



## OPEN ACCESS

## EDITED BY

Sang-im Lee,  
Daegu Gyeongbuk Institute of Science and  
Technology (DGIST), Republic of Korea

## REVIEWED BY

Sonja C Vernes,  
University of St Andrews, United Kingdom  
Mirjam Knörnschild,  
Leibniz Institut für Evolutions und  
Biodiversitätsforschung, Germany

## \*CORRESPONDENCE

Eliot A. Brenowitz  
✉ eliotb@uw.edu

RECEIVED 02 April 2023

ACCEPTED 14 June 2023

PUBLISHED 30 June 2023

## CITATION

Brenowitz EA and Beecher MD (2023) An  
ecological and neurobiological perspective  
on the evolution of vocal learning.  
*Front. Ecol. Evol.* 11:1193903.  
doi: 10.3389/fevo.2023.1193903

## COPYRIGHT

© 2023 Brenowitz and Beecher. This is an  
open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that  
the original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# An ecological and neurobiological perspective on the evolution of vocal learning

Eliot A. Brenowitz<sup>1,2\*</sup> and Michael D. Beecher<sup>1,2</sup>

<sup>1</sup>Animal Behavior Program, Department of Psychology, University of Washington, Seattle, WA, United States,

<sup>2</sup>Department of Biology, University of Washington, Seattle, WA, United States

Vocal production learning (VPL) involves the use of auditory experience to guide the production of novel signals or to modify pre-existing signals. It allows animals to develop signals that are more complex and/or more flexible than innately developed signals. It has evolved rarely in vocal animals, widespread only in three avian and four mammalian taxa. The evolution of VPL was accompanied by innovations of the vocal motor neural circuitry. VPL is rare because of its various costs. Ecology, social spacing, and social fluidity can favor the evolution of VPL. It is striking that most taxa with VPL evolved in visually limited habitats, where sound is the only effective channel for communicating over distance from sender to receiver. Selective factors that favor the ability to produce complex and/or flexible signals would act predominantly on acoustic signals, and favor the evolution of VPL. Learning may be the only practical way to develop a signal complex enough to encode different types of information for assessment by receivers in animals that rely on acoustic communication, or to modify signals as local social factors dictate.

## KEYWORDS

animal communication, bat, birdsong, dolphin, evolution of vocal learning, Humpback Whale, language, songbird

## 1 Introduction

Vocal learning is a distinctive phenomenon, inherently interesting to humans given our unique proficiency with language (Hauser et al., 2002; Beecher, 2021). Other than humans, the ability to learn to produce species-typical vocalizations has been demonstrated to be widespread only in three avian taxa (oscine passerines or songbirds, parrots and hummingbirds), and four mammalian taxa (cetaceans, pinnipeds, bats and elephants). In most taxa that use acoustic communication, animals develop signals normally without environmental input such as hearing the vocal signals of conspecific adults. We refer to such signals, and their underlying developmental program, as “innate.”

There is an extensive literature on the evolution of vocal learning in humans and animals, which attempts to explain why it has evolved so rarely, and examining whether there are evolutionary antecedents to vocal learning in nonhuman primates and other mammalian taxa. Most of this literature focuses on what qualifies as vocal learning

(e.g., Janik and Slater, 2000; Fischer, 2017; Wirthlin et al., 2019; Martins and Boeckx, 2020; Tyack, 2020; Janik and Knörnschild, 2021; Vernes et al., 2021), the evolution and taxonomic distribution of vocal learning (Nottebohm, 1972; Fitch, 2000; Petkov and Jarvis, 2012; Nowicki and Searcy, 2014; Jarvis, 2019; Corballis, 2020; Vernes and Wilkinson, 2020), and convergent neural and genetic mechanisms of vocal learning (Gahr, 2000; Jarvis, 2007; White, 2010; Sober and Brainard, 2012; Condro and White, 2014; Chen et al., 2016; Roberts et al., 2017; Rodenas-Cuadrado et al., 2018; Jarvis, 2019; Choe and Jarvis, 2021). Surprisingly, however, little attention has been devoted to the critical roles of environment and ecology as selective factors and constraints in the evolution of vocal learning (but see Janik and Slater, 2000; Fischer, 2017; Cheney and Seyfarth, 2018; Hedwig et al., 2021). Our goal in this paper is to provide an integrative framework for understanding the evolution of vocal learning that includes consideration of 1) The neural substrates that enable animals to modify vocal motor behavior in response to environmental input; 2) The costs of vocal learning that constrain its evolution; 3) Ecological factors that select for vocal learning; and 4) the adaptive behavioral advantages of vocal learning.

## 2 Vocal production learning

Vernes et al. (2021) and Tyack (2016) have reviewed the various forms of vocal learning found in higher vertebrates. We agree with them that vocal learning is not binary, but varies in expression within and between taxa. We believe, however, that there are limits to what should be considered vocal learning. Ambiguity has arisen in this field from inconsistent standards for what qualifies as learning, including at one extreme, sources of signal plasticity such as growth, maturation, and response to immediate sensory feedback (as in turn-taking and chorusing). Expanding on Vernes et al., we broadly define vocal learning as modification of an individual's vocal signaling behavior in response to experience. This learning can take one of two forms. In vocal production learning, an individual develops a novel signal or modifies a pre-existing one on the basis of input from the vocalizations of other individuals and auditory feedback from their own vocalizations. In vocal usage learning, the individual learns to employ a pre-existing signal (which itself may be learned or innate) in a new context (Janik and Slater, 2000) (Vernes et al., 2021). In this paper we focus exclusively on vocal production learning (abbreviated VPL hereafter) since it has evolved in relatively few acoustically signaling animals, and its evolution has proven challenging to explain (Fitch, 2000; Nowicki and Searcy, 2014).

VPL is most clearly demonstrated when animals modify their vocalizations to match a novel acoustic signal to which they have not been previously exposed. With captive animals, experimental studies can be conducted in which subjects are trained to produce sounds not in their normal species-typical repertoire, such as heterospecific vocalizations (e.g., Baptista and Petrino, 1984; Stoeger et al., 2012; Stansbury and Janik, 2019), conspecific signal variants such as different geographic dialects (e.g., Marler and Tamura, 1964; Mennill et al., 2018), or experimentally

manipulated conspecific signals (Marler and Peters, 1988; Rose et al., 2004). In animals where it is not feasible or ethical to bring them into captivity, VPL can be seen when members of a social group modify their vocalizations to mimic a novel conspecific signal to which they have been exposed, or individuals learn to accurately mimic novel heterospecific signals (as in avian vocal mimics such as mockingbirds and mynah birds). In this review, we will focus on cases where vocal learning appears to be common or widespread within the taxon, i.e., in three avian taxa (songbirds, parrots and hummingbirds), and four mammalian taxa (cetaceans, pinnipeds, bats and elephants). Cases that appear to be exceptional within a larger taxon (e.g., humans within primates, bellbirds within suboscines (Kroodsma et al., 2013)) are ideal cases for testing the hypotheses we will develop in this paper about special ecological selection pressures or existing preadaptations that favor the evolution of VPL, because in such cases one can contrast the exceptional case with all its close non-VPL relatives. We return to this point in the Discussion.

In most mammalian taxa with acoustic communication, signals develop innately (Janik and Slater, 1997; Vernes et al., 2021). Given our shared evolutionary ancestry with nonhuman primates, there has long been a teleological presumption that primates must show some form of VPL (Lameira et al., 2022). There are claims that observations of developmental plasticity or individual variation in the calls of some nonhuman primates constitute evidence of VPL (Lemasson et al., 2011; Wich et al., 2012; Zürcher et al., 2021), but these reports are not supported by any of the above criteria. Decades of study have failed to produce convincing evidence of VPL in any primate species (Hammerschmidt and Fischer, 2008; Fischer, 2017; Cheney and Seyfarth, 2018). Social nonhuman primates can signal in the olfactory, visual, auditory, and tactile domains and may thus be able to engage in complex communication behavior using innately developed vocalizations (Mitoyen et al., 2019). Nonhuman primates and other higher vertebrates do exhibit vocal usage learning in which they learn to associate an innate signal with a novel social or ecological context (Janik and Slater, 2000).

The failure to find VPL in nonhuman primates is seen by some investigators as being inconsistent with our shared evolutionary history, and has led to suggestions that changes in the frequency and/or temporal structure of calls observed with growth and maturation in some primates are examples of production learning (e.g., Takahashi et al., 2015; Lameira et al., 2022). Other investigators have suggested that VPL, usage learning, and various forms of non-learning dependent vocal plasticity such as ontogenetic changes in call timing or frequency, turn-taking, and chorusing all fall along a continuum of vocal learning, perhaps serving as evolutionary and mechanistic antecedents to VPL seen in humans and songbirds (Petkov and Jarvis, 2012; Jarvis, 2019; Wirthlin et al., 2019; Martins and Boeckx, 2020; Bruno et al., 2021). These broad suggestions do not stand up well under close scrutiny, however (e.g., Martins and Boeckx, 2020). There is no direct evidence to support the idea that the ability of some nonhuman primates and other animals to make small changes in call structure with maturation or experience is dependent on similar mechanisms as VPL (Fischer et al., 2015). In species where

individuals coordinate calling, as with turn-taking in Naked Mole Rats (*Heterocephalus glaber*) (Yosida et al., 2007; Pika et al., 2018), overlapping calls in *Smilisca* tree frogs (Ryan, 1986), and jamming of competitors' calls in bats (Corcoran and Conner, 2014), we agree with Vernes et al. (2021) that “there are some vertebrates where temporal vocal coordination can clearly be attributed to a central nervous system oscillator that is responsive to call perception, rather than a learned mechanism...” These types of coordinated calling may be regulated by non-learning dependent neural mechanisms such as acoustically evoked inhibition of vocal motor output, and timing input from brain regions that integrate auditory and motor information (Endepols and Walkowiak, 1999; Kelley and Bass, 2010; Banerjee and Vallentin, 2022). One need not invoke learning to explain rapid and transient responses to sensory input from conspecifics, a phenomenon observed in all acoustically signaling animals, including insects, fish, amphibians, and reptiles.

The nature of the genetic-developmental program underlying VPL varies within and between taxa. At one extreme are genetic-developmental programs which guide young humans and chicks in many songbird species to selectively attend to and memorize species-typical vocalizations, a process Peter Marler described as “an instinct to learn” (Marler, 1991; Kuhl, 2004; Vouloumanos and Werker, 2007). The memorized signals guide subsequent sensorimotor learning as the youngster gradually translates auditory memories to a motor program through auditory feedback from its initially poorly structured vocalizations. If human infants and songbird chicks are not exposed to conspecific auditory models during an early sensitive period, then as adults they typically produce poorly structured vocalizations that bear little resemblance to normal conspecific signals (Vouloumanos and Werker, 2007). Among songbirds an example is the White-crowned Sparrow (*Zonotrichia leucophrys*), the subject of Marler's early studies of song learning (Marler, 1997) (<https://youtu.be/7fCBTMMcyuI>). A more elaborate form of learning from external auditory models is seen in animals that mimic the signals of other species, such as Superb Lyrebirds (*Menura novaehollandiae*) (Dalziell et al., 2022) (<https://youtu.be/XUvVskyQTtE>). Grey Seals (*Halichoerus grypus*) (Stansbury and Janik, 2019) and Asian Elephants (*Elephas maximus*) (Stoeger et al., 2012) can copy human speech sounds.

At the other extreme of developmental programs are taxa where species-typical vocalizations appear to develop innately, i.e., without environmental input such as hearing conspecific adults. In these species, early exposure to and memorization of a conspecific acoustic model is not required to develop effective species-normal vocalizations, nor does early social isolation or deafening prevent it. Innate signals appear to be the typical case in most animal taxa, and most mammalian taxa in particular, including nonhuman primates (Cheney and Seyfarth, 2018; Vernes et al., 2021), and in subsocial antbird and flycatcher species (Kroodtsma, 1984; Kroodtsma, 1989; Kroodtsma and Konishi, 1991; Touchton et al., 2014).

