *124*(1), 598–608. https://doi.org/10.1121/1.2924123

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bayart, F., Hayashi, K. T., Faull, K. F., Barchas, J. D., & Levine, S. (1990). Influence of maternal proximity on behavioral and physiological responses to separation in infant rhesus monkeys (*Macaca mulatta*). *Behavioral Neuroscience*, *104*(1), 98–107. https://doi.org/10.1037/0735-7044.104.1.98

Bercovitch, F. B., Hauser, M. D., & Jones, J. H. (1995). The endocrine stress response and alarm vocalizations in rhesus macaques. *Animal Behaviour*, *49*(6), 1703–1706. https://doi.org/10.1016/0003-3472(95)90093-4

Blumstein, D. T., & Chi, Y. Y. (2012). Scared and less noisy: Glucocorticoids are associated with alarm call entropy. *Biology Letters*, *8*(2), 189–192. https://doi.org/10.1098/rsbl.2011.0832

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.51). Retrieved January 9, 2018, from http://www.praat.org/

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *7*(1), 1–20. https://doi.org/10.1098/rsbl.2011.0832

Briefer, E. F. (2018). Vocal contagion of emotions in non-human animals. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873), 20172783. https://doi.org/10.1098/rspb.2017.2783

Briefer, E. F., Linhart, P., Policht, R., Špinka, M., Leliveld, L. M. C., Düpjan, S., … Tallet, C. (2019). Vocal expression of emotional valence in pigs across multiple call types and contexts. *PeerJ Preprints*. https://doi.org/10.7287/peerj.preprints.27934v1

Briefer, E. F., Maigrot, A. L., Mandel, R., Freymond, S. B., Bachmann, I., & Hillmann, E. (2015). Segregation of information about emotional arousal and valence in horse whinnies. *Scientific Reports*, *4*, 1–11. https://doi.org/10.1038/srep09989

Briefer, E. F., Tettamanti, F., & McElligott, A. G. (2015). Emotions in goats: Mapping physiological, behavioural and vocal profiles. *Animal Behaviour*, *99*, 131–143. https://doi.org/10.1016/J.ANBEHAV.2014.11.002

Briefer, E. F., Vizier, E., Gygax, L., & Hillmann, E. (2019). Expression of emotional valence in pig closed-mouth grunts: Involvement of both source- and filter-related parameters. *The Journal of the Acoustical Society of America*, *145*(5), 2895–2908. https://doi.org/10.1121/1.5100612

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Mächler, M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.3929/ethz-b-000240890

Center for Conservation Bioacoustics. (2014). Raven Pro: Interactive sound analysis software (version 1.5) [Computer software]. Retrieved January 7, 2020, from http://ravensoundsoftware.com/

Charmandari, E., Tsigos, C., & Chrousos, G. (2005). Endocrinology of the stress response. *Annual Review of Physiology*, *67*(1), 259–284. https://doi.org/10.1146/annurev.physiol.67.040403.120816

Cheney, D. L. (1977). The acquisition of rank and the development of reciprocal alliances among free-ranging immature baboons. *Behavioral Ecology and Sociobiology*, *2*(3), 203–318. https://doi.org/10.1007/BF00299742

Collins, K. T., McGreevy, P. D., Wheatley, K. E., & Harcourt, R. G. (2011). The influence of behavioural context on Weddell seal (*Leptonychotes weddellii*) airborne mother-pup vocalisation. *Behavioural Processes*, *87*(3), 286–290. https://doi.org/10.1016/j.beproc.2011.06.005

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology*, *22*(2), 142–146. https://doi.org/10.1016/J.CUB.2011.11.053

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Marry. https://doi.org/10.1037/h0076058

de Waal, F. B. M. (2011). What is an animal emotion? *Annals of the New York Academy of Sciences*, *1224*(1), 191–206. https://doi.org/10.1111/j.1749-6632.2010.05912.x

de Waal, F. B. M., & Bauers, K. A. (1991). “Coo” vocalizations in stumptailed macaques: A controlled functional analysis. *Behaviour*, *119*(1–2), 143–160. https://doi.org/10.1163/156853991x00418

de Waal, F. B. M., & Van Hooff, J. A. R. A. M. (1981). Side-directed communication and agonistic interactions in chimpanzees. *Behaviour*, *77*(3), 164–198. https://doi.org/10.1163/156853981X00211

Devenport, L., Knehans, A., Sundstrom, A., & Thomas, T. (1989). Corticosterone’s dual metabolic actions. *Life Sciences*, *45*(15), 1389–1396. https://doi.org/10.1016/0024-3205(89)90026-X

Egnor, S. E. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, *27*(11), 649–654. https://doi.org/10.1016/J.TINS.2004.08.009

Engqvist, L. (2005). The mistreatment of covariate interaction terms in linear model analyses of behavioural and evolutionary ecology studies. *Animal Behaviour*, *70*(4), 967–971. https://doi.org/10.1016/j.anbehav.2005.01.016

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*(4), 917–925. https://doi.org/10.1016/J.ANBEHAV.2010.01.005

Faragó, T., Takács, N., Miklósi, Á., & Pongrácz, P. (2017). Dog growls express various contextual and affective content for human listeners. *Royal Society Open Science*, *4*(5), 170134. https://doi.org/10.1098/rsos.170134

Fichtel, C., Hammerschmidt, K., & Jürgens, U. (2001). On the vocal expression of emotion. A multi-parametric analysis of different states of aversion in the squirrel monkey. *Behavior*, *138*, 97–116. https://doi.org/doi:10.1163/15685390151067094

Fichtel, C., & Manser, M. (2010). Vocal communication in social groups. In P. Kappeler (Ed.), *Animal Behaviour: Evolution and Mechanisms* (pp. 29–54). Heidelberg, Heidelberg: Springer Berlin. https://doi.org/10.1007/978-3-642-02624-9

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, 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: Biological Sciences*, *284*(1859), 20170990. https://doi.org/10.1098/rspb.2017.0990

Filippi, P., Hoeschele, M., Spierings, M., & Bowling, D. L. (2019). Temporal modulation in speech, music, and animal vocal communication: Evidence of conserved function. *Annals of the New York Academy of Sciences*, nyas.14228. https://doi.org/10.1111/nyas.14228

Fischer, J. (2017). Primate vocal production and the riddle of language evolution. *Psychonomic Bulletin and Review*, *24*(1), 72–78. https://doi.org/10.3758/s13423-016-1076-8

Fischer, J., Noser, R., & Hammerschmidt, K. (2013). Bioacoustic field research: A primer to acoustic analyses and playback experiments with primates. *American Journal of Primatology*, *75*(7), 643–663. https://doi.org/10.1002/ajp.22153

Fischer, J., & Price, T. (2016). Meaning, intention, and inference in primate vocal communication. *Neuroscience and Biobehavioral Reviews*, *82*, 22–31. https://doi.org/10.1016/j.neubiorev.2016.10.014

Fischer, J., Wadewitz, P., & Hammerschmidt, K. (2016). Structural variability and communicative complexity in acoustic communication. *Animal Behaviour*, *134*, 229–237. https://doi.org/10.1016/j.anbehav.2016.06.012

Fitch, W. T., de Boer, B., Mathur, N., & Ghazanfar, A. A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, *2*(12), e1600723. https://doi.org/10.1126/sciadv.1600723

Fitch, W. T., & Hauser, M. D. (2006). Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic communication* (pp. 65–137). New York: Springer. https://doi.org/10.1007/0-387-22762-8_3

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, *63*(3), 407–418. https://doi.org/10.1006/anbe.2001.1912

Friel, M., Kunc, H. P., Griffin, K., Asher, L., & Collins, L. M. (2019). Positive and negative contexts predict duration of pig vocalisations. *Scientific Reports*, *9*(1), 2062. https://doi.org/10.1038/s41598-019-38514-w

Fugate, J. M. B., Gouzoules, H., & Nygaard, L. C. (2008). Recognition of rhesus macaque (*Macaca mulatta*) noisy screams: Evidence from conspecifics and human listeners. *American Journal of Primatology*, *70*(6), 594–604. https://doi.org/10.1002/ajp.20533

Fulkerson, W. J., & Tang, B. Y. (1979). Ultradian and circadian rhythms in the plasma concentration of cortisol in sheep. *Journal of Endocrinology*, *81*(1), 135–141. https://doi.org/10.1677/joe.0.0810135

Gogoleva, S. S., Volodina, E. V., Volodin, I. A., Kharlamova, A. V., & Trut, L. N. (2010). The gradual vocal responses to human-provoked discomfort in farmed silver foxes. *Acta*

Ethologica, *13*(2), 75–85. https://doi.org/10.1007/s10211-010-0076-3

Gottlieb, D. H., & Capitanio, J. P. (2013). Latent variables affecting behavioral response to the Human Intruder Test in infant rhesus macaques (*Macaca mulatta*). *American Journal of Primatology*, *75*(4), 314–323. https://doi.org/10.1002/ajp.22107

Gouzoules, H. (2005). Monkeying around with symbolism: Are vocalizations simple symbols...or more like cymbals? In L. Namy (Ed.), *Symbol use and symbolic representation: Developmental and comparative perspectives.* (pp. 245–265). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Gouzoules, H., & Gouzoules, S. (1989). Design-features and developemental modificantion of pigtail macaque, *Macaca nemestrina*, agonistic screams. *Animal Behaviour*, *37*, 383–401.

Gouzoules, H., & Gouzoules, S. (1995). Recruitment screams of pigtail monkeys (*Macaca nemestrina*): Ontogenetic perspectives. *Behaviour*, *132*(5), 431–450. https://doi.org/10.1163/156853995X00658

Gouzoules, H., & Gouzoules, S. (2000). Agonistic screams differ among four species of macaques: The significance of motivation-structural rules. *Animal Behaviour*, *59*(3), 501–512. https://doi.org/10.1006/anbe.1999.1318

Gouzoules, H., Gouzoules, S., & Marler, P. (1985). External reference and affective signaling in mammalian vocal communication. In G. Zivin (Ed.), *The development of expressive behavior: Biology environment interactions* (pp. 77–101). New York: Academic Press.

Gouzoules, H., Gouzoules, S., & Tomaszycki, M. (1998). Agonistic screams and the classification of dominance relationships: Are monkeys fuzzy logicians? *Animal Behaviour*, *55*(1), 51–60. https://doi.org/10.1006/anbe.1997.0583

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams:

Representational signalling in the recruitment of agonistic aid. *Animal Behaviour*, *32*(1), 182–193. https://doi.org/10.1016/S0003-3472(84)80336-X

Green, S. (1975). Variation in vocal pattern with social situation in the Japanese monkey (*Macaca fuscata*): A field study. In L. A. Rosenblum (Ed.), *Primate behavior* (pp. 1–102). New York: Academic Press.

Hamel, A. F., Lutz, C. K., Coleman, K., Worlein, J. M., Peterson, E. J., Rosenberg, K. L., … Meyer, J. S. (2017). Responses to the Human Intruder Test are related to hair cortisol phenotype and sex in rhesus macaques (*Macaca mulatta*). *American Journal of Primatology*, *79*(1), 1–10. https://doi.org/10.1002/ajp.22526

Hammerschmidt, K., Newman, J. D., & Champoux, M. (2000). Changes in rhesus macaque “coo” vocalizations during early development. *Ethology*, *886*, 873–886.

Hansen, E. W. (1976). Selective responding by recently separated juvenile rhesus monkeys to the calls of their mothers. *Developmental Psychobiology*, *9*(1), 83–88. https://doi.org/10.1002/dev.420090111

Hauser, M. D. (1991). Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations. *Ethology*, *89*(1), 29–46. https://doi.org/10.1111/j.1439-0310.1991.tb00291.x

Hauser, M. D., Evans, C. S., & Marler, P. (1993). The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. *Animal Behaviour*, *45*(3), 423–433. https://doi.org/10.1006/anbe.1993.1054

Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology*, *4*(3), 194–205. https://doi.org/10.1093/beheco/4.3.194

Jovanovic, T., & Gouzoules, H. (2001). Effect of nonmaternal restraint on the vocalizations of infant rhesus monkeys (*Macaca mulatta*). *American Journal of Primatology*, *53*(1), 33–45.

