

The Evolution of Human Vocal Emotion

Gregory A. Bryant 

Department of Communication, Center for Behavior, Evolution, and Culture, University of California, Los Angeles, USA

Abstract

Vocal affect is a subcomponent of emotion programs that coordinate a variety of physiological and psychological systems. Emotional vocalizations comprise a suite of vocal behaviors shaped by evolution to solve adaptive social communication problems. The acoustic forms of vocal emotions are often explicable with reference to the communicative functions they serve. An adaptationist approach to vocal emotions requires that we distinguish between evolved signals and byproduct cues, and understand vocal affect as a collection of multiple strategic communicative systems subject to the evolutionary dynamics described by signaling theory. We should expect variability across disparate societies in vocal emotion according to culturally evolved pragmatic rules, and universals in vocal production and perception to the extent that form–function relationships are present.

Keywords

emotion, evolution, signaling, vocal affect

Emotional communication is central to social life for many animals. Beginning with Darwin (1872), comparative analyses have documented clear similarities in the structures of affective expressions across species, including facial and vocal emotions. More recently, researchers in neuroscience and related fields have described the phylogenetically conserved mechanisms implementing emotional vocal signaling across mammals, including humans. But relatively less is known about the specific functional roles that vocal emotions play in human and nonhuman animal social life. By vocal emotions, I refer to the modulation of acoustic properties of vocalizations associated with affective communication, including emotional nonlinguistic utterances (e.g., crying and laughter) and affective prosody (i.e., pitch, loudness, speech rate, and voice quality), both of which interact with verbal content in speech. Vocal signals are adaptations, meaning that they are evolved solutions to recurrent adaptive problems of communication, shaped by natural selection. Historically, scholars of emotional expression have largely ignored questions of adaptation, but increasingly, researchers examining human vocal emotion from a biological perspective have started focusing more on communicative function (e.g., Altenmüller, Schmidt, & Zimmermann, 2013; Bachorowski & Owren, 2003; Briefer, 2012; Bryant & Barrett, 2007, 2008; Cosmides, 1983; Filippi et al., 2017; Keltner & Gross, 1999; Laukka & Elfenbein, 2012; Owren & Rendall,

2001; Pisanski, Cartei, McGettigan, Raine, & Reby, 2016; Sauter, Eisner, Ekman, & Scott, 2010; Scherer, 1984; Shariff & Tracy, 2011). Here, I approach human vocal emotion from an adaptationist perspective, and point to some current issues that researchers might consider in their examinations of this ubiquitous communicative behavior.

Emotional Expressions

Emotions can be construed as superordinate programs that coordinate physiological and psychological components to guide behavior in functionally organized ways (Al-Shawaf, Conroy-Beam, Asao, & Buss, 2016; Cosmides & Tooby, 2000). Affective expression is one important subcomponent often included in the activation of an emotional program, and affective vocalizations represent the most phylogenetically ancient channel of emotional communication (Bass, Gilland, & Baker, 2008). In the study of humans, a long research tradition has explored many systematic connections between acoustic structure in vocal behavior and broad emotional categories (e.g., Banse & Scherer, 1996; Juslin & Laukka, 2003; Murray & Arnott, 1993; Russell, Bachorowski, & Fernández-Dols, 2003). In the last decade or so, there has been a fair amount of research demonstrating that listeners across distant cultures can recognize several emotion categories with some success (Bryant & Barrett, 2008; Cordaro,

Keltner, Tshering, Wangchuk, & Flynn, 2016; Cordaro et al., 2018; Gendron, Roberson, van der Vyver, & Barrett, 2014; Pell, Monetta, Paulmann, & Kotz, 2009; Sauter et al., 2010; Scherer, Banse, & Wallbott, 2001; Thompson & Balkwill, 2006). Even moderate consistency across widely varying human societies suggests a systematic mapping between vocal acoustics and affective states (Elfenbein & Ambady, 2002).

This should come as little surprise given the biological basis of mechanisms underlying vocal behavior. Most generally, voice production in humans involves the independent contributions of vocal source dynamics and supralaryngeal filtering, described by source-filter theory (Fant, 1960; Titze, 1994). The voice source involves controlled air flow from the lungs through the glottis (housed in the larynx), which is converted to sound through the oscillating action of the vocal folds. Air flow causes vibration regimes in the vocal folds that result in a generally tonal, harmonically rich sound. The voiced sound then travels through the supraglottal vocal tract (the top of the larynx to the end of the mouth and nasal cavities) and is subject to a filtering process. Resonating frequencies resulting from this supralaryngeal filter are called formants, which in different configurations form the basis of vowel sounds. These production components result in many measurable sound parameters that are linked to systematic percepts in human listeners, including emotional information (Kreiman & Sidtis, 2011; Taylor, Charlton, & Reby, 2016). Brain mechanisms responsible for vocal emotional production are evolutionarily conserved, meaning that selection has maintained functional organization, and we can identify homologies across current species indicating their likely presence in a shared ancestor (Ackermann, Hage, & Ziegler, 2014; Jurgens, 2002; Owren, Amoss, & Rendall, 2011). Consequently, we see the same kinds of perceptible acoustic features in emotional vocalizations across quite different species (e.g., Briefer, 2012; Filippi et al., 2017).

These facts about the nature of vocal production suggest that many aspects of vocal communication are likely to manifest themselves universally across humans, and to some extent across mammalian species. But humans are particularly cultural beings, and consequently, many behaviors that are rooted in biological systems can appear variably as a function of cultural evolutionary processes. Moreover, people exhibit an extraordinary ability to nonverbally communicate subtle affective distinctions in their voices and faces. For example, recent work found that vocalizers can convey up to 24 different distinct emotions in short vocal bursts, albeit with gradients between the verbal label categories (Cowen, Elfenbein, Laukka, & Keltner, 2018). Overall, this presents theorists with a complicated scenario where biological constraints interact with cultural evolution. This can be true for a number of behaviors, including communicative behaviors, in nonhuman species as well. For example, humpback whale song has been shown to change due to horizontal transmission (e.g., Garland et al., 2011), a phenomenon in these animals described as cultural “revolution” (Noad, Cato, Bryden, Jenner, & Jenner, 2000). But no evidence yet exists outside of humans for cumulative cultural evolution (Kempe, Lycett, & Mesoudi, 2014).