Between these two extremes are many interesting cases, where development can be described as partly learned and partly innate. For example, in some songbird species, a bird deprived of conspecific song models will develop songs that are still somewhat species-typical, i.e., that have many but not all of the

characteristics of good species song. An example is the Song Sparrow (*Melospiza melodia*). Young sparrows raised in isolation from adult conspecific song develop songs with a combination of normal and abnormal features (Kroodtsma, 1977; Marler and Sherman, 1985; Searcy et al., 1985). When played to wild male Song Sparrows on territories or to females implanted with estradiol, isolate song evokes a stronger response than heterospecific song but a weaker response than natural conspecific song (Searcy et al., 1985). In other songbird species, a bird raised in isolation will develop perfectly good species songs, but different songs than if it had heard model songs. Examples are Grey Catbirds (*Dumetella carolinensis*) and Sedge Warblers (*Acrocephalus schoenobaenus*). When raised in social and acoustic isolation in captivity, these birds can invent large repertoires of songs with many species-typical acoustic features that vary considerably between individuals (Kroodtsma et al., 1997; Leitner et al., 2002).

In another intermediate case, young animals in some species initially develop species-typical vocalizations innately, but subsequently use auditory experience to modify these already-developed signals to converge on those of other individuals in their group or population (Vernes et al., 2021). For example, pups of several bat species develop effective isolation calls innately, but appear to use auditory experience subsequently to modify the frequency structure of their call to converge with calls of their mother and siblings (Esser and Schmidt, 1989; Esser, 1994; Knörnschild et al., 2012). Knörnschild and colleagues showed that call convergence in Greater Sac-winged Bats (*Saccopteryx bilineata*) occurs independently of relatedness among pups, and is not driven by maturational effects. In contrast to the extreme consequences of social isolation or deafening seen in humans and most songbirds, social isolation of Egyptian Fruit Bat (*Rousettus aegyptiacus*) pups delays, but does not prevent, development of normal adult calls (Prat et al., 2017), and early deafening in Pale Spear-nosed Bats (*Phyllostomus discolor*) produces only relatively small deficits in signal structure (Lattenkamp et al., 2021). Taken together, these observations indicate that the initial development of species-typical calls of young bats is not dependent on exposure to external auditory models and auditory feedback, but that these calls subsequently may converge on the calls of group members (Knörnschild, 2014).

VPL is generally regarded as a complex adaptation. As a complex trait, the odds against vocal production arising *de novo* from random mutations are high, and so it is not surprising that it is found in so few animal taxa. Complex traits are more likely to evolve when there are existing preadaptations of its critical components. Echolocation in cetaceans and bats may have been a key preadaptation for VPL in these taxa, when ecological and social factors favored the ability to develop signals that are more complex in structure and/or more flexible than can be encoded genetically in innate signals. Whereas in most mammals, vocalizations function in innate, reflexive expressions of affective states (Marler, 1980; Jürgens, 2002; Owen et al., 2011), in echolocation vocalizations are used instrumentally as environmental situations dictate. As with VPL, successful echolocation requires voluntary control over the vocal production mechanism and the ability to modulate vocalization rapidly in response to auditory feedback (Ulanovsky

and Moss, 2008). Echolocation is thought to be an ancestral trait in both cetaceans and bats (Geisler 2014; Churchill et al., 2016; Park et al., 2016; Jones and Teeling 2006; Wang et al., 2017), and likely to have evolved before (or along with) VPL in these taxa. Seals also may echolocate; they are reported to produce click-like sounds under water and it has been suggested that they use these sounds to locate prey, though the evidence is largely indirect (Renouf and Davis, 1982; Cziko et al., 2020).

### 3 Neural pathways and preadaptations for vocal production learning

There is a striking evolutionary conservation across acoustically signaling vertebrate taxa of the organization of the neural circuitry that regulates the voluntary production of vocalizations at the level of the caudal hindbrain and rostral spinal cord (Bass et al., 2008). Bass and colleagues propose that this shared vocal circuitry across species originates from a common developmental origin in rhombomere 8. Two evolutionary modifications of this ancestral circuitry have accompanied the evolution of vocal production learning (VPL). First, there is a posterior projection from the forebrain to vocal motor regions in the brainstem. Second, a novel anterior forebrain pathway has emerged, integrating auditory input and pre-motor output to regulate VPL (Jarvis, 2019). These neural innovations are observed in the brains of songbirds, which have been most extensively studied, and in the few vocal learning species of mammals in which the vocal circuitry has been investigated (Jürgens, 2009). In songbirds, auditory information is transmitted from the analogue of the primary auditory cortex to both the anterior and posterior vocal circuits, and neurons in each pathway exhibit auditory responses to conspecific song (reviewed in Brenowitz and Woolley, 2004). Another distinctive feature found only in songbirds is the presence of sex steroid receptors in neurons of the forebrain vocal control nuclei (Brenowitz, 2019). These neural adaptations appear to be lacking or rudimentary in species with innate vocal development. However, our confidence in making this statement is limited by the scarcity of comparative information on vocal neural circuitry across different groups of mammals. There is a pressing need for more extensive analysis encompassing a wider range of mammalian taxa.

Non-human primates and chickens, both of which do not learn their vocalizations, lack a direct projection from the forebrain to vocal motor neurons in the brainstem (Roberts et al., 2008; Cerkevich et al., 2022); voluntary calling in these groups is regulated by midbrain vocal pre-motor regions (Kuypers, 1958; Wild, 1997). In non-human primates, lesions of the motor cortex, as well as of the cerebellum and ventrolateral thalamus, which project to motor cortex, or of the putamen which receives cortical projections, have no effect on their vocalizations (reviewed by Jürgens, 2009). It is clear that there are distinct differences in the neural pathways that regulate the production of learned vocal behavior versus innate vocalization (Jürgens, 2002; Petkov and Jarvis, 2012). The presence of a direct projection from the

forebrain to brainstem vocal motor neurons may be a necessary adaptation for using auditory feedback to learn new vocal motor patterns. Claims of VPL in non-human primates would be bolstered by demonstrating the existence of such neural projections in their brains. Conversely, reports of VPL in the absence of this fundamental brain organization would undermine the suggestion that vocal plasticity in non-human primates served as an evolutionary precursor to human language learning (Petkov and Jarvis, 2012; Petkov and Wilson, 2012).

Jarvis and colleagues (2007; 2012) proposed that there is a posterior vocal pathway in human brains in which the face motor cortex projects to nucleus ambiguus and that this circuit is analogous to the posterior pathway in songbird brains. The primary auditory cortex projects to Wernicke's area, which is essential for language comprehension, as well as to other cortical regions. Wernicke's area has a bidirectional connection with Broca's area, which is necessary for speech production (Matsumoto et al., 2004; Jarvis, 2007); this connection provides a neural substrate for integration between vocal motor production and auditory feedback.

The neural circuits that regulate echolocation in bats and cetaceans may have served as a preadaptation for the neural control of VPL. Both echolocation and vocal learning require an integration between vocal motor production and auditory feedback, and the ability to voluntarily modify vocal signals in response to sensory feedback. The neural circuits for echolocation in bats have been studied intensively, albeit mostly from the perspective of auditory processing (Rubsamen and Schweizer, 1986; Gooler and O'Neill, 1987; Fenzl and Schuller, 2005). Neurons in the anterior cingulate cortex, motor cortex, and pyramidal motor systems project directly, as well as indirectly by intermediate nuclei, to vocal motor neurons in nucleus ambiguus in the brainstem (Huffman and Henson, 1990; Fenzl and Schuller, 2005; Metzner and Schuller, 2010; Halley et al., 2022). Input from the motor cortex serves motor coordination of learned vocal patterns, while input from the anterior cingulate cortex controls the voluntary initiation and suppression of vocalizations (Jürgens, 2002). In echolocating mammals, the existence of cortical projections to brainstem vocal motor neurons may have served as a substrate for the modification of forebrain circuitry for the regulation of VPL. The auditory system provides input to the vocal motor pathway for echolocation at multiple levels of the brain (Huffman and Henson, 1990; Ulanovsky and Moss, 2008). Exploring similarities between the neural circuits for echolocation and VPL in a wider range of echolocating mammals is a productive direction for future research.

### 4 Costs of vocal production learning

VPL, like any trait, has its costs and benefits. For a trait to evolve, its benefits must exceed its costs. While the potential benefits of vocal learning have been discussed often in the literature, the potential costs of vocal learning have generally not been considered. To redress this imbalance, we discuss potential costs here. Considering these costs may help us understand why VPL has evolved in so few animal taxa, and will inform our discussion of



ecological and environmental influences on signal evolution that follows.

The potential costs of VPL include: 1. Reliance on memorization to acquire a model for the development of species-typical vocalizations creates a risk of copy errors; a young animal might imprint on the signal of a different species in speciose habitats such as tropical rainforests, or on the vocalization of a different conspecific geographic dialect near zones of contact (Lachlan and Slater, 1999). Learning an atypical signal may prevent an individual from attracting a mate as an adult. 2. The prolonged process of learning to produce species-typical vocalizations seen in humans and some bird species increases the risk of adverse effects such as not finding appropriate tutors, hearing loss, poor nutrition (Nowicki et al., 1998), or exposure to heterospecific models. 3. The learning period delays the time until the user can benefit from producing the learned signals (Lachlan and Slater, 1999). 4. VPL requires forebrain circuitry that is lacking or rudimentary in animals with innate signal development, as well as direct projections from the forebrain to brainstem vocal motor regions, and integration between auditory and vocal motor neural circuits (Jarvis, 2019). Developing and maintaining dedicated vocal learning and production circuitry increases metabolic demand in the brain (Wennstrom et al., 2001; Von Eugen et al., 2022).

These costs of learning have resulted in the evolution of constraints on learning to limit those costs: 1. There are crude innate filters in the brains of juvenile songbirds and humans that focus attention selectively on conspecific vocalizations and serve to constrain VPL to reduce the risk of copy errors. 2. In some cases VPL is even more tightly constrained by being limited to the modification of innate calls guided by auditory feedback, as in several bat species (Knörnschild et al., 2012; Prat et al., 2015; Lattenkamp et al., 2021). 3. In many songbird species, there is only a limited sensitive period during which young birds can memorize conspecific song. While the timing varies between species and taxa, it generally coincides with the life history stage when juveniles are most likely to be exposed to adult conspecific tutors (Beecher and Brenowitz, 2005). Extending the period of learning plasticity beyond puberty in the first year (i.e., open-ended song learning) seems to occur mostly in birds species that need to learn large repertoires (Robinson et al., 2019), as well as species that need to modify their songs in response to unpredictable changes in the composition of social groups, as will be discussed in a later section. In species that only need to learn one or a few songs, extending song plasticity into adulthood may be too costly in time, energy required to maintain the song learning circuits in a fully grown state (Wennstrom et al., 2001), and increased risk of copying inappropriate songs. There is a sensitive period for language acquisition in young humans too (Kuhl, 2004; Vouloumanos and Werker, 2007; White et al., 2013). 4. In some songbird clades, comparative phylogenetic analyses suggest that the number of song types a bird sings (i.e., its repertoire size) may be smaller in derived than ancestral species (Irwin, 1988; Cardoso et al., 2007; Byers and Kroodsma, 2009). Larger repertoires involve more complex learning, and there may be selection to learn fewer songs to decrease this demand as new species are derived within a phylogenetic lineage. In summary, the existence of these various

innate mechanisms that channel or constrain learning illuminate the costs of learning to produce signals necessary for socialization and reproduction.

Birds are under extreme pressure to reduce energetic demands imposed by the brain, due to factors including small body size, flight, and maintaining a higher body temperature than mammals (Von Eugen et al., 2022). Seasonal plasticity of the forebrain song learning and production circuits in oscine birds may be an adaptation to reduce the energetic cost of developing and maintaining these regions (Wennstrom et al., 2001). The song control circuitry in adult birds dramatically regresses at the end of each breeding season and regenerates at the start of the next breeding season (Brenowitz, 2008). This is the most pronounced form of naturally occurring brain plasticity observed in any vertebrate.

Given the potential costs discussed above, we should not expect VPL to be widespread among animals, especially in its most complex form that requires early memorization of a conspecific model and a prolonged period of translating the model to a motor program guided by auditory feedback. The great majority of animal taxa that communicate acoustically do so successfully with innate signals, presumably including all calling insects, marine invertebrates, fish, amphibians, and reptiles as well as the majority of mammals and many bird species. We believe that considering these costs helps us understand why VPL has *not* evolved in most animal taxa that use acoustic signals. In the next section we adopt an ecological perspective which helps to explain why VPL *has* evolved in some taxa in spite of its costs.

## 5 The ecology of vocal communication and learning

Animals can potentially communicate in any sensory modality in which senders are able to produce a signal and for which receivers have receptors. The most common sensory channels for communication in vertebrates are olfactory, visual and auditory. Each channel has advantages and constraints for communication (Table 1). Considering these factors in relation to the types of information conveyed by signals, the behavioral and social contexts in which they are used, and the habitat in which signaling occurs, is essential to understanding the evolution of vocal learning. Previous discussions of vocal learning, however, have largely failed to consider ecological influences (for exceptions see Nottebohm, 1972; Janik and Slater, 2000; Cheney and Seyfarth, 2018).