Jovanovic, T., & Maestripieri, D. (2010). Effects of early traumatic experience on vocal expression of emotion in young female rhesus macaques. *Developmental Psychobiology*, *52*(8), 794–801. https://doi.org/10.1002/dev.20475

Kalin, N. H., Larson, C., Shelton, S. E., & Davidson, R. J. (1998). Asymmetric frontal brain activity, cortisol, and behavior associated with fearful temperament in rhesus monkeys. *Behavioral Neuroscience*, *112*(2), 286–292. https://doi.org/10.1037/0735-7044.112.2.286

Kalin, N. H., & Shelton, S. E. (1989). Defensive behaviors in infant rhesus monkeys: Environmental cues and neurochemical regulation. *Science*, *243*(4899), 1718–1721. https://doi.org/10.1126/science.2564702

Kalin, N. H., Shelton, S. E., Rickman, M., & Davidson, R. J. (1998). Individual differences in freezing and cortisol in infant and mother rhesus monkeys. *Behavioral Neuroscience*, *112*(1), 251–254. https://doi.org/10.1037//0735-7044.112.1.251

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Le Prell, C. G., Hauser, M. D., & Moody, D. B. (2002). Discrete or graded variation within rhesus monkey screams? Psychophysical experiments on classification. *Animal Behaviour*, *63*(1), 47–62. https://doi.org/10.1006/anbe.2001.1888

LeDoux, J. E., & Brown, R. (2017). A higher-order theory of emotional consciousness. *Proceedings of the National Academy of Sciences*, *114*(10), E2016-E2025. https://doi.org/10.1073/pnas.1619316114

Lehoczki, F., Szamosvölgyi, Z., Miklósi, Á., & Faragó, T. (2019). Dogs' sensitivity to strange pup separation calls: Pitch instability increases attention regardless of sex and experience. *Animal Behaviour*, *153*, 115–129. https://doi.org/10.1016/J.ANBEHAV.2019.05.010

Leliveld, L. M. C., Düpjan, S., Tuchscherer, A., & Puppe, B. (2017). Vocal correlates of emotional reactivity within and across contexts in domestic pigs (*Sus scrofa*). *Physiology and Behavior*, *181*, 117–126. https://doi.org/10.1016/j.physbeh.2017.09.010

Lemasson, A., Remeuf, K., Rossard, A., & Zimmermann, E. (2012). Cross-taxa similarities in affect-induced changes of vocal behavior and voice in arboreal monkeys. *PLoS ONE*, *7*(9), e45106. https://doi.org/10.1371/journal.pone.0045106

Lenth, R. V. (2016). Least-Squares Means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01

Lesnoff, M., & Lacelot, R. (2012). aod: Analysis of Overdispersed Data. R package version 1.3.1. Retrieved August 26, 2019, from https://cran.r-project.org/web/packages/aod.html

Liao, D. A., Zhang, Y. S., Cai, L. X., & Ghazanfar, A. A. (2018). Internal states and extrinsic factors both determine monkey vocal production. *Proceedings of the National Academy of Sciences*, *115*(15), 3978–3983. https://doi.org/10.1073/pnas.1722426115

Lingle, S., Wyman, M. T., Kotrba, R., Teichroeb, L. J., & Romanow, C. A. (2012). What makes a cry a cry? A review of infant distress vocalizations. *Current Zoology*, *58*(5), 698–726. https://doi.org/10.1093/czoolo/58.5.698

Linhart, P., Ratcliffe, V. F., Reby, D., & Špinka, M. (2015). Expression of emotional arousal in two different piglet call types. *PLoS ONE*, *10*(8), 1–13. https://doi.org/10.1371/journal.pone.0135414

Maestripieri, D., Jovanovic, T., & Gouzoules, H. (2000). Crying and infant abuse in rhesus monkeys. *Child Development*, *71*(2), 301–309. https://doi.org/10.1111/1467-8624.00145

Maigrot, A.-L., Hillmann, E., & Briefer, E. F. (2018). Encoding of emotional valence in wild boar (*Sus scrofa*) calls. *Animals*, *8*(6), 85. https://doi.org/10.3390/ani8060085

McCormack, K., Newman, T. K., Higley, J. D., Maestripieri, D., & Sanchez, M. M. (2009). Serotonin transporter gene variation, infant abuse, and responsiveness to stress in rhesus macaque mothers and infants. *Hormones and Behavior*, *55*(4), 538–547. https://doi.org/10.1016/j.yhbeh.2009.01.009

Meise, K., Keller, C., Cowlishaw, G., & Fischer, J. (2011). Sources of acoustic variation: Implications for production specificity and call categorization in chacma baboon (*Papio ursinus*) grunts. *The Journal of the Acoustical Society of America*, *129*(3), 1631–1641. https://doi.org/10.1121/1.3531944

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

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855–869. https://doi.org/10.1086/283219

Nesse, R. M. (1990). Evolutionary explanations of emotions. *Human Nature*, *1*(3), 261–289. https://doi.org/10.1007/BF02733986

Norcross, J. L., & Newman, J. D. (1993). Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls. *American Journal of Primatology*, *30*(1), 37–54. https://doi.org/10.1002/ajp.1350300104

Ordóñez-Gómez, J. D., Santillan-Doherty, A. M., & Hammerschmidt, K. (2019). Acoustic variation of spider monkey (*Ateles geoffroyi*) contact calls is related to caller isolation and affects listeners’ responses. *PLOS ONE*, *14*(4), e0213914. https://doi.org/10.1371/journal.pone.0213914

Owren, M. J. (2008). GSU Praat tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods*, *40*(3), 822–829. https://doi.org/10.3758/BRM.40.3.822

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, *73*(6), 530–544. https://doi.org/10.1002/ajp.20913

Panksepp, J. (2011). The basic emotional circuits of mammalian brains: Do animals have affective lives? *Neuroscience and Biobehavioral Reviews*, *35*(9), 1791–1804. https://doi.org/10.1016/j.neubiorev.2011.08.003

Patel, E. R., & Owren, M. J. (2007). Acoustics and behavioral contexts of “gecker” vocalizations in young rhesus macaques (Macaca mulatta). *The Journal of the Acoustical Society of America*, *121*(1), 575–585. https://doi.org/10.1121/1.2390662

Paul, E. S., & Mendl, M. T. (2018). Animal emotion: Descriptive and prescriptive definitions and their implications for a comparative perspective. *Applied Animal Behaviour Science*, *205*, 202–209. https://doi.org/10.1016/j.applanim.2018.01.008

Pfefferle, D., Hammerschmidt, K., Mundry, R., Ruiz-Lambides, A. V., Fischer, J., & Widdig, A. (2016). Does the structure of female rhesus macaque coo calls reflect relatedness and/or familiarity? *PLoS ONE*, *11*(8), e0161133. https://doi.org/10.1371/journal.pone.0161133

Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific Reports*, *5*, 1–11. https://doi.org/10.1038/srep13220

Puppe, B., Schön, P. C., Tuchscherer, A., & Manteuffel, G. (2005). Castration-induced vocalisation in domestic piglets, *Sus scrofa*: Complex and specific alterations of the vocal quality. *Applied Animal Behaviour Science*, *95*(1–2), 67–78. https://doi.org/10.1016/j.applanim.2005.05.001

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org/

Ralph, C. R., & Tilbrook, A. J. (2016). Invited Review: The usefulness of measuring glucocorticoids for assessing animal welfare. *Journal of Animal Science*. https://doi.org/10.2527/jas2015-9645

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *The Journal of the Acoustical Society of America*, *113*(6), 3390–3402. https://doi.org/10.1121/1.1568942

Rendall, D., Owren, M. J., & Rodman, P. S. (1998). The role of vocal tract filtering in identity cueing in rhesus monkey (*Macaca mulatta*) vocalizations. *The Journal of the Acoustical Society of America*, *103*(1), 602–614. https://doi.org/10.1121/1.421104

Rendall, D., Rodman, P. S., & Emond, R. E. (1996). Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Animal Behaviour*, *51*(5), 1007–1015. https://doi.org/10.1006/anbe.1996.0103

Riede, T. (2010). Elasticity and stress relaxation of rhesus monkey (*Macaca mulatta*) vocal folds. *Journal of Experimental Biology*, *213*, 2924–2932. https://doi.org/10.1242/jeb.044404

Romero, L. M. (2004). Physiological stress in ecology: Lessons from biomedical research. *Trends in Ecology & Evolution*, *19*(5), 249–255.

https://doi.org/10.1016/J.TREE.2004.03.008

Rowell, T. E., & Hinde, R. A. (1962). Vocal communication by the rhesus monkey *Macaca mulatta*. *Proceedings of the Royal Society of London*, *138*, 279–294.

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

Sanchez, M. M., McCormack, K. M., & Howell, B. R. (2015). Social buffering of stress responses in nonhuman primates: Maternal regulation of the development of emotional regulatory brain circuits. *Social Neuroscience*, *10*(5), 512–526. https://doi.org/10.1080/17470919.2015.1087426

Scarantino, A., & Clay, Z. (2015). Contextually variable signals can be functionally referential. *Animal Behaviour*, *100*, e1-8. https://doi.org/10.1016/j.anbehav.2014.08.017

Scarantino, A., & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review*, *3*(4), 444–454. https://doi.org/10.1177/1754073911410745

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: A new take on ultimate and proximate influences in vocal production. *Biological Reviews*, *93*(4), 2071–2082. https://doi.org/10.1111/brv.12437

Schehka, S., & Zimmermann, E. (2009). Acoustic features to arousal and identity in disturbance calls of tree shrews (*Tupaia belangeri*). *Behavioural Brain Research*, *203*(2), 223–231. https://doi.org/10.1016/j.bbr.2009.05.007

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE*, *8*(10), e76674. https://doi.org/10.1371/journal.pone.0076674

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research.

Psychological Bulletin. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

Schrader, L., & Todt, D. (1998). Vocal quality is correlated with levels of stress hormones in domestic pigs. *Ethology*, *104*(10), 859–876. https://doi.org/10.1111/j.1439-0310.1998.tb00036.x

Schwartz, J. W., & Gouzoules, H. (2019). Decoding human screams: Perception of emotional arousal from pitch and duration. *Behaviour*, *156*(13–14), 1283–1307. https://doi.org/10.1163/1568539X-00003566

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (in press). Evolving views on cognition in animal communication: Contributions from scream research. *Animal Behavior and Cognition.*

Sèbe, F., Poindron, P., Andanson, S., Chandeze, H., Delval, E., Despres, G., … Boissy, A. (2012). Bio-acoustic analyses to assess emotion in animals: Acoustic patterns are linked to behavioural, cardiac and hormonal responses of ewes to the separation from their lambs. *Bioacoustics*, *21*(1), 54.

Seyfarth, R. M., & Cheney, D. L. (2003a). Meaning and emotion in animal vocalizations. *Annals of the New York Academy of Sciences*, *1000*, 32–55. https://doi.org/10.1196/annals.1280.004

Seyfarth, R. M., & Cheney, D. L. (2003b). Signalers and receivers in animal communication. *Annual Review of Psychology*, *54*(1), 145–173. https://doi.org/10.1146/annurev.psych.54.101601.145121

Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and Language*, *115*(1), 92–100. https://doi.org/10.1016/j.bandl.2009.10.003

Seyfarth, R. M., & Cheney, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour*, *124*, 339–346. https://doi.org/10.1016/j.anbehav.2016.05.020

Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K., & Hammerschmidt, K. (2010). The central importance of information in studies of animal communication. *Animal Behaviour*, *80*(1), 3–8. https://doi.org/10.1016/j.anbehav.2010.04.012

Siebert, K., Langbein, J., Schön, P. C., Tuchscherer, A., & Puppe, B. (2011). Degree of social isolation affects behavioural and vocal response patterns in dwarf goats (*Capra hircus*). *Applied Animal Behaviour Science*, *131*(1–2), 53–62. https://doi.org/10.1016/j.applanim.2011.01.003

Simons, R. C., & Bielert, C. F. (1973). An experimental study of vocal communication between mother and infant monkeys (*Macaca nemestrina*). *American Journal of Physical Anthropology*, *38*(2), 455–461. https://doi.org/10.1002/ajpa.1330380248

Stoeger, A. S., Charlton, B. D., Kratochvil, H., & Fitch, W. T. (2011). Vocal cues indicate level of arousal in infant African elephant roars. *The Journal of the Acoustical Society of America*, *130*(3), 1700–1710. https://doi.org/10.1121/1.3605538

Sugiura, H. (2007). Effects of proximity and behavioral context on acoustic variation in the coo calls of Japanese macaques. *American Journal of Primatology*, *69*, 1412–1424. https://doi.org/10.1002/ajp.20447

Szipl, G., Ringler, E., Spreafico, M., & Bugnyar, T. (2017). Calls during agonistic interactions vary with arousal and raise audience attention in ravens. *Frontiers in Zoology*, *14*(1), 1–13. https://doi.org/10.1186/s12983-017-0244-7

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Tomaszycki, M. L., Davis, J. E., Gouzoules, H., & Wallen, K. (2001). Sex differences in infant rhesus macaque separation-rejection vocalizations and effects of prenatal androgens. *Hormones and Behavior*, *39*(4), 267–276. https://doi.org/10.1006/HBEH.2001.1659

Turner, A. I., Rivalland, E. T. A., Clarke, I. J., & Tilbrook, A. J. (2010). Stressor specificity of sex differences in hypothalamo-pituitary-adrenal axis activity: cortisol responses to exercise, endotoxin, wetting, and isolation/retraint stress in gonadectomized male and female sheep. *Endocrinology*, *151*(9), 4324–4331. https://doi.org/10.1210/en.2010-0234

Venables, W., & Ripley, B. (2013). *Modern applied statistics with S-PLUS*. Springer.

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires - Hard vs. soft classification approaches. *PLoS ONE*, *10*(4), e0125785. https://doi.org/10.1371/journal.pone.0125785

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology*, *21*(5), 195–205. https://doi.org/10.1002/evan.21319

Wheeler, B. C., & Fischer, J. (2015). The blurred boundaries of functional reference: A response to Scarantino & Clay. *Animal Behaviour*, *100*, 9–13. https://doi.org/10.1016/j.anbehav.2014.11.007

Yamaguchi, C., Izumi, A., & Nakamura, K. (2010). Time course of vocal modulation during isolation in common marmosets (*Callithrix jacchus*). *American Journal of Primatology*, *72*(8), 681–688. https://doi.org/10.1002/ajp.20824

Yeon, S. C., Kim, Y. K., Park, S. J., Lee, S. S., Lee, S. Y., Suh, E. H., … Lee, H. J. (2011). Differences between vocalization evoked by social stimuli in feral cats and house cats.