Signals and Cues

The origins of emotional vocal signals can only be understood in the communicative context of senders and receivers. Animal signaling of any kind involves the strategic production of behavioral acts or structures that affect receivers by design, and these signals work effectively because of coevolved responses by receivers, generally for the mutual benefit of both parties (Maynard Smith & Harper, 2003). But not all information transmitted between organisms constitutes signaling, leaving important empirical questions for researchers examining any communicative interactions (Scott-Phillips, 2008). Thus, it is important to distinguish between adaptive signals, byproduct cues, and deceptive coercion (i.e., costly influence on a receiver). Applying this evolutionary framework to vocal affect in humans and nonhuman animals is a fundamental theoretical and empirical problem for researchers studying communicative behavior.

Form and Function

The physical structure of any signal is inherently shaped by its communicative function. For vocal emotion signals specifically, this principle has been useful for understanding nonhuman animal vocal communication (Morton, 1977; Owren & Rendall, 2001), as well as human vocalizations (e.g., Bryant, 2013; Bryant & Barrett, 2007; Cosmides, 1983; Fernald, 1992). The acoustic effects of signals on receivers can be direct, as in the function of interrupting behavior through the use of a loud prohibitive utterance; or indirect, such as the pairing of a threat display with an actual attack, which subsequently in future interactions requires only the display to be effective (Owren & Rendall, 2001). Additionally, this perspective allows us to understand evolutionary convergences in why signals are sometimes structured similarly due to overlaps in the adaptive problems they solve. For example, a fear scream produced by a scared individual might share perceptible acoustic features with an infant wailing (e.g., Sobin & Alpert, 1999). They both tend to be relatively high in average pitch, pitch variability, and loudness. But the signals have these particular features for somewhat different evolutionary reasons. Fear screams evolved from alarm signals: loud, abrupt, and acoustically chaotic displays penetrate noisy environments and are difficult for animals to habituate to—ideal qualities for alerting related group members and inducing fear in them. Infant wailing, which is also typically motivated by a complex of negative emotions, and is a form of an alarm signal (Marler, 1955), shares with screams the design feature of being difficult to habituate to. Its aversive spectral characteristics, however, might be present in order to promote behavior in others that causes crying cessation (Zeifman, 2001). We will return to the example of crying in what follows.

Depending on the context, a variety of selection mechanisms can be responsible for the evolutionary establishment of a given affective signaling system. For example, alarm calling associated with fear has been explained in many species as resulting

from kin selection dynamics where the potential costs to the signaler are outweighed by the benefits conferred to closely related receivers (e.g., Sherman, 1977). But alarm calls have also been shown to be sometimes well designed to signal directly to predators as a deterrent (Zuberbühler, Jenny, & Bshary, 1999). Laughter associated with various positive emotions evolved from play vocalizations (Bryant, 2020; Provine, 2001), and can function as an honest signal of an intent to affiliate through the dynamics of reciprocal altruism (Bryant et al., 2016; Bryant et al., 2018). Angry vocalizations might enhance honest indicators of formidability and thus can be understood as a means to negotiate conflicts through mutual assessment to avoid a physical confrontation (Sell et al., 2010). These examples illustrate how different affective vocal signals might have evolved due to quite different evolutionary processes—the particular acoustic forms are shaped by distinct communicative functions.

An important component in the evolution of signaling is the conflict of interest between senders and receivers. When conflict of interest is high, arms race dynamics can lead to exaggerated structural features in signals (Krebs & Dawkins, 1984). For instance, the subtle predictive cue of a dog moving its lip to avoid self-injury prior to a biting attack can evolve into a baring teeth threat display that communicates affective intentions, thus reducing the likelihood of needing to engage in a physical conflict. Because signals of formidability are often mediated by proximate physiological mechanisms of anger (Sell, Tooby, & Cosmides, 2009), we can see effects of these mechanisms in the signal, including arousal-linked features such as high pitch and loudness, and possibly acoustic nonlinearities such as rapid frequency jumps and deterministic chaos (Fitch, Neubauer, & Herzel, 2002). But these features might be largely reproducible by individuals who have neither the motivation nor the ability to physically back up what the signal invites (i.e., they are deceptive), making it in the interest of receivers to check reliability. This is the process of ritualization—over evolutionary time, more convincing signals will enhance efficacious features with concurrent pressure on receivers to occasionally call the bluff, and so on. The resulting signal structure, in the case of signaling an intent to attack, will be a highly noticeable aggressive display. Exaggerated displays can evolve for other reasons as well, including cases where conflicts of interest are relatively low, such as alarm calls and crying described next.

Since Darwin (1872), many evolutionarily oriented researchers have examined the vocal correlates of emotions in humans and nonhumans (for reviews, see Altenmüller et al., 2013; Briefer, 2012), with a typical emphasis on the connections between emotional internal states and directly associated acoustic configurations in vocalizations. But when an animal produces an emotional expression, what might the function be? A common approach in research with people is to consider many affective displays as automatic indicators of senders' internal states (i.e., cues; e.g., Ekman, 1997). From a game theoretic perspective, this alone is not a stable signaling strategy. Stability here is defined as the state where, given reasonably fixed conditions, alternative strategies by either senders or receivers will

not be positively selected, as the players have entered into the most mutually beneficial interactive pattern (Maynard Smith, 1982). All else being equal, individuals should limit the amount of information they broadcast that is not associated with benefits to them (aggregated over evolutionary time, not for any given encounter) to avoid being manipulated. Alternatively, regularly produced displays are likely to be strategic signals designed to affect the behavior of target receivers in systematic ways, rather than broadcast cues of internal states to be exploited (Fridlund, 1994; Maynard Smith & Harper, 2003). That is not to say that internal states are not involved in the signaling production—indeed, emotional signaling has underlying proximate mechanisms implementing it, and there are specific domains of emotion expression where honestly signaling internal states can be evolutionarily stable. For example, the physiology underpinning vocalized distress signals in young animals can reveal a highly aroused and fearful internal state quite honestly, but by design, the signals typically elicit rapid adaptive responses in adult caregivers. As this example illustrates, a functional approach requires that we do not focus solely on the proximate internal and external variables that lead to the behavior, but also examine the consequences of typical and systematic responses to those signals.

