Perspective

A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait

Morgan Wirthlin,[1,*] Edward F. Chang,[2] Mirjam Knörnschild,[3] Leah A. Krubitzer,[4,5] Claudio V. Mello,[6] Cory T. Miller,[7] Andreas R. Pfenning,[1] Sonja C. Vernes,[8,9] Ofer Tchernichovski,[10] and Michael M. Yartsev[11,*]

[1]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA, USA
[2]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA
[3]Museum for Natural History Berlin, Invalidenstr. 43, 10115 Berlin, Germany
[4]Center for Neuroscience, University of California, Davis, CA, USA
[5]Department of Psychology, University of California, Davis, CA, USA
[6]Department of Behavioral Neuroscience, OHSU, Portland, OR, USA
[7]Cortical Systems and Behavior Laboratory, Neurosciences Graduate Program, University of California San Diego, La Jolla, CA, USA
[8]Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
[9]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, the Netherlands
[10]Department of Psychology, Hunter College, The City University of New York, New York, NY, USA
[11]Helen Wills Neuroscience Institute and Department of Bioengineering, UC Berkeley, Berkeley, CA, USA
*Correspondence: mwirthlin@cmu.edu (M.W.), myartsev@berkeley.edu (M.M.Y.)
https://doi.org/10.1016/j.neuron.2019.09.036

Vocal learning is a behavioral trait in which the social and acoustic environment shapes the vocal repertoire of individuals. Over the past century, the study of vocal learning has progressed at the intersection of ecology, physiology, neuroscience, molecular biology, genomics, and evolution. Yet, despite the complexity of this trait, vocal learning is frequently described as a binary trait, with species being classified as either vocal learners or vocal non-learners. As a result, studies have largely focused on a handful of species for which strong evidence for vocal learning exists. Recent studies, however, suggest a continuum in vocal learning capacity across taxa. Here, we further suggest that vocal learning is a multi-component behavioral phenotype comprised of distinct yet interconnected modules. Discretizing the vocal learning phenotype into its constituent modules would facilitate integration of findings across a wider diversity of species, taking advantage of the ways in which each excels in a particular module, or in a specific combination of features. Such comparative studies can improve understanding of the mechanisms and evolutionary origins of vocal learning. We propose an initial set of vocal learning modules supported by behavioral and neurobiological data and highlight the need for diversifying the field in order to disentangle the complexity of the vocal learning phenotype.

Evolving Definitions of Vocal Learning

The behavioral trait typically studied as "vocal learning" involves "the influence of auditory information, including feedback, on vocal development" (Nottebohm, 1972). A leading framework from Janik and Slater (1997, 2000) distinguishes between (1) comprehension learning, in which an individual can learn to associate an acoustic signal with an action or an outcome (e.g., a dog can learn to sit after hearing the word "sit"); (2) vocal usage learning, in which an individual can learn to produce an innate vocalization in a novel context (e.g., a dog can learn to bark when prompted); and (3) vocal production learning, in which an individual's vocalizations are modified in some way as a result of social interaction with other individuals (as if a fictive dog could learn to repeat the word "speak," a vocalization not part of this species' natural repertoire).

A lingering problem with working definitions of vocal production learning is that it is not entirely clear what level of vocal production plasticity qualifies a species to cross the "threshold" to be considered a vocal production learner. How inclusive should be the set of species that possess the vocal learning trait? In some taxa, all species are thought to be vocal production

learners (songbirds and parrots) (Nottebohm, 1972). In other taxa, strong evidence for vocal production learning exists for only a handful of species within a hugely diverse clade (hummingbirds, cetaceans, bats, and pinnipeds) (Janik and Slater, 1997; Jarvis, 2004). In further cases, the current evidence for vocal learning rests on more limited or anecdotal evidence (non-human primates, elephants, mice, rock hyraxes, and suboscine bellbirds) (Arriaga et al., 2012; Fischer, 2017; Kershenbaum et al., 2012; Saranathan et al., 2007; Stoeger and Manger, 2014). In the inbred laboratory mouse, for example, one study has suggested that there may be evidence for limited vocal plasticity (Arriaga et al., 2012), while other studies demonstrate that vocalizations are entirely innate (Hammerschmidt et al., 2012, 2015; Mahrt et al., 2013) and that variability in vocal output across behavioral states may be a byproduct of changes in locomotor activity (Blumberg, 1992). This uneven evidential landscape presents a conundrum for researchers interested in performing comparative analyses of vocal learning behavior.

Further developing on the concept of vocal production learning, the continuum hypothesis was proposed (Arriaga and Jarvis, 2013; Petkov and Jarvis, 2012). The continuum
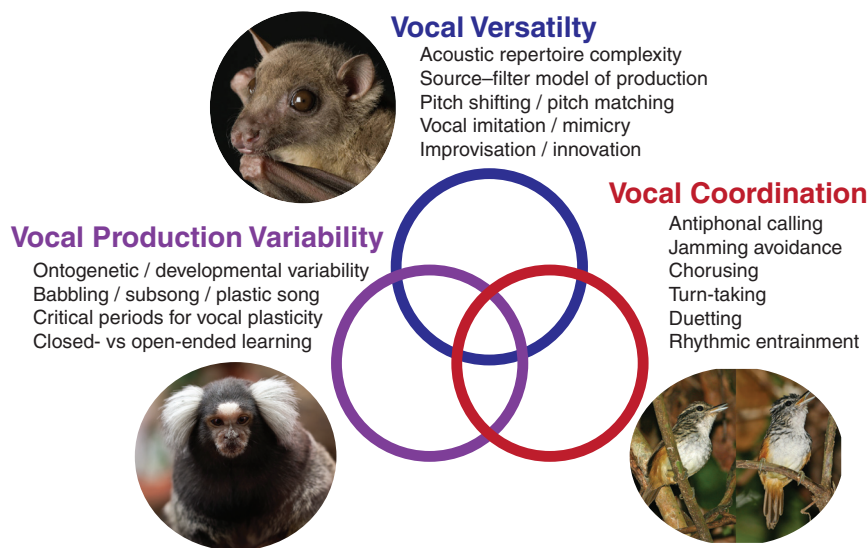
Vocal Versatilty
Acoustic repertoire complexity
Source–filter model of production
Pitch shifting / pitch matching
Vocal imitation / mimicry
Improvisation / innovation

Vocal Coordination
Antiphonal calling
Jamming avoidance
Chorusing
Turn-taking
Duetting
Rhythmic entrainment

Vocal Production Variability
Ontogenetic / developmental variability
Babbling / subsong / plastic song
Critical periods for vocal plasticity
Closed- vs open-ended learning

Figure 1. Vocal Learning as the Intersection of Several Component Modules
As a starting point for further discussion of the modular framework for vocal learning, we propose three modules, listed here with related concepts from the broader vocal learning literature. Species discussed further in this review in relation to each module include the common marmoset (*Callithrix jacchus*, photo credit Leszek Leszczyński) for vocal production variability; warbling antbirds (*Hypocnemis sp.*, photo credit Joseph Tobias) for vocal coordination; and the Egyptian fruit bat (*Rousettus aegypticus*, photo credit Joel Sartore) for vocal versatility.

by the mosaic model of brain evolution (Barton and Harvey, 2000; Moore and DeVoogd, 2017), in which individual behavioral modules can be independently targeted by evolutionary selection. That is, it may be possible for a species to have a high degree of complexity in some modules but not others. We suggest that even if a species is a "poor" vocal learner on average (when considering all modules combined), it might still be a preferred system for studying the neurobiology of a single module in which it excels. For example, the transition from highly variable juvenile vocalizations (e.g., babbling) to stereotyped adult vocalizations can be observed both in excellent canonical vocal learners such as songbirds (e.g., zebra finch) (Tchernichovski et al., 2001) and also in some traditionally defined vocal non-learners (e.g., the Japanese quail) (Derégnaucourt et al., 2009). Both species show a developmental transition from variable to stereotyped calls. However, the quail does so without any imitative learning while the songbird acquires novel sounds learned from its tutors (Derégnaucourt et al., 2009). Here, the lack of other features in the quail that are traditionally attributed to the vocal learning trait is in fact an advantage, allowing studies to focus on a single important component of vocal development in isolation of others. Similarly, at the neural level, a species specialized for a given module is advantageous in that it provides a focused assessment of the neural mechanisms underlying it. The quail, for example, does not possess the brain circuitry necessary for vocal imitation yet still retains the neural components necessary for driving variable juvenile vocalizations. Thus, it can be used as a precise system for studying that specific module in isolation from the behavioral to the neural level.