Behavioural Processes, *87*(2), 183–189. https://doi.org/10.1016/j.beproc.2011.03.003

Zei Pollermann, B., & Archinard, M. (2002). Acoustic patterns of emotions. In E. Keller, G. Bailly, A. Monaghan, J. Terken, & M. Huckvale (Eds.), *Improvements in speech synthesis* (pp. 237–245). Chichester, England: John Wiley & Sons.

Zimmermann, E., Leliveld, L. M. C., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: A comparative approach to mammalian voices. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of emotional communication: From sounds in nonhuman mammals to speech and music in men* (pp. 117–132). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199583560.001.0001

CHAPTER III: HUMANS ACCURATELY PERCEIVE EMOTIONAL AROUSAL FROM ACOUSTIC VARIATION WITHIN TWO RHESUS MACAQUE (*MACACA MULATTA*) VOCALIZATION TYPES

Abstract

The hypothesis that the ways in which humans express emotions are evolutionarily conserved dates back to Darwin. One approach to studying the evolutionary history of emotional vocal communication is to test the capacity to perceive the emotional significance of the vocalizations of other species. Few studies have examined human perception of emotion from acoustic variation within a heterospecific vocalization type, and even fewer have done so with vocalizations from a species with which most listeners are unfamiliar. Using a binary forced choice task, we tested perception of the emotional intensity represented in coos and screams of infant and juvenile female rhesus macaques (*Macaca mulatta*) by 113 human listeners without, and 12 listeners with, experience (as researchers or care technicians) with this species. Each stimulus pair contained one high- and one low-arousal vocalization, as measured at the point of recording by stress hormone levels for coos and the degree of intensity of agonism for screams. For coos as well as screams, both inexperienced and experienced participants accurately identified the high-arousal vocalization at significantly above-chance rates. Experience was associated with significantly greater accuracy with scream stimuli but not coo stimuli, and with a tendency to indicate screams as reflecting greater emotional intensity than coos. Neither empathy, human emotion recognition, nor attitudes toward animal welfare showed any relationship with responses. Participants were sensitive to the fundamental frequency, noisiness, and duration of vocalizations; some of these tendencies likely facilitated accurate perceptions, perhaps due to evolutionary homologies in the physiology of arousal and vocal production between humans and macaques. Overall, our findings support a view of evolutionary continuity in emotional vocal communication.

Introduction

The hypothesis that the ways in which humans express emotions are evolutionarily conserved, and shared with other species due to common descent, dates back to Darwin (1872). Comparing the acoustic correlates of emotional states—specifically, emotional arousal (activation of the sympathetic nervous system in association with intense emotions; Mendl, Burman, & Paul, 2010; Russell, 2003)—in human speech and nonlinguistic vocalizations to those in animal calls suggests broad cross-species similarities, and thus supports a view of phylogenetic continuity in vocal emotion expression (Briefer, 2012; Zimmermann, Leliveld, & Schehka, 2013). Specifically, the fundamental frequency (F0; commonly perceived as pitch) of vocalizations appears to consistently correlate positively with emotional arousal across taxa, likely due to tensing of the vocal fold muscles increasing the rate of oscillation (Briefer, 2012; Riede, 2010; Scherer, 1986; Titze, 1994). Other acoustic correlates of emotional arousal include F0 modulation, vocal duration, and noisiness (Briefer, 2012). The effects of arousal on these acoustic parameters likely reflect evolutionary homologies in the mechanisms of arousal and vocal production.

Communication is inherently two-faceted, involving signal production by senders as well as perception by receivers. The hypothesis that vocal emotion expression shows evolutionarily homologous similarities across species predicts that listeners should be sensitive to the acoustic cues to the emotional states of conspecific and heterospecific vocalizers alike, depending in part on the phylogenetic distance between the two species (though see Filippi, Congdon, et al., 2017). One approach to testing this prediction is to examine human perceptions of emotion from heterospecific vocalizations (Belin et al., 2008; Faragó et al., 2014; Faragó, Takács, Miklósi, & Pongrácz, 2017; Filippi, Congdon, et al., 2017; Filippi, Gogoleva, Volodina, Volodin, & de

Boer, 2017; Kelly et al., 2017; Linnankoski, Laakso, Aulanko, & Leinonen, 1994; Maruščáková et al., 2015; McComb, Taylor, Wilson, & Charlton, 2009; Nicastro & Owren, 2003; Pongrácz, Molnár, Miklósi, & Csányi, 2005; Scheumann, Hasting, Kotz, & Zimmermann, 2014; Tallet, Špinka, Maruščáková, & Šimeček, 2010). These studies have generally confirmed human sensitivity to emotional information available in heterospecific vocalizations, and thus support Darwin's (1872) hypothesis that significant aspects of human emotional expression are evolutionarily homologous with those of other species.

Listeners have tended to base perceptions of emotional arousal in heterospecific vocalizations predominantly on the mean F0 (Faragó et al., 2014, 2017; Filippi, Congdon, et al., 2017; Kelly et al., 2017; Maruščáková et al., 2015; Nicastro & Owren, 2003; Pongrácz et al., 2005), consistent with findings of studies examining perception of emotion in human vocalizations (Faragó et al., 2014; Juslin & Laukka, 2003; Sauter, Eisner, Calder, & Scott, 2010; Schwartz & Gouzoules, 2019). Other significant acoustic parameters have included vocal duration (Faragó et al., 2014, 2017; Maruščáková et al., 2015; Nicastro & Owren, 2003) and noisiness (Faragó et al., 2014; Filippi, Gogoleva, et al., 2017). Insofar as these acoustic parameters naturally vary according to vocalizer arousal (Briefer, 2012), these perceptual tendencies are generally credited with facilitating accurate perceptions of emotion from heterospecific vocalizations (for an exception, see Kelly et al., 2017).

A complicating factor is that most studies in this area have included multiple vocalization types as stimuli, making it difficult to know whether listeners were reacting to acoustic variation *within* a single vocalization type, or *between* vocalization types (for more on this distinction, see Fischer, Wadewitz, & Hammerschmidt, 2016). Distinct vocalization types typically serve disparate social functions and potentially reflect differing cognitive processes, whereas acoustic variation within a vocalization type more reliably relates to the vocalizer's emotional state (Schamberg, Wittig, & Crockford, 2018; Schwartz, Engelberg, & Gouzoules, in press). Few studies have investigated human perception of emotion from acoustic variation within a vocalization type of another species (for exceptions, see Faragó et al., 2017; Filippi, Congdon, et al., 2017; Filippi, Gogoleva, et al., 2017; Nicastro & Owren, 2003; Pongrácz et al., 2005).

Most of these studies (Faragó et al., 2017; Nicastro & Owren, 2003; Pongrácz et al., 2005) used vocalizations of domestic cats (*Felis catus*) or dogs (*Canis lupis familiaris*). These, of course, are domesticated animals with which many humans are familiar—and indeed, which might have undergone evolutionary changes in emotional expression to improve communication with humans (Pongrácz, Molnár, & Miklósi, 2010; though see Filippi, Gogoleva, et al., 2017). Thus these studies' findings might not extend to the vocalizations of undomesticated and less familiar species. Indeed, several studies have demonstrated an association between experience with a species and increased accuracy in recognizing the emotional significance of its vocalizations (Faragó et al., 2017; McComb et al., 2009; Nicastro & Owren, 2003; Scheumann et al., 2014; Tallet et al., 2010 though see Pongrcáz et al., 2005). This raises the question of the degree to which human accuracy in heterospecific emotion recognition tasks reflects evolutionary homology, as opposed to a capacity to learn to recognize any individual species' unique emotional expressions through exposure regardless of relatedness. One might expect the degree of evolutionary homology in vocal emotion expression to depend on the phylogenetic distance between taxa (though see Filippi, Congdon, et al., 2017), making nonhuman primates—our closest relatives—an important group in which to test human emotion perception.

The present study examined human perceptions of emotional arousal from acoustic variation within vocalization types of rhesus macaques (*Macaca mulatta*), a cercopithecid primate whose lineage is estimated to have diverged from that of humans c. 30 million years ago (Steiper & Young, 2008). Although rhesus macaques are common in research institutions and in urban areas in South Asia, many humans have little to no direct experience observing or interacting with them, presenting an opportunity to test the degree to which humans can perform fine-grained discriminations within call types of an undomesticated, unfamiliar species. The rhesus macaque vocal repertoire contains a variety of vocalization types including coos—tonal vocalizations associated with social functions ranging from signaling the presence of food to reuniting separated kin (Bayart, Hayashi, Faull, Barchas, & Levine, 1990; Hauser, 1991; Hauser & Marler, 1993)—and screams—high-F0 vocalizations generally uttered during agonistic interactions, functioning to recruit aid from allies, typically matrilineal kin (H. Gouzoules, 2005; H. Gouzoules, Gouzoules, & Tomaszycki, 1998; S. Gouzoules, Gouzoules, & Marler, 1984). A recent study that compared stress hormone levels and coo vocalizations uttered in a formal behavioral test (Human Intruder Test; Kalin & Shelton, 1989), and also compared screams produced during naturally during social interactions that varied in the degree of agonism received by the vocalizer, suggested that both vocalization types exhibit within-type acoustic variation that correlates with the emotional arousal of the caller (Chapter II). Specifically, arousal was associated with decreases in duration and increases in the mean F0, F0 range, and jitter (cycle-to-cycle variation in F0) of coos, and in increases the mean F0 of screams, along with some other related parameters. The present study examined whether human listeners are sensitive to the emotional significance of variation among coos and screams, and assessed the role of experience, other participant characteristics, and acoustic properties.

In two brain imaging studies, humans have failed to accurately classify rhesus macaque vocalizations according to emotional valence (Belin et al., 2008; Fritz et al., 2018), though the different vocalizations did elicit different cerebral responses, suggesting the possibility of unconscious processing (Belin et al., 2008). However, each study compared perceptions of different vocalization types, not of acoustic variants of the same vocalization type. Humans have shown an ability to recognize the individual identity of rhesus macaques based on their noisy screams alone (Fugate, Gouzoules, & Nygaard, 2008), indicating a sensitivity to acoustic variation within a single vocalization type in this species, though not necessarily how it relates to emotion. Based on this, and on previous research showing sensitivity to emotional cues in the vocalizations of other species (including the closely related Barbary macaque, *Macaca sylvanus*; Filippi, Congdon, et al., 2017), we hypothesized that participants would accurately perceive differences in emotional arousal within pairs of coos and of screams. We also examined perceptions of arousal differences between the two call types.

We assessed the effects of experience working as researchers and technicians with rhesus macaques and other animals, to test the hypothesis that familiarity affects sensitivity to the emotional significance of other species' vocalizations (Nicastro & Owren, 2003; Parsons, LeBeau, Kringelbach, & Young, 2019; Tallet et al., 2010). We also examined the effects of empathy (Cambridge Behaviour Scale; Baron-Cohen & Wheelwright, 2004), human emotion recognition (Cambridge Face-Voice Battery; Golan, Baron-Cohen, & Hill, 2006; Golan, Baron-Cohen, Hill, & Rutherford, 2007), and attitudes toward animal welfare (Animal Attitudes Scale; Herzog, Grayson, & McCord, 2015). Although one might hypothesize that such characteristics might correlate with the ability to recognize emotion from heterospecific vocalizations, previous research has provided little support for this notion (Maruščáková et al., 2015). We predicted that mean F0 would play a role in participants' perceptions of arousal, in line with previous research on human emotion perception (Filippi, Congdon, et al., 2017; Kelly et al., 2017; Sauter et al., 2010; Schwartz & Gouzoules, 2019) and with the positive relationship between arousal and the mean F0 of coos and screams (Chapter II). We also examined the role of noisiness and duration of vocalizations, since these have been shown to affect listeners' perceptions of the emotional states of animals (Faragó et al., 2014, 2017; Filippi, Gogoleva, et al., 2017; Maruščáková et al., 2015; Nicastro & Owren, 2003), and F0 range and jitter, which correlate with emotional arousal in rhesus macaque coos (Chapter II).

Methods

Ethical Statement

All animal procedures were approved by the Emory University Institutional Animal Care and Use Committee in accordance with the Animal Welfare Act and the U.S. Department of Health and Human Services "Guide for Care and Use of Laboratory Animals." Procedures involved in sourcing vocalizations used as stimuli for this study took place as part of the Stress, Obesity, and Diabetes research project at the Yerkes National Primate Research Center (YNPRC) (Principle Investigators: M. M. Sanchez & M. E. Wilson), independent of this study of vocal perception. Human testing procedures were approved by and conducted in compliance with the Emory University Institutional Review Board (IRB00102796).

Participants

One hundred thirteen undergraduate students at Emory University and volunteers from the Atlanta area comprised our "inexperienced" sample; each indicated no experience working with or observing rhesus macaques. Emory undergraduate participants were recruited via an online portal, and received credit in an introductory psychology course; other participants contacted the researchers after seeing fliers advertising the study posted around Emory’s campus, and were not compensated in any way. Twelve individuals employed by or affiliated with the YNPRC comprised our “experienced” sample, each reporting a non-zero number of months working with rhesus macaques (mean ± SD = 111 ± 76 months). These participants volunteered after hearing about the study through an email announcement or word-of-mouth from other Yerkes employees.