Dezecache, Mercier, and Scott-Phillips (2013) introduced the term *emotional vigilance* to describe the mechanisms that shape how receivers respond to emotional signals. One way that signals can become evolutionarily stable is if there is a potential cost of producing disingenuous versions (Biernaskie, Perry, & Grafen, 2018). If receivers are vigilant—meaning they can detect false versions of a signal and then act in a manner that is against the interests of the sender—there will be pressure on senders to typically produce genuine signals. In the case of affective displays, there would often be positive selection for generating a signal that provides honest information regarding the underlying physiology of the sender and the likely pattern of upcoming behavior connected to the signal (i.e., it is predictive). But there are also dynamics that limit the extent to which receivers should trust such signals because of the risks of being exploited. This approach reconciles historically opposing views on emotion signaling by outlining the circumstances by which revealing internal states benefits both senders and receivers.

Crying offers an excellent example to explore the signal/cue distinction in more depth. Cry-like signals and whimpering are evident across all of the great apes (Hauser, 1996), suggesting that a common ancestor produced a similar vocalization > 25 MYA. Human crying has functional similarities to chimpanzee whimpering (Pusey, 1983) as well as bonobo screams (Oller et al., 2019), and was likely derived from a distress scream or whimpering signal produced by the common ancestor of humans and chimpanzees. But different species take unique evolutionary trajectories and homologous signals will diverge in form and function. The distinctive characteristics of a particular signal are often due to periods of ritualization in which signal properties are enhanced to serve their specific communicative functions (Tinbergen, 1952). Human crying has evolved particularly extreme acoustic features that coevolved with systematic

caregiver response patterns (Zeifman, 2001). The characteristic sound of crying is no accident. Affective arousal and valence induced by a variety of conditions for an animal can lead to vocalized exhalations tied to breathing—byproducts of the physiology associated with the given circumstances (Newman, 2007). For instance, the experience of distress can result in hyperventilation and tension release that could occasionally result in vocalized noises that include phonation. In the case of young offspring, one can easily imagine how this might elicit attention from a caregiver. Infants more likely to exhibit such behavior in circumstances of distress or other discomfort fare better than infants who do not exhibit such behavior, all else being equal. This lays the groundwork for a ritualization process whereby vocalizations reliably associated with discomfort can develop acoustic features that make the crying more noticeable, turning the behavior into an adaptation (i.e., a signal) that is shaped by the responses it elicits in targeted receivers (e.g., caretakers, relatives).

When caregivers share genes with the infant, it is in their interest to respond with some regularity to crying (Lummaa, Vuorisalo, Barr, & Lehtonen, 1998). Over time, selection can shape signals to elicit responses in situations that go well beyond the initial eliciting conditions that enhanced fitness, leading to widened criteria for what triggers the signaling behavior, and complementary flexibility in the responsive reactions. But it is not in the interest of receivers in this scenario to respond indifferently—there should be emotional vigilance to avoid being exploited by overzealous signalers. An arms race ensues that is shaped by the dynamics of the specific interactive properties. In other words, the costs associated with signaling are weighed against the benefits of eliciting responses, but the costs include the trade-offs managed by the receivers where it will be in their interest to limit the degree to which a response is given. The same logic applies to adult crying as well, but with the added complexity of volitional signaling (see following lines).

A conflict of interest between criers and cry responders will result in a signal that is hard to ignore and aversive to receivers. Crying certainly has these sound qualities, typically manifesting as loud, abrupt, high-frequency bursts, often with nonlinear features such as broadband noise and rapid shifts in pitch and loudness (Soltis, 2004; Zeifman, 2001). Selection has caused crying to have these features, and simultaneously shaped perceptual sensitivity to the sound. But what else does crying reveal? Here the distinction between signals and cues can be obvious and important. Imagine an infant with a lung infection whose cries contain distinct spectral features that reveal the illness. Clearly there is no selection for infants to signal illness, but the effects of the infection are impossible to mask in vocalizations. Or imagine that an infant is below average in his strength and size, resulting in cries that reveal weak respiration or other physical deficits. Again, there is no stable signaling strategy for transmitting that information, but the information is available nonetheless. There are clear benefits for caregivers to identify health issues from readily available cues, so we should expect perceptual adaptations for detecting health status in others. Overall, the interactive dynamics must be parsed carefully to separate production adaptations from perceptual ones. The emotional com-

ponents of crying behavior include the systematic effect of proximate mechanisms leading to its activation, which evolved because of the mutually beneficial responses to it. But other aspects of the crying can provide additional information beyond the affective signaling that play a role in the evolutionary stability of crying strategies (Furlow, 1997; also see commentaries in Soltis, 2004).

An important corollary of this approach to crying is that kin interaction is required for the signaling system to initially evolve. The immediate interaction is heavily asymmetrical in that caregivers are paying costs to attend to crying infants who benefit significantly, with receiver benefits only coming later as a function of the infant's development to maturity and subsequent reproduction. In a population of unrelated individuals, crying would not evolve because the genes that cause the behavior would not propagate relative to genes that inhibited such behavior. In terms of population genetics, genes for ignoring crying would invade and take over a population of unrelated responders. Many animal signaling systems have been described successfully from this perspective, with alarm signaling being the most common and theoretically well developed (Maynard Smith & Harper, 2003). The cost to an individual alarm caller over time must be outweighed by the inclusive fitness benefits of making the call, resulting in an evolved motivation by all individuals to produce alarm calls in kin groups (Sherman, 1977).