Our modular framework encourages a move toward a more holistic and inclusive view of this complex trait that incorporates a greater diversity of species. As with any behavior, vocal learning does not exist in a vacuum, but rather reflects the ways in which each species is uniquely adapted to its physical and social environment and constrained by its particular evolutionary history (Krubitzer and Prescott, 2018). To properly account for this evolutionary context, comparison of a diverse set of vocal learning taxa is necessary to distinguish mechanistic traits shared broadly across vocal learners, which could reflect fundamental mechanisms for vocal learning (Pfenning et al., 2014), from lineage-specific traits, which could reflect alternative mechanisms or unique, taxon-specific elaborations of the vocal
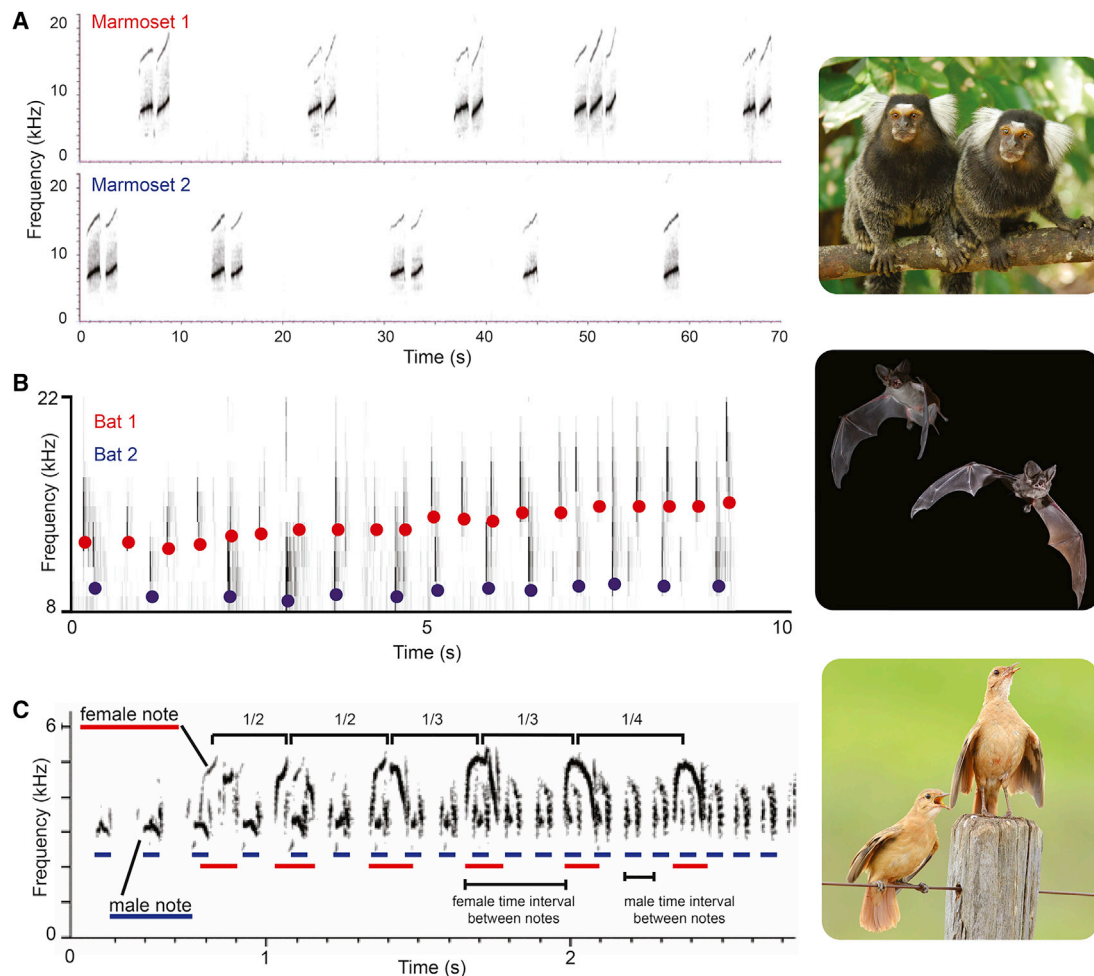
hypothesis states that with respect to vocal production learning, species can be ranked according to their vocal learning ability on a continuous scale ranging from vocal non-learners, through limited and moderate learners, to complex and high vocal learners. The most rudimentary or limited form of vocal learning is stated to be the ability to modify the amplitude and temporal structure of an innate call. Placed above this is the capacity for vocal improvisation, which may include spectral modifications, and at the top are species capable of spectro-temporal modification of their vocalizations in order to imitate (or mimic) an acquired acoustic target (Arriaga and Jarvis, 2013). By allowing for the existence of a gray area between vocal learning and non-learning, the continuum hypothesis opened the field to investigation of a broader range of species that would not meet the strict criteria of Janik and Slater for vocal production learning, but which might nevertheless provide valuable insight into the broader mechanisms for vocal control. However, by proposing a fixed set of behavioral and neurobiological criteria by which species advance along a unidimensional continuum from low to high, the continuum hypothesis, as originally formulated, makes the implicit assumption that there is a single trajectory along which the evolution of vocal production learning proceeds. As we discuss here, it is possible that certain features of vocal production, but not others, could evolve independently to be influenced by learning in different taxa rather than collectively moving toward a single end state. Expanding on the insights of the continuum hypothesis, we suggest that a more comprehensive account of the diversity of vocal learning capacities among taxa could provide a more complete and unified understanding of the vocal learning phenotype.

A Modular Hypothesis of Vocal Learning
We propose that rather than a binary or a single continuous trait, vocal learning represents a multidimensional behavioral trait (Beecher and Brenowitz, 2005; Lattenkamp and Vernes, 2018; Wirthlin, 2016) comprised of discrete sub-traits, which we term modules (Figure 1). Our use of the term module here is inspired

Figure 2. Vocal Coordination

(A) In common marmoset antiphonal calling (figure adapted from Miller and Wren Thomas, 2012, photo credit Tina Gunhold-de Oliveira), adults engage in ongoing bouts of rhythmically coordinated calling (or turn-taking).

(B) Bats use the acoustic reflections of echolocation calls to navigate and detect prey. While flying and foraging in a group context, the echolocation of conspecifics threatens to "jam" the signal of a caller. To avoid this, the European free-tailed bat (*Tadarida teniotis*, photo credit Paolo Taranto) dynamically shifts the timing and pitch of their calls to avoid overlap (blue and red dots indicate the minimum frequency of echolocation calls of two individual bats, adapted from Ulanovsky et al., 2004).

(C) During vocal duetting, rufous horneros (*Furnarius rufus*, photo credit Bruno Rennó) tightly coordinate their vocal production. These duets are refined over time between partners to match a highly structured pattern with subsecond temporal precision (adapted from Laje and Mindlin, 2003).

learning phenotype (Brenowitz and Beecher, 2005). As a starting point for further deliberation, we propose three initial modules: (1) vocal coordination, (2) vocal production variability, and (3) vocal versatility (Figure 1). We do not claim that these modules encompass the full diversity of vocal learning-related phenotypes or mechanisms, but we believe they capture several of the key known components of this complex trait. In our discussion, we aim to highlight the utility of the modular approach in diversifying the study of vocal learning.

Module 1: Vocal Coordination

The first proposed module refers to the ability of an animal to flexibly modify the temporal production of its vocal output. This may range from, at its most basic and non-learned state, reflexive vocal production in response to a stimulus, to predictive vocal production coordinated with the actions of a partner, to synchronized rhythmic entrainment in which the ongoing temporal vocal coordination becomes self-guided (Figure 2). In many respects, this module complements the concept of vocal usage learning as defined by Janik and Slater (2000), wherein a species may learn to produce a learned or non-learned vocalization in a new context. However, vocal coordination as defined here is presented as a spectrum that would also include vocal usage that is non-learned on one end, to instances where the temporal parameters of a vocalization itself may be learned, which would qualify as vocal production learning under Janik and Slater (2000).

Vocal Coordination Abilities across Species

In its simplest expression, vocal coordination takes the form of a reflexive response to a stimulus, such as the call-and-response behavior exhibited by multiple species, with very

limited evidence for plasticity (Ghazanfar and Takahashi, 2014). More complex forms of vocal coordination can emerge as a response to different forms of "jamming" in the acoustic environment. In the example of call alternation, multiple vocalizing individuals may adjust the phase and tempo of their calls in response to neighboring callers in order to avoid signal interference. This has been extensively studied in anurans, whose breeding success depends on the precise timing of calls to maximize the chances of their own calls being heard by a potential mate, or to actively interrupt the calls of their neighbors (Awbrey, 1978; Grafe, 1996; Zelick and Narins, 1983). A similar phenomenon is observed in group-foraging bats, where individuals can dynamically coordinate their echolocation calls during flight to avoid jamming interference (Ulanovsky and Moss, 2008). This may be achieved through diverse mechanisms, including temporal synchronization of calls (this module) as well as shifting in the frequency of calls to avoid acoustic overlap (falling under the vocal versatility module, discussed later) (Obrist, 1995; Ulanovsky et al., 2004). Here there are interesting mechanistic parallels with jamming avoidance in the weakly electric fish that present an example of non-vocal temporal coordination and where the neurobiological underpinnings have been extensively elucidated (Bell and Maler, 2005; Heiligenberg, 1977). Importantly, in the electric fish, the frog, and the bat, the signals are thought to be innate, yet their temporal structure can be flexibly adjusted in response to the social and acoustic environment.

Evidence from songbirds suggests that vocal coordination may represent a distinct module of vocal behavior even in more adept vocal learners. Although in zebra finches only the male sings, both sexes produce innate social calls that can be rhythmically coordinated (Benichov et al., 2016a). In an acoustic jamming challenge in which the normal timing of a bird's calls was interrupted by a vocal robot, the birds displayed rapid rhythmic adaptation to avoid their calls being disrupted, demonstrating vocal coordination that is not simply responsive, but predictive of future calls (Benichov et al., 2016a). Finally, lesions targeting the primary song production motor pathway (comprised of HVC, proper name, as well as RA, the robust nucleus of the arcopallium) impaired the birds' ability to predictively coordinate the precise timing of these innate calls. This effect was found not only in male zebra finches, but also in female zebra finches—who don't sing—suggesting that a neuroanatomical locus for the precise timing of these calls resides within the avian song system (Benichov et al., 2016a).

Several animal species, including some traditionally considered vocal production non-learners, can synchronize the timing of vocal exchanges with a partner with great degrees of precision, e.g., antiphonal calling, vocal turn-taking, and duetting (Benichov et al., 2016b; Pika et al., 2018). In mammals, neotropical singing mice (*Scotinomys* sp.) engage in countersinging behavior that is tightly temporally coordinated, comprised of simple notes that are identifiable from birth (Campbell et al., 2014). During countersinging exchanges, the ability to initiate vocal behavior in response to the proper cue (vocal offset of the countersinging partner) is impaired by reversible inhibition of the orofacial primary motor cortex (Okobi et al., 2019), suggesting a cortical basis for the control of this behavior, despite the spectral parameters of the notes not being learned. Among

primates, vocal turn-taking is evident across all the taxonomic groups (Levinson, 2016). For example, mate-paired gibbons sing elaborate and extensive duets comprised of complex sequences of individually simple calls (Clarke et al., 2006; Elliott, 1981). While the calls themselves appear to be innate, the proper timing of their duet sequence takes months to develop (Maples et al., 1989). Importantly, certain aspects of this behavior are learned during development: in marmosets and humans, turn-taking behavior is learned over the first few months of life (Chow et al., 2015). Furthermore, turn-taking is not exclusive to the vocal domain, being evident also in human sign language as well as the gestural turn-taking observed in great apes (Levinson, 2016). This vocal and gestural coordination in primates has been proposed as a possible precursor to human speech (Ghazanfar and Takahashi, 2014; Levinson, 2016).