Informed consent was obtained from all individual participants included in the study. Each participant reported their gender (**inexperienced:** 73% female, 27% male, 0% other; **experienced:** 67% female, 33% male, 0% other), age (**inexperienced** mean ± SD = 20.2 ± 4.2 years; **experienced** = 34.2 ± 8.2 years), and native language (**inexperienced:** 67% English [including 6% natively bilingual], 19% Mandarin/Chinese, 4% Spanish, 3% Korean, 9% other languages [< 2% each]; **experienced:** 100% English [including one natively bilingual participant]). All but one of the experienced participants indicated professional work with other taxa in addition to rhesus macaques, including chimpanzees and sooty mangabeys (housed at Yerkes), pigtailed macaques (formerly housed at Yerkes), other great apes, lemurs, dogs, and cats. A small minority of inexperienced participants (N = 21) indicated professional work with animals, including dogs, cats, and/or rodents. Participants were screened for hearing impairments, and disclosure of an impairment that might interfere with perception of auditory stimuli was considered an exclusionary criterion.

Stimuli

The stimuli used in this study comprised pairs of rhesus macaque vocalizations—either two coos, two screams, or one coo and one scream (hereafter, mixed)—separated by one second of silence. These vocalizations were collected from infant and juvenile female rhesus macaques living in naturalistic social groups within large outdoor enclosures at the YNPRC Field Station in Lawrenceville, GA, using procedures described at length in Chapter II. Briefly, coos of 39 macaques were recorded during Human Intruder Tests (Kalin & Shelton, 1989) consisting of three 10-minute stages—Alone (the macaque was alone in a roomy cage within a testing room), Profile (a technician wearing a lifelike mask sat motionless at a 90-degree angle to the cage), and Stare (the technician sat facing the cage and directing gaze at the macaque but otherwise remained motionless). A Sony DCR-SR85 video camera rig (Tokyo, Japan) (onboard Sony electret condenser microphone) captured video and audio recordings of the tests, including any vocalizations uttered by macaques. Audio was digitized in a 48-kHz, 32-bit .WAV format. Plasma cortisol concentrations were measured from blood sampled obtained immediately before and immediately after each test; the difference between these two measurements served as a rough indicator of each macaque's physiological stress reaction to the test. Screams of 18 macaques (a subset of the 39 from which coos were obtained) were recorded during naturally occurring social interactions within the macaques' home enclosures, using a Sennheiser ME66 directional microphone with MZW66 windshield (Wennebostel, Germany) and a Marantz PMD671 solid state recorder (Tokyo, Japan). Recordings were digitized as 44.1-kHz, 16- or 24-bit .WAV files. Each scream bout recording was accompanied by a note of the vocalizer identity and whether the vocalizer received agonistic physical contact prior to the scream(s).

Vocalizations were extracted from larger recordings and prepared for use as stimuli using Adobe Audition CC (Adobe Systems, San Jose, CA). 48-kHz audio from the Human Intruder Test coo recordings was down-sampled to 44.1 kHz and dithered from 32- to 16-bit depth, and 24-bit scream recordings were dithered to 16-bit depth. The Human Intruder testing room contained a HoMedics SS-2000 noise generator (Commerce Township, MI, USA) set to the “waterfall” setting (brown noise) at high volume, to create a controlled and constant auditory background for testing. Prior to coo extraction, audio was prepared using the "Noise Reduction (process)" function in Audition: a noise profile was generated from a 5-s sample of uninterupted noise occurring immediately after a coo (the closest to the middle of the HIT), and then the whole sound file underwent reduction of noise fitting that profile. The function was set to reduce noise by 50% and by 20 dB, resulting in substantial reduction but not elimination of background noise (Chapter II). The generally very high F0 of the screams (2-10 kHz), permitted removal of background noise at lower frequencies; this was done by selecting noise at lower frequencies using the marquee and lasso selection tools in Audition, and reducing the sound level of the selection. After noise reduction/removal, all vocalizations underwent RMS amplitude normalization (following best practices outlined by Owren & Bachorowski 2007) in Audition, using the equal loudness contour (which weights frequency amplitudes according to their perceptual loudness). As a result, most of the vocalizations’ sound envelopes occupied approximately -12 dB, while the remaining background noise of the coos occupied a sound level of approximately -42 dB. Final waveforms and spectrograms of individual vocalizations were inspected to ensure that they contained no clipping or other distortions.

To construct the coo stimulus pairs, 6-month-old macaques were paired into dyads consisting of one high-arousal macaque and one low-arousal macaque, defined by a difference in cortisol reactivity (i.e., increase in plasma cortisol concentration from before until after the test) of $\geq$ 1.5 SD. No dyad exhibited a difference in baseline cortisol concentrations of > 1 SD. Every dyad meeting these criteria was used to create stimulus pairs, resulting in 28 dyads consisting of 13 individuals. Individual differences rather than arousal levels inevitably accounted for some acoustic differences among coos within stimulus pairs; in principle, acoustic variation due to individual characteristics should be random and should not be expected to have confounded the results. For each individual, three coos were sampled from the Stare stage: the closest to the 2.5-minute mark, the closest to the 5-minute mark, and the closest to the 7.5-minute mark. Then for each macaque dyad, three coo stimulus pairs were constructed: one pairing the two 2.5-minute coos, one pairing the two 5-minute coos, and one pairing the two 7.5-minute coos. This resulted in a total of 84 coo stimulus pairs, each a combination of 39 individual coos. Each coo in each stimulus pair was preceded and followed by a fade-in and fade-out of the background noise lasting 0.3 s.

Each scream stimulus pair consisted of two screams from a single macaque during two separate social interactions—one scream uttered after experiencing agonistic physical contact (high arousal) and one with no contact (low-arousal). To construct these stimuli, tonal screams (S. Gouzoules et al., 1984) were sampled from the 8 macaques from whom tonal screams were recorded during at least one interaction with contact and at least one without. We selected one tonal scream from each bout (i.e., multiple screams emitted in response to a single agonistic interaction) on the basis of recording quality (no clipping, uninterrupted by other vocalizations or loud sounds, e.g., a monkey striking a fence or enrichment structure during a display). All possible combinations of one high-arousal scream and one low-arousal scream from the same individual were exploited to construct stimuli, resulting in 39 stimulus pairs representing 33 screams. Numbers of screams per macaque ranged from 2-9, resulting in 1-18 scream stimulus pairs per macaque.

Each mixed stimulus pair included one scream and one coo; no individual vocalization was repeated across these mixed stimuli. Each of the 33 screams was randomly paired with a coo—either from a subset of the coos comprising the coo stimulus pairs, or randomly chosen from a larger corpus—resulting in 33 mixed stimulus pairs. To account for any potential confounding effects of the presence of background noise in the coo recordings versus absence thereof in the scream recordings, we added pure noise, sampled from the Human Intruder Test (coo) recordings, to the scream recordings to match the frequency profile and sound level of the coo recordings, and included 0.3-s fade-ins and fade-outs before and after the screams to match the coos.

The order of the two vocalizations comprising each stimulus pair was alternated so that an equal number of stimulus pairs contained a high-arousal vocalization first and a low-arousal vocalization second, and that each individual vocalization appeared first equally often as it appeared second. Where applicable, the one second of silence between the two sound files began at the offset of the noise fade-out of the first vocalization, and ended at the onset of the noise fade-in of the second vocalization. The stimulus pairs were re-inspected by ear to ensure that the RMS-amplitude normalization employed earlier had in fact resulted in equal perceptual loudness. In a minority of cases where one vocalization was obviously louder than the other, the amplitude of one was adjusted manually in Audition.

Procedure

Testing took place in the Bioacoustics Laboratory at Emory University (inexperienced participants, and 3 of 12 experienced participants), or in an office in the research building at the YNPRC Field Station (9 of 12 experienced participants). After providing their informed consent and demographic information, inexperienced participants completed two pen-and-paper surveys: the 40-item Cambridge Behavioral Scale (Baron-Cohen & Wheelwright, 2004), which yields a score between 0 and 80 with higher scores indicating greater empathy, and the 10-item brief Animal Attitudes Scale (Herzog et al., 2015), which yields a score between 0 and 50 with higher scores indicating greater concern for animal welfare. Due to time constraints, experienced participants did not complete these surveys.

Subsequent procedures took place on a Dell Optiplex 755 desktop computer (Round Rock, TX) running E-Prime 2.0 software (Psychology Software Tools, Inc., Pittsburgh, PA). Participants were instructed that they would hear a series of pairs of vocalizations and that their task was to choose, for each pair, the vocalization that they thought reflected a more intense emotion. They were given the examples of calmness or boredom as low-intensity emotions, and fear or stress as examples of high-intensity emotions. We used the terminology of intense emotions rather than “arousal” in order to circumvent participants’ potential unfamiliarity with the formal concept of arousal. Arousal has been described as the intensity of an emotional state (e.g., Briefer 2012; though see Russell 2003), and previous research has shown a strong positive correlation between listener ratings of the ‘intensity’ and ‘activation’ (i.e., arousal) of vocal stimuli (Laukka, Juslin, & Bresin, 2005).

Each participant listened to every stimulus pair through headphones (Beyerdynamic DT 770 Pro, Beyerdynamic GmbH & Co., Heilbronn, Germany, or Sony MDR-CD180, Tokyo, Japan). On each trial, they indicated whether they perceived the first or second vocalization to be more emotionally intense by pressing a ‘1’ or ‘2’ key on a peripheral serial response box (model 200a, Psychology Software Tools, Inc., Pittsburgh, PA). Each response triggered a two-second pause followed by the onset of the subsequent stimulus pair. The task was broken into four blocks of stimulus pairs: screams (N = 39 stimulus pairs), mixed (N = 33), one half of the coo stimuli (N = 42), and the other half of the coo stimuli (N = 42). These four blocks were presented in a random order, as were the stimulus pairs within each block. After each block, participants were offered the option to take a break lasting as long as they needed, ending when the participant pressed a key indicating they were ready to continue.

The subset of inexperienced participants reporting English as a first language then completed the Cambridge Mindreading Face-Voice (CAM) battery (Golan et al., 2006, 2007). This 100-item task was presented on the same computer; in it, participants view or listen to human emotional facial or vocal expressions and, for each, choose one of four emotion labels to best describe the expression. It yields a score between 0 and 100, with higher scores indicating greater recognition of human emotional facial and vocal expressions. The CAM battery requires a high level of English language proficiency, hence we limited testing to participants who reported English as a first language. Experienced participants did not complete the CAM battery, due to their job-related time constraints. Finally, after completing all other procedures, experienced participants provided written answers to qualitative questions about their knowledge of rhesus macaque vocal communication and their thought process during the task.

Acoustic Analyses

Measurements of mean F0, min F0, max F0, mean harmonic-noise ratio (HNR, with low values indicating a noisier call), jitter, and duration were obtained in Praat 6.0.29 (Boersma & Weenink, 2013). Spectrograms were generated using fast-Fourier transform with a Gaussian window shape and 2-ms DFT size. Each vocalization was highlighted manually, looking at the spectrogram and waveform and listening to the vocalization, to obtain its duration while excluding any reverberation. F0 measurements were then obtained from the highlighted portion using the Quantify Source command in the GSU Praat Tools package Version 1.9 (Owren, 2008). The default settings for this command were used, with the exception that the pitch ceiling was set to 3000 Hz for coos and 10000 Hz for screams to account for the F0 range of these vocalizations. The command uses Praat's To Pitch autocorrelation function to estimate a F0 contour, which the user then inspects and can manually correct if necessary (e.g., octave correction, removal of any unvoiced segments) (Owren, 2008). We used the relative average perturbation jitter measurement. F0 range was calculated as the difference between max and min F0. Mean ± SD acoustic parameter values were as follows: **coos**: mean F0 = 995 ± 169 Hz, F0 range = 340 ± 214 Hz, mean HNR = 13.5 ± 4.32 dB, jitter = $9.07\text{x}10^{-5} \pm 1.60\text{x}10^{-4}$, duration = 0.580 ± 0.145 s; **screams**: mean F0 = 4879 ± 1447 Hz, F0 range = 3533 ± 2029 Hz, mean HNR = 0.78 ± 3.31 Hz, jitter = $2.39\text{x}10^{-5} \pm 4.18\text{x}10^{-5}$, duration = 0.504 ± 0.211 s. For each stimulus pair, the relative difference in mean F0, F0 range, mean HNR, duration, and jitter between the two vocalizations was calculated as the value for the second vocalization minus the value for the first vocalization, divided by the value for the first vocalization; similarly, the relative difference between the high- and low-arousal vocalization was calculated as the value for the high-arousal vocalization minus the value for the low-arousal vocalization, divided by the latter. Three screams were too noisy to estimate the F0 contour, resulting in four scream stimulus pairs missing relative differences on acoustic parameters other than duration. Responses to these stimulus pairs were excluded from the statistical analyses of the effects of acoustics.

Statistical Analyses

All statistical analyses were conducted in R Statistical Environment (R Core Team, 2018) and used an alpha of 0.05. To assess the overall accuracy of participants' responses, each participant was attributed one Accuracy Score per stimulus type (coos, screams, mixed), equal to proportion of stimulus pairs to which the participant responded correctly. For mixed stimulus pairs, selecting the scream was treated as "correct" for the purposes of analysis. We used one-sample t tests (t.test command in R) to determine whether the two participant groups—experienced and inexperienced—exhibited significantly greater Accuracy Scores than the 50% accuracy predicted by chance, for each of the three stimulus types. These tests were one-tailed for coo stimuli and scream stimuli (to determine whether participant accuracy was greater than predicted by chance), and two-tailed for mixed stimuli (to determine whether participants selected one vocalization type significantly more often than the other).