The example of crying as an evolved emotional signal provides guidelines to understanding any vocal emotion. Distinguishing between signals and cues can be quite difficult empirically. One strategy is to look for evidence of design (Williams, 1966). Functionally organized systems contain structural features that were shaped by selection to solve adaptive problems, and the specific engineering that manifests is not likely to have occurred by chance. Close examination of the physical properties of vocal emotional signals and the systematic behavioral reactions to them can be reverse-engineered in an assessment of how well those properties appear to solve a given adaptive problem of communication. The general empirical approach integrates psychophysics (i.e., the analysis of how stimulus properties affect perception and cognition), cross-species comparative analysis, and evolutionary signaling theory, together which allows researchers to examine how signal forms are integrally shaped by their evolved communicative functions (see Pisanski & Bryant, 2019, for a review of research from this perspective).

Comparative analysis is necessary for understanding the phylogenetic history of any trait, including vocal emotions, and allows us to recognize species-specific evolutionary developments in signal properties. There must be some original behavior from which any affective signal was derived (which can be another signal), and then a social interactive pattern that contributed to the fitness-enhancing features of producing and systematically responding to the signal. There will be other factors that additionally play into the evolutionarily stable strategies that emerge, including byproduct aspects of emotional signals that inform receivers beyond their proper functioning, as well as constraints imposed by the social environment in which the signal evolved. To make matters more complex, volitional control

of affective vocalizations can open up additional niches in which the signal might be effective (see following lines). Finally, the dynamics of deceptive signals will come into play, as certain levels of dishonesty can be maintained if the benefits of honest signaling are robust (Johnstone & Grafen, 1993). But there are currently no formal models of deception in specifically human domains of vocal emotion, including crying, laughter, or other affective vocalizations.

Dual Pathway Vocal Production

Human vocal communication is zoologically unique because of dual pathway vocal production (Ackermann et al., 2014; Jurgens, 2002; Owren et al., 2011). An evolutionarily conserved vocal emotion system common to all mammal species is supplemented in humans by a speech articulation system with direct connections from motor cortex to laryngeal control musculature (Ackermann, 2008; Owren et al., 2011). Recently, scholars have pointed out that a variety of volitional control mechanisms likely exist in nonhuman primates, suggesting that the underlying neural circuitry is more complex than originally assumed, and that multiple dimensions of classification are needed for proper comparative analysis (e.g., Gruber & Grandjean, 2017; Lameira, 2017; Loh, Petrides, Hopkins, Procyk, & Amiez, 2017). Humans, however, are the only species with speech, which is assimilated with a variety of cognitive systems and is the foundation for linguistic communication. Different proposals have attempted to describe the integration of affective and linguistic vocal expression, as the two systems interface in ways that are not currently well understood. Specific affective vocalization types interrupt speech resulting in an inability to talk when laughing or crying, for example. But affective prosodic patterning operates in tandem with linguistic prosody, resulting in a hierarchically structured production stream with clear emotional signaling components. Moreover, people are able to modulate their voices in a variety of ways to accommodate different social contexts and motivational states (see Pisanski et al., 2016, for a recent review). Phonological, morphological, and syntactic systems have consequences for many prosodic phenomena (Cutler, Dahan, & van Donselaar, 1997), resulting in recursive structure within linguistic levels of intonational phonology in production (Ladd, 1986) and brain processing associated with perception (e.g., Friederici & Alter, 2004). Some models of prosody have posited interactions between different prosodic systems that map onto the dual pathway framework (e.g., Fujisaki, 1988; Sammler, Grosbras, Anwander, Bestelmeyer, & Belin, 2015), and related work shows that linguistic and affective prosodic production can trade off as a function of emotional intent (McRoberts, Studdert-Kennedy, & Shankweiler, 1995).

Dual pathway vocal production has important implications for evolutionary theories of emotional signaling. The decoupling of affective motivation from the production of the vocal signal opens up a niche for manipulation unlike most nonhuman primate vocal production. If an individual can produce a convincing facsimile of an affective vocalization, then that individual stands a chance at gaining the benefits of that signal

while paying minimal, if any, costs. Strategic dynamics of communicative social interaction then cause selection for vigilance on the part of receivers to not be unduly manipulated (Krebs & Dawkins, 1984). For example, a fake laugh might mislead a potential cooperater in an interaction to face defection (Bryant & Aktipis, 2014), a fake crier could motivate advantageous caregiver attention (Nakayama, 2010), and faked pleasure during sex could garner benefits from partners (Brewer & Hendrie, 2011). Selection for detecting such manipulative signals can result in an arms race that pits refinements in accurate signal reproduction against sensitivity in detection systems (Bryant & Aktipis, 2014; Bryant et al., 2018; Krebs & Dawkins, 1984).

But most volitional affective vocal production, including the deliberate nonverbal vocalizations just described, is likely functioning during social interaction in varying pragmatic ways—cooperative in both the Gricean sense and the biological sense (Pinker, Nowak, & Lee, 2008). For example, volitional laughter can signal comprehension (Flamson & Bryant, 2013; O'Donnell-Trujillo & Adams, 1983), conversational turn-taking (Gavioli, 1995), and verbal play (Bryant, 2011; Holt, 2016). Other vocalizations can be analyzed similarly—we groan, grunt, and shriek during conversational interactions, motivated by ostensive intent, but not necessarily because of genuine emotional triggering. One interesting outcome of the production combination of speech and volitional affective vocalizations is that nonverbal signals can be inserted into speech, such as laughter and cries embedded into spoken syllables (e.g., Nwokah, Hsu, Davies, & Fogel, 1999). This is unlike spontaneous (i.e., genuine) affective vocalizations that necessarily interrupt speech. While generally due to benign intent, embeddedness could plausibly reveal deceptive manipulation in certain contexts, such as in cases where individuals are feigning emotional reactions to gain some benefit from the receiver (e.g., fake laughing to pretend real emotional engagement). People might be sensitive to, and sometimes suspicious of, the presence of linguistic phenomena during the production of volitional vocal emotions (e.g., laughter or crying embedded into speech syllables), especially when there is some perceived risk of being manipulated. Relatedly, on the production side, we might expect a relatively lower rate of these types of phenomena in individuals with skills in deceptive behavior.