Among birds, coordinated vocal exchanges are observed in all three taxa traditionally considered vocal production learners—songbirds, parrots, and hummingbirds—but also in vocal non-learners such as swans, owls, and woodpeckers (Farabaugh, 1982; Hall, 2009). For example, extensive and highly flexible temporal vocal coordination is also observed in suboscine Passerines, the sister taxon of songbirds (i.e., oscine Passerines). The suboscine clade consists of more than 1,000 species that are widely regarded as being vocal non-learners on the basis of deafening and hand-rearing experiments performed in just 3 closely related species of a single family, Tyrannidae (Kroodsma, 1984, 1989; Kroodsma and Konishi, 1991). In other suboscine families, there is suggestive evidence for learning in the temporal domain: male long-tailed manakins (suboscine family Pipridae) develop coordinated duet songs over time as part of their cooperative lek-breeding displays (Trainer et al., 2002). Mate-paired horneros (family Furnariidae) sing a highly structured duet with a complex rhythmic time signature, in which the male and female must match and coordinate their tempo with that of their partner's on a subsecond level (Diniz et al., 2018; Laje and Mindlin, 2003). Pair-living Peruvian warbling antbirds (family Thamnophilidae) not only develop coordinated duets but also display remarkable temporal vocal flexibility: females will respond to the signals of unpaired female sexual rivals by jamming their own male partner's song, which the males will attempt to circumvent by adjusting the tempo of their own song in order to avoid jamming interference (Tobias and Seddon, 2009).

Neural Mechanisms of Vocal Coordination

The ability to vocalize, in its least derived state, is controlled by hindbrain (brainstem) circuits. The brainstem includes primary motor neurons that innervate the musculature of the vocal apparatus. Similar brainstem circuits for vocalization are found in all vocalizing vertebrates examined thus far, ranging from vocalizing fish and frogs to humans (Bass et al., 2008; Brahic and Kelley, 2003; Jürgens, 2009; Wild et al., 1997). In the basal condition, patterns of vocal behavior are generated entirely by this conserved brainstem circuitry and are generated under specific hormonal, emotional, or motivational states conveyed via projections from the hypothalamus and/or amygdala (Cohen and Cheng, 1982; Jürgens, 2009; Sisneros et al., 2004).

In species that display more complex temporal vocal coordination, higher forebrain structures project onto these hindbrain circuits to synchronize vocal output with greater facility. As a zebra

finch sings, neuronal activity in the primary motor output to the brainstem (nucleus RA) is tightly time-locked to specific song motifs (Leonardo and Fee, 2005). The temporal encoding of this ensemble activity derives from descending projections from premotor nucleus HVC (Hahnloser et al., 2002; Nottebohm et al., 1982). The role of HVC in temporal control over song has been demonstrated by cooling experiments, in which focal cooling of HVC leads to an overall slowing of song production, further implicating this nucleus in the temporal aspects of learned song (Long and Fee, 2008). Further, recordings in HVC of wild songbirds demonstrate that HVC encodes the temporal pattern not only for their own part in a coordinated duet but for that of their partner's as well (Fortune et al., 2011; Hoffmann et al., 2019).

Cortical structures have been specifically implicated in the temporal coordination of vocal output in humans. As in the zebra finch, in humans focal cooling of speech-associated premotor Broca's area leads to a specific slowdown in overall speech timing without significant loss in speech quality, while cooling of speech-associated ventral motor cortex results in a breakdown in the spectral quality of speech, with little effect on timing (Long et al., 2016). In the case of human speech, strokes in various speech areas can result in "acquired arrhythmia," the selective disruption of the ability to reproduce a rhythm or entrain to a beat, even in cases where rhythmic perception and melodic production are unaffected (Fries and Swihart, 1990; Mavlov, 1980). Electrophysiological recordings during speech indicate that a high degree of temporal coordination is present within the laryngeal/orofacial sensorimotor cortex (Bouchard et al., 2013; Dichter et al., 2018). Although no single higher-order motor area for temporal processing has been definitively established, available evidence suggests critical involvement of Broca's area and the supplemental motor area (SMA), with additional roles for the basal ganglia and cerebellum (Kotz and Schwartze, 2010, 2011).

Taken together, these examples suggest that the vocal coordination module comprises a set of vocal behaviors, including jamming avoidance, vocal turn-taking, and complex duetting. Vocal coordination may involve varying degrees of learning, which are not imitative per se, but which demonstrate clear mechanistic overlaps with imitative vocal learning: both are guided by feedback, both can be acquired during a sensitive period of development, and both may be controlled by overlapping brain centers.

Module 2: Vocal Production Variability and Feedback

The second module relates to how a given species dynamically changes the acoustic variability of its vocal output, either during critical periods of development or in adulthood (Doupe and Kuhl, 1999; Tchernichovski et al., 2001; Tramontin and Brenowitz, 2000). During development, vocal patterns may become more structured and less variable over time, as in the case with other motor behaviors. In adults, variability in vocal output might change from moment to moment with behavioral state. At both timescales, feedback may play a critical role in shaping vocal behavior. Here too, we suggest that expanding investigation to a wider diversity of species beyond the canonical vocal learners may provide valuable insight.

Canonical vocal learning species begin their vocal development with a highly variable phase (e.g., vocal overproduction

and babbling), which has been studied extensively in human infants and songbirds (Doupe and Kuhl, 1999) and more recently in bats and marmosets (Elowson et al., 1998; Knörnschild et al., 2006; Pistorio et al., 2006). One role of babbling is to promote vocal exploration, which allows the animal to learn how to use its vocal apparatus to produce a variety of vocal sounds. In vocal learners, learning to "control the instrument" is thought to be an early stage in the process of matching its vocal production to an external model (as in song learning). The later consolidation of vocal sounds is guided by auditory and social feedback. Feedback can shape vocalizations at multiple levels: most essential is the role of auditory feedback from the animal's own vocalizations (Fehér et al., 2009; Konishi, 1965). Auditory feedback allows for self-evaluation of production errors (Gadagkar et al., 2016) and may be influenced by an external auditory input, sometimes referred to as a sensory template (Margoliash, 2002). Finally, vocal production variability may be guided by higher-level social selective pressures. For example, social feedback from a parent, or from an adult conspecific, may reinforce the production of a call by its young in the correct context (Figure 3). It appears, however, that changes in vocal variability, either during development or from moment to moment, are not unique to vocal learning species, as discussed below.

Vocal Variability with and without Feedback

In its most rudimentary form, vocal development involves an initial phase of juvenile vocal variability, which largely disappears by adulthood without an apparent influence of any auditory or social feedback (Egnor and Hauser, 2004; Janik and Slater, 1997; Kroodsma, 1982). However, the ability to dynamically control variability during vocal production is not limited to vocal learners, having been observed in several taxa traditionally considered vocal non-learners (Elowson et al., 1998; Pistorio et al., 2006). For example, stages of juvenile vocal variability resembling the dynamics by which subsong utterances are transformed into adult syllables and motifs in songbirds have been observed in quail (Derégnaucourt et al., 2009; Tchernichovski et al., 2001). In juvenile Japanese quail, the precise spectral and temporal trajectory by which noisy initial calls develop into adult crows varies greatly between individuals, and yet the outcome of vocal development is highly similar across individuals. Although we do not understand why juvenile vocal variability exists in some vocal non-learners, it illustrates how a trait may be associated with imitative vocal learning but be insufficient on its own to drive it. Importantly, it further provides an opportunity for studying the transition from vocal variability to consolidation of structure in animal models in which external feedback does not influence vocal development.

In other species, external sources do influence vocal variability but do not influence the vocalizations that ultimately develop. For example, parental interactions influence vocal production variability in juvenile marmosets, who exhibit a phase of highly variable output over the course of vocal development (Elowson et al., 1998; Pistorio et al., 2006). However, although these parental interactions correlate with the developmental rate of calls, they do not influence the final outcome of vocal development, i.e., the acoustic structure of the adult vocalizations (Gultekin and Hage, 2017).
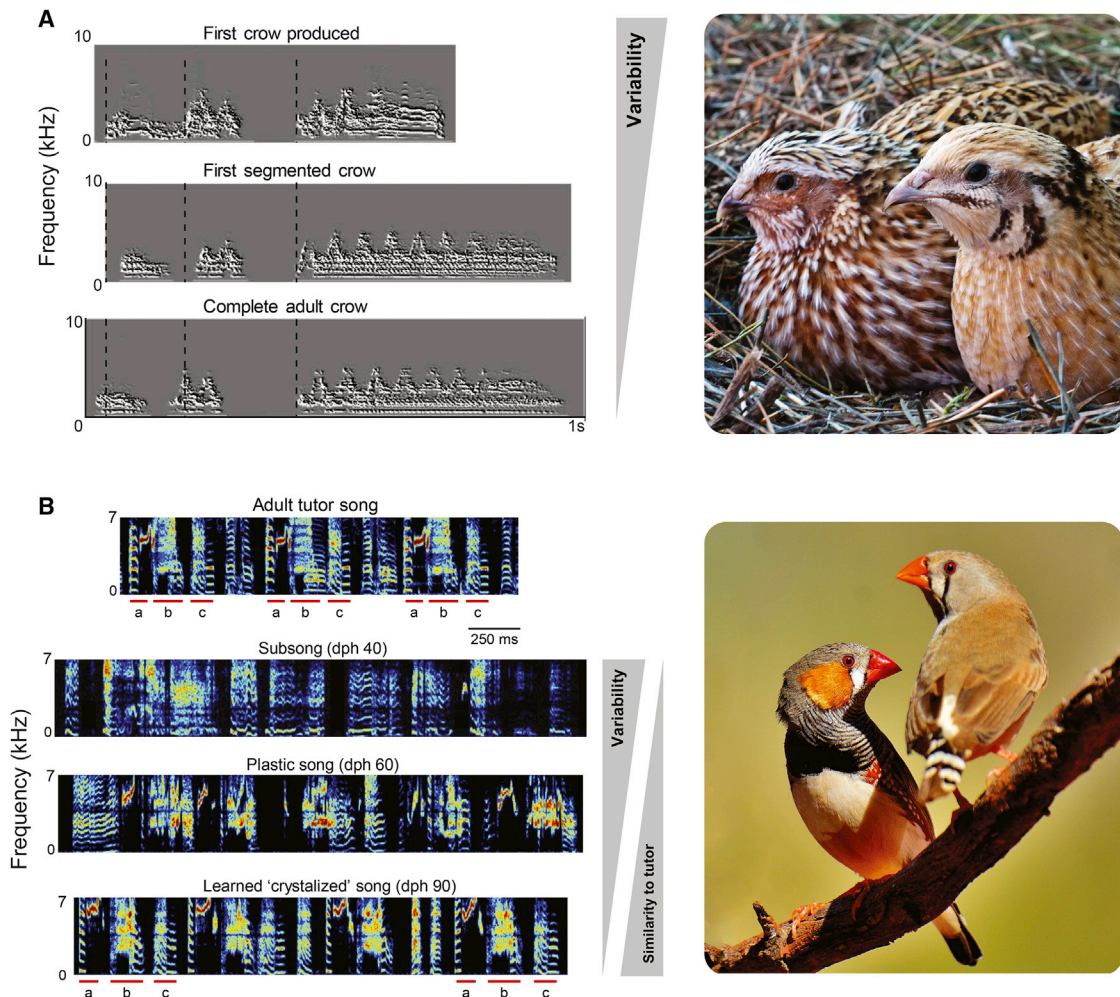
Figure 3. Vocal Production Variability and Feedback
During the plastic period of vocal development, diverse sources contribute to variability in vocal output and the neural motor signal that underlies it.
(A) Juvenile Japanese quail exhibit an initial period of vocal variability that wanes over the course of the development of their adult crow (adapted from Derégnaucourt et al., 2009, photo credit Roger McLaughlin).
(B) Juvenile zebra finches similarly experience an initial period of high vocal variability (subsong), which gradually (via plastic song) crystallizes into the adult song type. In both species, vocal variability decreases over development, but only in the finch does this decrease in vocal variability coincide with increasing similarity to the song of the tutor, a process that requires auditory feedback (adapted from Fee and Goldberg, 2011, photo credit Laurie Boyle). In some respects, vocal learning can thus be defined as the process by which internally guided developmental processes (as in the quail) become flexible to external influences (as in the finch). Abbreviations: dph, days post-hatch.