### 5.1 Sensory channels for communication

Chemical signals offer high species specificity, persistence, and receiver “privacy” due to the need for specialized receptors. They are constrained by slow transmission and the difficulty of localizing the sender over distance (Alberts, 1992). Chemical cues are therefore generally limited to short ranges (Wilson and Bossert, 1963; Wisenden, 2008).

TABLE 1 Communication in different sensory channels.

	Chemical	Visual	Auditory
<i>Persistence</i>	Yes	Limited	No
<i>Ability to modulate signal structure or amplitude rapidly</i>	No	Yes	Yes
<i>Transmission range</i>	Short to medium	Short to medium, habitat-dependent	Long
<i>Sender locatability</i>	Difficult	Easy	Medium
<i>Private channel</i>	Yes	Limited	No, except with narrow frequency band signal and peripheral auditory filtering by receiver
<i>Time constraint on use</i>	None	Daylight only except for bioluminescent signals	None

Visual signals transmit instantaneously, allowing for rapid modulation of signal structure. They are easily localized when in the line of sight and can be graded in expression. Visual signals, however, are limited to daylight conditions (except for bioluminescent signals), can be obstructed by vegetation or turbid water, and offer little receiver privacy. Visual signals are therefore only effective over short to medium distances in forested or turbid aquatic habitats.

Acoustic signals have several advantages over the other modalities. They can be effective over long ranges, can be used in different environments and day and night, and allow rapid variation in spectral and temporal structure. The wide range of frequencies that can be produced by avian and mammalian sound production structures, along with the ability to rapidly modulate both frequency and amplitude, make acoustic signals flexible and potentially rich in information content. There are several constraints on the auditory channel. Sound intensity decreases by half for every doubling of distance from the sender due to geometrical spreading of wave fronts, limiting the effective range over which the signal can be detected above background noise (i.e., active space). Additionally, signal structure can suffer non-linear degradation over distance due to factors such as atmospheric absorption, refraction, turbulence, and reverberations off vegetation (Morton, 1975; Marten et al., 1977; Richards and Wiley, 1980; Brenowitz, 1986). In water, sound transmission is less affected by frequency attenuation and temporal degradation. Ambient noise can mask signals, making detection less reliable (Brenowitz, 1982; Wiley, 2017). Animals can mitigate noise masking by emphasizing frequencies with lower ambient noise levels and “tuning” their peripheral auditory system to signal frequencies (Wilczynski and Capranica, 1984; Ryan and Brenowitz, 1985). Localization of the sender in vertebrates requires complex neural computational processing of binaural auditory information (Konishi, 1986). Acoustic signals are susceptible to eavesdropping by competitors, predators, and parasites that are sensitive to the signal frequencies (White et al., 2022). Auditory signals may be more vulnerable to eavesdropping than olfactory or visual signals since sound detection does not rely on specialized receptors like chemical signals, and sound transmission is generally less directional than light transmission. Despite these constraints, acoustic communication has distinct

advantages in terms of long-distance transmission, usability in visually obstructed habitats at anytime of day, and suitability for rapid signaling interactions in dynamic social environments (Table 1). The wide range of frequencies producible by avian and mammalian sound production structures, along with rapid modulation of frequency and amplitude, allows acoustic signals to be flexible in structure and therefore have the potential for high-information content.

## 5.2 Habitat, spacing, and communication

Habitat and typical spacing patterns are important factors in the evolution of vocal learning. It is striking that most of the taxa in which individuals are known to learn to produce vocalizations (i.e., songbirds, parrots, hummingbirds, pinnipeds, cetaceans and bats) evolved in visually limited habitats with turbulent media, where sound is the only effective channel for communicating over the typical long distances between senders and receivers. Selective factors that favor the ability to modify the structure of signals used to communicate with conspecifics over distance in these habitats would therefore necessarily act predominantly or exclusively on acoustic signals, and favor the evolution of VPL.

The greatest species diversity of songbirds, hummingbirds, and parrots occurs in forested habitats, especially in the tropics (MacArthur and MacArthur, 1961). Songbirds and hummingbirds both evolved during the Eocene (55–40 MYA), in Australia and Eurasia, respectively (Barker et al., 2004; McGuire et al., 2014). Eocene-era Australia was wetter and supported “a luxurious forested biome” (Reichgelt et al., 2022), and much of Eurasia was covered with subtropical evergreen forest (Utescher and Mosbrugger, 2007). Parrots evolved about 74 MYA (late Cretaceous) in Australasia when it was part of the Gondwanan land mass (Wright et al., 2008). During this period Australia had a cool, wet climate and was heavily vegetated with coniferous forests (Huber et al., 2018). It is therefore likely that songbirds, hummingbirds, and parrots all first evolved in densely vegetated habitats in which visual communication was limited except at close quarters. Subsequent evolutionary radiation led to exploitation of forested habitats with high species diversity in all three taxa.

The major taxa with VPL rely on acoustic signals to communicate over distances in visually limited habitats. Songbirds produce songs for territorial advertisement and mate attraction over 100 m or more (Brenowitz, 1982; Stouffer, 2007; Catchpole and Slater, 2008). It has been estimated that the songs of male Greater Sac-winged Bat choruses have an effective range of 100 m or more (Smotherman et al., 2016; Knörnschild et al., 2017). Hummingbird neighbors may be separated by distances on the order of tens of meters (Hixon et al., 1983). Vocalizations in parrots are often used for mate recognition and group cohesion over kilometres (Berg et al., 2011; Berg et al., 2012; Rühmann et al., 2019). Marine mammals use sounds to communicate over considerable distances (Janik, 2005). The ocean is well suited to transmitting sounds over the long distances that may separate senders and receivers. Refraction of sound waves from layers of water that differ in temperature can lead to sound transmission with little attenuation within a channel over thousands of kilometres (Munk et al., 1994). Individual Humpback Whales (*Megaptera novaeangliae*) that belong to a local breeding population may be separated by tens of kilometres, and Blue whales (*Balaenoptera musculus*) by hundreds of kilometres (Sirović et al., 2007; Dunlop, 2018). Members of social cohorts of Common Bottle-nosed Dolphins (*Tursiops truncatus*) may be separated by tens of meters (Chereskin et al., 2022), and different groups by tens of kilometres. Dolphin signature whistles may be detectable up to 25 km from the sender (Janik, 2000).

Over these distances, chemical and visual signals will not reliably reach intended receivers in birds and marine mammals (e.g., King et al., 2013). Air and water are turbulent media which can complicate or eliminate directional information for chemical signals, and this limits their effective range in these environments to a few body lengths from the sender at best. Vegetation in avian habitats can obstruct a receiver's view of the sender. Cetaceans and seals occupy aquatic habitats in which light is absorbed by water; long wavelengths of light are completely absorbed by 20 m in clear water, more rapidly in turbid ocean water (Adolfson and Berghage, 1974), less than the distances typically separating senders and receivers. Echolocating bats forage at night and many species roost in dark caves during the day. The acoustic channel is therefore best suited to communicating over distance to receivers in these environments and conditions (Brenowitz, 1986).

In summary, in habitats where chemical or visual signals cannot be reliably transmitted to distant receivers, vocal signaling will be favored. But some additional factors are required to explain the advantage of *learning* these vocal signals. After all, many taxa that occupy the same habitats as vocal learners are able to communicate successfully using innately developed acoustic signals. We discuss factors favoring the evolution of VPL in the next section.

### 5.3 Information content of signals

Animals communicate information about multiple states and events (Table 2), including about identity (individual, group, species), the external world (predators, food sources), their motivational state, and their individual quality in assessment (mating and agonistic) contexts. For animals such as songbirds

TABLE 2 Types of signals in intraspecific animal communication systems.

1. Information about the external world	a. Food signals
	b. Predator (warning/alarm)
2. Information about identity	a. Species
	b. Individual
	c. Group
	d. Kinship
	e. Status
3. Signals that indicate motivational state and/or synchronize social interactions	a. Agonistic contexts
	b. Courtship & mating contexts
	c. Parent-offspring communication
	d. Intra-group (cohesion) signals
	e. Inter-group (spacing) signal
4. Assessment [overlaps with all else]	a. Mate attraction context
	b. Agonistic context
	c. Predation context

and cetaceans that can only use the sound channel to communicate over long distances in visually limited habitats, acoustic signals may have to include multiple types of information. For example, the song of a male White-crowned Sparrow conveys information on individual, dialect (local group), and species identity as well as being a territorial keep-out signal and potentially a mate advertisement signal (see Figure 1 and <https://youtu.be/7fCBTMMcyuI>) (Nelson and Poessel, 2007). In the many species where both males and females sing, song can also convey the sender's sex.

When the acoustic channel is the only one available to communicate critical information over the distances that typically separate sender and receiver, VPL may be the only practical means of developing acoustic signals complex enough to encode multiple types of information (Marler, 1960). This hypothesis can be evaluated by comparing the songs of birds in the two Passeriform suborders, the oscines and suboscines. As discussed above, songs in most (likely all) oscine species are learned, while song is thought to develop innately in most suboscine antbird and flycatcher species (Kroodsma, 1984; Seddon, 2005; Touchton et al., 2014). Individual song variation, individual recognition *via* song, precise copying of auditory models, song repertoires, and geographic variation in songs are the rule in oscines, but are absent or rare in suboscine antbird and flycatcher species (Lindell, 1998; Bard et al., 2002; Seddon, 2005; Kroodsma, 2011). In general suboscine species produce songs with relatively simple structure, consisting of a small number of whistled notes, with little evidence of the pronounced frequency modulations and “two-voiced” syllables produced independently by the two sides of the syrinx that are characteristic of oscine songs (Suthers and Zollinger, 2004; Goller, 2022) (see Figure 2). Suboscine birds do not have repertoires of songs, unlike most oscines (Beecher and Brenowitz, 2005; Touchton et al., 2014). Mimicry of heterospecific vocalizations does not occur in suboscines but is widespread among oscine birds (Touchton

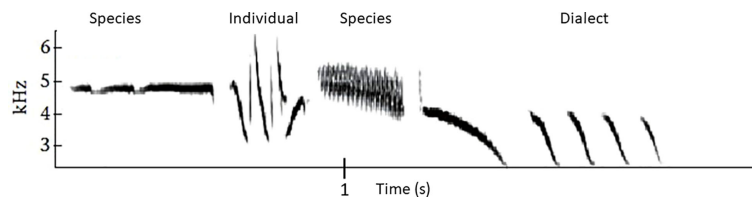


FIGURE 1 Sound spectrogram of White-crowned Sparrow song showing the segregation of different types of information to different song components, based on Soha and Marler (2000) and Nelson and Poesel (2007). (Spectrogram modified with permission from Nelson and Poesel, 2007).

et al., 2014; Dalziell et al., 2015). Sufficient individual variation in song structure to allow reliable discrimination between neighbors and strangers is uncommon among suboscines (Stoddard et al., 1991; Bard et al., 2002; Kroodsmma, 2011). Suboscine flycatcher and antbird songs generally show little geographic variation, even over thousands of kilometres in species with wide geographic distributions (Johnson, 1980; Lindell, 1998; Sedgwick, 2001; Kroodsmma, 2011), unlike the songs of oscines which can show distinct geographic dialects even over short distances (Marler and Tamura, 1962). This lack of variation precludes most suboscines from using song to convey membership in a local group.