To determine the effects of experience working professionally with rhesus macaques on responses, we ran three binomial, logit-link generalized linear mixed models (GLMM)—one for each stimulus type. The outcome variable was whether or not an individual response was correct, and the binomial experience variable (whether the participant had professional work experience with rhesus macaques) was entered as a fixed effect. To account for non-independence of observations among responses from a single participant, or to a single stimulus pair, participant and stimulus pair were entered as crossed random intercept effects. GLMM were likewise used to determine the effects of other participant characteristics on responses; this analysis was limited to the inexperienced participants' responses since the experienced participants did not complete the surveys or the CAM battery. We again ran one GLMM per stimulus type. The outcome variable for each model and random effects structure were the same; fixed effects included Empathy Quotient, Animal Attitudes Score, CAM Score, and whether a participant had worked with any animal species (other than rhesus macaques) in a professional capacity. Finally, four more GLMM were used to assess the effects of acoustic variables on participant responses—one per participant group for coo and scream stimuli (not mixed stimuli). The random effects structure was the same as the other models; the outcome variable was whether the participant pressed “1” or “2”, irrespective of which response was correct. Relative differences in mean F0, F0 range, mean HNR, jitter, and duration between the two vocalizations in a stimulus pair were entered as fixed effects. To avoid overfitting, nonsignificant fixed effects were removed from each model. All GLMM were fit by maximum likelihood (Laplace Approximation) using the glmer function in the lme4 package in R (Bates, Mächler, Bolker, & Walker, 2015). Expected correlations of fixed effect coefficients were checked to avoid multicollinearity. Each nonsignificant fixed effect was individually re-entered into the model to assess its level of non-significance (Barr, Levy, Scheepers, & Tily, 2013) and check against Type II errors.

Results

One-sample t tests indicated that Accuracy Scores (percentage correct) of inexperienced as well as experienced participants were significantly higher than predicted by chance (50%) for coo stimuli, though effect sizes were small (**inexperienced:** mean = 57.8, 95% CI > 56.7, t = 12.98, p < 0.0001; **experienced:** mean = 54.5, 95% CI > 50.8, t = 2.19, p = 0.025); the same was true for scream stimuli (**inexperienced:** mean = 52.2, 95% CI > 50.8, t = 2.53, p = 0.006; **experienced:** mean = 59.4, 95% CI > 55.4, t = 4.26, p < 0.001) (Figure 3.1). GLMM indicated that experienced participants responded correctly significantly more often than inexperienced participants on scream stimuli (Table 3.1, Figure 3.1). Overall, inexperienced participants did not show a significant tendency to select coos over screams or vice versa in response to mixed stimuli (mean = 49.5, 95% CI[46.1, 52.9], t = -0.289, p = 0.773), in contrast to experienced participants, who showed a strong preference for screams as more emotionally intense (mean =

81.8, 95% CI[68.3, 95.4], $t = 5.16$, $p < 0.001$) (Figure 3.1). GLMM indicated that this contrast represented a significant difference between inexperienced and experienced participants (Table 3.1). In a questionnaire administered after the task, every experienced participant indicated knowledge of the names of the vocalizations they had heard, and 6 reported knowledge of the contexts in which coos and screams occur. 6 experienced participants reported an explicit strategy of selecting screams over coos in mixed trials, and 4 of those 6 indicated that this strategy was because they perceived the agonistic contexts associated with screams to be more emotionally intense than contexts associated with coos. Finally, among inexperienced participants, those who had professional work experience with a non-primate animal responded significantly less accurately to coo stimuli than those who had not worked with any animals (Table 3.1); we did not find any other significant effects of participant characteristics on responses.

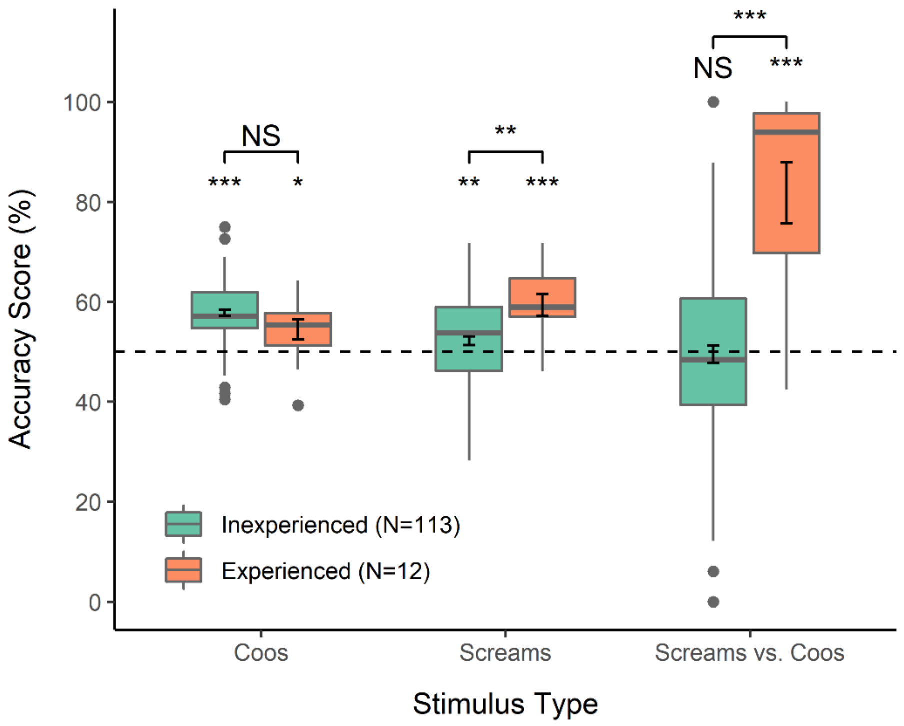

Figure 3.1. Accuracy Scores of inexperienced and experienced participants when selecting the vocalization reflecting a more intense emotion, from between two coos, two screams, or one of each. For mixed stimuli, selecting the scream was considered "accurate" for the purposes of analysis and visualization. Error bars within boxes show standard error of the mean. Asterisks above brackets indicate effect of experience; asterisks above individual columns indicate difference from chance (represented by the dotted line). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 3.1. Effects of participant characteristics on odds of correct responses or, for mixed stimuli, odds of selecting a scream over a coo. Characteristics other than experience with rhesus macaques were tested only among participants inexperienced with primates.

Coo Stimuli	Est. coeff.	SE	p value
Experience with rhesus macaques	-0.213	0.125	0.089
Empathy Quotient	0.001	0.093	0.813
Animal Attitudes Score	0.007	0.008	0.325
CAM Score	-0.001	0.004	0.881
Experience with non-primates	-0.226	0.093	0.015
Scream Stimuli			
Experience with rhesus macaques	0.368	0.141	0.009
Empathy Quotient	0.003	0.004	0.503
Animal Attitudes Score	0.002	0.008	0.798
CAM Score	< 0.001	0.004	0.940
Experience with non-primates	0.004	0.110	0.968
Mixed Stimuli			
Experience with rhesus macaques	2.428	0.379	< 0.001
Empathy Quotient	-0.008	0.009	0.409

Animal Attitudes Score	-0.012	0.018	0.514
CAM Score	0.001	0.010	0.937
Experience with non-primates	-0.380	0.241	0.116

On coo trials, inexperienced and experienced participants alike most often selected as more emotionally intense the coo that was higher in mean F0, lower in HNR (i.e., more noisy), or longer in duration, while inexperienced participants also showed a slight tendency to select the coo showing a wider F0 range (Table 3.2; Figure 3.2a). Congruent with this, the higher-arousal coo in a pair was, on average, higher in mean F0 and F0 range, and lower in HNR, though some of these average relative differences were minute and relative differences among all parameters ranged from negative to positive (Table 3.2). Plotting the coo pairs by relative mean F0 difference and average participant response (Figure 3.2a) shows that, for most (though by no means all) of the pairs in which the second coo represented greater arousal than the first (orange/light points), the second coo was also higher in F0 (i.e., relative F0 difference > 0), and was more often selected by participants (i.e., proportion of '2' responses > 0.5). A strong positive correlation is also visible between relative duration difference and proportion of '2' responses, but this does not appear to have translated into correct responses; if anything, it seems that in the majority of coo pairs, the coo representing greater arousal was shorter in duration (i.e., more orange/light points falling at relative duration difference < 0) (Figure 3.2a), though the average duration difference between high- and low-arousal coos was close to zero (Table 3.2).

Table 3.2. Descriptives for relative differences in acoustic parameters between high- and low-arousal vocalizations within stimulus pairs, and effects of acoustics on responses.

		Inexperienced			**Experienced**		
Coo Stimuli	Mean ± SD	Est. coeff.	SE	p value	Est. coeff.	SE	p value
Mean F0	0.078 ± 0.274	2.818	0.480	< 0.001	3.316	0.534	< 0.001
F0 Range	0.866 ± 1.728	0.149	0.074	0.043	0.126	0.094	0.184
Mean HNR	-0.152 ± 0.373	-0.933	0.175	< 0.001	-1.033	0.209	< 0.001
Jitter	2.000 ± 8.273	0.069	0.080	0.389	-0.013	0.094	0.892

Duration	0.011 ± 0.374	1.936	0.310	< 0.001	2.917	0.420	< 0.001
Scream Stimuli							
Mean F0	0.282 ± 0.385	-0.272	0.429	0.525	2.061	0.552	< 0.001 [a]
F0 Range	0.811 ± 1.438	0.315	0.106	0.003	0.056	0.372	0.710
Mean HNR	-3.013 ± 7.821	0.002	0.004	0.685	-0.004	0.007	0.587
Jitter	1.693 ± 3.968	0.005	0.025	0.843	< 0.001	0.031	0.996
Duration	0.779 ± 0.917	0.088	0.178	0.619	1.227	0.224	< 0.001 [a]

[a] Effects estimated individually due to collinearity.

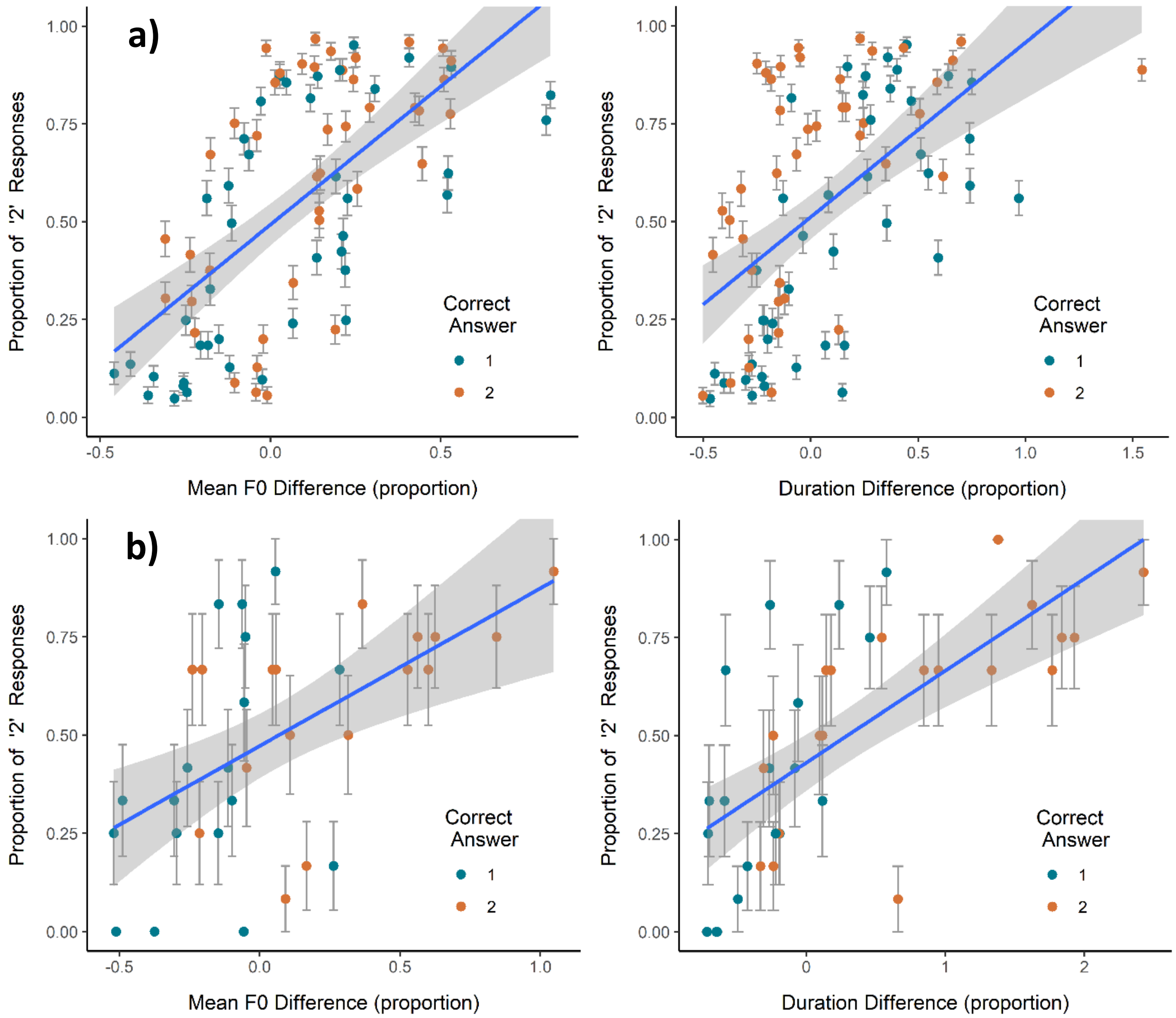

Figure 3.2. Effects of relative differences in F0 or duration between the second and first vocalization in a pair on responses to (a) coo trials (all participants), and (b) scream trials (experienced participants only). Each point represents one stimulus pair; color represents whether the first or second vocalization in the pair represented greater arousal.