Among the simplest forms in which feedback adaptively influences vocal behavior is in the modulation of juvenile begging calls. In many birds and mammals, infants engage in vocal begging to solicit parental food provisioning. Although begging calls are innate, competing nestlings and littermates alter the variability of their calls, and at least in some cases, this variability is guided by social and auditory feedback. This includes canonical vocal learning taxa such as songbirds (David, 1988; Ligout et al., 2016; Price et al., 1996; Tuero et al., 2015) and parrots (Stamps et al., 1989) but also taxa with little or no evidence for vocal learning, including cuckoos (Davies et al., 1998), pigeons (Mondloch, 1995), pigs (Iacobucci et al., 2015; Weary and Fraser, 1995), meerkats (Manser and Avey, 2000), and pygmy marmosets (Elowson et al., 1998). In songbirds, nestlings are able to modify their begging calls to more closely match the acoustic structure of their parents' song, the capacity for which is reduced with lesions to the developing song system, suggesting that variable nest begging behavior could represent a prerequisite for the evolution of song learning (Liu et al., 2009).

The influence of auditory feedback on vocal variability extends across varying timescales and stages of development: whereas "closed-ended" learners undergo only a single sensitive period of exploratory vocal variability in their lifetimes, so-called "open-ended" learners maintain the ability to re-initiate the process in adulthood (Beecher and Brenowitz, 2005; Brenowitz and Beecher, 2005). This maintenance or reopening of vocal plasticity does not necessarily occur as a result of social learning, but rather in some cases is a result of changes in neuroendocrine
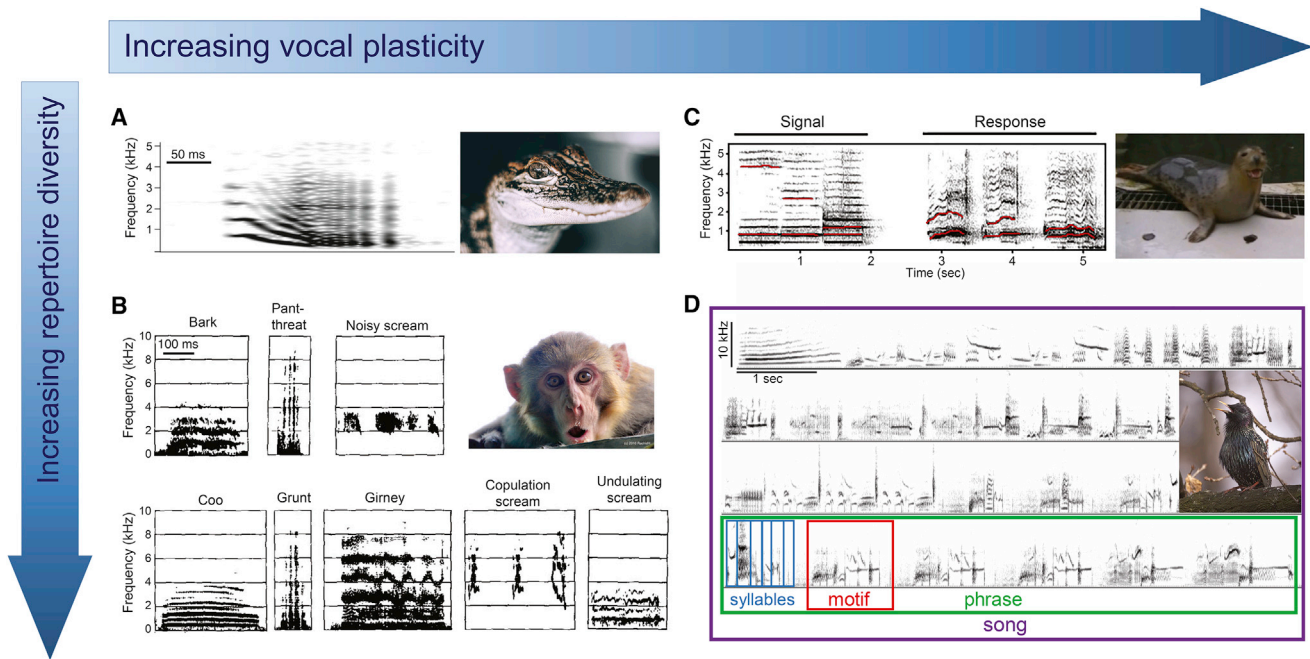
Figure 4. Vocal Versatility
This complex module includes two features: vocal diversity (vertical axis) and vocal plasticity (horizontal axis). Diversity of vocal repertoire is low in vocalizing fish and crocodilians (A, adapted from Riede et al., 2011, photo credit Jackson Jost), with a small set of calls produced under a limited set of circumstances. Other species may present a more diverse acoustic repertoire with a range of categorically discrete call types, as in the case of the macaque monkey (B, adapted from Hauser et al., 1993, photo credit Rachid Homsany). Vocal plasticity may involve small modifications to pre-existing vocalizations to match a template, as in the case of a harbor seal trained to imitate a sequence of human vowel formants (C, adapted and photo credit from Stansbury and Janik, 2019). In its most advanced instantiation, species may present the ability to mimic other species, novel sounds, and complex series of syllables, motifs, and phrases, as in the common starling (D, adapted from Ball et al., 2006, photo credit Paweł Więcek).

factors, as in the case of the seasonal singing of many species of songbirds and vocalizing fish (Forlano et al., 2015; Tramontin and Brenowitz, 2000). Even in species that exhibit only a single critical period for vocal plasticity, experimental disruption of auditory feedback leads to a dynamic increase in vocal variability of a previously fixed vocalization, revealing an ongoing role for this module in the experience-dependent maintenance of adult vocalizations (Tschida and Mooney, 2012). At one extreme end of this module one might include species like some thrushes and nightingales that develop and maintain a high degree of vocal variability in their repertoire by incorporating new elements that do not appear to be guided by an external source, a process often referred to as innovation or improvisation (Hughes et al., 2002; Johnson, 2006).

Neural Mechanisms of Vocal Variability
Interestingly, exploratory variability is not a passive feature of neural systems, but rather one that is dynamically regulated to facilitate motor learning (Tchernichovski et al., 2001; Wu et al., 2014). During juvenile song learning in zebra finches, the forebrain song nucleus LMAN—a component of the anterior forebrain loop required for song learning (Ölveczky et al., 2005)—acts as a driver of vocal behavioral variability by injecting variable, pseudo-random patterns of neuronal spiking into the premotor output nucleus RA (Kao et al., 2005). Over the course of learning, this variable output is refined by the influence of auditory feedback-based error correction, which ultimately enables the consolidation of learned adult song (Andalman and Fee,

2009; Gadagkar et al., 2016; Tschida and Mooney, 2012; Warren et al., 2011). In mammals, similar cortico-striatal-thalamocortical loops enable flexible motor behavior learning (Dhawale et al., 2017; Kreitzer and Malenka, 2008; Wise et al., 1996) and could conceivably play a similar role in the context of vocal production variability.

In sum, vocal production variability may change developmentally, with or without feedback, and across varying life stages. In songbirds, vocal variability is supported by a distinct neurobiological substrate. This observation motivates a comparative approach to identify the diverse mechanisms underlying vocal variability across species.

Module 3: Vocal Versatility
Lastly, we propose a two-axis module which we term vocal versatility. This module is defined as the combination of two related traits: vocal diversity, which is the repertoire size of vocal sounds a species is capable of producing, and vocal plasticity, the degree to which this repertoire can be modified with experience (Figure 4). Vocal versatility thus ranges from species that produce only a few innate calls, such as humming fish or crocodilians, to species with advanced interspecific mimicry and the ability to imitate novel sounds not part of the species' natural repertoire, as exemplified by many songbirds and parrots. Likewise, this two-dimensional framework also includes species with diverse vocal repertoires that are largely innate, such as the domesticated chicken and macaque monkey, as well as species

with fairly limited natural diversity of vocal sounds that are nevertheless amenable to learned modification, such as harbor seals. In our description of this module, we discuss the relationship between the peripheral structures for sound production and the neural circuits that control them, as well as how the potential co-evolution of these two processes could subserve the evolution of vocal versatility.