Assessment is one of the key features of an animal communication system (Table 2). Well-studied examples are signals that convey an animal’s health and vigour and persuade an opponent to withdraw from a confrontation or a potential partner to mate, or dissuade a predator from pursuit. The more detectable quality levels that a sender can encode in a display signal, the more informative the signal is to receivers. In cases like these, there is strong selection for increasing the levels of variation in the signal. In many songbirds, the variation contained in a bird’s song repertoire goes far beyond what is needed to indicate species and individual identity or membership in a particular group, and is used

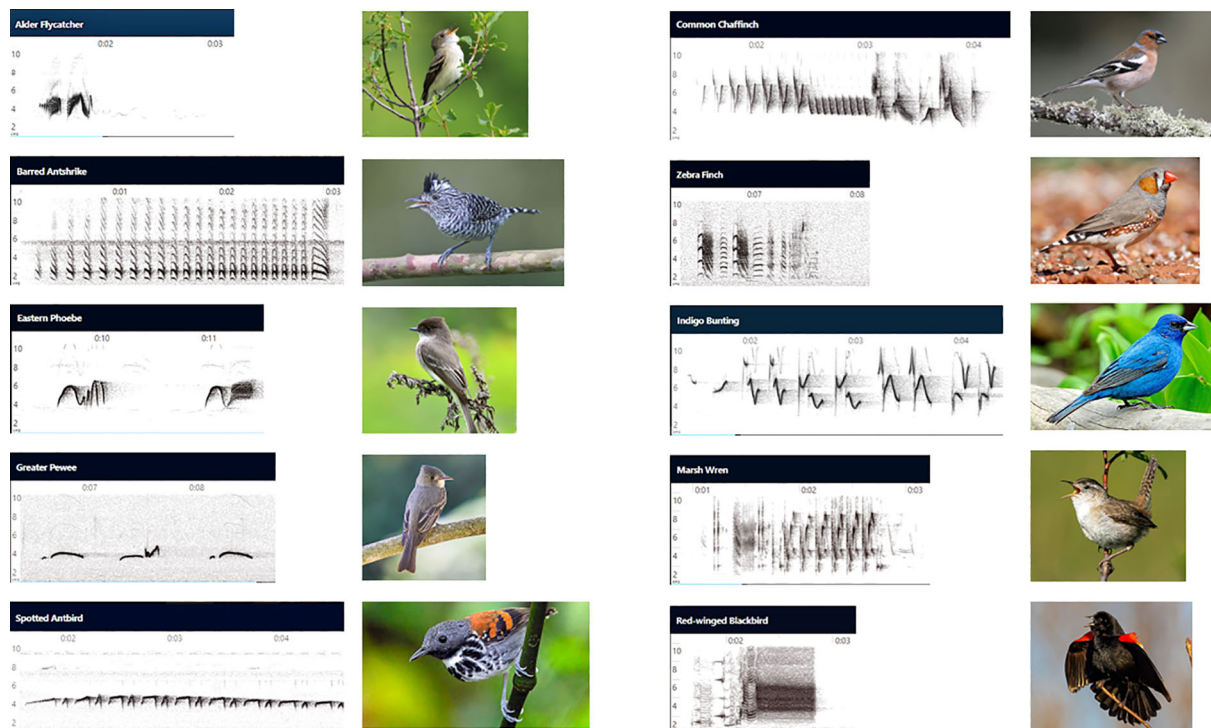


FIGURE 2 Innately developed songs are less complex than learned songs. Sound spectrograms for innate songs of five suboscine species and photographs of each species are shown in the two left columns, and spectrograms of learned songs of five oscine species and species photographs are shown in the right two columns. Signal frequency in kHz is shown at the left, and elapsed time in sec is shown at the top of each spectrogram. Each of the five suboscine species has been shown to develop normal conspecific song when young birds were raised in acoustic isolation, deafened, or tutored only with heterospecific song (Kroodsmma, 1984; Kroodsmma, 1989; Kroodsmma and Konishi, 1991; Touchton et al., 2014). Each of the five oscine species was shown to fail to develop normal conspecific song when young were raised in acoustic isolation (Thorpe, 1958; Rice and Thompson, 1968; Immelmann, 1969; Marler et al., 1972; Kroodsmma and Canady, 1985; Brenowitz et al., 1994). Note that suboscine songs consist of at most two syllable types, often repeated. Oscine songs consist of several different syllable types arranged with complex syntactical structure. Spectrograms and photographs were obtained with permission at <https://ebird.org/home>. Readers may listen to these songs at the website. (Spectrogram images were lightly edited in Adobe Photoshop to enhance the visual contrast between song traces and background noise.)



instead to impress potential mates or opponents. Several non-exclusive hypotheses have been offered to explain how this variation might benefit the sender, including facilitating female choice and male–male competition (also see [Searcy and Nowicki, 2019](#)). For example, female Satin Bowerbirds (*Ptilonorhynchus violaceus*) choose mates in part based on the number and accuracy of their mimicry of the songs and calls of other species ([Coleman et al., 2007](#)). Neighboring territorial male Song Sparrows use shared and unshared song types in vocal exchanges to mediate graded agonistic interactions in a complex manner ([Beecher et al., 2000](#)). Although the status of several competing functional hypotheses has not been fully resolved, they all agree on one point: what is assessed is the sender’s ability to learn to produce complex vocal signals.

Adequate variation to encompass these different types of messages in vocalizations is possible only when learning through auditory experience plays a large role in the normal development of signals. It is thus not surprising that complex vocal signaling is widespread in the oscine passerines but absent (or rare) in the innate songs of most suboscine flycatchers and antbirds. Similar considerations apply to learned and innate signals in other taxa.

## 5.4 Sociality and communication

In animals that rely on acoustic signals to communicate membership in a local breeding population or social group, and/or individual identity within a group, the fluidity of group membership, of coalitions within groups, and the spatial distribution of extended groups, can be important factors favoring the ability to modify signal structure and thus select for VPL ([Nottebohm, 1972](#); [Poole et al., 2005](#)). Mobile animals (e.g., birds, bats, seals, cetaceans) that are not restricted to their natal area by kin relationships or communal breeding may disperse over distance and join groups of unfamiliar conspecifics that use novel signals to convey membership. Other animals, like African Savannah Elephants (*Loxodonta africana africana*), live in social groups that change in size and composition over time depending on seasonal patterns of food availability and other environmental and social factors ([Hedwig et al., 2021](#)). The composition of social groups in these taxa may be fluid, with changes in membership due to immigration, emigration, and fission/fusion patterns ([Aureli et al., 2008](#)). If individuals benefit from producing signals that communicate membership in a new group or coalition, innate signals would not be flexible enough to modify these signals over relatively short time periods to allow convergence with a shared group vocalization, as in Savannah Elephants ([Poole et al., 2005](#)), Bottle-nosed dolphins ([Watwood et al., 2004](#)), and Greater Spear-nosed Bats (*Phyllostomus hastatus*) ([Boughman, 1998](#)), or to diverge from the calls of other individuals in the local area to facilitate individual recognition as in Green-rumped Parrotlets (*Forpus passerinus*) ([Berg et al., 2012](#)). VPL may be selected for in these circumstances, especially when members of groups are typically spaced over distances where olfactory and visual cues are not reliable.

There are numerous examples of learned signal convergence. Black-capped Chickadees (*Parus atricapillus*) in experimentally

combined flocks converge on shared call notes, with a significant decrease in variance among birds in the frequency of a key note within seven days ([Mammen and Nowicki, 1981](#); [Nowicki, 1989](#)). Male Village Indigobirds (*Vidua chalybeata*) in a communal display area (i.e., a lek) converge on the songs of the male with the highest breeding success ([Payne, 1985](#)). The song types shared by local males change from year to year, and some song types regularly go extinct due to turnover in group membership and changes in social status. Innate signals would not allow modification of song structure to match local social conditions. Greater Sac-winged Bat pups learn adult territorial song by imitating their harem male ([Knörnschild et al., 2010](#)). In Humpback Whales in the Southern Hemisphere there is population-wide conformity to one song pattern and every few years each member of the population completely replaces their song in so-called cultural “revolutions”; it has been suggested that these changes are due to social or cultural learning ([Allen et al., 2018](#), but see [Mercado, 2022](#)). Developing and changing these complex songs would not be possible with innate signals. Similar considerations apply to animals like Green-rumped Parrotlets ([Berg et al., 2012](#)) that learn to produce signals that diverge from those of others in the local area.

Vocalizations in nonhuman primates that live in social groups evolved to serve different behavioral functions than those of avian and mammalian vocal production learners. Calls in social primates are used to “facilitate social interactions by reducing uncertainty about the signaller’s intentions and likely behavior” ([Cheney and Seyfarth, 2018](#)). Species identity seems to be conveyed primarily by visual cues in diurnally active primates ([Pokorny and Waal, 2009](#); [Dahl et al., 2010](#); [Hirata et al., 2010](#)), and individual and kin identity at close range can be determined from the combination of visual, acoustic, and chemical cues ([Sliwa et al., 2011](#); [Henkel and Setchell, 2018](#)). A small repertoire of innate call types that can be graded in rate, duration and amplitude, and combined with other call types, provides sufficient flexibility and information content for individuals to assess each other’s likely behavior during social interactions, especially when combined with chemical and graded visual displays. The identity of the caller, the context in which it calls, the history of interactions with that individual, and its decision to call all convey more useful information to the receiver than the structure of the call itself. Under these conditions, the potential costs of VPL may outweigh the benefits. Selective pressures might thus center on increasing the cognitive capacity to keep track of the complex history of interactions with the different members of a social group, and to associate the current context of calling with this history. Nonhuman primates have therefore been selected to learn to use innate vocal signals in novel social or environmental contexts (i.e., vocal usage learning) ([Fischer, 2017](#); [Cheney and Seyfarth, 2018](#); [Vernes et al., 2021](#)).

## 5.5 Summary

The discussion above can be summarized as follows.

1. In visually limited habitats, the auditory channel is the only effective channel for communication over long distances.

2. Innate acoustic signals are limited in variety and structural complexity which limits the number of different types of messages and the amount of information that can be conveyed in a signal.
3. Species that can use only the auditory channel to communicate over distances that typically separate sender and receiver, and that gain from the ability to modify signals to converge with or diverge from those of other group members (dolphins, seals, bats, elephants), or benefit from increasing signal complexity (songbirds, whales), will be under selection to evolve VPL. Innate signals do not allow a sender the flexibility to modify signals as local social factors dictate, nor do they permit expanding the variation in the signal set in contexts where it is adaptive for senders to signal their quality to receivers.

## 6 Functional explanations of vocal production learning

Much attention has been devoted to understanding the adaptive value of VPL (reviewed by [Nottebohm, 1972](#); [Lachlan and Slater, 1999](#); [Slater and Janik, 2010](#)). Various functional hypotheses have been proposed, but few have been rigorously tested. [Nowicki and Searcy \(2014\)](#) evaluated five major hypotheses and concluded that comparative analyses most strongly support the proposal that VPL evolved to increase the number and complexity of vocalizations in response to either sexual selection driven by female mating preferences or kin selection to share information with relatives. While these hypotheses might explain VPL in some taxa of songbirds (but see [Soma and Garamszegi, 2011](#)), whales, and singing bats ([Knörnschild et al., 2010](#)) that learn complex signals, they do not account for other taxa that learn relatively simple calls used for group and/or individual recognition. These hypotheses also likely do not explain the evolution of language in hominids ([Cheney and Seyfarth, 2018](#)).

It is not surprising that no one or even two hypotheses can explain the evolution of VPL in all taxa. Phylogenetic analyses indicate that it is most parsimonious to conclude that VPL evolved independently in the different mammalian clades, and also likely independently in hummingbirds, and parrots and songbirds. Songbirds and parrots are sister clades and it is not yet clear whether their common ancestor learned to vocalize while the suboscine antbirds and flycatchers lost this ability, as opposed to independent evolution of learning in songbirds and parrots ([Nowicki and Searcy, 2014](#)). Each independent evolution of VPL by a species was the response to a suite of selective factors unique to that species, potentially including its mating and social systems, the presence or absence of extra-pair mating, female mate choice behavior, habitat, spacing behavior, life history patterns, phylogenetic history, and the presence or absence of preadaptations. Once VPL has evolved in a species, there can be modification of one or more aspects of learning within a species as populations disperse to new locations, and with rapid speciation following radiation to new habitats as occurred in Passeriform birds

([Barker et al., 2004](#)). Within the monophyletic songbird lineage, there is extensive diversity in how many songs are learned, when they are learned, from whom they are learned, and their behavioral function ([Beecher and Brenowitz, 2005](#); [Brenowitz and Beecher, 2005](#)). This diversification can make it difficult to identify a single adaptive explanation for the original evolution of VPL in the root ancestral species of a clade, especially if that species no longer exists.

While there may not be any single unitary function that can explain the evolution of all VPL, it is possible to propose general advantages that will favor its evolution. Vocal production learning allows animals to develop signals that are more complex in structure and/or more flexible than can be encoded genetically in innate signals. Increased signal complexity may be adaptive when ecological constraints limit animals to the auditory channel for encoding different types of information in signals used over distances that separate senders and receivers, as in songbirds, parrots, and whales. Complex signals are also beneficial in behavioral contexts where receivers use signals to assess the quality of senders as in Satin Bowerbirds and Superb Lyrebirds. The ability to modify signals with reference to environmental information can be adaptive when they are used to convey membership in a social system with unpredictable membership, as in Savannah Elephants. Plasticity can also be advantageous when individuals benefit from altering signals to match those of adjacent conspecifics as in bats and dolphins, or to diverge from the signals of neighbors as in Green-rumped Parrotlets. VPL may evolve independently in all of these contexts, complicating any attempt at a unitary evolutionary synthesis.

