Defining the multidimensional phenotype: new opportunities to integrate the behavioral ecology and behavioral neuroscience of vocal learning

Timothy F. Wright[1,2], **Elizabeth P. Derryberry**[3]

[2]Dept of Biology, New Mexico State Univ, Las Cruces, NM 88005 USA

[3]Dept of Ecology and Evolutionary Biology, Univ of Tennessee, Knoxville, TN 37996 USA

Abstract

Vocal learning has evolved independently in several lineages. This complex cognitive trait is commonly treated as binary: species either possess or lack it. This view has been a useful starting place to examine the origins of vocal learning, but is also incomplete and potentially misleading, as specific components of the vocal learning program – such as the timing, extent and nature of what is learned – vary widely among species. In our review we revive an idea first proposed by Beecher and Brenowitz (2005) by describing six dimensions of vocal learning: (1) **which** vocalizations are learned, (2) **how much** is learned, (3) **when** it is learned, (4) **who** it is learned from, (5) **what** is the extent of the internal template, and (6) **how** is the template integrated with social learning and innovation. We then highlight key examples of functional and mechanistic work on each dimension, largely from avian taxa, and discuss how a multi-dimensional framework can accelerate our understanding of why vocal learning has evolved, and how brains became capable of this important behaviour.

Keywords

behavioral ecology; behavioral neuroscience; brain evolution; call; cognition; comparative method; integrative biology; neurogenetics; song; trait evolution; vocal production learning

Introduction: vocal learning

Language is a complex cognitive trait unique to humans (Fisher and Marcus, 2006). We develop this trait through a combination of genetic influences and social learning in a process known as *vocal production learning,* also termed *imitative learning*, in which novel acoustic signals are acquired via imitation of conspecifics (Jarvis, 2019). Although our closest living relatives, the non-human primates, do show other forms of learning, such as auditory learning (learning to perceive or react to a vocalization differently as a result of

experience) or vocal usage learning (learning to use innate vocalizations in a new context), they are with rare exception unable to learn to produce novel vocalizations (Fisher and Marcus, 2006; Jarvis, 2019). However, vocal production learning does occur in much more distantly related lineages, such as songbirds (Marler and Tamura, 1962) and whales (Janik, 2014). This observation begs the questions of why and how vocal production learning has evolved repeatedly and independently in multiple lineages of birds and mammals, including humans.

Studies that leverage this pattern of repeated evolution give insight into the selective advantages of this complex cognitive trait and the neurogenetic mechanisms that give rise to it. For example, comparative studies of taxa with vocal learning, like songbirds and parrots, and their non-learning relatives yield a number of hypotheses for the benefits of vocal production learning (Catchpole and Slater, 2008). Species that learn to vocalize may be better than non-learners at producing vocalizations that transmit effectively in their habitat (Catchpole and Slater, 2008), recognizing others such as neighbours (Beecher and Brenowitz, 2005) or group members (Sewall et al., 2016), sharing information in kin groups (Nowicki and Searcy, 2014), producing elaborate vocal repertoires that serve to repel competitors (Burt et al., 2001), or permit the assessment of the quality of potential mates (Catchpole and Slater, 2008). Similarly, comparisons of vocal learners to non-learners have also revealed fundamental differences in neuroarchitecture and gene expression patterns, providing mechanistic hypotheses for how birds learn to vocalize. Differences in neuroarchitecture include the presence of cortical-striatal-basal ganglia loops in learning taxa like songbirds and humans, and the absence of such loops in non-learners like chickens and macaques (Pfenning et al., 2014). There is also growing evidence of convergent specializations in gene expression between learners in the nuclei composing the loops, suggesting a common set of gene pathways involved. For instance, the robust nuclei of archopallium (RA) in songbirds and the laryngeal motor cortex in humans show strong similarities in gene expression patterns, whereas Area X in songbirds and the anterior striatum in humans share a different set of co-expressed genes (Pfenning et al., 2014; Whitney et al., 2014). Thus, substantial progress has been made in understanding why and how vocal production learning evolved by treating it as a binary trait that species either possess or lack entirely.