Peripheral Components of Vocal Versatility: Source and Filter

Broadly across vocalizing vertebrates, a primary vocal organ (source) provides the soundwaves that serve as the basis for vocal output (the larynx in frogs and most mammals, the phonic lips in dolphins, or syrinx in birds). As the soundwaves propagate through the vocal tract (the trachea, air sacs, pharynx, and/or nasal cavity), they are "filtered" by a set of supralaryngeal structures (e.g., tongue, lips, jaw, palate, beak, trunk) that can dynamically change the shape of the vocal tract and its resonance frequencies. The source-filter theory distinguishes these physical structures based on their role in shaping vocal output as well as by their anatomical separation (Fant, 1960; Taylor and Reby, 2010). Importantly, it provides a helpful framework for conceptualizing how different species may employ highly different peripheral mechanisms for influencing vocal versatility. For example, songbirds are specialized for controlling the source (i.e., their syringes, but also parts of the filter system; see Elie and Theunissen, 2016). In humans, acoustic output is extensively shaped by filtering (e.g., using our lips and tongue). Moreover, alternative source and filter mechanisms can be combined to achieve similar ends. For instance, one elephant has been observed to imitate some human speech sounds by mechanically manipulating his oral cavity using his trunk, demonstrating a highly divergent example of spectral filtering (Stoeger et al., 2012). Despite this diversity in vocal production mechanisms, there are common threads across species, particularly in the neuronal mechanisms of vocal plasticity. We organize our description of this module to focus separately on the source and filter structures and their distinct neurobiological mechanisms in each section.

Source: Evolution of Structure Complexity and Forebrain Control

At the source level, several species have evolved unique adaptations in their primary vocal organ to enable more flexible and elaborate forms of vocal output. The syringeal anatomy of parrots and songbirds is more complex compared to non-learning birds such as ducks and ibises (Ames, 1971; Warner, 1972). Songbirds have further evolved a "superfast" syringeal musculature as well as the ability to independently control sound production in each of the two sides of the syrinx. This in turn allows for the production of more complex vocalizations, with the degree of anatomical complexity possibly also correlating with the species' capacity for vocal imitation (Ames, 1971; Suthers, 1990; Uchida et al., 2010; Warner, 1972; Zollinger and Suthers, 2004). To a lesser degree, some diversity in source complexity is also observed across mammals. For example, the larynges of some bats have also been found to contain superfast muscles (Elemans et al., 2011). However, in most other vocal production learning mammals, there is little evidence for increasing anatomical complexity at the source level. Yet, many animals have found creative alternate means of achieving vocal diversity by utilizing

different anatomical properties of the source, including nonlinear acoustic processes. Such is the case for the vocalizations of African wild dogs and pant-hoots of chimpanzees, where harmonic calls are transformed into a graded series of increasingly complex and noisy calls through relatively simple additive application of respiratory pressure (Riede et al., 2004; Wilden et al., 1998).

However, changes in the anatomy of the vocal organ alone are unlikely to be sufficient to enable more advanced forms of vocal versatility. Recent support for this notion comes from anatomical studies in non-human primates. While in the past it has been claimed that the human "descended larynx" is uniquely evolved to enable complex speech (Lieberman, 1984), more recent work has disputed this claim, suggesting that even the vocal tract of macaques may be "speech-ready," missing only the necessary neural structures for learning rather than the physiological capacity for more complex vocal production (Fitch et al., 2016). Thus, in humans as in other vocal learners, it is believed that at the level of the central nervous system, higher brain structures must gain control over the brainstem, resulting in increased source control, to facilitate higher vocal versatility.

Experimental work in humans and songbirds has illuminated many of the neurobiological mechanisms underlying forebrain control over vocal versatility (Dichter et al., 2018; Simonyan and Horwitz, 2011; Sober et al., 2008). Nevertheless, further work is needed to elucidate parallels and points of divergence between findings in avian and mammalian systems. One potential approach is to search for shared mechanisms underlying the capacity for modifying vocal pitch, a trait shared by several avian and mammalian species including songbirds (Sober and Brainard, 2012), parrots (Osmanski and Dooling, 2009), cetaceans (Moore and Pawloski, 1990), pinnipeds (Stansbury and Janik, 2019), bats (Genzel et al., 2019; Prat et al., 2015, 2017), and New World primates (Hotchkin et al., 2015; Zhao et al., 2019). Identifying common neural mechanisms that facilitate this capacity would first require detailed delineation of forebrain structures that enable the adaptive modification of pitch. However, this challenge is becoming more tractable with novel genetic and anatomical tracing approaches for identifying homologous brain structures and cell types across taxa (Briscoe and Ragsdale, 2018; Pfenning et al., 2014).

Filter: Evolution of Structure Complexity and Forebrain Control

While major research focus has been devoted to laryngeal and syringeal control, we know relatively little about the diverse mechanisms mediating articulatory filtering in vocalizing animals. Considering the crucial role of supralaryngeal structures (such as the tongue) in enabling spectral shaping during vocal production, this presents a major gap in our knowledge. Among primates, direct motor cortex projections to the hypoglossal nucleus (which innervates the tongue) appear to be absent in saddle-back tamarin, be present to a degree in squirrel monkey and rhesus macaque, and be very dense in humans (Jürgens and Alipour, 2002). This increase in density also correlates with an overall increase in the degree of tongue involvement during vocalizations in these species (Jürgens and Alipour, 2002). Although the tongues of our closest relatives play a relatively minor role in their vocal versatility, parrots display highly dynamic control over their tongue, which plays a prominent role in the
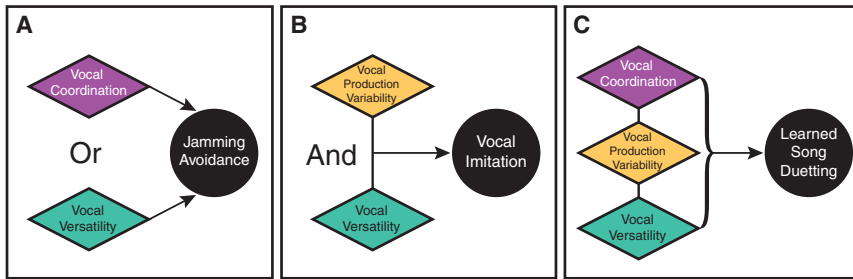
Figure 5. The Modular Framework for Vocal Learning
Three hypothetical scenarios demonstrate how modules may be combined to produce differing behavioral outcomes.
(A) Acoustic jamming avoidance can be achieved either by synchronizing call timing (vocal coordination) or by adaptive pitch shifting (vocal versatility).
(B) Vocal imitation requires both production variability and vocal versatility.
(C) Learned song duetting requires vocal coordination in addition to vocal imitation. Although modeling the full complexity of the naturalistic behaviors these hypothetical scenarios represent would require far more nuanced models, this provides a template for how a modular framework can be used to build and test hypotheses for how more general behavioral modules may be shared and recombined to produce a diversity of behavioral outcomes.

production of their extensive natural vocal repertoire as well as their expert imitative abilities (Ohms et al., 2012; Patterson and Pepperberg, 1994). Although parrots have been shown to have more extensive forebrain circuitry for learned vocal behavior than songbirds (Chakraborty et al., 2015), the extent to which this circuitry plays a role in the control of vocal articulatory structures beyond the syrinx remains unexplored. Among bats, the Egyptian fruit bat (*Rousettus aegyptiacus*) similarly displays an exceptional degree of control over the tongue: whereas most bats' echolocation calls are produced from the larynx, in *R. aegyptiacus* they are exclusively tongue based (Lee et al., 2017; Yovel et al., 2011). These bats exhibit a capacity for long-term and persistent vocal plasticity of their social communication calls across multiple spectral parameters (Genzel et al., 2019; Prat et al., 2015, 2017), further supporting a potential link between the degree of control over the articulatory filtering structures and the capacity for vocal plasticity.

Coordination of Source and Filter

Many animal vocalizations involve some degree of coordination between various source and articulatory filtering mechanisms (as well as respiratory output), which is reflected in their increasingly integrated representation at higher levels of vocal control circuits (Jürgens, 2002). However, as the precise mechanisms for the neural control of source and filter structures are seldom examined in combination, the possible mechanisms for integrating and controlling them remain largely unexplored. In humans, electrophysiological recordings during speech production reveal that the larynx and articulatory structures each have distinct somatotopic representations and temporally coordinated activity in the sensorimotor cortex (Bouchard et al., 2013). Based on the vocal production learner convergences observed for laryngeal/syringeal control, it is likely that other vocal learners could possess similar shared neurobiological specializations for the coordinated cortical control of the articulatory structures that enable vocal versatility. We suggest that comparative work across a diverse set of species to address this possibility represents a rich domain for future neurobiological inquiry.

Integration of Modules

We suggest that understanding the neurobiological underpinnings and evolution of vocal learning will require identifying the full set of modules that are brought together to subserve the emergence of the more complex vocal learning phenotype.

However, some discussion of how modules may be considered both as discrete as well as interdependent components of the broader behavioral phenotype is warranted.

Here, the reader may consider the analogy of a baker preparing a cake, requiring some combination of flour, eggs, butter, sugar, and so forth. Just as each one of these ingredients may be involved in many other recipes, so may a module involving motor variability and feedback be involved in both song learning and learning of a novel manual dexterity task (Wu et al., 2014). Likewise, the baker may be a master or amateur at combining these ingredients, just as increasing neurobiological integration of these modules may be evolutionarily necessary to produce the most accomplished, canonical examples of vocal learning. The evolution of a novel behavior need not necessarily involve the evolution of an entire system *de novo*. Rather, pre-existing modules—along with the neural circuits, cell types, and/or molecular pathways that subserve them—may be repurposed in combination with other modules to make something new.

As a demonstration of the modular framework, we highlight the previously discussed example of acoustic jamming avoidance during echolocation in bats foraging simultaneously (Figure 2B). While this function could be achieved through purely temporal synchronization (vocal coordination module), it may also be accomplished by shifting the peak frequency of echolocation calls to avoid interference with conspecifics (vocal versatility module) (Ratcliffe et al., 2004; Ulanovsky et al., 2004) (Figure 5A). Alternatively, other behavioral functions might require combination of modules. For instance, vocal imitation requires vocal versatility, but also the ability to shift from high to low vocal variability over the course of learning (Fee and Goldberg, 2011) (Figure 5B). More complex forms of vocal learning may require greater integration of larger sets of modules: for instance, the virtuosic, four-part synchronized chorusing of plain-tailed wrens (*Thryothorus euophrys*) (Mann et al., 2006), as in medieval polyphony or Baroque counterpoint in humans, requires vocal variability and feedback, vocal versatility, and temporal coordination between singers for the behavior to be successfully executed.