## 7 Discussion

Animals that communicate with acoustic signals begin life with some genetic specification of the direction that development of the acoustic signaling system will take. In most species, development of species-typical vocalizations occurs without the need for exposure to such vocalizations or other appropriate experience; we have referred to the signal and its development in these cases as “innate”. In other species, signal development proceeds similarly, but subsequently the signal can be modified by experience with appropriate auditory models. In still other species, the genetic-developmental program serves to guide the choice of an appropriate auditory model, or to channel auditory learning in some way towards species-appropriate signals. The balance between innate and learned contributions to vocal development varies within and between taxa, and reflects the relative costs and benefits of learning for a given species, as well as the ecological and social constraints that it faces when signaling.

In taxa that learn to produce complex signals, like songbirds, juveniles typically memorize external models of conspecific sounds and use auditory feedback to guide the sensorimotor process of matching their vocal motor production to the external models. In taxa that learn to produce relatively simple contact calls, such as in young bats, individuals may modify their vocal production to resemble (and in some cases, to diverge from) the calls of other individuals that they hear on an ongoing basis. (It is not yet known whether any mammals that can modify calls as adults need to

memorize external species-typical auditory models as juveniles in order to produce those calls.) Once evolved within a clade, the form and function of learned signals can be modified in response to the specific ecological factors encountered by different species.

VPL seems most likely to evolve in taxa that meet one or more of the following conditions. 1. The species occupies a habitat where individuals must rely predominantly or exclusively on the acoustic channel to communicate over distances that typically separate senders and receivers. 2. Individuals gain from using auditory input to modify the structure of their signals to converge with or diverge from those of conspecifics in the same area, as in the contexts of signaling group membership (convergence) or individual identity (divergence). 3. Individuals live in social groups where membership is fluid over time, requiring the ongoing ability to modify vocal identity signals, as seems to occur in Savannah Elephants (Poole et al., 2005; Stoeger and Manger, 2014), Village Indigobirds (Payne, 1985), and southern Humpback Whales (Allen et al., 2018). Individuals in these species are relatively long-lived (i.e., 10+ years), which may increase the selective pressure to be able to update acoustic identity signals as group membership changes over their lifespans. 4. Individuals gain from developing signals that allow receivers to accurately assess the sender's health and vigour to attract a mate and/or deter a rival (Table 2). VPL will evolve when these benefits outweigh the potential costs of this complex sensorimotor process. We expect VPL to evolve only rarely, when one or more of the selective advantages is present.

Language development in humans is a *sui generis* example of VPL. It involves a sensorimotor process analogous to other forms of VPL, but also includes a considerable cognitive component. Children learn grammatical and syntactical correctness rules by listening to mature speakers, and generalize them to novel phrases. Their vocabulary expands rapidly through fast-mapping (Nicolaidis, 2006). Gestural languages such as American Sign Language include a large cognitive component (Supalla et al., 2014). In general, human language learning is so much more complex than even the most impressive animal examples of VPL, including even the large song repertoires of some songbirds, that trying to lump them together as instances of template-based "vocal production learning" is of questionable value.

The burden of evidence does not support the suggestion that human language arose from simpler VPL in nonhuman primates. An alternative hypothesis is that limitations on the maximum potential information content of visual gestural signals favored the evolution of vocal learning (Prieur et al., 2020). It has been widely speculated that early hominids relied primarily on these visual displays for social communication. As social complexity rapidly increased with hominid cultural evolution, the selection for increased complexity of communication may have favored the elaboration of acoustic signaling, given the greater flexibility of vocal production mechanisms compared with gestural signals (Cheney and Seyfarth, 2018). Genetic programming of acoustic signals may not have been able to keep pace with cultural evolution of social interactions, coordinated foraging, and tool use, and this might have favored the evolution of VPL in humans. Learned spoken language complements visual gestures to produce a multimodal communication system of considerable complexity and high potential information content in modern humans.

There are many unanswered questions and grey areas in the study of the evolution of vocal learning (Vernes et al., 2021). Carefully controlled developmental studies in a wider range of animals are urgently needed. Training studies should be used where feasible and ethical. We believe that such studies may well demonstrate VPL in species in taxa outside those in which it has been found thus far, where ecological and social factors favoring its evolution occur. If a given newly-demonstrated VPL species is an exception for its larger taxon, then it provides an excellent test of our hypotheses about ecological selection pressures or existing preadaptations favoring VPL. In searching for new examples of VPL, anecdotal just-so stories should be avoided. Proposed neural and genetic mechanisms of vocal learning should be examined and experimentally tested more widely. We look forward to seeing continued research on the evolution of vocal learning.

## Author contributions

EB and MB conceived the idea and wrote the paper. Both authors contributed to the article and approved the submitted version.

## Funding

National Institutes of Health grant R01 NS103973 to corresponding author Brenowitz provided partial support during preparation of this paper.

## Acknowledgments

We thank Dr. William Searcy and two reviewers for helpful comments. We are grateful to Vanessa Claire Powell and Matthew Medler of the Macaulay Library of the Cornell Lab of Ornithology for their assistance in providing access to media used in this paper. We thank D. St-Jacques, S. Caballero Carrera, S. Martin, A. Zahm, G. Heaton, M. Kraml, P. Hawrylyshyn, D. Danko, J. Livaudais, B. Imhoff, and C. Wiley for allowing us to use their splendid bird photos.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Adolfson, J., and Berghage, T. (1974). *Perception and performance under water* (Oxford, England: John Wiley & Sons).
- Alberts, A. C. (1992). Constraints on the design of chemical communication systems in terrestrial vertebrates. *Am. Nat.* 139, S62–S89. doi: 10.1086/285305
- Allen, J. A., Garland, E. C., Dunlop, R. A., and Noad, M. J. (2018). Cultural revolutions reduce complexity in the songs of humpback whales. *Proc. R. Soc. B: Biol. Sci.* 285, 20182088. doi: 10.1098/rspb.2018.2088
- Aureli, F., schaffner, C. m., Boesch, C., Bearder, S. k., Call, J., Chapman, C. a., et al. (2008). Fission-fusion dynamics: new research frameworks. *Curr. Anthropology* 49, 627–654. doi: 10.1086/586708
- Banerjee, A., and Vallentin, D. (2022). Convergent behavioral strategies and neural computations during vocal turn-taking across diverse species. *Curr. Opin. Neurobiol.* 73, 102529. doi: 10.1016/j.conb.2022.102529
- Baptista, L. F., and Petrinovich, L. (1984). Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Anim. Behav.* 32, 172–181. doi: 10.1016/S0003-3472(84)80335-8
- Bard, S. C., Hau, M., Wikelski, M., and Wingfield, J. C. (2002). Vocal distinctiveness and response to conspecific playback in the spotted antbird, a Neotropical suboscine. *Condor* 104, 387–394. doi: 10.1093/condor/104.2.387
- Barker, F. K., Cibois, A., Schikler, P., Feinstein, J., and Cracraft, J. (2004). Phylogeny and diversification of the largest avian radiation. *Proc. Natl. Acad. Sci.* 101, 11040–11045. doi: 10.1073/pnas.0401892101
- Bass, A. H., Gilland, E. H., and Baker, R. (2008). Evolutionary origins for social vocalization in a vertebrate hindbrain–spinal compartment. *Science* 321, 417–421. doi: 10.1126/science.1157632
- Beecher, M. D. (2021). Why are no animal communication systems simple languages? *Front. Psychol.* 12. doi: 10.3389/fpsyg.2021.602635
- Beecher, M. D., and Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends Ecol. Evol.* 20, 143–149. doi: 10.1016/j.tree.2005.01.004
- Beecher, M. D., Campbell, S. E., Burt, J. M., Hill, C. E., and Nordby, J. C. (2000). Song-type matching between neighbouring song sparrows. *Anim. Behav.* 59, 21–27. doi: 10.1006/anbe.1999.1276
- Berg, K. S., Delgado, S., Cortopassi, K. A., Beissinger, S. R., and Bradbury, J. W. (2012). Vertical transmission of learned signatures in a wild parrot. *Proc. R. Soc. B: Biol. Sci.* 279, 585–591. doi: 10.1098/rspb.2011.0932
- Berg, K. S., Delgado, S., Okawa, R., Beissinger, S. R., and Bradbury, J. W. (2011). Contact calls are used for individual mate recognition in free-ranging green-rumped parrotlets, *forpus passerinus*. *Anim. Behav.* 81, 241–248. doi: 10.1016/j.anbehav.2010.10.012
- Boughman, J. W. (1998). Vocal learning by greater spear-nosed bats. *Proceedings: Biol. Sci.* 265, 227–233. doi: 10.1098/rspb.1998.0286
- Brenowitz, E. A. (1982). The active space of red-winged blackbird song. *J. Comp. Physiol.* 147, 511–522. doi: 10.1007/BF00612017
- Brenowitz, E. A. (1986). Environmental influences on acoustic and electric animal communication. *Brain Behav. Evol.* 28, 32–42. doi: 10.1159/000118690
- Brenowitz, E. A. (2008). “Plasticity of the song control system in adult birds,” in *Neuroscience of birdsong*. Eds. H. P. Zeigler and P. Marler. (Cambridge: Cambridge University Press), 332–349.
- Brenowitz, E. A. (2019). “Hormones and animal communication,” in *Oxford Encyclopedia of neuroendocrine and autonomic systems*. Ed. R. J. Nelson. (New York and Oxford: Oxford University Press).
- Brenowitz, E. A., and Beecher, M. D. (2005). Song learning in birds: diversity and plasticity, opportunities and challenges. *Trends Neurosci.* 28, 127–132. doi: 10.1016/j.jtins.2005.01.004
- Brenowitz, E. A., Nalls, B., Kroodmsa, D. E., and Horning, C. (1994). Female marsh wrens do not provide evidence of anatomical specializations of song nuclei for perception of male song. *J. Neurobiology* 25, 197–208. doi: 10.1002/neu.480250210
- Brenowitz, E. A., and Woolley, S. M. (2004). “The avian song control system: a model for understanding changes in neural structure and function,” in *Plasticity in the auditory system*. Eds. T. Parks, E. W. Rubel, A. Popper and R. Fay. (New York: Springer-Verlag), 228–284.
- Bruno, J. H., Jarvis, E. D., Liberman, M., and Tchernichovski, O. (2021). Birdsong learning and culture: analogies with human spoken language. *Annu. Rev. Linguistics* 7, 449–472. doi: 10.1146/annurev-linguistics-090420-121034
- Byers, B. E., and Kroodmsa, D. E. (2009). Female mate choice and songbird song repertoires. *Anim. Behav.* 77, 13–22. doi: 10.1016/j.anbehav.2008.10.003
- Cardoso, G. C., Atwell, J. W., Ketterson, E. D., and Price, T. D. (2007). Inferring performance in the songs of dark-eyed juncos (*Junco hyemalis*). *Behav. Ecol.* 18, 1051–1057. doi: 10.1093/beheco/arm078
- Catchpole, C. K., and Slater, P. J. B. (2008). *Bird song: biological themes and variations* (Cambridge, UK: Cambridge University Press).
- Cerkevich, C. M., Rathelot, J.-A., and Strick, P. L. (2022). Cortical basis for skilled vocalization. *Proc. Natl. Acad. Sci.* 119, e2122345119. doi: 10.1073/pnas.2122345119
- Chen, Y., Matheson, L. E., and Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proc. Natl. Acad. Sci.* 113, 6641–6646. doi: 10.1073/pnas.1522306113
- Cheney, D. L., and Seyfarth, R. M. (2018). Flexible usage and social function in primate vocalizations. *Proc. Natl. Acad. Sci.* 115, 1974–1979. doi: 10.1073/pnas.1717572115
- Chereskin, E., Connor, R. C., Friedman, W. R., Jensen, F. H., Allen, S. J., Sorensen, P. M., et al. (2022). Allied male dolphins use vocal exchanges to “bond at a distance”. *Curr. Biol.* 32, 1657–1663.e1654. doi: 10.1016/j.cub.2022.02.019
- Churchill, M., Martinez-Caceres, M., De muizon, C., Mnieckowski, J., and Geisler, J. (2016). The origin of high-frequency hearing in whales. *Curr. Biol.* 26, 2144–2149. doi: 10.1016/j.cub.2016.06.004
- Choe, H. N., and Jarvis, E. D. (2021). The role of sex chromosomes and sex hormones in vocal learning systems. *Hormones Behav.* 132, 104978. doi: 10.1016/j.yhbeh.2021.104978
- Coleman, S. W., Patricelli, G. L., Coyle, B., Siani, J., and Borgia, G. (2007). Female preferences drive the evolution of mimetic accuracy in male sexual displays. *Biol. Lett.* 3, 463–466. doi: 10.1098/rsbl.2007.0234
- Condro, M. C., and White, S. A. (2014). Recent advances in the genetics of vocal learning. *Comp. Cogn. Behav. Rev.* 9, 75–98. doi: 10.3819/ccbr.2014.90003
- Corballis, M. C. (2020). Crossing the Rubicon: behaviorism, language, and evolutionary continuity. *Front. Psychol.* 11, 653. doi: 10.3389/fpsyg.2020.00653
- Corcoran, A. J., and Conner, W. E. (2014). Bats jamming bats: food competition through sonar interference. *Science* 346, 745–747. doi: 10.1126/science.1259512
- Cziko, P. A., Munger, L. M., Santos, N. R., and Terhune, J. M. (2020). Weddell seals produce ultrasonic vocalizations. *J. Acoustical Soc. America* 148, 3784–3796. doi: 10.1121/10.0002867
- Dahl, C. D., Logothetis, N. K., Bühlhoff, H. H., and Wallraven, C. (2010). The Thatcher illusion in humans and monkeys. *Proc. R. Soc. B: Biol. Sci.* 277, 2973–2981. doi: 10.1098/rspb.2010.0438
- Dalziel, A. H., Welbergen, J. A., Iqic, B., and Magrath, R. D. (2015). Avian vocal mimicry: a unified conceptual framework. *Biol. Rev.* 90, 643–668. doi: 10.1111/brv.12129
- Dalziel, A. H., Welbergen, J. A., and Magrath, R. D. (2022). Male Superb lyrebirds mimic functionally distinct heterospecific vocalizations during different modes of sexual display. *Anim. Behav.* 188, 181–196. doi: 10.1016/j.anbehav.2022.04.002
- Dunlop, R. A. (2018). The communication space of humpback whale social sounds in wind-dominated noise. *J. Acoustical Soc. America* 144, 540–551. doi: 10.1121/1.5047744
- Endepols, H., and Walkowiak, W. (1999). Influence of descending forebrain structures on processing of acoustic signals and audiomotor integration in the anuran midbrain. *Eur. J. Morphology* 37, 182–184. doi: 10.1076/ejom.37.2.182.4753
- Esser, K. H. (1994). Audio-vocal learning in a non-human mammal: the lesser spear-nosed bat *Phyllostomus discolor*. *Neuroreport* 5, 1718–1720. doi: 10.1097/00001756-199409080-00007
- Esser, K.-H., and Schmidt, U. (1989). Mother-infant communication in the lesser spear-nosed bat *Phyllostomus discolor* (Chiroptera, Phyllostomidae) [and] evidence for acoustic learning. *Ethology* 82, 156–168. doi: 10.1111/j.1439-0310.1989.tb00496.x
- Fenzl, T., and Schuller, G. (2005). Echolocation calls and communication calls are controlled differentially in the brainstem of the bat *Phyllostomus discolor*. *BMC Biol.* 3, 17. doi: 10.1186/1741-7007-3-17
- Fischer, J. (2017). Primate vocal production and the riddle of language evolution. *Psychonomic Bull. Rev.* 24, 72–78. doi: 10.3758/s13423-016-1076-8
- Fischer, J., Wheeler, B. C., and Higham, J. P. (2015). Is there any evidence for vocal learning in chimpanzee food calls? *Curr. Biol.* 25, R1028–R1029. doi: 10.1016/j.cub.2015.09.010
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends Cogn. Sci.* 4, 258–267. doi: 10.1016/S1364-6613(00)01494-7
- Gahr, M. (2000). Neural song control system of hummingbirds: comparison to swifts, vocal learning (Songbirds) and nonlearning (Suboscines) passerines, and vocal learning (Budgerigars) and nonlearning (Dove, owl, gull, quail, chicken) nonpasserines. *J. Comp. Neurol.* 426, 182–196. doi: 10.1002/1096-9861(20001016)426:2<182::AID-CNE2>3.0.CO;2-M
- Geisler, J. H., Colbert, M. W., and Carew, J. L. (2014). A new fossil species supports an early origin for toothed whale echolocation. *Nature* 508, 383–386. doi: 10.1038/nature13086
- Goller, F. (2022). Vocal athletics—from birdsong production mechanisms to sexy songs. *Anim. Behav.* 184, 173–184. doi: 10.1016/j.anbehav.2021.04.009
- Gooler, D. M., and O'Neill, W. E. (1987). Topographic representation of vocal frequency demonstrated by microstimulation of anterior cingulate cortex in the echolocating bat, *Pteronotus parnellii parnellii*. *J. Comp. Physiol. A* 161, 283–294. doi: 10.1007/BF00615248
- Halley, A. C., Baldwin, M. K. L., Cooke, D. F., Englund, M., Pineda, C. R., Schmid, T., et al. (2022). Coevolution of motor cortex and behavioral specializations associated