Universals and Emotion Categories

A substantial body of research in humans and nonhuman animals has explored the extent to which acoustic structural properties of communicative signals map onto specific categories of emotion (Ekman, 1972), including a subset of studies examining vocalizations and music (for a review, see Juslin & Laukka, 2003). The history of this work is extensive and beyond the scope of this article, but debates continue regarding a couple of key questions. One issue is whether emotions are best understood as basic categories (e.g., anger, sadness, happiness) or if a dimensional approach is more appropriate (e.g., high vs. low arousal, and positive vs. negative valence; Ekman, 1992; Russell, 1980). Other theories integrate aspects of these approaches, such as how

cognitive appraisal drives dynamic, multicomponent information processing resulting in specific, context-sensitive vocal production and perception (e.g., Scherer, 1986). As described next, these approaches each touch on important elements in emotional communication and can be integrated through an evolutionary framework.

A somewhat related question concerns the extent to which basic emotions are universal, including emotion states and affective signaling in the face and voice (Jack, Garrod, Yu, Caldara, & Schyns, 2012; Keltner, Sauter, Tracy, & Cowen, 2019; Russell et al., 2003). Recently there has been a surge of interest in carving out emotional categories more finely (e.g., Cowen et al., 2018). Consequently, universality can be construed as existing on a continuum, with universals being confirmed empirically, for example, as a function of the ubiquity of a translatable verbal label, among other factors. To some extent, these debates exist as historical artifacts, and, through the lens of evolutionary theory, are in many ways superfluous. How universal does universal have to be? What does one counterexample actually mean? And what exactly is it that is universal—is it the underlying emotion construct, the structural features of affective signals, or the ability of perceivers to correctly identify emotions in some task, or all of these? There are many good reasons to expect that none of these questions can be answered completely in the affirmative. But these issues are quite independent from whether the underlying emotion signaling systems are evolved, or rooted in biology.

When examining signaling behaviors as solutions to particular adaptive communication problems, questions of universality, while interesting and potentially informative, become secondary. By analogy, consider color perception. We know that humans have reliably developing and invariant sensitivity to variations in wavelengths and spectral properties of light waves due to cellular organization that corresponds to percepts of color (Palmer, 1999). But judgments of color must rely on language and categorization. A similar situation exists for emotion research. We know there is a reliably developing emotion system that includes subcomponents for affective signaling, but measures of the signals must also be filtered through verbal labelling schemes. Researchers in actual color perception, or emotion signaling, should be less preoccupied about the effects of language and more directly concerned with the design features of the respective systems, including the range and variation of the physical signals that trigger the activation of typical perceptual and behavioral responses. Of course, there are important questions lying at the interface of perception and language, but debates concerning the degree of universal perception of a particular language-based category will not get us much closer to properly describing the underlying architecture of the proposed evolved system.

The extent to which a form–function relationship is apparent to listeners (i.e., how vocalizations affect the behavior of target audience members as a function of their physical properties) is likely the best predictor of how successfully a signal might be recognized universally across cultures (Bryant & Barrett, 2007). The question of whether emotions can be conceptualized as comprising basic categories that correspond to verbal labels,

and are hierarchically integrated in some fashion, does not have to be at odds with other approaches that emphasize the dynamic and dimensional nature of emotion. Basic emotions correspond loosely to specific adaptive problems, and the recent refinements of these categories also likely reflect the domain-specific characteristics of various affective signals. But dimensional approaches capture important variation and have been useful for nonhuman primate research as well as work with people. There are clear universals in how vocal affect manifests itself in human communication, and there are also clear variations in display rules (i.e., culturally evolved, pragmatically based patterns of affective expression) and how people understand affect and intention—the role of language and culture in how we talk about and conceptualize emotions has its place in the broader scheme of emotion research. But I argue that the unifying framework is an evolutionary approach that emphasizes evolved form–function relationships between the physical properties of signals and the adaptive response patterns in receivers.

Another important issue involves the way emotions are typically examined empirically. There are several constraints on affective vocalizations that are difficult to accommodate in the lab, leading to problems of ecological validity. Scripted and performed displays produced in largely decontextualized situations will result in affective tokens that are generally quite removed from the spontaneous counterparts we see in real life. Instead, such displays are generated from volitional vocal production governed by different underlying neural machinery, with the production and perception processes subject to a variety of cultural factors, including language and display rules. These considerations alone introduce serious limitations on the structural regularities we might see, and consequently will affect perceivers' ability to identify them accurately. But what about the perceivers' tasks? Certainly, making repeated judgments of vocal clips out of context only triggers the appropriate detection systems to a limited extent, and instead listeners are unconsciously considering many extraneous variables during the task, such as inferring researchers' intent, scrutinizing esoteric dimensions of the stimuli, engaging in a task like a typical distracted subject, and so on (Bryant, 2012). For example, brain imaging research on the perception of spontaneous and volitional laughter found that anterior medial prefrontal cortex associated with mind reading was activated while hearing volitional laughter but not spontaneous laughter (McGettigan et al., 2013). This illustrates how acted stimuli can trigger cognitive systems for recognizing the strategic motivations of producers. These kinds of limitations in experimental contexts will negatively impact people's ability to perform similarly as they might in real-life interactions.