The open framework of our hypothesis, an essential property of modular systems, acknowledges that at present we know very little about the capacity for vocal learning of the vast majority of extant animal species. We expect, and indeed encourage, that further modules will be proposed as additional behavioral, neurobiological, molecular, and genomic data are collected,

and that existing modules may be disputed, discretized, or synthesized as new evidence allows. Thus, the modular framework does not represent an attempt to redefine or classify distinct categories of vocal learning—such helpful frameworks have already been provided (Beecher and Brenowitz, 2005; Brenowitz and Beecher, 2005; Janik and Slater, 2000). Rather, we present a complementary perspective wherein vocal learning may be considered as a unique combination of sub-traits, which may each individually be learned or non-learned, and which may individually be involved in many other behavioral processes as well. This allows us to investigate different mechanistic aspects of vocal learning in a broad context, across different species, different behaviors, and different biological functions.

Conclusions

We formulate a framework for the study of vocal learning, in which the behavior is proposed to be a complex trait comprised of distinct but interdependent subcomponents, which we term modules. Within this framework, we suggest three potential modules: (1) vocal coordination, (2) vocal production variability and feedback, and (3) vocal versatility. We propose that incorporating species along the different axes of distinct modules will facilitate a better mechanistic and evolutionary understanding of the building blocks which together comprise the vocal learning trait. The three modules discussed here benefit from positioning species along these individual axes, starting from the origin (or most rudimentary level) and advancing toward those champion species that best exemplify each module. The broadly comparative work that will be necessary to assess species across these modules is greatly facilitated by rapid advancement in modern genomic tools and computational approaches (Berger et al., 2018; Bernstein et al., 2010; Buenrostro et al., 2015; Koepfli et al., 2015; Kowalczyk et al., 2019; Teeling et al., 2018; Yue et al., 2014; Zhu et al., 2018) that can enable linking complex traits like vocal learning to their underlying neurobiological processes. Ultimately, we believe modular and multidimensional frameworks will allow for more efficient comparison of traits between human and animal models, as we may ask not which species represents the singular best model for vocal learning, but rather which particular aspects of vocal learning can best be studied across a diverse plurality of model systems.

REFERENCES

Ames, P.L. (1971). The morphology of the syrinx in passerine birds, *Volume 37* (CT: Peabody Museum of Natural History, Yale University New Haven).

Andalman, A.S., and Fee, M.S. (2009). A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. Proc. Natl. Acad. Sci. USA *106*, 12518–12523.

Arriaga, G., and Jarvis, E.D. (2013). Mouse vocal communication system: are ultrasounds learned or innate? Brain Lang. *124*, 96–116.

Arriaga, G., Zhou, E.P., and Jarvis, E.D. (2012). Of mice, birds, and men: the mouse ultrasonic song system has some features similar to humans and song-learning birds. PLoS ONE *7*, e46610.

Awbrey, F.T. (1978). Social Interaction among Chorusing Pacific Tree Frogs, *Hyla regilla*. Copeia *1978*, 208–214.

Ball, G., Sockman, K., Duffy, D., and Gentner, T. (2006). A neuroethological approach to song behavior and perception in European starlings: interrelationships among testosterone, neuroanatomy, immediate early gene expression, and immune function. Adv. Stud. Behav. *36*, 59–122.

Barton, R.A., and Harvey, P.H. (2000). Mosaic evolution of brain structure in mammals. Nature *405*, 1055–1058.

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.

Beecher, M.D., and Brenowitz, E.A. (2005). Functional aspects of song learning in songbirds. Trends Ecol. Evol. *20*, 143–149.

Bell, C.C., and Maler, L. (2005). Central Neuroanatomy of Electrosensory Systems in Fish. In Electroreception, T.H. Bullock, C.D. Hopkins, A.N. Popper, and R.R. Fay, eds. (New York, NY: Springer New York), pp. 68–111.

Benichov, J.I., Benezra, S.E., Vallentin, D., Globerson, E., Long, M.A., and Tchernichovski, O. (2016a). The Forebrain Song System Mediates Predictive Call Timing in Female and Male Zebra Finches. Curr. Biol. *26*, 309–318.

Benichov, J.I., Globerson, E., and Tchernichovski, O. (2016b). Finding the Beat: From Socially Coordinated Vocalizations in Songbirds to Rhythmic Entrainment in Humans. Front. Hum. Neurosci. *10*, 255.

Berger, M.J., Wenger, A.M., Guturu, H., and Bejerano, G. (2018). Independent erosion of conserved transcription factor binding sites points to shared hindlimb, vision and external testes loss in different mammals. Nucleic Acids Res. *46*, 9299–9308.

Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R., et al. (2010). The NIH Roadmap Epigenomics Mapping Consortium. Nat. Biotechnol. *28*, 1045–1048.

Blumberg, M.S. (1992). Rodent ultrasonic short calls: locomotion, biomechanics, and communication. J. Comp. Psychol. *106*, 360–365.

Bouchard, K.E., Mesgarani, N., Johnson, K., and Chang, E.F. (2013). Functional organization of human sensorimotor cortex for speech articulation. Nature *495*, 327–332.

Brahic, C.J., and Kelley, D.B. (2003). Vocal circuitry in *Xenopus laevis*: telencephalon to laryngeal motor neurons. J. Comp. Neurol. *464*, 115–130.

Brenowitz, E.A., and Beecher, M.D. (2005). Song learning in birds: diversity and plasticity, opportunities and challenges. Trends Neurosci. *28*, 127–132.

Briscoe, S.D., and Ragsdale, C.W. (2018). Homology, neocortex, and the evolution of developmental mechanisms. Science *362*, 190–193.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol. *109*, 1–9.

Campbell, P., Pasch, B., Warren, A.L., and Phelps, S.M. (2014). Vocal ontogeny in neotropical singing mice (*Scotinomys*). PLoS ONE *9*, e113628.

Chakraborty, M., Walløe, S., Nedergaard, S., Fridel, E.E., Dabelsteen, T., Pakkenberg, B., Bertelsen, M.F., Dorrestein, G.M., Brauth, S.E., Durand, S.E., and

Jarvis, E.D. (2015). Core and Shell Song Systems Unique to the Parrot Brain. PLoS ONE *10*, e0118496.

Chow, C.P., Mitchell, J.F., and Miller, C.T. (2015). Vocal turn-taking in a non-human primate is learned during ontogeny. Proc. Biol. Sci. *282*, 20150069.

Clarke, E., Reichard, U.H., and Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. PLoS ONE *1*, e73.

Cohen, J., and Cheng, M.F. (1982). Effects of testosterone metabolites and estrogen in the midbrain control of courtship behavior in the male ring dove (*Streptopelia risoria*). Neuroendocrinology *34*, 64–74.

David, J.T.H. (1988). Supply and Demand in Tree Swallow Broods: A Model of Parent-Offspring Food-Provisioning Interactions in Birds. Am. Nat. *131*, 175–202.

Davies, N.B., Kilner, R.M., and Noble, D.G. (1998). Nestling cuckoos, *Cuculus canorus*, exploit hosts with begging calls that mimic a brood. Proc. R. Soc. Lond. B Biol. Sci. *265*, 673–678.

Derégnaucourt, S., Saar, S., and Gahr, M. (2009). Dynamics of crowing development in the domestic Japanese quail (*Coturnix coturnix japonica*). Proc. Biol. Sci. *276*, 2153–2162.

Dhawale, A.K., Smith, M.A., and Ölveczky, B.P. (2017). The Role of Variability in Motor Learning. Annu. Rev. Neurosci. *40*, 479–498.

Dichter, B.K., Breshears, J.D., Leonard, M.K., and Chang, E.F. (2018). The Control of Vocal Pitch in Human Laryngeal Motor Cortex. Cell *174*, 21–31.e9.

Diniz, P., da Silva Júnior, E.F., Webster, M.S., and Macedo, R.H. (2018). Duetting behavior in a Neotropical ovenbird: sexual and seasonal variation and adaptive signaling functions. J. Avian Biol. *49*, jav-01637.

Doupe, A.J., and Kuhl, P.K. (1999). Birdsong and human speech: common themes and mechanisms. Annu. Rev. Neurosci. *22*, 567–631.

Egnor, S.E.R., and Hauser, M.D. (2004). A paradox in the evolution of primate vocal learning. Trends Neurosci. *27*, 649–654.

Elemans, C.P.H., Mead, A.F., Jakobsen, L., and Ratcliffe, J.M. (2011). Super-fast muscles set maximum call rate in echolocating bats. Science *333*, 1885–1888.

Elie, J.E., and Theunissen, F.E. (2016). The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. Animal Cognition *19*, 285–315.

Elliott, H.H. (1981). Video Analysis of Siamang (*Hylobates syndactylus*) Songs. Behaviour *76*, 128–151.

Elowson, A.M., Snowdon, C.T., and Lazaro-Perea, C. (1998). Infant 'Babbling' in a Nonhuman Primate: Complex Vocal Sequences with Repeated Call Types. Behaviour *135*, 643–664.

Fant, G. (1960). Acoustic theory of speech production (The Hague: Mouton).

Farabaugh, S. (1982). The ecological and social significance of duetting. In Acoustic Communication in Birds: Song Learning & Its Consequences, D.E. Kroodsma and E.H. Miller, eds. (New York: Academic Press), pp. 85–124.

Fee, M.S., and Goldberg, J.H. (2011). A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. Neuroscience *198*, 152–170.

Fehér, O., Wang, H., Saar, S., Mitra, P.P., and Tchernichovski, O. (2009). *De novo* establishment of wild-type song culture in the zebra finch. Nature *459*, 564–568.

Fischer, J. (2017). Primate vocal production and the riddle of language evolution. Psychon. Bull. Rev. *24*, 72–78.

Fitch, W.T., de Boer, B., Mathur, N., and Ghazanfar, A.A. (2016). Monkey vocal tracts are speech-ready. Sci. Adv. *2*, e1600723.

Forlano, P.M., Sisneros, J.A., Rohmann, K.N., and Bass, A.H. (2015). Neuroendocrine control of seasonal plasticity in the auditory and vocal systems of fish. Front. Neuroendocrinol. *37*, 129–145.

Fortune, E.S., Rodríguez, C., Li, D., Ball, G.F., and Coleman, M.J. (2011). Neural mechanisms for the coordination of duet singing in wrens. Science *334*, 666–670.