- with flight and echolocation in bats. *Curr. Biol.* 32, 2935–2941.e2933. doi: 10.1016/j.cub.2022.04.094
- Hammerschmidt, K., and Fischer, J. (2008). “Constraints in primate vocal production,” in *The evolution of communicative flexibility: complexity, creativity, and adaptability in human and animal communication*. Eds. D. K. Oller and U. Griebel. (Cambridge, MA: The MIT Press), 93–119.
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* 298, 1569–1579. doi: 10.1126/science.298.5598.1569
- Hedwig, D., Poole, J., and Granli, P. (2021). Does social complexity drive vocal complexity? insights from the two African elephant species. *Animals* 11 3071. doi: 10.3390/ani11113071
- Henkel, S., and Setchell, J. M. (2018). Group and kin recognition via olfactory cues in chimpanzees (Pan troglodytes). *Proc. R. Soc. B: Biol. Sci.* 285, 20181527.
- Hirata, S., Fuwa, K., Sugama, K., Kusunoki, K., and Fujita, S. (2010). Facial perception of conspecifics: chimpanzees (Pan troglodytes) preferentially attend to proper orientation and open eyes. *Anim. Cogn.* 13, 679–688. doi: 10.1007/s10071-010-0316-y
- Hixon, M. A., Carpenter, F. L., and Paton, D. C. (1983). Territory area, flower density, and time budgeting in hummingbirds: an experimental and theoretical analysis. *Am. Nat.* 122, 366–391. doi: 10.1086/284141
- Huber, B. T., Hobbs, R. W., and Bogus, K. A. (2018). Expedition 369 preliminary report: Australia Cretaceous climate and tectonics. *Natl. Sci. Foundation*.
- Huffman, R. F., and Henson, O. W. Jr. (1990). The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus. *Brain Res. Brain Res. Rev.* 15, 295–323. doi: 10.1016/0165-0173(90)90005-9
- Immelmann, K. (1969). “Song development in the zebra finch and other estrildid finches,” in *Bird vocalizations*. Ed. R. A. Hinde. (Cambridge: Cambridge UP), 61–77.
- Irwin, R. E. (1988). The evolutionary importance of behavioural development: the ontogeny and phylogeny of bird song. *Anim. Behav.* 36, 814–824. doi: 10.1016/S0003-3472(88)80164-7
- Janik, V. M. (2000). Source levels and the estimated active space of bottlenose dolphin (Tursiops truncatus) whistles in the Moray firth, Scotland. *J. Comp. Physiol. A* 186, 673–680. doi: 10.1007/s003590000120
- Janik, V. M. (2005). “Underwater acoustic communication networks in marine mammals,” in *Animal communication networks*. Ed. P. K. McGregor. (Cambridge: Cambridge University Press), 390–415.
- Janik, V. M., and Knörnschild, M. (2021). Vocal production learning in mammals revisited. *Philos. Trans. R. Soc. B: Biol. Sci.* 376, 20200244. doi: 10.1098/rstb.2020.0244
- Janik, V. M., and Slater, P. J. B. (1997). “Vocal learning in mammals,” in *Advances in the study of behavior*. Eds. P. J. B. Slater, J. S. Rosenblatt, C. T. Snowdon and M. Milinski. (Cambridge, MA: Academic Press), 59–99.
- Janik, V. M., and Slater, P. J. B. (2000). The different roles of social learning in vocal communication. *Anim. Behav.* 60, 1–11. doi: 10.1006/anbe.2000.1410
- Jarvis, E. D. (2007). Neural systems for vocal learning in birds and humans: a synopsis. *J. Ornithology* 148, 35–44. doi: 10.1007/s10336-007-0243-0
- Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science* 366, 50–54. doi: 10.1126/science.aax0287
- Johnson, N. K. (1980). Character variation and evolution of sibling species in the empidonax difficilis-flavescens complex (Aves, tyrannidae). *Wilson Bull.* 93, 412–414.
- Jones, G., and Teeling, E. C. (2006). The evolution of echolocation in bats. *Trends Ecol. Evol.* 21, 149–156. doi: 10.1016/j.tree.2006.01.001
- Jürgens, U. (2002). Neural pathways underlying vocal control. *Neurosci. Biobehav. Rev.* 26, 235–258. doi: 10.1016/S0149-7634(01)00068-9
- Jürgens, U. (2009). The neural control of vocalization in mammals: a review. *J. Voice* 23, 1–10. doi: 10.1016/j.jvoice.2007.07.005
- Kelley, D. B., and Bass, A. H. (2010). Neurobiology of vocal communication: mechanisms for sensorimotor integration and vocal patterning. *Curr. Opin. Neurobiol.* 20, 748–753. doi: 10.1016/j.conb.2010.08.007
- King, S. L., Sayigh, L. S., Wells, R. S., Fellner, W., and Janik, V. M. (2013). Vocal copying of individually distinctive signature whistles in bottlenose dolphins. *Proc. R. Soc. B: Biol. Sci.* 280. doi: 10.1098/rspb.2013.0053
- Knörnschild, M. (2014). Vocal production learning in bats. *Curr. Opin. Neurobiol.* 28, 80–85. doi: 10.1016/j.conb.2014.06.014
- Knörnschild, M., Blüml, S., Steidl, P., Eckenweber, M., and Nagy, M. (2017). Bat songs as acoustic beacons - male territorial songs attract dispersing females. *Sci. Rep.* 7, 13918. doi: 10.1038/s41598-017-14434-5
- Knörnschild, M., Nagy, M., Metz, M., Mayer, F., and Von Helversen, O. (2010). Complex vocal imitation during ontogeny in a bat. *Biol. Lett.* 6, 156–159. doi: 10.1098/rsbl.2009.0685
- Knörnschild, M., Nagy, M., Metz, M., Mayer, F., and Von Helversen, O. (2012). Learned vocal group signatures in the polygynous bat saccopteryx bilineata. *Anim. Behav.* 84, 761–769. doi: 10.1016/j.anbehav.2012.06.029
- Konishi, M. (1986). Centrally synthesized maps of sensory space. *Trends Neurosci.* 9, 163–168. doi: 10.1016/0166-2236(86)90053-6
- Kroodsma, D. E. (1977). A re-evaluation of song development in the song sparrow. *Anim. Behav.* 25, 390–399. doi: 10.1016/0003-3472(77)90013-6
- Kroodsma, D. E. (1984). Songs of the alder flycatcher (Empidonax alnorum) and willow flycatcher (Empidonax traillii) are innate. *Auk* 101, 13–24. doi: 10.1093/auk/101.1.13
- Kroodsma, D. E. (1989). Male Eastern phoebes (Sayornis phoebe; tyrannidae, passeriformes) fail to imitate songs. *J. Comp. Psychol.* 103 227–232. doi: 10.1037/0735-7036.103.3.227
- Kroodsma, D. (2011). Neither individually distinctive songs nor “lek signatures” are demonstrated in subsocial screaming pihas. *Auk* 128, 789–790. doi: 10.1525/auk.2011.128.4.789
- Kroodsma, D. E., and Canady, R. A. (1985). Differences in repertoire size, singing behavior, and associated neuroanatomy among marsh wren populations have a genetic basis. *Auk* 102, 439–446. doi: 10.1093/auk/102.3.439
- Kroodsma, D., Hamilton, D., Sánchez, J. E., Byers, B. E., Fandiño-Mariño, H., Stemple, D. W., et al. (2013). Behavioral evidence for song learning in the subsocial bellbirds (Procnias spp.; cotingidae). *Wilson J. Ornithology* 125, 1–14. doi: 10.1676/12-033.1
- Kroodsma, D. E., Houlihan, P. W., Fallon, P. A., and Wells, J. A. (1997). Song development by grey catbirds. *Anim. Behav.* 54, 457–464. doi: 10.1006/anbe.1996.0387
- Kroodsma, D. E., and Konishi, M. (1991). A subsocial bird (eastern phoebe, sayornis phoebe) develops normal song without auditory feedback. *Anim. Behav.* 42, 477–487. doi: 10.1016/S0003-3472(05)80047-8
- Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nat. Rev. Neurosci.* 5, 831–843. doi: 10.1038/nrn1533
- Kuypers, H. G. (1958). Corticobulbar connexions to the pons and lower brain-stem in man: an anatomical study. *Brain* 81, 364–388. doi: 10.1093/brain/81.3.364
- Lachlan, R. F., and Slater, P. J. B. (1999). The maintenance of vocal learning by gene-culture interaction: the cultural trap hypothesis. *Proc. R. Soc. London. Ser. B: Biol. Sci.* 266, 701–706. doi: 10.1098/rspb.1999.0692
- Lameira, A. R., Santamaria-Bonfil, G., Galeone, D., Gamba, M., Hardus, M. E., Knott, C. D., et al. (2022). Sociality predicts orangutan vocal phenotype. *Nat. Ecol. Evolution* 6, 644–652. doi: 10.1038/s41559-022-01689-z
- Lattenkamp, E. Z., Linnenschmidt, M., Mardus, E., Vernes, S. C., Wiegand, L., and Schutte, M. (2021). The vocal development of the pale spear-nosed bat is dependent on auditory feedback. *Philos. Trans. R. Soc. B: Biol. Sci.* 376, 20200253. doi: 10.1098/rstb.2020.0253
- Leitner, S., Nicholson, J., Leisler, B., Devoogd, T. J., and Catchpole, C. K. (2002). Song and the song control pathway in the brain can develop independently of exposure to song in the sedge warbler. *Proc. R. Soc. London. Ser. B: Biol. Sci.* 269, 2519–2524. doi: 10.1098/rspb.2002.2172
- Lemasson, A., Ouattara, K., Petit, E. J., and Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evol. Biol.* 11, 362. doi: 10.1186/1471-2148-11-362
- Lindell, C. (1998). Limited geographic variation in the vocalizations of a Neotropical furnariid, synallaxis albescens. *Wilson Bull.* 110, 368–374.
- Macarthur, R. H., and Macarthur, J. W. (1961). On bird species diversity. *Ecology* 42, 594–598. doi: 10.2307/1932254
- Mammen, D. L., and Nowicki, S. (1981). Individual differences and within-flock convergence in chickadee calls. *Behav. Ecol. Sociobiology* 9, 179–186. doi: 10.1007/BF00302935
- Marler, P. R. (1980). Primate Vocalization: Affective or Symbolic?. In: T. A. Sebeok and J. Umiker-Sebeok(eds) *Speaking of Apes. Topics in Contemporary Semiotics*. Springer, Boston, MA. doi: 10.1007/978-1-4613-3012-7\_13
- Marler, P. R. (1960). “Bird songs and mate selection,” in *Animal Sounds and Communication* Ed. W. E. Lanyon and W. N. Tavolga. (Washington, D.C.), 348–367.
- Marler, P. (1991). “The instinct for vocal learning: songbirds,” in *Plasticity of development*. Ed. S. E. Brauth, W. S. Hall and R. J. Dooling (Cambridge, MA, US: MIT Press), 107–125.
- Marler, P. (1997). Three models of song learning: evidence from behavior. *J. Neurobiol.* 33, 501–516. doi: 10.1002/(SICI)1097-4695(19971105)33:5<501::AID-NEU2>3.0.CO;2-8
- Marler, P., Mundinger, P., Waser, M., and Lutjen, A. (1972). Effects of acoustical stimulation and deprivation on song development in red-winged blackbirds. *Anim. Behav.* 20, 586–606. doi: 10.1016/S0003-3472(72)80024-1
- Marler, P., and Peters, S. (1988). The role of song phonology and syntax in vocal learning preferences in the song sparrow, melospiza melodia. *Ethology formerly Z. fur Tierpsychologie* 77, 125–149.
- Marler, P., and Sherman, V. (1985). Innate differences in singing behaviour of sparrows reared in isolation from adult conspecific song. *Anim. Behav.* 33, 57–71. doi: 10.1016/S0003-3472(85)80120-2
- Marler, P., and Tamura, M. (1962). Song “dialects” in three populations of white-crowned sparrows. *Condor* 64, 368–377. doi: 10.2307/1365545
- Marler, P., and Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in a sparrow. *Science* 146, 1483–1486. doi: 10.1126/science.146.3650.1483
- Marler, P. (1980). “Primate vocalization: Affective or symbolic?,” in *Speaking of apes: A critical anthology of two-way communication with man*. Eds. T. A. Sebeok and J. Umiker-Sebeok (Boston, MA: Springer US), 221–229. doi: 10.1007/978-1-4613-3012-7\_13
- Marten, K., Quine, D., and Marler, P. (1977). Sound transmission and its significance for animal vocalization: II. tropical forest habitats. *Behav. Ecol. Sociobiology* 2, 291–302. doi: 10.1007/BF00299741