The binary view, however, disregards the extreme diversity of learning programs found within vocal production learners. Just a few minutes listening to a dawn chorus on a spring day will reveal a tremendous diversity in learning programs. One obvious aspect of this variation is repertoire size, or the number of different vocalizations produced (Catchpole and Slater, 2008). Adult great tits, *Parus major*, have small repertoires, often fewer than five song types (Krebs et al., 1978), whereas adult male nightingales, *Luscinia megahynchos*, often have repertoires of hundreds of different songs (Todt, 1971). Less immediately obvious is the fact that both of these species are capable of learning throughout their lifetime (McGregor and Krebs, 1989; Todt and Böhner, 1994), whereas others in the chorus, such as the chaffinch, *Fringilla coelebs*, have a more restricted period of song learning (Lachlan and Slater, 2003). This remarkable diversity in learning programs led Brenowitz and Beecher (Beecher and Brenowitz, 2005; Brenowitz and Beecher, 2005) to argue that the binary trait of vocal learning might be better considered as a multi-dimensional trait in which different

songbird species have evolved to different points on each of several dimensions, such as repertoire size and timing of learning. For at least a decade there was limited adoption of this perspective, or of the integration it was intended to promote, in part because how to quantify these different dimensions had not been fully articulated, and in part because Baker and Brenowitz published their ideas in two separate papers, one targeting evolutionary biologists and the other neurobiologists. Recently there have again been calls to develop a more modular view of this complex trait (Lattenkamp and Vernes, 2018; Wirthlin et al., 2019). Here we expand on those calls by reframing and extending Beecher and Brenowitz's framework to identify six quantifiable dimensions (Figure 1) that capture the vocal learning phenotype across a broader array of species and signal types. We argue that this multi-dimensional perspective will accelerate and integrate functional and mechanistic studies of vocal production learning and thereby improve our understanding of its evolutionary origins across birds and mammals.

A multi-dimensional definition of vocal learning is a powerful framework providing both short- and long-term advantages. In the short term, it explicitly recognizes the existing diversity of learning programs and provides a common framework for quantifying and studying this diversity. In the longer term, widespread adoption of a multi-dimensional definition will facilitate parallel data collection on the vocal learning phenotype across taxonomic groups, enabling richer comparative research (Beecher and Brenowitz, 2005). By more precisely defining the vocal learning phenotype, this framework allows us to test whether different dimensions are under different selective forces and whether they are governed by different mechanisms. This perspective should facilitate new hypotheses and collaborations that will move the field towards a better integration of evolutionary and mechanistic explanations for the diversity of learning patterns. Below, we define the different dimensions of the vocal learning phenotype and illustrate the advantages of using this approach with select examples from the recent literature. Although our examples are drawn exclusively from studies of songbirds and parrots, these dimensions should, in theory, apply across all taxa with vocal learning, and we hope that this framework will stimulate additional work on these issues in other taxa.

How many dimensions?

We propose that vocal learning can be deconstructed into six independent dimensions (Figure 1). The first dimension is **which vocalizations are learned**, by which we mean which functional classes of the vocal repertoire are acquired through learning. The second dimension is **how many vocalizations are learned**, or alternatively, what is the size of the learned repertoire. The third dimension is **when are vocalizations learned**, meaning at what period(s) of an individual's life is the learning program active. The fourth dimension is **who are vocalizations learned from,** where potential learning models are distinguished by social relationships to the learning individual (e.g. parents, siblings, territorial neighbors, social group members, other species). The fifth dimension is **what internal template is possessed** by an individual that guides and restricts learning. The sixth dimension is **the degree of integration of the internal template with social learning and innovation** during vocal learning. This last dimension captures variation among individuals and species in the relative importance of these three components in determining the final form of a vocal signal.