Fries, W., and Swihart, A.A. (1990). Disturbance of rhythm sense following right hemisphere damage. Neuropsychologia *28*, 1317–1323.

Gadagkar, V., Puzerey, P.A., Chen, R., Baird-Daniel, E., Farhang, A.R., and Goldberg, J.H. (2016). Dopamine neurons encode performance error in singing birds. Science *354*, 1278–1282.

Genzel, D., Desai, J., Paras, E., and Yartsev, M.M. (2019). Long-term and persistent vocal plasticity in adult bats. Nat. Commun. *10*, 3372.

Ghazanfar, A.A., and Takahashi, D.Y. (2014). The evolution of speech: vision, rhythm, cooperation. Trends Cogn. Sci. *18*, 543–553.

Grafe, T.U. (1996). The function of call alternation in the African reed frog (*Hyperolius marmoratus*): precise call timing prevents auditory masking. Behav. Ecol. Sociobiol. *38*, 149–158.

Gultekin, Y.B., and Hage, S.R. (2017). Limiting parental feedback disrupts vocal development in marmoset monkeys. Nat. Commun. *8*, 14046.

Hahnloser, R.H.R., Kozhevnikov, A.A., and Fee, M.S. (2002). An ultra-sparse code underlies the generation of neural sequences in a songbird. Nature *419*, 65–70.

Hall, M.L. (2009). A Review of Vocal Duetting in Birds. In Advances in the Study of Behavior, M. Naguib, K. Zuberbuehler, N.S. Clayton, and V.M. Janik, eds. (Academic Press), pp. 67–121.

Hammerschmidt, K., Reisinger, E., Westekemper, K., Ehrenreich, L., Strenzke, N., and Fischer, J. (2012). Mice do not require auditory input for the normal development of their ultrasonic vocalizations. BMC Neurosci. *13*, 40.

Hammerschmidt, K., Whelan, G., Eichele, G., and Fischer, J. (2015). Mice lacking the cerebral cortex develop normal song: insights into the foundations of vocal learning. Sci. Rep. *5*, 8808.

Hauser, M.D., Evans, C.S., and Marler, P. (1993). The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. Anim. Behav. *45*, 423–433.

Heiligenberg, W. (1977). Principles of electrolocation and jamming avoidance in electric fish: a neuroethological approach (New York: Springer-Verlag).

Hoffmann, S., Trost, L., Voigt, C., Leitner, S., Lemazina, A., Sagunsky, H., Abels, M., Kollmansperger, S., Maat, A.T., and Gahr, M. (2019). Duets recorded in the wild reveal that interindividually coordinated motor control enables cooperative behavior. Nat. Commun. *10*, 2577.

Hotchkin, C.F., Parks, S.E., and Weiss, D.J. (2015). Noise-Induced Frequency Modifications of Tamarin Vocalizations: Implications for Noise Compensation in Nonhuman Primates. PLoS ONE *10*, e0130211.

Hughes, M., Hultsch, H., and Todt, D. (2002). Imitation and Invention in Song Learning in Nightingales (*Luscinia megarhynchos* B., Turdidae). Ethology *108*, 97–113.

Iacobucci, P., Colonnello, V., D'Antuono, L., Cloutier, S., and Newberry, R.C. (2015). Piglets call for maternal attention: Vocal behaviour in *Sus scrofa domesticus* is modulated by mother's proximity. Appl. Anim. Behav. Sci. *171*, 88–93.

Janik, V.M., and Slater, P.J.B. (1997). Vocal Learning in Mammals. Adv. Stud. Behav. *26*, 59–100.

Janik, V.M., and Slater, P.J.B. (2000). The different roles of social learning in vocal communication. Anim. Behav. *60*, 1–11.

Jarvis, E.D. (2004). Learned birdsong and the neurobiology of human language. Ann. N Y Acad. Sci. *1016*, 749–777.

Johnson, S.L. (2006). Do American robins acquire songs by both imitating and inventing? Wilson J. Ornithol. *118*, 341–352.

Jürgens, U. (2002). Neural pathways underlying vocal control. Neurosci. Biobehav. Rev. *26*, 235–258.

Jürgens, U. (2009). The neural control of vocalization in mammals: a review. J. Voice *23*, 1–10.

Jürgens, U., and Alipour, M. (2002). A comparative study on the cortico-hypoglossal connections in primates, using biotin dextranamine. Neurosci. Lett. *328*, 245–248.

Kao, M.H., Doupe, A.J., and Brainard, M.S. (2005). Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. Nature *433*, 638–643.

Kershenbaum, A., Ilany, A., Blaustein, L., and Geffen, E. (2012). Syntactic structure and geographical dialects in the songs of male rock hyraxes. Proc. Biol. Sci. *279*, 2974–2981.

Knörnschild, M., Behr, O., and von Helversen, O. (2006). Babbling behavior in the sac-winged bat (*Saccopteryx bilineata*). Naturwissenschaften *93*, 451–454.

Koepfli, K.-P., Paten, B., and O'Brien, S.J.; Genome 10K Community of Scientists (2015). The Genome 10K Project: a way forward. Annu. Rev. Anim. Biosci. *3*, 57–111.

Konishi, M. (1965). The role of auditory feedback in the control of vocalization in the white-crowned sparrow. Z. Tierpsychol. *22*, 770–783.

Kotz, S.A., and Schwartze, M. (2010). Cortical speech processing unplugged: a timely subcortico-cortical framework. Trends Cogn. Sci. *14*, 392–399.

Kotz, S.A.E., and Schwartze, M. (2011). Differential input of the supplementary motor area to a dedicated temporal processing network: functional and clinical implications. Front. Integr. Nuerosci. *5*, 86.

Kowalczyk, A., Meyer, W.K., Partha, R., Mao, W., Clark, N.L., and Chikina, M. (2019). RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics, btz468.

Kreitzer, A.C., and Malenka, R.C. (2008). Striatal plasticity and basal ganglia circuit function. Neuron *60*, 543–554.

Kroodsma, D.E. (1982). Learning and the ontogeny of sound signals in birds. In Acoustic communication in birds, D.E. Kroodsma, E.H. Miller, and H. Ouellet, eds. (New York: Academic Press), pp. 1–23.

Kroodsma, D.E. (1984). Songs of the Alder Flycatcher (*Empidonax alnorum*) and Willow Flycatcher (*Empidonax traillii*) Are Innate. Auk *101*, 13–24.

Kroodsma, D.E. (1989). Male eastern phoebes (*Sayornis phoebe*; Tyrannidae, Passeriformes) fail to imitate songs. J. Comp. Psychol. *103*, 227–232.

Kroodsma, D.E., and Konishi, M. (1991). A suboscine bird (eastern phoebe, *Sayornis phoebe*) develops normal song without auditory feedback. Anim. Behav. *42*, 477–487.

Krubitzer, L.A., and Prescott, T.J. (2018). The Combinatorial Creature: Cortical Phenotypes within and across Lifetimes. Trends Neurosci. *41*, 744–762.

Laje, R., and Mindlin, G.B. (2003). Highly structured duets in the song of the South American Hornero. Phys. Rev. Lett. *91*, 258104.

Lattenkamp, E.Z., and Vernes, S.C. (2018). Vocal learning: a language-relevant trait in need of a broad cross-species approach. Curr. Opin. Behav. Sci. *21*, 209–215.

Lee, W.-J., Falk, B., Chiu, C., Krishnan, A., Arbour, J.H., and Moss, C.F. (2017). Tongue-driven sonar beam steering by a lingual-echolocating fruit bat. PLoS Biol. *15*, e2003148.

Leonardo, A., and Fee, M.S. (2005). Ensemble coding of vocal control in birdsong. J. Neurosci. *25*, 652–661.

Levinson, S.C. (2016). Turn-taking in Human Communication – Origins and Implications for Language Processing. Trends Cogn. Sci. *20*, 6–14.

Lieberman, P. (1984). The biology and evolution of language (Cambridge, MA: Harvard University Press).

Ligout, S., Dentressangle, F., Mathevon, N., and Vignal, C. (2016). Not for Parents Only: Begging Calls Allow Nest-Mate Discrimination in Juvenile Zebra Finches. Ethology *122*, 193–206.

Liu, W.C., Wada, K., and Nottebohm, F. (2009). Variable food begging calls are harbingers of vocal learning. PLoS ONE *4*, e5929.

Long, M.A., and Fee, M.S. (2008). Using temperature to analyse temporal dynamics in the songbird motor pathway. Nature *456*, 189–194.

Long, M.A., Katlowitz, K.A., Svirsky, M.A., Clary, R.C., Byun, T.M., Majaj, N., Oya, H., Howard, M.A., 3rd, and Greenlee, J.D.W. (2016). Functional Segrega-

tion of Cortical Regions Underlying Speech Timing and Articulation. Neuron *89*, 1187–1193.

Mahrt, E.J., Perkel, D.J., Tong, L., Rubel, E.W., and Portfors, C.V. (2013). Engineered deafness reveals that mouse courtship vocalizations do not require auditory experience. J. Neurosci. *33*, 5573–5583.

Mann, N.I., Dingess, K.A., and Slater, P.J.B. (2006). Antiphonal four-part synchronized chorusing in a Neotropical wren. Biol. Lett. *2*, 1–4.

Manser, M.B., and Avey, G. (2000). The effect of pup vocalisations on food allocation in a cooperative mammal, the meerkat (*Suricata suricatta*). Behav. Ecol. Sociobiol. *48*, 429–437.

Maples, E.G., Jr., Haraway, M.M., and Hutto, C.W. (1989). Development of coordinated singing in a newly formed siamang pair (*Hylobates syndactylus*). Zoo Biol. *8*, 367–378.

Margoliash, D. (2002). Evaluating theories of bird song learning: implications for future directions. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *188*, 851–866.

Mavlov, L. (1980). Amusia due to rhythm agnosia in a musician with left hemisphere damage: a non-auditory supramodal defect. Cortex *16*, 331–338.

Miller, C.T., and Wren Thomas, A. (2012). Individual recognition during bouts of antiphonal calling in common marmosets. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *198*, 337–346.

Mondloch, C.J. (1995). Chick hunger and begging affect parental allocation of feedings in pigeons. Anim. Behav. *49*, 601–613.