- Martins, P. T., and Boeckx, C. (2020). Vocal learning: beyond the continuum. *PLoS Biol.* 18, e3000672. doi: 10.1371/journal.pbio.3000672
- Matsumoto, R., Nair, D. R., Lapresto, E., Najm, I., Bingaman, W., Shibasaki, H., et al. (2004). Functional connectivity in the human language system: a cortico-cortical evoked potential study. *Brain* 127, 2316–2330. doi: 10.1093/brain/awh246
- McGuire, J. A., Witt, C. C., Remsen, J. V. Jr., Corl, A., Rabosky, D. L., Altschuler, D. L., et al. (2014). Molecular phylogenetics and the diversification of hummingbirds. *Curr. Biol.* 24, 910–916. doi: 10.1016/j.cub.2014.03.016
- Mennill, D. J., Doucet, S. M., Newman, A. E. M., Williams, H., Moran, I. G., Thomas, I. P., et al. (2018). Wild birds learn songs from experimental vocal tutors. *Curr. Biol.* 28, 3273–3278.e3274. doi: 10.1016/j.cub.2018.08.011
- Mercado, E. (2022). The humpback's new songs: diverse and convergent evidence against vocal culture via copying in humpback whales. *Anim. Behav. Cogn.* 9, 196–206. doi: 10.26451/abc.09.02.03.2022
- Metzner, W., and Schuller, G. (2010). "Chapter 9.4 - vocal control in echolocating bats," in *Handbook of behavioral neuroscience*. Ed. S. M. Brudzynski. (Oxford: Elsevier), 403–415.
- Mitoyen, C., Quigley, C., and Fusani, L. (2019). Evolution and function of multimodal courtship displays. *Ethology* 125, 503–515. doi: 10.1111/eth.12882
- Morton, E. S. (1975). Ecological sources of selection on avian sounds. *Am. Nat.* 109, 17–34. doi: 10.1086/282971
- Munk, W. H., Spindel, R. C., Baggeroer, A., and Birdsall, T. G. (1994). The heard island feasibility test. *J. Acoustical Soc. America* 96, 2330–2342. doi: 10.1121/1.410105
- Nelson, D. A., and Poesel, A. (2007). Segregation of information in a complex acoustic signal: individual and dialect identity in white-crowned sparrow song. *Anim. Behav.* 74, 1073–1084. doi: 10.1016/j.anbehav.2007.01.018
- Nicolaidis, K. (2006). "Speech development," in *Encyclopedia of language & linguistics*, 2nd ed. Ed. K. Brown. (Oxford: Elsevier), 722–736.
- Nottebohm, F. (1972). The origins of vocal learning. *Am. Nat.* 106, 116–140. doi: 10.1086/282756
- Nowicki, S. (1989). Vocal plasticity in captive black-capped chickadees: the acoustic basis and rate of call convergence. *Anim. Behav.* 37, 64–73. doi: 10.1016/0003-3472(89)90007-9
- Nowicki, S., Peters, S., and Podos, J. (1998). Song learning, early nutrition and sexual selection in Songbirds. *Am. Zoologist* 38, 179–190. doi: 10.1093/icb/38.1.179
- Nowicki, S., and Searcy, W. A. (2014). The evolution of vocal learning. *Curr. Opin. Neurobiol.* 28, 48–53. doi: 10.1016/j.conb.2014.06.007
- Owren, M. J., Amoss, R. T., and Rendall, D. (2011). Two organizing principles of vocal production: implications for nonhuman and human primates. *Am. J. Primatol.* 73, 530–544. doi: 10.1002/ajp.20913
- Park, T., Fitzgerald, E. M., and Evans, A. R. (2016). Ultrasonic hearing and echolocation in the earliest toothed whales. *Biol. Lett.* 12, doi: 10.1098/rsbl.2016.0060
- Payne, R. B. (1985). Behavioral continuity and change in local song populations of village indigobirds *vidua chalybeata*. *Z. für Tierpsychologie* 70, 1–44. doi: 10.1111/j.1439-0310.1985.tb00498.x
- Petkov, C., and Jarvis, E. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evolutionary Neurosci.* 4, doi: 10.3389/fevo.2012.00012
- Petkov, C. L., and Wilson, B. (2012). On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2077–2088. doi: 10.1098/rstb.2012.0073
- Pika, S., Wilkinson, R., Kendrick, K. H., and Vernes, S. C. (2018). Taking turns: bridging the gap between human and animal communication. *Proc. R. Soc. B Biol. Sci.* 285, doi: 10.1098/rspb.2018.0598
- Pokorny, J. J., and Waal, F. B. M. D. (2009). Monkeys recognize the faces of group mates in photographs. *Proc. Natl. Acad. Sci.* 106, 21539–21543. doi: 10.1073/pnas.0912174106
- Poole, J. H., Tyack, P. L., Stoeger-Horwath, A. S., and Watwood, S. (2005). Elephants are capable of vocal learning. *Nature* 434, 455–456. doi: 10.1038/434455a
- Prat, Y., Azoulay, L., Dor, R., and Yovel, Y. (2017). Crowd vocal learning induces vocal dialects in bats: playback of conspecifics shapes fundamental frequency usage by pups. *PLoS Biol.* 15, e2002556. doi: 10.1371/journal.pbio.2002556
- Prat, Y., Taub, M., and Yovel, Y. (2015). Vocal learning in a social mammal: demonstrated by isolation and playback experiments in bats. *Sci. Adv.* 1, e1500019. doi: 10.1126/sciadv.1500019
- Prieur, J., Barbu, S., Blois-Heulin, C., and Lemasson, A. (2020). The origins of gestures and language: history, current advances and proposed theories. *Biol. Rev.* 95, 531–554. doi: 10.1111/brv.12576
- Reichgelt, T., Greenwood, D. R., Steinig, S., Conran, J. G., Hutchinson, D. K., Lunt, D. J., et al. (2022). Plant proxy evidence for high rainfall and productivity in the Eocene of Australia. *Paleoceanography Paleoclimatology* 37, e2022PA004418. doi: 10.1029/2022PA004418
- Renouf, D., and Davis, M. B. (1982). Evidence that seals may use echolocation. *Nature* 300, 635–637. doi: 10.1038/300635a0
- Rice, J. O. H., and Thompson, W. L. (1968). Song development in the indigo bunting. *Anim. Behav.* 16, 462–469. doi: 10.1016/0003-3472(68)90041-9
- Richards, D. G., and Wiley, R. H. (1980). Reverberations and amplitude fluctuations in the propagation of sound in a forest: implications for animal communication. *Am. Nat.* 115, 381–399. doi: 10.1086/283568
- Roberts, T. F., Hisey, E., Tanaka, M., Kearney, M. G., Chattree, G., Yang, C. F., et al. (2017). Identification of a motor-to-auditory pathway important for vocal learning. *Nat. Neurosci.* 20, 978. doi: 10.1038/nn.4563
- Roberts, T. F., Klein, M. E., Kubke, M. F., Wild, J. M., and Mooney, R. (2008). Telencephalic neurons monosynaptically link brainstem and forebrain premotor networks necessary for song. *J. Neurosci.* 28, 3479–3489. doi: 10.1523/JNEUROSCI.0177-08.2008
- Robinson, C. M., Snyder, K. T., and Creanza, N. (2019). Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning. *eLife* 8, e44454. doi: 10.7554/eLife.44454.051
- Rodenas-Cuadrado, P. M., Mengede, J., Baas, L., Devanna, P., Schmid, T. A., Yartsev, M., et al. (2018). Mapping the distribution of language related genes FoxP1, FoxP2, and CntnaP2 in the brains of vocal learning bat species. *J. Comp. Neurol.* 526, 1235–1266. doi: 10.1002/cne.24385
- Rose, G. J., Goller, F., Gritton, H. J., Plamondon, S. L., Baugh, A. T., and Cooper, B. G. (2004). Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* 432, 753–758. doi: 10.1038/nature02992
- Rubsamen, R., and Schweizer, H. (1986). Control of echolocation pulses by neurons of the nucleus ambiguus in the rufous horseshoe bat, *Rhinolophus rouxi*. *J. Comp. Physiol. A* 159, 689–699. doi: 10.1007/BF00612041
- Rühmann, J., Soler, M., Pérez-Contreras, T., and Ibáñez-Álamo, J. D. (2019). Territoriality and variation in home range size through the entire annual range of migratory great spotted cuckoos (*Clamator glandarius*). *Sci. Rep.* 9, 6238. doi: 10.1038/s41598-019-41943-2
- Ryan, M. J. (1986). Synchronized calling in a treefrog (*Smilisca sila*). short behavioral latencies and implications for neural pathways involved in call perception and production. *Brain Behav. Evol.* 29, 196–206. doi: 10.1159/000118681
- Ryan, M. J., and Brenowitz, E. A. (1985). The role of body size, phylogeny, and ambient noise in the evolution of bird song. *Am. Nat.* 126, 87–100. doi: 10.1086/284398
- Searcy, W. A., Marler, P., and Peters, S. S. (1985). Songs of isolation-reared sparrows function in communication, but are significantly less effective than learned songs. *Behav. Ecol. Sociobiology* 17, 223–229. doi: 10.1007/BF00300140
- Searcy, W. A., and Nowicki, S. (2019). Birdsong learning, avian cognition and the evolution of language. *Anim. Behav.* 151, 217–227. doi: 10.1016/j.anbehav.2019.01.015
- Seddon, N. (2005). Ecological adaptation and species recognition drives vocal evolution in neotropical subsocial birds. *Evolution* 59, 200–215, 216. doi: 10.1111/j.0014-3820.2005.tb00906.x
- Sedgwick, J. A. (2001). Geographic variation in the song of willow flycatchers: differentiation between empidonax traillii adastus and e. t. extimus. *Auk* 118, 366–379. doi: 10.1093/auk/118.2.366
- Sirović, A., Hildebrand, J. A., and Wiggins, S. M. (2007). Blue and fin whale call source levels and propagation range in the southern ocean. *J. Acoustic Soc. America* 122, 1208–1215. doi: 10.1121/1.2749452
- Slater, P. J. B., and Janik, V. M. (2010). "Vocal learning," in *Encyclopedia of animal behavior*. Eds. M. D. Breed and J. Moore. (Oxford: Academic Press), 551–557.
- Sliwa, J., Duhamel, J.-R., Pascalis, O., and Wirth, S. (2011). Spontaneous voice and face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc. Natl. Acad. Sci.* 108, 1735–1740. doi: 10.1073/pnas.1008169108
- Smotherman, M., Knörnschild, M., Smarsh, G., and Bohn, K. (2016). The origins and diversity of bat songs. *J. Comp. Physiol. A Neuroethol Sens Neural Behav. Physiol.* 202, 535–554. doi: 10.1007/s00359-016-1105-0
- Sober, S. J., and Brainard, M. S. (2012). Vocal learning is constrained by the statistics of sensorimotor experience. *Proc. Natl. Acad. Sci.* 109, 21099–21103. doi: 10.1073/pnas.1213622109
- Soha, J. A., and Marler, P. (2000). A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Anim. Behav.* 60, 297–306. doi: 10.1006/anbe.2000.1499
- Soma, M., and Garamszegi, L. Z. (2011). Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. *Behav. Ecol.* 22, 363–371. doi: 10.1093/beheco/arq219
- Stansbury, A. L., and Janik, V. M. (2019). Formant modification through vocal production learning in Gray seals. *Curr. Biol.* 29, 2244–2249.e2244. doi: 10.1016/j.cub.2019.05.071
- Stoddard, P. K., Beecher, M. D., Horning, C. L., and Campbell, S. E. (1991). Recognition of individual neighbors by song in the song sparrow, a species with song repertoires. *Behav. Ecol. Sociobiology* 29, 211–215. doi: 10.1007/BF00166403
- Stoeger, A. S., and Manger, P. (2014). Vocal learning in elephants: neural bases and adaptive context. *Curr. Opin. Neurobiol.* 28, 101–107. doi: 10.1016/j.conb.2014.07.001
- Stoeger, A. S., Mietchen, D., Oh, S., De Silva, S., Herbst, C. T., Kwon, S., et al. (2012). An Asian elephant imitates human speech. *Curr. Biol.* 22, 2144–2148. doi: 10.1016/j.cub.2012.09.022
- Stouffer, P. C. (2007). Density, territory size, and long-term spatial dynamics of a guild of terrestrial insectivorous birds near Manaus, Brazil. *Auk* 124, 291–306. doi: 10.1093/auk/124.1.291
- Supalla, T., Hauser, P. C., and Bavelier, D. (2014). Reproducing American sign language sentences: cognitive scaffolding in working memory. *Front. Psychol.* 5, 859. doi: 10.3389/fpsyg.2014.00859