On scream trials, inexperienced participants showed a slight tendency to select the scream exhibiting a wider F0 range, while experienced participants tended to select the scream exhibiting a higher mean F0 and longer duration (Table 3.2; Figure 3.2b). The effects of differences in mean F0 and duration of scream stimuli on responses were negatively correlated, i.e., inclusion of one effect reduced the estimated size of the other, indicating collinearity (coefficient of correlation of fixed effects = -0.746). As with responses to coos, each of these effects of acoustic differences on responses to scream stimuli was in the direction consistent with the mean relative acoustic difference between high- and low-arousal screams (Table 3.2). Examining Figure 3.2b, it appears that the tendency shown by experienced participants to select the higher-F0 and/or longer scream led them to respond accurately more often than not (i.e., more green/dark points in the lower-left corner and more orange/light points in the upper-right corner).

Discussion

The hypothesis of evolutionary homology in emotional vocal communication between humans and other species is traceable to Darwin (1872) and predicts that humans should be able to accurately perceive the emotional significance of heterospecific vocalizations—specifically, to discern this based on acoustic variation within (not only between) vocalization types. One might hypothesize a priori that more closely related species should share more homologous similarities in vocal emotion expression, in which case human perception of emotion from nonhuman primate vocalizations should be especially accurate (though see Filippi, Congdon, et al., 2017). We tested this prediction by having listeners with and some without job-related experience with rhesus macaques listen to coos and screams of this species, and choose the vocalization they perceived to reflect greater emotional intensity. Inexperienced and experienced participants alike exhibited above-chance accuracy, relative to actual differences in emotional arousal between coos, as measured by increases in stress hormone levels, and between screams, as inferred from differences in the intensity of agonism received by vocalizers. Overall, these findings support the hypothesis that aspects of emotional vocal communication are shared between humans and rhesus macaques.

Accuracy and Experience

Effect sizes were modest (mean accuracy less than 60%), though comparable with the findings of similar studies (Filippi, Congdon, et al., 2017; Filippi, Gogoleva, et al., 2017). That said, it is important to note the fine-grained nature of the discriminations we asked participants to perform. Whereas many similar studies have sampled vocalizations from relatively disparate contexts, e.g., play vs. agonism (Faragó et al., 2017) or isolation vs. reunion vs. nursing vs. veterinary procedures (Maruščáková et al., 2015), our coos were all sampled from an identical controlled experimental behavioral test; the only factors differentiating the contexts from which high- and low-arousal coos were sampled were the characteristics and background of an individual macaque, shaping its subjective experience of the test, as reflected in stress hormone levels. Insofar as separation from the mother and the rest of the group (albeit for only 30 minutes) probably evoked a high-arousal, negative emotional state in every six-month-old macaque, the emotional difference between coos within a pair was probably relatively small. The same reasoning applies to screams; almost every rhesus macaque scream occurs in a context associated with a relatively high-arousal, negative emotional state. Thus, while we operationally labelled vocalizations “high- and low-arousal,” these labels are relative; it might in fact be more accurate to think of the arousal levels represented among our stimuli as ranging from “somewhat high” to “very high.” In light of that, it is striking that participants were able to achieve accuracy that was significantly above chance. Our inexperienced participants’ sensitivity to the emotional significance of rhesus macaque vocalizations might be innate, or it might come about through the developmental process of learning emotional communication from human-human interactions and/or interactions with familiar/domesticated species, or some combination of both.

While even inexperienced participants performed with above-chance accuracy, we did find some effects of experience with rhesus macaques, supporting the hypothesis that familiarity with a species can improve sensitivity to emotional communication by that species (Nicastro & Owren, 2003; Parsons et al., 2019; Tallet et al., 2010). Experienced participants exhibited slightly (but statistically significantly) greater accuracy with scream stimulus pairs, and far more often chose the scream over the coo in mixed stimulus pairs, though experience had no effect on accuracy with coo stimulus pairs. The reasons behind this contrast between coos and screams are uncertain. One possible explanation is that humans might typically have more experience perceiving cues to emotion in vocalizations exhibiting the F0 range of coos (generally 400-3000 Hz) than that of rhesus macaque screams (generally 2-10 kHz), since the former is closer to the human vocal range. Emotion perception from screams might therefore have been less intuitive for inexperienced human listeners, making experience a more significant factor. That said, tonal screams comprise only one of several rhesus macaque scream classes, warranting caution in extrapolating to screams more generally. Interpreting the effect of experience on responses to mixed stimuli is more straightforward: several participants reported prior knowledge of the socioecological contexts associated with coos and screams, and indicated that they drew on this knowledge to judge screams as more emotionally intense than coos. Inexperienced participants, in contrast, presumably lacked such a basis of prior knowledge on which to form a consistent explicit strategy. The effect of experience on responses to mixed stimuli was by far the strongest effect of any participant characteristic; its effects on accuracy with scream pairs was modest by comparison.

Surprisingly, among participants inexperienced with rhesus macaques, having formally worked with (non-primate) animals was associated with decreased odds of responding correctly to coo stimuli. The reasons for this are unclear. Other participant characteristics—empathy, human emotion recognition, and animal attitudes—failed to show any relationship with responses, in line with one previous study (Maruščáková et al., 2015).

Acoustic Cues to Emotional Arousal

Vocal F0 is a well-established correlate of emotional arousal across mammals including humans, and humans have been shown to perceive higher-F0 vocalizations from a variety of species (including conspecific vocalizations) as more emotionally intense (e.g., Briefer, 2012; Faragó et al., 2014, 2017; Filippi, Congdon, et al., 2017; Juslin & Laukka, 2003; Kelly et al., 2017; Maruščáková et al., 2015; Nicastro & Owren, 2003; Pongrácz et al., 2005; Sauter et al., 2010; Scherer, 2003; Schwartz & Gouzoules, 2019). This trend is probably due in part to phylogenetically widespread mechanisms of arousal and vocal production: arousal increases tension in the vocal fold muscles, resulting in faster oscillation, i.e., higher F0 (Briefer, 2012; Riede, 2010; Scherer, 1986; Titze, 1994). In the present study, inexperienced and experienced participants alike tended to indicate the higher-F0 coo in a pair as more emotionally intense. Experienced, but not inexperienced, participants appeared to base scream selections partly on F0, as well, perhaps contributing to their greater accuracy. These findings are consistent not only with the literature in general but also with the findings of a recent study on rhesus macaque vocal production, which demonstrated a positive correlation between F0 and arousal among coos as well as screams (Chapter II). Thus vocal F0 appears to be a significant indicator of emotional arousal in humans and rhesus macaques (and other mammalian species) alike, probably due to evolutionary homologies in the physiology and anatomy of vocal production. In the present study as in previous ones (Kelly et al., 2017), humans' sensitivity to F0 seems to have led participants to respond correctly in some cases and incorrectly in other cases, but the former appears to have been more common than the latter (Figure 3.2). Indeed, higher-arousal vocalizations were indeed higher in mean F0 on average than lower-arousal vocalizations (Table 3.2).

Given the significant role of F0 in within-vocalization-type discriminations, it is puzzling that inexperienced participants did not consistently indicate high-F0 screams as reflecting a more intense emotion than lower-F0 coos. Just as F0 and other acoustic parameters seem to correlate more strongly with arousal within a vocalization type than between vocalization types from a vocal production standpoint (Schamberg et al., 2018), our participants appeared to rely on F0 as a cue to emotional intensity more for within- than between-type discrimination. This observation is consistent with the "identification-attribution model" of nonlinguistic vocal perception, which postulates that classification or identification of vocalization types is cognitively and perhaps neurally distinct from the attribution of emotional states to vocalizations (Anikin, Bååth, & Persson, 2018; Sperduti, Guionnet, Fossati, & Nadel, 2014; Spunt & Lieberman, 2012).

In addition to F0, inexperienced and experienced participants both tended to select a coo when it was longer in duration or more noisy (lower HNR) than the alternative, consistent with previous findings on human perception of emotional arousal (Faragó et al., 2014; Filippi, Gogoleva, et al., 2017; Schwartz & Gouzoules, 2019). Unlike with F0, these findings do not match those reported in Chapter II with respect to objective acoustic correlates of emotional arousal in rhesus macaque coos; arousal showed no significant relationship with HNR, and showed a negative relationship with duration. That said, among coo stimulus pairs used in the present study, the higher-arousal coo was noisier on average; the effects of noisiness on participants' decisions might thus have led them to respond correctly on some trials. Participants' tendency to select the longer coo likely played no role in their overall accuracy, instead leading them to the correct and incorrect responses an approximately equal proportion of the time (Figure 3.2a). Experienced participants also tended to select the longer scream; while Chapter II reported no significant correlation between arousal and scream duration (perhaps due in part to the inclusion of another scream class, pulsed screams, which are by their nature constrained in duration), among our stimuli high-arousal screams were longer on average, and this may have played a role in facilitating experienced participants' relatively high accuracy on scream trials (Figure 3.2b). That said, the effects of scream F0 and duration on responses appeared to be collinear, making them difficult to disambiguate. Finally, inexperienced participants showed a preference for the coo or scream in a pair that showed a wider F0 range, though these effect sizes were small.

Overall, consistencies between our findings and those of other studies suggest that human listeners might use similar acoustic cues to assess the emotional arousal represented by acoustic variants within a vocalization type, whether it be a human nonlinguistic vocalization (Sauter et al., 2010; Schwartz & Gouzoules, 2019), a vocalization of a species with which the listener is familiar (Faragó et al., 2014, 2017; Nicastro & Owren, 2003), or a vocalization of an unfamiliar species (Filippi, Congdon, et al., 2017; Kelly et al., 2017). The role of familiarity and experience in determining the salience of different acoustic cues might depend somewhat on the vocalization type in question.

Emotional Arousal and Valence

A remaining issue is the distinction between emotional arousal and valence, i.e., the spectrum from negative to positive states (Mendl et al., 2010; Russell, 2003). Participants were not instructed on the distinction between emotional valence and arousal, and thus it is possible that some of them might have interpreted “a more intense emotion” to mean a state that is more negative as well as more aroused. However, the term “intensity” is often used interchangeably with “arousal,” and previous research has shown that listeners provide similar ratings for both terms with respect to vocal stimuli (Laukka et al., 2005). In truth, the stimulus pairs might have represented differences in emotional valence in addition to arousal: higher concentrations of stress hormones might reflect a more negative state than lower concentrations, and contact aggression might evoke a more negative emotional reaction than non-contact aggression. As in many other studies of animal emotional vocal communication (reviewed in Briefer, 2012), the present one does not permit us to discern the precise roles of emotional arousal and valence. With that said, the mechanisms by which emotional arousal affects the voice are more straightforward than for valence (Briefer, 2012), and F0—a well-established correlate of emotional arousal but not valence—strongly influenced our participants’ responses. For these reasons, we remain confident that emotional arousal played a role in our findings.

Conclusions

The present study demonstrated that listeners with and without job-related experience with rhesus macaques have the capacity to discern fine-grained distinctions in the emotional arousal level represented in the coos and screams of that species. Experience was associated with limited improvement in this ability, but the most obvious effect of experience was seen in between-vocalization-type discriminations, where semantic knowledge about the socioecology of coos and screams appeared to influence responses. Thus our findings are consistent with the hypothesis that humans' capacity for accurate emotional perception of rhesus macaque vocalizations stems in part from the evolutionary ancestry shared by the two species. Specifically, it appears that listeners utilized acoustic cues that reliably indicate emotional arousal in humans and rhesus macaques alike due to homology in the physiology and anatomy of arousal and vocal production. Rhesus macaques and other nonhuman primates are closely related to humans relative to many other taxa whose vocalizations have yet to be studied with respect to human emotion perception; our findings can only be interpreted with respect to evolution occurring over the last 30 million years, the time since the human and rhesus macaque lineages diverged (Steiper & Young, 2008). Future research examining vocal emotion expression in more phylogenetically distant species, and human perception thereof, will further improve our understanding of the evolutionary history of emotional communication.

References

Anikin, A., Bååth, R., & Persson, T. (2018). Human non-linguistic vocal repertoire: Call types and their meaning. *Journal of Nonverbal Behavior*, *49*(2), 53–80. https://doi.org/10.1007/s10919-017-0267-y

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bayart, F., Hayashi, K. T., Faull, K. F., Barchas, J. D., & Levine, S. (1990). Influence of maternal proximity on behavioral and physiological responses to separation in infant rhesus monkeys (*Macaca mulatta*). *Behavioral Neuroscience*, *104*(1), 98–107. https://doi.org/10.1037/0735-7044.104.1.98

Belin, P., Fecteau, S., Charest, I., Nicastro, N., Hauser, M. D., & Armony, J. L. (2008). Human cerebral response to animal affective vocalizations. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1634), 473–481. https://doi.org/10.1098/rspb.2007.1460

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.51). Retrieved January 9, 2018, from http://www.praat.org/

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *7*(1), 1–20. https://doi.org/10.1098/rsbl.2011.0832

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Marry. https://doi.org/10.1037/h0076058

Faragó, T., Andics, A., Devecseri, V., Kis, A., Gacsi, M., & Miklosi, A. (2014). Humans rely on the same rules to assess emotional valence and intensity in conspecific and dog vocalizations. *Biology Letters*, *10*(1), 20130926. https://doi.org/10.1098/rsbl.2013.0926

Faragó, T., Takács, N., Miklósi, Á., & Pongrácz, P. (2017). Dog growls express various contextual and affective content for human listeners. *Royal Society Open Science*, *4*(5), 170134. https://doi.org/10.1098/rsos.170134

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, 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: Biological Sciences*, *284*(1859), 20170990. https://doi.org/10.1098/rspb.2017.0990

Filippi, P., Gogoleva, S. S., Volodina, E. V., Volodin, I. A., & de Boer, B. (2017). Humans identify negative (but not positive) arousal in silver fox vocalizations: Implications for the adaptive value of interspecific eavesdropping. *Current Zoology*, *63*(4), 445–456. https://doi.org/10.1093/cz/zox035

Fischer, J., Wadewitz, P., & Hammerschmidt, K. (2016). Structural variability and communicative complexity in acoustic communication. *Animal Behaviour*, *134*, 229–237. https://doi.org/10.1016/j.anbehav.2016.06.012

Fritz, T., Mueller, K., Guha, A., Gouws, A., Levita, L., Andrews, T. J., & Slocombe, K. E.