These are not easy problems to overcome. Researchers must find the appropriate trade-off between ecological validity in stimulus materials and proper experimental control. Additionally, we must be careful in developing our dependent measures. For stimulus materials, the use of spontaneous vocalizations is often possible, whether they are extracted from natural conversations or videos online (e.g., Anikin & Persson, 2017). If the source material is generated spontaneously, we have better assurance that the tokens reflect typical production in the world, as opposed to actors who play into stereotypes and other contrived

performances (Bryant & Fox Tree, 2005). Spontaneous vocal stimulus materials can be systematically manipulated acoustically (e.g., PSOLA), thus affording tight experimental control combined with natural vocal production variation (e.g., Bryant & Aktipis, 2014). But even potentially more important than the stimulus materials is the task that participants must complete when they engage with a study. The use of rating scales is a convenient method to get at people's judgments, and in some cases is the only viable alternative to examine some phenomenon. But when possible, introducing a task where a judgment involves identifying an actual correct answer is more likely to tap into the psychology of interest. Consider our earlier analogy of color perception. If we asked people around the world to rate how blue a color is, or just identify the color verbally, we would get tremendous variation, and might even come to the conclusion, based on the pattern of responses, that people are perceiving it differently. But if we instead used an ABX task where two different hues of a color are presented, then a third presentation matched one of the two previous ones and participants had to decide whether it was A or B, we could see dramatically different results. The ABX task has a correct answer in that X actually matches either A or B. Requiring the identification of a correct answer is likely to reduce cultural variation in task performance, and also is more in line with the assumptions of the statistical methods commonly used in the social sciences (i.e., there are more likely to be normal distributions in people's ability to access a correct answer than in people's opinions).

Conclusion

Emotional vocalizations are shaped by natural selection to affect receivers in systematic ways, and coevolve with response patterns. But cues of emotion activation can be used by perceivers to predict behavior, and there are likely to be perceptual adaptations to detect them. The acoustic structures of affective vocal signals are evolutionarily designed to have particular communicative effects, solving specific adaptive problems for senders and receivers. These adaptive problems map loosely onto traditional basic emotion categories, but research could benefit by switching the emphasis from finer categorization schemes based on verbal labels to more ecologically informed analyses of communicative function. An integration of current approaches would involve acknowledging (a) affect programs differentiate to some extent according to basic categories, (b) dimensional approaches to emotion have predictive utility in understanding the structural features of signals by capturing fundamental aspects of their adaptive design features (e.g., measuring arousal), and (c) ecological context is crucial for inferential programs in determining the effect of emotional signals such that specific information is not usually encoded in the signals.

Signaling strategies can be formalized through game theoretic analyses of repeated interactions within social groups. These analyses reveal that senders should not broadcast internal states without some expected utility, and the introduction of volitional signals decoupled from genuine emotional triggers adds a level of complexity to human vocal signaling that is often rather different from nonhuman communication. But as for nonhuman vocal communication, context is paramount. The physical structure of

affective signals in any modality is just one source of contextualized information that usually requires rich inferential processing to be effective. A complete account of the evolutionary origins of vocal affect requires theorists to move beyond comparative analyses of vocalizations and take seriously the need to properly map the functionally organized nature of affective vocal production and perception.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Gregory A. Bryant  <https://orcid.org/0000-0002-7240-4026>

References

- Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, *31*(6), 265–272.
- Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *37*(6), 529–546.
- Al-Shawaf, L., Conroy-Beam, D., Asao, K., & Buss, D. M. (2016). Human emotions: An evolutionary psychological perspective. *Emotion Review*, *8*(2), 173–186.
- Altenmüller, E., Schmidt, S., & Zimmermann, E. (Eds.). (2013). *The evolution of emotional communication: From sounds in nonhuman mammals to speech and music in man*. Oxford, UK: Oxford University Press.
- Anikin, A., & Persson, T. (2017). Nonlinguistic vocalizations from online amateur videos for emotion research: A validated corpus. *Behavior Research Methods*, *49*(2), 758–771.
- Bachorowski, J. A., & Owren, M. J. (2003). Sounds of emotion: Production and perception of affect-related vocal acoustics. *Annals of the New York Academy of Sciences*, *1000*(1), 244–265.
- Banase, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*, 614–636.
- Bass, A. H., Gilland, E. H., & Baker, R. (2008). Evolutionary origins for social vocalization in a vertebrate hindbrain-spinal compartment. *Science*, *321*(5887), 417–421.
- Biernaskie, J. M., Perry, J. C., & Grafen, A. (2018). A general model of biological signals, from cues to handicaps. *Evolution Letters*, *2*(3), 201–209.
- Brewer, G., & Hendrie, C. A. (2011). Evidence to suggest that copulatory vocalizations in women are not a reflexive consequence of orgasm. *Archives of Sexual Behavior*, *40*(3), 559–564.
- Briefer, E. F. (2012). Vocal expression of emotions in mammals: Mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1–20.
- Bryant, G. A. (2011). Verbal irony in the wild. *Pragmatics and Cognition*, *19*(2), 291–309.
- Bryant, G. A. (2012). Is verbal irony special? *Language and Linguistics Compass*, *6*(11), 673–685.
- Bryant, G. A. (2013). Animal signals and emotion in music: Coordinating affect across groups. *Frontiers in Psychology*, *4*, 1–13. <https://doi.org/10.3389/fpsyg.2013.00990>
- Bryant, G. A. (2020). Evolution, structure, and functions of human laughter. In K. Floyd & R. Weber (Eds.), *Handbook of communication science and biology* (pp. 63–77). New York, NY: Routledge.