- Suthers, R. A., and Zollinger, S. A. (2004). Producing song: the vocal apparatus. *Ann. New York Acad. Sci.* 1016, 109–129. doi: 10.1196/annals.1298.041
- Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., et al. (2015). The developmental dynamics of marmoset monkey vocal production. *Science* 349, 734–738. doi: 10.1126/science.aab1058
- Thorpe, W. H. (1958). The learning of song patterns by birds, with especial references to the song of the chaffinch. *Ibis* 100, 535–570. doi: 10.1111/j.1474-919X.1958.tb07960.x
- Touchton, J. M., Seddon, N., and Tobias, J. A. (2014). Captive rearing experiments confirm song development without learning in a tracheophone suboscine bird. *PLoS One* 9, e95746. doi: 10.1371/journal.pone.0095746
- Tyack, P. L. (2016). “Vocal learning and auditory-vocal feedback,” in *Vertebrate sound production and acoustic communication*. Eds. R. A. Suthers, W. T. Fitch, R. R. Fay and A. N. Popper. (Cham: Springer International Publishing), 261–295.
- Tyack, P. L. (2020). A taxonomy for vocal learning. *Philos. Trans. R. Soc. Lond B Biol. Sci.* 375, 20180406. doi: 10.1098/rstb.2018.0406
- Ulanovsky, N., and Moss, C. F. (2008). What the bat's voice tells the bat's brain. *Proc. Natl. Acad. Sci.* 105, 8491–8498. doi: 10.1073/pnas.0703550105
- Utescher, T., and Mosbrugger, V. (2007). Eocene Vegetation patterns reconstructed from plant diversity [1/amp]mdash; a global perspective. *Palaeogeography Palaeoclimatology Palaeoecol.* 247, 243–271. doi: 10.1016/j.palaeo.2006.10.022
- Vernes, S. C., Kriengwatana, B. P., Beeck, V. C., Fischer, J., Tyack, P. L., Ten Cate, C., et al. (2021). The multi-dimensional nature of vocal learning. *Philos. Trans. R. Soc. B: Biol. Sci.* 376, 20200236. doi: 10.1098/rstb.2020.0236
- Vernes, S. C., and Wilkinson, G. S. (2020). Behaviour, biology and evolution of vocal learning in bats. *Philos. Trans. R. Soc. B: Biol. Sci.* 375, 20190061. doi: 10.1098/rstb.2019.0061
- Von Eugon, K., Endepols, H., Drzezza, A., Neumaier, B., Güntürkün, O., Backes, H., et al. (2022). Avian neurons consume three times less glucose than mammalian neurons. *Curr. Biol.* 32, 4306–4313. doi: 10.1016/j.cub.2022.07.070
- Vouloumanos, A., and Werker, J. F. (2007). Listening to language at birth: evidence for a bias for speech in neonates. *Dev. Sci.* 10, 159–164. doi: 10.1111/j.1467-7687.2007.00549.x
- Wang, Z., Zhu, T., Xue, H., Fang, N., Zhang, J., Zhang, L., et al. (2017). Prenatal development supports a single origin of laryngeal echolocation in bats. *Nat. Ecol. Evol.* 1, 21. doi: 10.1038/s41559-016-0021
- Watwood, S. L., Tyack, P. L., and Wells, R. S. (2004). Whistle sharing in paired male bottlenose dolphins, *tursiops truncatus*. *Behav. Ecol. Sociobiology* 55, 531–543. doi: 10.1007/s00265-003-0724-y
- Wennstrom, K. L., Reeves, B. J., and Brenowitz, E. A. (2001). Testosterone treatment increases the metabolic capacity of adult avian song control nuclei. *J. Neurobiol.* 48, 256–264. doi: 10.1002/neu.1055
- White, S. A. (2010). Genes and vocal learning. *Brain Lang.* 115, 21–28. doi: 10.1016/j.bandl.2009.10.002
- White, E. J., Hutka, S. A., Williams, L. J., and Moreno, S. (2013). Learning, neural plasticity and sensitive periods: implications for language acquisition, music training and transfer across the lifespan. *Front. Syst. Neurosci.* 7, 90–90. doi: 10.3389/fnsys.2013.00090
- White, T. E., Latty, T., and Umbers, K. D. L. (2022). The exploitation of sexual signals by predators: a meta-analysis. *Proc. R. Soc. B: Biol. Sci.* 289, 20220444. doi: 10.1098/rspb.2022.0444
- Wich, S. A., Krützen, M., Lameira, A. R., Nater, A., Arora, N., Bastian, M. L., et al. (2012). Call cultures in orang-utans? *PLoS One* 7, e36180. doi: 10.1371/journal.pone.0036180
- Wilczynski, W., and Capranica, R. R. (1984). The auditory system of anuran amphibians. *Prog. Neurobiol.* 22, 1–38. doi: 10.1016/0301-0082(84)90016-9
- Wild, J. M. (1997). Neural pathways for the control of birdsong production. *J. Neurobiol.* 33, 653–670. doi: 10.1002/(SICI)1097-4695(19971105)33:5<653::AID-NEU11>3.0.CO;2-A
- Wiley, R. H. (2017). How noise determines the evolution of communication. *Anim. Behav.* 124, 307–313. doi: 10.1016/j.anbehav.2016.07.014
- Wilson, E. O., and Bossert, W. H. (1963). Chemical communication among animals. *Recent Prog. Hormone Res.* 19, 673–716.
- Wirthlin, M., Chang, E. F., Knörnschild, M., Krubitzer, L. A., Mello, C. V., Miller, C. T., et al. (2019). A modular approach to vocal learning: disentangling the diversity of a complex behavioral trait. *Neuron* 104, 87–99. doi: 10.1016/j.neuron.2019.09.036
- Wisenden, B. D. (2008). Active space of chemical alarm cue in natural fish populations. *Behaviour* 145, 391–407. doi: 10.1163/156853908783402920
- Wright, T. F., Schirtzinger, E. E., Matsumoto, T., Eberhard, J. R., Graves, G. R., Sanchez, J. J., et al. (2008). A multilocus molecular phylogeny of the parrots (Psittaciformes): support for a gondwanan origin during the cretaceous. *Mol. Biol. Evol.* 25, 2141–2156. doi: 10.1093/molbev/msn160
- Yosida, S., Kobayasi, K. I., Ikebuchi, M., Ozaki, R., and Okanoya, K. (2007). Antiphonal vocalization of a subterranean rodent, the naked mole-rat (*Heterocephalus glaber*). *Ethology* 113, 703–710. doi: 10.1111/j.1439-0310.2007.01371.x
- Zürcher, Y., Willems, E. P., and Burkart, J. M. (2021). Trade-offs between vocal accommodation and individual recognisability in common marmoset vocalizations. *Sci. Rep.* 11. doi: 10.1038/s41598-021-95101-8