(2018). Human behavioural discrimination of human, chimpanzee and macaque affective vocalisations is reflected by the neural response in the superior temporal sulcus. *Neuropsychologia*, *111*, 145–150. https://doi.org/10.1016/j.neuropsychologia.2018.01.026

Fugate, J. M. B., Gouzoules, H., & Nygaard, L. C. (2008). Recognition of rhesus macaque (*Macaca mulatta*) noisy screams: Evidence from conspecifics and human listeners. *American Journal of Primatology*, *70*(6), 594–604. https://doi.org/10.1002/ajp.20533

Golan, O., Baron-Cohen, S., & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger Syndrome. *Journal of Autism and Developmental Disorders*, *36*(2), 169–183. https://doi.org/10.1007/s10803-005-0057-y

Golan, O., Baron-Cohen, S., Hill, J. J., & Rutherford, M. D. (2007). The "Reading the Mind in the Voice" test-revised: A study of complex emotion recognition in adults with and without autism spectrum conditions. *Journal of Autism and Developmental Disorders*, *37*(6), 1096–1106. https://doi.org/10.1007/s10803-006-0252-5

Gouzoules, H. (2005). Monkeying around with symbolism: Are vocalizations simple symbols...or more like cymbals? In L. Namy (Ed.), *Symbol use and symbolic representation: Developmental and comparative perspectives.* (pp. 245–265). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Gouzoules, H., Gouzoules, S., & Tomaszycki, M. (1998). Agonistic screams and the classification of dominance relationships: Are monkeys fuzzy logicians? *Animal Behaviour*, *55*(1), 51–60. https://doi.org/10.1006/anbe.1997.0583

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour*, *32*(1), 182–193. https://doi.org/10.1016/S0003-3472(84)80336-X

Hauser, M. D. (1991). Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations. *Ethology*, *89*(1), 29–46. https://doi.org/10.1111/j.1439-0310.1991.tb00291.x

Hauser, M. D., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology*, *4*(3), 194–205. https://doi.org/10.1093/beheco/4.3.194

Herzog, H., Grayson, S., & McCord, D. (2015). Brief measures of the animal attitude scale. *Anthrozoos*, *28*(1), 145–152. https://doi.org/10.2752/089279315X14129350721894

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/0033-2909.129.5.770

Kalin, N. H., & Shelton, S. E. (1989). Defensive behaviors in infant rhesus monkeys: Environmental cues and neurochemical regulation. *Science*, *243*(4899), 1718–1721. https://doi.org/10.1126/science.2564702

Kelly, T., Reby, D., Levréro, F., Keenan, S., Gustafsson, E., Koutseff, A., & Mathevon, N. (2017). Adult human perception of distress in the cries of bonobo, chimpanzee, and human infants. *Biological Journal of the Linnean Society*, *120*(4), 919–930. https://doi.org/10.1093/biolinnean/blw016

Laukka, P., Juslin, P. N., & Bresin, R. (2005). A dimensional approach to vocal expression of emotion. *Cognition and Emotion*, *19*(5), 633–653. https://doi.org/10.1080/02699930441000445

Linnankoski, I., Laakso, M., Aulanko, R., & Leinonen, L. (1994). Recognition of emotions in macaque vocalizations by children and adults. *Language and Communication*, *14*(2), 183–

192. https://doi.org/10.1016/0271-5309(94)90012-4

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*(2), 121–131. https://doi.org/10.1037/a0038870

McComb, K., Taylor, A. M., Wilson, C., & Charlton, B. D. (2009). The cry embedded within the purr. *Current Biology*, *19*(13), 507–508. https://doi.org/10.1016/j.cub.2009.05.033

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

Nicastro, N., & Owren, M. J. (2003). Classification of domestic cat (*Felis catus*) vocalizations by naive and experienced human listeners. *Journal of Comparative Psychology*, *117*(1), 44–52. https://doi.org/10.1037/0735-7036.117.1.44

Owren, M. J. (2008). GSU Praat tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods*, *40*(3), 822–829. https://doi.org/10.3758/BRM.40.3.822

Owren, M. J., & Bachorowski, J.-A. (2007). Measuring emotion-related vocal acoustics. In J. Coan & J. Allen (Eds.), *The handbook of emotion elicitation and assessment* (pp. 239–266). Oxford: Oxford University Press.

Parsons, C. E., LeBeau, R. T., Kringelbach, M. L., & Young, K. S. (2019). Pawsitively sad: Pet-owners are more sensitive to negative emotion in animal distress vocalizations. *Royal Society Open Science*, *6*(8), 181555. https://doi.org/10.1098/rsos.181555

Pongrácz, P., Molnár, C., & Miklósi, Á. (2010). Barking in family dogs: An ethological approach. *The Veterinary Journal*, *183*(2), 141–147. https://doi.org/10.1016/J.TVJL.2008.12.010

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*(2), 136–144. https://doi.org/10.1037/0735-7036.119.2.136

R Core Team. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.r-project.org/

Riede, T. (2010). Elasticity and stress relaxation of rhesus monkey (*Macaca mulatta*) vocal folds. *Journal of Experimental Biology*, *213*, 2924–2932. https://doi.org/10.1242/jeb.044404

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

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: A new take on ultimate and proximate influences in vocal production. *Biological Reviews*, *93*(4), 2071–2082. https://doi.org/10.1111/brv.12437

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.99.2.143

Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication*, *40*(1), 227–256. https://doi.org/10.1016/S0167-6393(02)00084-5

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*(3), e91192. https://doi.org/10.1371/journal.pone.0091192

Schwartz, J. W., & Gouzoules, H. (2019). Decoding human screams: Perception of emotional arousal from pitch and duration. *Behaviour*, *156*(13–14), 1283–1307. https://doi.org/10.1163/1568539X-00003566

Schwartz, J. W., Engelberg, J. W. M., & Gouzoules, H. (in press). Evolving views on cognition in animal communication: Contributions from scream research. *Animal Behavior and Cognition.*

Sperduti, M., Guionnet, S., Fossati, P., & Nadel, J. (2014). Mirror Neuron System and Mentalizing System connect during online social interaction. *Cognitive Processing*, *15*(3), 307–316. https://doi.org/10.1007/s10339-014-0600-x

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

Steiper, M. E., & Young, N. M. (2008). Timing primate evolution: Lessons from the discordance between molecular and paleontological estimates. *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(4), 179–188. https://doi.org/10.1002/evan.20177

Tallet, C., Špinka, M., Maruščáková, I., & Šimeček, P. (2010). Human perception of vocalizations of domestic piglets and modulation by experience with domestic pigs (*Sus scrofa*). *Journal of Comparative Psychology*, *124*(1), 81–91. https://doi.org/10.1037/a0017354

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Zimmermann, E., Leliveld, L. M. C., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: A comparative approach to mammalian voices. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of emotional communication: From sounds in nonhuman mammals to speech and music in men* (pp. 117–132). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199583560.001.0001

CHAPTER IV: GENERAL DISCUSSION

A complete understanding of animal behavior requires integration of both proximate and ultimate perspectives (Bateson & Laland, 2013; Laland, Odling-Smee, Hoppitt, & Uller, 2013; Laland, Sterelny, Odling-Smee, Hoppitt, & Uller, 2011; Nesse, 2013; Tinbergen, 1963). This synthesis should entail not only a recognition of the need for both the proximate and ultimate levels of study, but additional exploration of what proximate mechanisms mean for ultimate evolutionary processes, and vice versa. At the most basic level of evolutionary theory, hypothesizing that a trait is adaptive (i.e., functions to increase an individual's reproductive fitness, and evolved through direct natural selection due to the same positive fitness impacts; Williams, 1966) requires the proximate-level assumption that variation in the trait is heritable (due to genetic variation) to some degree. Research into the proximate mechanisms underlying traits can thus help address ultimate-level questions regarding their evolutionary history and general processes in evolution.

In the General Introduction of this dissertation, I argued that the study of vocal emotion expression (the ways in which acoustic characteristics of vocalizations change depending on the emotional state of the vocalizer) and the evolution thereof holds great potential to elucidate broader principles regarding the relationships between proximate mechanisms and evolutionary processes. One takeaway was that while vocal usage (i.e., whether a vocalization occurs, and which vocalization type) is often influenced by emotion, context perception and other cognitive processes can also play a role; acoustic variation *within* vocalization types appears to relate strongly to emotional states, especially emotional arousal (Schamberg, Wittig, & Crockford, 2018; Schwartz, Engelberg, & Gouzoules, in press). I reviewed how arousal and other components of emotions relate to within-type acoustic variation, and explored the potential significance of this for general processes in vocal evolution, touching on natural selection, phylogenetic inertia, correlated response to selection, and evolutionary constraints. I posited a model of the evolution of vocal communication in which emotion-related acoustic changes and other outputs of emotional states might, to the degree that they are integrated with each other, collectively exhibit evolutionary change in a direction consistent with the sum of all selection pressures acting on all integrated outputs.

One objective of this dissertation is to contribute toward testing the predictions of hypotheses about general processes in the evolution of vocal acoustics, and vocal communication more broadly. The hypothesis that patterns of vocal emotion expression been conserved through phylogenetic inertia predicts that the acoustic correlates of emotions will be broadly similar across species (Briefer, 2012; Filippi et al., 2017; Zimmermann, Leliveld, & Schehka, 2013). The hypothesis that any given acoustic change occurring in a given context might exhibit a correlated response to selection pressures acting on other acoustic changes, and/or other integrated outputs of an emotional state, predicts that acoustic parameters of vocalizations are linked to emotional states and integrated with each other at the proximate level (a precondition for these kinds of evolutionary dynamics). The primary goal of the research reported in this dissertation was to test these predictions by assessing the degree of consistency in the vocal expression of emotional arousal across contexts and vocalization types within rhesus macaques, and between rhesus macaques and humans. The two studies comprising this dissertation indicate a similar correlation between emotional arousal and vocal F0 in two different rhesus macaque vocalization types occurring in disparate socioecological contexts, and demonstrate a capacity among naïve as well as experienced humans to accurately assess the levels of emotional arousal represented in these vocalizations. These findings are broadly consistent with the hypotheses outlined above, while some caveats and nuances to the results raise the possibility that direct selection might be capable of altering patterns of vocal emotion expression in some vocalization types.

Summary of Findings and Proximate Perspectives

Chapter II described a study examining the correlations between emotional arousal and acoustic parameters of two rhesus macaque vocalization types, coos and screams (two subclasses—tonal and pulsed), recorded during a brief behavioral test involving temporary social isolation or a naturally occurring agonistic social interaction, respectively. Emotional arousal was associated with increases in the overall F0 of coos as well as screams in these two contexts. This similarity is especially striking given the different methods used to measure arousal for the two vocalization types (stress hormone elevations for coos and agonistic intensity for screams). At the proximate level, this trend probably reflects the physiological effects of emotional arousal on the mechanisms of vocal production, namely, tensing of vocal muscles resulting in increased rate of oscillation of the vocal folds (Briefer, 2012; Riede, 2010; Titze, 1994); this process appears to occur irrespective of whether the vocalization in question is a coo or a scream, and whether the emotional arousal response was elicited by a behavioral test or an agonistic interaction. Thus this process exhibits both "stimulus generalizability"—multiple different contexts yield a similar emotional arousal response and change in vocal production—and emotional "pleiotropy"—the same emotional arousal response results in changes in the acoustic properties of multiple different vocalization types (sensu Anderson & Adolphs, 2014).

There were also some differences in patterns of vocal emotion expression between coos and screams, and caveats to the relationship between arousal and F0 within each. The relationship between arousal and coo F0 appeared to dampen over the course of development;

this result could reflect a true interaction between the effects of age and arousal on vocal acoustics, and/or it could be due to the greater arousal response shown by the youngest age group (6 months of age). Among screams, the relationships between arousal and a couple of acoustic parameters differed by scream class. Note that only tonal and pulsed screams were analyzed, and these have been shown to elicit weaker responses than noisy and arched screams (corresponding to severe aggression and rank challenge, respectively) (Gouzoules, Gouzoules, & Marler, 1984). It is conceivable that additional acoustic parameters might vary in their relationship to arousal across a wider sample of rhesus macaque screams. These nuances in results around coos and screams suggest the relationship between arousal and vocal F0, though generally consistent, is not absolute and can vary. While coos showed significant relationships between arousal and F0 modulation (positive) and duration (negative), screams did not (although it is possible that a different array of scream classes might have shown different results). These contrasts represent variability in patterns of emotion expression across vocalization types and contexts, and probably have to do with the differing mechanisms of production of coos and screams, though exactly how remains an open question.