- Bryant, G. A., & Aktipis, C. A. (2014). The animal nature of spontaneous human laughter. *Evolution and Human Behavior*, 35(4), 327–335.
- Bryant, G. A., & Barrett, H. C. (2007). Recognizing intentions in infant-directed speech: Evidence for universals. *Psychological Science*, 18(8), 746–751.
- Bryant, G. A., & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, 8(1–2), 135–148.
- Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., . . . De Smet, D. (2016). Detecting affiliation in colughter across 24 societies. *Proceedings of the National Academy of Sciences of the USA*, 113(17), 4682–4687.
- Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Amir, D., Chávez, B., . . . Fux, M. (2018). The perception of spontaneous and volitional laughter across 21 societies. *Psychological Science*, 29(9), 1515–1525.
- Bryant, G. A., & Fox Tree, J. E. (2005). Is there an ironic tone of voice? *Language and Speech*, 48(3), 257–277.
- Cordaro, D. T., Keltner, D., Tshering, S., Wangchuk, D., & Flynn, L. M. (2016). The voice conveys emotion in ten globalized cultures and one remote village in Bhutan. *Emotion*, 16(1), 117–128.
- Cordaro, D. T., Sun, R., Keltner, D., Kamble, S., Huddar, N., & McNeil, G. (2018). Universals and cultural variations in 22 emotional expressions across five cultures. *Emotion*, 18(1), 75–83.
- Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 864–881.
- Cosmides, L., & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (2nd ed., pp. 91–115). New York: Guilford Press.
- Cowen, A. S., Elfenbein, H. A., Laukka, P., & Keltner, D. (2018). Mapping 24 emotions conveyed by brief human vocalization. *American Psychologist*. Advance online publication. <https://doi.org/10.1037/amp0000399>
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40(2), 141–201.
- Darwin, C. (1872). *The expression of the emotions in man and animals*. London, UK: John Murray.
- Dezechache, G., Mercier, H., & Scott-Phillips, T. C. (2013). An evolutionary approach to emotional communication. *Journal of Pragmatics*, 59, 221–233.
- Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), *Nebraska Symposium on Motivation (Vol. 19)*, pp. 207–282. Lincoln: University of Nebraska Press.
- Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion*, 6(3–4), 169–200.
- Ekman, P. (1997). Should we call it expression or communication? *Innovation: The European Journal of Social Science Research*, 10(4), 333–344.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128, 208–235.
- Fant, G. (1960). *Acoustic theory of speech production*. The Hague, The Netherlands: Mouton.
- Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 391–428). Oxford, UK: Oxford University Press.
- Filippi, P., Congdon, J. V., Hoang, J., Bowling, D. L., Reber, S. A., Pašukonis, A., . . . Newen, A. (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). <https://doi.org/10.1098/rspb.2017.0990>
- 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.
- Flamson, T. J., & Bryant, G. A. (2013). Signals of humor: Encryption and laughter in social interaction. In M. Dynel (Ed.), *Developments in linguistic humour theory* (pp. 49–74). Amsterdam, The Netherlands: John Benjamins.
- Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego, CA: Academic Press.
- Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language*, 89(2), 267–276.
- Fujisaki, H. (1988). A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In O. Fujimura (Ed.), *Vocal physiology: Voice production, mechanisms, and functions* (Vol. 2, pp. 347–355). Philadelphia, PA: Lippincott Williams & Wilkins.
- Furlow, B. F. (1997). Human neonatal cry quality as an honest signal of fitness. *Evolution and Human Behavior*, 18, 175–193.
- Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., . . . Noad, M. J. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current Biology*, 21(8), 687–691.
- Gavioli, L. (1995). Turn-initial versus turn-final laughter: Two techniques for initiating remedy in English/Italian bookshop service encounters. *Discourse Processes*, 19(3), 369–384.
- Gendron, M., Roberson, D., van der Vyver, J. M., & Barrett, L. F. (2014). Cultural relativity in perceiving emotion from vocalizations. *Psychological Science*, 25(4), 911–920.
- Gruber, T., & Grandjean, D. (2017). A comparative neurological approach to emotional expressions in primate vocalizations. *Neuroscience & Biobehavioral Reviews*, 73, 182–190.
- Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.
- Holt, E. (2016). Laughter at last: Playfulness and laughter in interaction. *Journal of Pragmatics*, 100, 89–102.
- Jack, R. E., Garrod, O. G., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences of the USA*, 109(19), 7241–7244.
- Johnstone, R. A., & Grafen, A. (1993). Dishonesty and the handicap principle. *Animal Behaviour*, 46, 759–764.
- Jurgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, 26, 235–238.
- Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129, 770–814.
- Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. *Cognition and Emotion*, 13(5), 467–480.
- Keltner, D., Sauter, D., Tracy, J., & Cowen, A. (2019). Emotional expression: Advances in basic emotion theory. *Journal of Nonverbal Behavior*, 43(2), 133–160.
- Kempe, M., Lycett, S. J., & Mesoudi, A. (2014). From cultural traditions to cumulative culture: Parameterizing the differences between human and nonhuman culture. *Journal of Theoretical Biology*, 359, 29–36.
- Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioral ecology: An evolutionary approach* (pp. 380–402). Oxford, UK: Blackwell Scientific.
- Kreiman, J., & Sidtis, D. (2011). *Foundations of voice studies: An interdisciplinary approach to voice production and perception*. Hoboken, NJ: John Wiley and Sons.
- Ladd, D. R. (1986). Intonational phrasing: The case for recursive prosodic structure. *Phonology*, 3, 311–340.
- Lameira, A. R. (2017). Bidding evidence for primate vocal learning and the cultural substrates for speech evolution. *Neuroscience & Biobehavioral Reviews*, 83, 429–439.
- Laukka, P., & Elfenbein, H. A. (2012). Emotion appraisal dimensions can be inferred from vocal expressions. *Social Psychological and Personality Science*, 3(5), 529–536.
- Loh, K. K., Petrides, M., Hopkins, W. D., Procyk, E., & Amiez, C. (2017). Cognitive control of vocalizations in the primate ventrolateral-dorsomedial frontal (VLF-DMF) brain network. *Neuroscience & Biobehavioral Reviews*, 82, 32–44.