Moore, J.M., and DeVoogd, T.J. (2017). Concerted and mosaic evolution of functional modules in songbird brains. Proc. Biol. Sci. *284*, 20170469.

Moore, P.W.B., and Pawloski, D.A. (1990). Investigations on the Control of Echolocation Pulses in the Dolphin (*Tursiops truncatus*). In Sensory Abilities of Cetaceans: Laboratory and Field Evidence, J.A. Thomas, R.A. Kastelein, and M.A. Boston, eds. (US: Springer), pp. 305–316.

Nottebohm, F. (1972). The Origins of Vocal Learning. Am. Nat. *106*, 116–140.

Nottebohm, F., Kelley, D.B., and Paton, J.A. (1982). Connections of vocal control nuclei in the canary telencephalon. J. Comp. Neurol. *207*, 344–357.

Obrist, M.K. (1995). Flexible bat echolocation: the influence of individual, habitat and conspecifics on sonar signal design. Behav. Ecol. Sociobiol. *36*, 207–219.

Ohms, V.R., Beckers, G.J.L., ten Cate, C., and Suthers, R.A. (2012). Vocal tract articulation revisited: the case of the monk parakeet. J. Exp. Biol. *215*, 85–92.

Okobi, D.E., Jr., Banerjee, A., Matheson, A.M.M., Phelps, S.M., and Long, M.A. (2019). Motor cortical control of vocal interaction in neotropical singing mice. Science *363*, 983–988.

Ölveczky, B.P., Andalman, A.S., and Fee, M.S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. PLoS Biol. *3*, e153.

Osmanski, M.S., and Dooling, R.J. (2009). The effect of altered auditory feedback on control of vocal production in budgerigars (*Melopsittacus undulatus*). J. Acoust. Soc. Am. *126*, 911–919.

Patterson, D.K., and Pepperberg, I.M. (1994). A comparative study of human and parrot phonation: acoustic and articulatory correlates of vowels. J. Acoust. Soc. Am. *96*, 634–648.

Petkov, C.I., and Jarvis, E.D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. Front. Evol. Neurosci. *4*, https://doi.org/10.3389/fnevo.2012.00012.

Pfenning, A.R., Hara, E., Whitney, O., Rivas, M.V., Wang, R., Roulhac, P.L., Howard, J.T., Wirthlin, M., Lovell, P.V., Ganapathy, G., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. Science *346*, 1256846.

Pika, S., Wilkinson, R., Kendrick, K.H., and Vernes, S.C. (2018). Taking turns: bridging the gap between human and animal communication. Proc. Biol. Sci. *285*, 20180598.

Pistorio, A.L., Vintch, B., and Wang, X. (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). J. Acoust. Soc. Am. *120*, 1655–1670.

Prat, Y., Taub, M., and Yovel, Y. (2015). Vocal learning in a social mammal: Demonstrated by isolation and playback experiments in bats. Science Advances *1*, e1500019.

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.

Price, K., Harvey, H., and Ydenberg, R. (1996). Begging tactics of nestling yellow-headed blackbirds, *Xanthocephalus xanthocephalus*, in relation to need. Anim. Behav. *51*, 421–435.

Ratcliffe, J.M., Hofstede, H.M.t., Avila-Flores, R., Fenton, M.B., McCracken, G.F., Biscardi, S., Blasko, J., Gillam, E., Orprecio, J., and Spanjer, G. (2004). Conspecifics influence call design in the Brazilian free-tailed bat, *Tadarida brasiliensis*. Can. J. Zool. *82*, 966–971.

Riede, T., Owren, M.J., and Arcadi, A.C. (2004). Nonlinear acoustics in pant hoots of common chimpanzees (Pan troglodytes): frequency jumps, subharmonics, biphonation, and deterministic chaos. Am. J. Primatol. *64*, 277–291.

Riede, T., Tokuda, I.T., and Farmer, C.G. (2011). Subglottal pressure and fundamental frequency control in contact calls of juvenile *Alligator mississippiensis*. J. Exp. Biol. *214*, 3082–3095.

Saranathan, V., Hamilton, D., Powell, G.V.N., Kroodsma, D.E., and Prum, R.O. (2007). Genetic evidence supports song learning in the three-wattled bellbird *Procnias tricarunculata* (Cotingidae). Mol. Ecol. *16*, 3689–3702.

Simonyan, K., and Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. Neuroscientist *17*, 197–208.

Sisneros, J.A., Forlano, P.M., Knapp, R., and Bass, A.H. (2004). Seasonal variation of steroid hormone levels in an intertidal-nesting fish, the vocal plainfin midshipman. Gen. Comp. Endocrinol. *136*, 101–116.

Sober, S.J., and Brainard, M.S. (2012). Vocal learning is constrained by the statistics of sensorimotor experience. Proc. Natl. Acad. Sci. USA *109*, 21099–21103.

Sober, S.J., Wohlgemuth, M.J., and Brainard, M.S. (2008). Central contributions to acoustic variation in birdsong. J. Neurosci. *28*, 10370–10379.

Stamps, J., Clark, A., Arrowood, P., and Kus, B. (1989). Begging Behavior in Budgerigars. Ethology *81*, 177–192.

Stansbury, A.L., and Janik, V.M. (2019). Formant Modification through Vocal Production Learning in Gray Seals. Curr. Biol. *29*, 2244–2249.e4.

Stoeger, A.S., and Manger, P. (2014). Vocal learning in elephants: neural bases and adaptive context. Curr. Opin. Neurobiol. *28*, 101–107.

Stoeger, A.S., Mietchen, D., Oh, S., de Silva, S., Herbst, C.T., Kwon, S., and Fitch, W.T. (2012). An Asian elephant imitates human speech. Curr. Biol. *22*, 2144–2148.

Suthers, R.A. (1990). Contributions to birdsong from the left and right sides of the intact syrinx. Nature *347*, 473–477.

Taylor, A.M., and Reby, D. (2010). The contribution of source–filter theory to mammal vocal communication research. J. Zool. *280*, 221–236.

Tchernichovski, O., Mitra, P.P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. Science *291*, 2564–2569.

Teeling, E.C., Vernes, S.C., Dávalos, L.M., Ray, D.A., Gilbert, M.T.P., and Myers, E.; Bat1K Consortium (2018). Bat Biology, Genomes, and the Bat1K Project: To Generate Chromosome-Level Genomes for All Living Bat Species. Annu. Rev. Anim. Biosci. *6*, 23–46.

Tobias, J.A., and Seddon, N. (2009). Signal jamming mediates sexual conflict in a duetting bird. Curr. Biol. *19*, 577–582.

Trainer, J.M., McDonald, D.B., and Learn, W.A. (2002). The development of coordinated singing in cooperatively displaying long-tailed manakins. Behav. Ecol. *13*, 65–69.

Tramontin, A.D., and Brenowitz, E.A. (2000). Seasonal plasticity in the adult brain. Trends Neurosci. *23*, 251–258.

Tschida, K., and Mooney, R. (2012). The role of auditory feedback in vocal learning and maintenance. Curr. Opin. Neurobiol. *22*, 320–327.

Tuero, D.T., Gloag, R., and Reboreda, J.C. (2015). Nest environment modulates begging behavior of a generalist brood parasite. Behav. Ecol. *27*, 204–210.

Uchida, A.M., Meyers, R.A., Cooper, B.G., and Goller, F. (2010). Fibre architecture and song activation rates of syringeal muscles are not lateralized in the European starling. J. Exp. Biol. *213*, 1069–1078.

Ulanovsky, N., and Moss, C.F. (2008). What the bat's voice tells the bat's brain. Proc. Natl. Acad. Sci. USA *105*, 8491–8498.

Ulanovsky, N., Fenton, M.B., Tsoar, A., and Korine, C. (2004). Dynamics of jamming avoidance in echolocating bats. Proc. Biol. Sci. *271*, 1467–1475.

Warner, R.W. (1972). The anatomy of the syrinx in passerine birds. J. Zool. *168*, 381–393.

Warren, T.L., Tumer, E.C., Charlesworth, J.D., and Brainard, M.S. (2011). Mechanisms and time course of vocal learning and consolidation in the adult songbird. J. Neurophysiol. *106*, 1806–1821.

Weary, D.M., and Fraser, D. (1995). Calling by domestic piglets: reliable signals of need? Anim. Behav. *50*, 1047–1055.

Wild, J.M., Li, D., and Eagleton, C. (1997). Projections of the dorsomedial nucleus of the intercollicular complex (DM) in relation to respiratory-vocal nuclei in the brainstem of pigeon (*Columba livia*) and zebra finch (*Taeniopygia guttata*). J. Comp. Neurol. *377*, 392–413.

Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (1998). Subharmonics, biphonation, and deterministic chaos in mammal vocalization. Bioacoustics *9*, 171–196.

Wirthlin, M. (2016). Evolutionary Neurogenomic Approaches Provide Insight into the Molecular Basis of Vocal Learning, Doctoral Dissertation (Oregon Health & Science University).

Wise, S.P., Murray, E.A., and Gerfen, C.R. (1996). The Frontal Cortex-Basal Ganglia System in Primates. Crit. Rev. Neurobiol. *10*, 317–356.

Wu, H.G., Miyamoto, Y.R., Gonzalez Castro, L.N., Ölveczky, B.P., and Smith, M.A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. Nat. Neurosci. *17*, 312–321.

Yovel, Y., Geva-Sagiv, M., and Ulanovsky, N. (2011). Click-based echolocation in bats: not so primitive after all. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *197*, 515–530.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al.; Mouse ENCODE Consortium (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zelick, R.D., and Narins, P.M. (1983). Intensity discrimination and the precision of call timing in two species of neotropical treefrogs. J. Comp. Physiol. *153*, 403–412.

Zhao, L., Rad, B.B., and Wang, X. (2019). Long-lasting vocal plasticity in adult marmoset monkeys. Proc. Biol. Sci. *286*, 20190817.

Zhu, Y., Sousa, A.M.M., Gao, T., Skarica, M., Li, M., Santpere, G., Esteller-Cucala, P., Juan, D., Ferrández-Peral, L., Gulden, F.O., et al. (2018). Spatio-temporal transcriptomic divergence across human and macaque brain development. Science *362*, https://doi.org/10.1126/science.aat8077.

Zollinger, S.A., and Suthers, R.A. (2004). Motor mechanisms of a vocal mimic: implications for birdsong production. Proc. Biol. Sci. *271*, 483–491.