Chapter III described a study testing the hypothesis that vocal emotion expression is sufficiently similar between humans and rhesus macaques to allow humans to accurately perceive emotional arousal from acoustic variation within rhesus macaque vocalization types. In support of that hypothesis, participants both with and without previous job-related experience with rhesus macaques showed modest but significantly above-chance accuracy when identifying the higher-arousal vocalization from among two coos or two screams. Experience was associated with a slight increase in accuracy on scream stimuli, though not on coo stimuli. Participants generally perceived coos to reflect greater arousal when they were higher-F0, noisier, or longer, while experienced participants (but not inexperienced participants) perceived higher-F0 and longer screams to be more aroused. This sensitivity to F0 is consistent with actual patterns of vocal emotion expression in rhesus macaques as reported in Chapter II; however, participants' sensitivity to the duration of coos, though consistent with how listeners perceive human screams (Schwartz & Gouzoules, 2019), contradicted the actual negative relationship between arousal and coo duration. Inexperienced participants' sensitivity to the emotional significance of rhesus macaque vocalizations might be innate, or it might come about through the developmental process of learning emotional communication from human-human interactions and/or interactions with familiar/domesticated species, or some combination of both.

Evolutionary Implications

This dissertation, taken together with other studies of vocal emotion expression, carries implications regarding three questions about the evolution of the relationship between arousal and F0 variation within vocalization types: (1) How many times and at what point(s) has this relationship arisen over the course of evolution? (2) What were the roles of various evolutionary processes and pressures in the initial evolution of this relationship? (3) What evolutionary processes have been responsible for maintaining this relationship over time? The possible answers to these questions are discussed below. My focus on F0 here is not intended to imply that F0 is necessarily prime with respect to emotion expression or other functions; many mammalian vocalizations lack a clear F0, including a subset of rhesus macaque screams as well as other vocalizations associated with extreme emotional arousal or valence across mammals (Fitch, Neubauer, & Herzel, 2002). Rather, the research reported in this dissertation more strongly supports conclusions regarding the relationship between arousal and F0, and the evolution thereof, than other acoustic parameters, warranting special consideration.

(1) Regarding the first question, the findings of Chapter II with respect to F0 are consistent with the known effects of arousal on vocal F0 in other mammalian species including humans (Briefer, 2012). Chapter III demonstrated that humans are sensitive to the emotional significance of variation in the F0 of rhesus macaque calls, as has been demonstrated with the calls of several other vertebrate species (Filippi et al., 2017). The results of Chapters II and III together suggest that some aspects of vocal emotion expression, especially the relationship between arousal and vocal F0, are homologous and have been conserved within both the human and rhesus macaque lineages since they diverged c. 30 million years ago (Steiper & Young, 2008). Based on results across mammals and some other taxa, the relationship between arousal and F0 probably arose early in mammalian evolution, or even earlier, and has been inherited not only in rhesus macaques and humans but also many other lineages (Briefer, 2012; Filippi et al., 2017; Zimmermann et al., 2013).

(2) Through what evolutionary processes and pressures did the relationship between emotional arousal and vocal F0 initially arise? The receiver-precursor model (Bradbury & Vehrencamp, 1998) would posit that direct selection stemming from communicative function played a role in that initial evolutionary process. As one hypothetical example, a higher-F0 version of a vocalization might be more salient to receivers, as indicated in humans in Chapter III and in an array of other studies (e.g., Filippi et al., 2017; Schwartz & Gouzoules, 2019). If this represents a pre-existing sensory bias, then individuals of a hypothetical ancestral mammalian species who emitted a higher-F0 version of a vocalization in situations in which the stakes were highest, i.e., when they were most aroused, might therefore have achieved increased

odds of a beneficial response from receivers, relative to senders who did not show this tendency. On the other hand, the sender-precursor model (Bradbury & Vehrencamp, 1998) would posit that increases in vocal F0 originated as a byproduct of a broader arousal response, serving as an initially functionless cue to the sender’s emotional state, from which receivers might then have evolved to obtain information.

The findings of Chapter II suggest that, in rhesus macaques, the F0 of a given vocalization, regardless of vocalization type, appears to be linked to arousal state of a vocalizer, hence vocal F0 can be viewed as integrated at the proximate level with the other effects of arousal on the body and behavior (including increased blood flow, mobilization of energy stores, alertness, various other acoustic properties of said vocalization type, and the F0 of other vocalization types), any of which might theoretically carry fitness consequences not only in the context in question but in other arousing contexts as well. If this was true of a hypothetical mammalian ancestor as it appears to be in rhesus macaques, then any evolutionary change in the physiology of emotion or the contexts that elicit emotional responses, stemming from selection pressures acting on any of these effects of arousal, might be expected to have yielded a correlated response in the F0 of any vocalization emitted in a high-intensity context. Therefore, it is plausible that a hypothetical mammalian ancestor might have begun to exhibit increased vocal F0 in more intense situations as a result of selection favoring the broader arousal response within those contexts, with vocal F0 serving initially as a cue to the sender’s arousal state, i.e., the sender-precursor model.

(3) How has the relationship between emotional arousal and vocal F0 been maintained over evolutionary time in rhesus macaques, humans, and other lineages? Addressing this question requires considering the possible communicative significance and fitness consequences

of this relationship. One possibility, again, is that it is beneficial for senders to emit higher-F0 vocalizations in more intense contexts. This could be the case if doing so exploits pre-existing sensory biases in receivers (i.e., the receiver-precursor model), and/or if receivers have evolved to respond to emotional F0 variation in ways that benefit senders (i.e., the sender-precursor model). In these scenarios, even if direct selection played no role in the initial evolution of the increase in vocal F0 observed in more intense contexts (if it arose as a cue, solely through selection favoring a broader arousal response, i.e., the sender-precursor model), vocal F0 might have been exapted to serve a communicative function, and subsequently maintained by positive or stabilizing selection. Alternatively, if senders experience no fitness benefits or costs to emitting higher-F0 vocalizations when more aroused, then phylogenetic inertia and the broader benefits of the arousal response would seem sufficient to explain why the positive relationship between vocal F0 and arousal has been conserved.

The relationship between arousal and vocal F0 need not necessarily be beneficial or even neutral to sender fitness in order to be conserved over evolutionary time; in theory, this relationship could be negative. For example, it is conceivable that masking one’s internal state from competitors might, in some situations, be beneficial, in which case selection might favor a decoupling of emotional arousal and vocal F0. This decoupling would be expected to evolve, unless it is prevented by evolutionary constraints (see Fitch & Hauser, 2006). To speculate about what such a constraint might hypothetically look like, for the sake of argument, it is conceivable that a hypothetical mutation decoupling arousal and vocal F0, so that a more aroused animal would not emit higher-F0 versions of a vocalization, might dampen arousal’s beneficial effects on the rest of the body, interfere with the respiratory function of the larynx, etc., and thus not be favored by selection. As long as this decoupling is thus constrained, within-type variation in vocal F0 would be expected to remain physiologically linked to other outcomes of arousal, and all of these outcomes of arousal should be expected to collectively respond to the sum of selection pressures acting on them. Thus any potentially negative fitness effects of communicating one's emotional state could conceivably be nullified by the positive effects of other outcomes of arousal, resulting in the conservation of increased vocal F0 in intense contexts over evolutionary time. This dissertation did not examine the fitness consequences of the relationship between arousal and vocal F0, but this seems like a fertile area for future research.

Unlike overall F0, other acoustic parameters including duration and F0 modulation showed less consistent relationships with emotional arousal between rhesus macaque coos and screams. The evolutionary implications of these contrasts are unclear. They could reflect evolved differences in patterns of emotion expression between these two vocalization types, perhaps relating to the different functions of each in their respective contexts, which would suggest selection is capable of some degree of tinkering with the relationships between emotional arousal and vocal acoustics in one vocalization type or context without necessarily yielding a correlated response in another vocalization type or context. Alternatively, differences the methods used to measure arousal (stress hormones for coos, agonistic intensity for screams) could have played a role in differences in vocal emotion expression between coos and screams. Similarities in vocal emotion expression between coos and screams cannot be explained by differences in method, hence this discussion has largely focused on those similarities.

Conclusions

Recently, prominent authors in the field of animal communication have called for more research into the mechanisms of vocal production and how vocal communication systems evolve

(Fischer, 2017; Fischer & Price, 2016; Owren, Amoss, & Rendall, 2011; Seyfarth & Cheney, 2017; Wheeler & Fischer, 2012, 2015). A central tenet of this dissertation is that vocal emotion expression research contributes to both of these goals. The studies reported herein add rhesus macaques to a growing list of species exhibiting similar patterns of vocal emotion expression, and suggest how the proximate nature and ultimate functions of emotions can affect the evolutionary trajectory of emotional behavior. They are consistent with view that the physiological components of emotions and the acoustic qualities of vocalizations are linked together at both the proximate and ultimate levels. Thus there is a need, when generating hypotheses about the evolution of a particular acoustic change within a vocalization type occurring in a particular context, to consider the whole, asking not only what might be the communicative significance of the particular acoustic change itself, but also what proximate psychological and physiological processes underlie it, the other functional outcomes of those processes, how selection might act on those, and how evolutionary change due to such selection pressures might affect vocal acoustics and vice versa. These questions hold the potential to lead to novel predictions, the testing of which will serve to further disambiguate the roles of natural selection and other processes in the evolution of vocal acoustics.

References

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

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in Ecology and Evolution*, *28*(12), 712–718. https://doi.org/10.1016/j.tree.2013.09.013

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication*. Sunderland, MA: Sinauer.

Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *7*(1), 1–20. https://doi.org/10.1098/rsbl.2011.0832

Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, 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: Biological Sciences*, *284*(1859), 20170990. https://doi.org/10.1098/rspb.2017.0990

Fischer, J. (2017). Primate vocal production and the riddle of language evolution. *Psychonomic Bulletin and Review*, *24*(1), 72–78. https://doi.org/10.3758/s13423-016-1076-8

Fischer, J., & Price, T. (2016). Meaning, intention, and inference in primate vocal communication. *Neuroscience and Biobehavioral Reviews*, *82*, 22–31. https://doi.org/10.1016/j.neubiorev.2016.10.014

Fitch, W. T., & Hauser, M. D. (2006). Unpacking "honesty": Vertebrate vocal production and the evolution of acoustic signals. In A. M. Simmons, R. R. Fay, & A. N. Popper (Eds.), *Acoustic communication* (pp. 65–137). New York: Springer. https://doi.org/10.1007/0-387-

22762-8_3

Fitch, W. T., Neubauer, J., & Herzel, H. (2002). Calls out of chaos: The adaptive significance of nonlinear phenomena in mammalian vocal production. *Animal Behaviour*, *63*(3), 407–418. https://doi.org/10.1006/anbe.2001.1912

Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: Representational signalling in the recruitment of agonistic aid. *Animal Behaviour*, *32*(1), 182–193. https://doi.org/10.1016/S0003-3472(84)80336-X

Laland, K. N., Odling-Smee, J., Hoppitt, W., & Uller, T. (2013). More on how and why: Cause and effect in biology revisited. *Biology & Philosophy*, *28*(5), 719–745. https://doi.org/10.1007/s10539-012-9335-1

Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is Mayr’s proximate-ultimate dichotomy still useful? *Science*, *334*(6062), 1512–1516. https://doi.org/10.1126/science.1210879

Nesse, R. M. (2013). Tinbergen’s four questions, organized: A response to Bateson and Laland. *Trends in Ecology & Evolution*, *28*, 681–682. https://doi.org/10.1016/j.tree.2013.10.008

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, *73*(6), 530–544. https://doi.org/10.1002/ajp.20913

Riede, T. (2010). Elasticity and stress relaxation of rhesus monkey (*Macaca mulatta*) vocal folds. *Journal of Experimental Biology*, *213*, 2924–2932. https://doi.org/10.1242/jeb.044404

Schamberg, I., Wittig, R. M., & Crockford, C. (2018). Call type signals caller goal: A new take on ultimate and proximate influences in vocal production. *Biological Reviews*, *93*(4), 2071–

2082. https://doi.org/10.1111/brv.12437

Schwartz, J. W., & Gouzoules, H. (2019). Decoding human screams: Perception of emotional arousal from pitch and duration. *Behaviour*, *156*(13–14), 1283–1307. https://doi.org/10.1163/1568539X-00003566

Seyfarth, R. M., & Cheney, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour*, *124*, 339–346. https://doi.org/10.1016/j.anbehav.2016.05.020

Steiper, M. E., & Young, N. M. (2008). Timing primate evolution: Lessons from the discordance between molecular and paleontological estimates. *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(4), 179–188. https://doi.org/10.1002/evan.20177

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift Für Tierpsychologie*, *20*, 410–433.

Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology*, *21*(5), 195–205. https://doi.org/10.1002/evan.21319

Wheeler, B. C., & Fischer, J. (2015). The blurred boundaries of functional reference: A response to Scarantino & Clay. *Animal Behaviour*, *100*, 9–13. https://doi.org/10.1016/j.anbehav.2014.11.007

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton, NJ: Princeton University Press. https://doi.org/10.1086/332352

Zimmermann, E., Leliveld, L. M. C., & Schehka, S. (2013). Toward the evolutionary roots of affective prosody in human acoustic communication: A comparative approach to mammalian voices. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *Evolution of* *emotional communication: From sounds in nonhuman mammals to speech and music in men* (pp. 117–132). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199583560.001.0001