- Lummaa, V., Vuorisalo, T., Barr, R. G., & Lehtonen, L. (1998). Why cry? Adaptive significance of intensive crying in human infants. *Evolution and Human Behavior, 19*, 193–202.
- Marler, P. (1955). Characteristics of some animal calls. *Nature, 176*, 6–8.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge, UK: Cambridge University Press.
- Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, UK: Oxford University Press.
- McGettigan, C., Walsh, E., Jessop, R., Agnew, Z. K., Sauter, D. A., Warren, J. E., & Scott, S. K. (2013). Individual differences in laughter perception reveal roles for mentalizing and sensorimotor systems in the evaluation of emotional authenticity. *Cerebral Cortex, 25*(1), 246–257.
- McRoberts, G. W., Studdert-Kennedy, M., & Shankweiler, D. P. (1995). The role of fundamental frequency in signaling linguistic stress and affect: Evidence for a dissociation. *Perception & Psychophysics, 57*(2), 159–174.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist, 111*, 855–869.
- Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*, 1097–1108.
- Nakayama, H. (2010). Development of infant crying behavior: A longitudinal case study. *Infant Behavior and Development, 33*(4), 463–471.
- Newman, J. D. (2007). Neural circuits underlying crying and cry responding in mammals. *Behavioural Brain Research, 182*(2), 155–165.
- Noad, M. J., Cato, D. H., Bryden, M. M., Jenner, M. N., & Jenner, K. C. S. (2000). Cultural revolution in whale songs. *Nature, 408*(6812). <https://doi.org/10.1038/35046199>
- Nwokah, E. E., Hsu, H. C., Davies, P., & Fogel, A. (1999). The integration of laughter and speech in vocal communication: A dynamic systems perspective. *Journal of Speech, Language, and Hearing Research, 42*(4), 880–894.
- O'Donnell-Trujillo, N., & Adams, K. (1983). Heheh in conversation: Some coordinating accomplishments of laughter. *Western Journal of Speech Communication, 47*, 175–191.
- Oller, D. K., Griebel, U., Iyer, S. N., Jhang, Y., Warlaumont, A. S., Dale, R., & Call, J. (2019). Language origins viewed in spontaneous and interactive vocal rates of human and bonobo infants. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.00729>
- 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.
- 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: Issues, News, and Reviews, 10*(2), 58–71.
- Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior, 33*(2), 107–120.
- Pinker, S., Nowak, M. A., & Lee, J. J. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences of the USA, 105*(3), 833–838.
- Pisanski, K., & Bryant, G. A. (2019). The evolution of voice perception. In N. S. Eidsheim & K. L. Meizel (Eds.), *Oxford handbook of voice studies* (pp. 269–300). New York, NY: Oxford University Press.
- Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: A window into the origins of human vocal control? *Trends in Cognitive Sciences, 20*(4), 304–318.
- Provine, R. R. (2001). *Laughter: A scientific investigation*. New York, NY: Penguin.
- Pusey, A. E. (1983). Mother-offspring relationships in chimpanzees after weaning. *Animal Behaviour, 31*(2), 363–377.
- Ruch, W., & Ekman, P. (2001). The expressive pattern of laughter. In A. Kaszniak (Ed.), *Emotion, qualia, and consciousness* (pp. 426–443). Singapore: World Scientific Publishing.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.
- Russell, J. A., Bachorowski, J. A., & Fernández-Dols, J. M. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*(1), 329–349.
- Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology, 25*(23), 3079–3085.
- 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 of the USA, 107*(6), 2408–2412.
- Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 293–317). Mahwah, NJ: Lawrence Erlbaum Associates.
- Scherer, K. R. (1986). Vocal affect expression: A review and model for future research. *Psychological Bulletin, 99*, 143–165.
- Scherer, K. R. (2003). Vocal communication of emotion: A review of research paradigms. *Speech Communication, 40*(1–2), 227–256.
- Scherer, K. R., Banse, R., & Wallbott, H. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76–92.
- 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). <https://doi.org/10.1371/journal.pone.0091192>
- Scott-Phillips, T. (2008). Defining biological communication. *Journal of Evolutionary Biology, 21*(2), 387–395.
- Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., . . . Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B: Biological Sciences, 277*(1699), 3509–3518.
- Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. *Proceedings of the National Academy of Sciences of the USA, 106*(35), 15073–15078.
- Shariff, A. F., & Tracy, J. L. (2011). What are emotion expressions for? *Current Directions in Psychological Science, 20*(6), 395–399.
- Sherman, P. W. (1977). Nepotism and the evolution of alarm calls. *Science, 197*, 1246–1253.
- Sobin, C., & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research, 28*(4), 347–365.
- Soltis, J. (2004). The signal functions of early infant crying. *Behavioral and Brain Sciences, 27*(4), 443–458.
- Taylor, A. M., Charlton, B. D., & Reby, D. (2016). Vocal production by terrestrial mammals: Source, filter, and function. In R. Suthers, W. Fitch, R. Fay, & A. Popper (Eds.), *Springer handbook of auditory research: Vertebrate sound production and acoustic communication* (Vol. 53, pp. 229–260). Cham, Switzerland: Springer.
- Thompson, W., & Balkwill, L. L. (2006). Decoding speech prosody in five languages. *Semiotica, 158*(1–4), 407–424.
- Tinbergen, N. (1952). Derived activities: Their causation, biological significance, origin and emancipation during evolution. *Quarterly Review of Biology, 27*, 1–32.
- Titze, I. R. (1994). *Principles of voice production*. Englewood Cliffs, NJ: Prentice Hall.
- Williams, G. C. (1966). *Adaptation and natural selection: A critique of some current evolutionary thought*. Princeton, NJ: Princeton University Press.
- Zeifman, D. M. (2001). An ethological analysis of human infant crying: Answering Tinbergen's four questions. *Developmental Psychobiology, 39*, 265–285.
- Zuberbühler, K., Jenny, D., & Bshary, R. (1999). The predator deterrence function of primate alarm calls. *Ethology, 105*(6), 477–490.