Beecher and Brenowitz (2005) proposed five dimensions: 1) when is song learned, 2) how many songs a bird learns, 3) copy fidelity, 4) role of early song experience, and 5) degree of canalization. They also include the caveat that responsiveness to social factors could be included as another dimension in song learning. Two of our dimensions (Dimensions 2&3: how much is learned and when it is learned) largely overlap with Beecher and Brenowiz's, but are broadened to encompass all learned vocalizations, not only song. Our other four dimensions contain some similar components but extend these dimensions, in order to a) broaden the framework beyond songbirds to include other vocal learning taxa and beyond song to include other learned vocalizations, b) define dimensions in ways that are more quantifiable and c) to explicitly incorporate the role of social interactions in the learning process. In Box 1 we outline different approaches to addressing the critical question of whether a given signal is learned in the first place. Below we discuss each of our six learning dimensions in turn, focusing on how variation in each dimension can be measured, what evidence exists of variation among individuals or species, what potential sources of selection could act on this variation, and what potential mechanisms might govern it (Table 1).

Dimension 1: Which vocalizations are learned?

Those studying avian vocal repertoires have long recognized a distinction between *songs* and *calls*. Songs are typically characterized as being acoustically complex, and, at least in oscine songbirds and hummingbirds, learned (but see (Spector, 1994) for extensive discussion of various definitions of song). In contrast, calls are characterized as acoustically simple, and in many species, develop without significant learning. However, there is increasing recognition that calls compose a functionally heterogeneous collection of signals, and that, in at least some taxa, some categories of calls are learned (Marler, 2004a) (Table 1). Examples include the rain calls of the chaffinch, *Fringilla coeleb*, (Baptista, 1990), flight calls of the red crossbill, *Loxia curvirostris* (Sewall, 2009), the chick-a-dee calls of the black-capped chickadee, *Poecile atricapillus* (Mischler et al., 2020), and the contact calls of many parrots (Wright and Dahlin, 2018). Which parts of the repertoire are learned differs among taxa and may even differ between the sexes. For example, in the zebra finch, *Taenopygia guttata*, males learn both their songs and distance calls, but not other calls like the 'tet' and 'stack', while females do not appear to learn any part of their repertoire (Slater and Jones, 1995; Ter Maat et al., 2014). Learning of calls and/or song does, however, occur in females of many other species, offering additional scope for variation in what is learned across and within species (Odom et al., 2014; Riebel et al., 2019; Sewall, 2009).

Given song's importance in both intra-sexual competition and mate attraction (Catchpole and Slater, 2008), many have posited that sexual selection has played a major role in the evolution of song, and, by extension, song learning (Nowicki and Searcy, 2014). More recently, those studying learned contact calls used for social recognition and group membership have proposed that vocal learning evolved to permit the flexible labelling of relationships within dynamic social systems (Sewall et al., 2016). Contrasts between song and call learning are also instructive at the mechanistic level. Zebra finches are the predominant model for investigating the mechanisms of song learning, and increasingly of call learning (Simpson and Vicario, 1990; Ter Maat et al., 2014), and studies of song development in this species have produced a rich understanding of the neural, hormonal,

genetic and epigenetic mechanisms that interact to allow learning. Fewer studies have investigated the mechanisms underlying call learning in this species, but to date those that have indicate that the same neural centers involved in learning song also govern the learning of calls (Simpson and Vicario, 1990; Ter Maat et al., 2014). Because the production of many unlearned calls uses the same pathway as the production of learned song and calls, the vocal production pathway may predate the evolution of learning itself (Ter Maat et al., 2014). In contrast, the discovery of parallel core and shell circuits connecting neural centers for vocal learning in parrots (Chakraborty and Jarvis, 2015) raises the intriguing possibility that learned and unlearned vocalizations, or even different parts of the learned repertoire, may have different neural substrates in this group. Determining whether the conservation of the same neural pathways for the production of learned and unlearned vocalizations or the recruitment of new ones is the more common pattern is an exciting avenue requiring bidirectional cross-talk between neuro- and evolutionary biologists (Jarvis, 2019).

Dimension 2: How many acoustically distinct vocalizations are learned?

This dimension considers the number of **acoustically distinct signals** that are in a species' repertoire (Table 1). While this dimension could also be considered in terms of functionally distinct signals, we suggest that considering the number of acoustically distinct signals most effectively captures the extent of variation among species or individuals in the size of the learned repertoire and best distinguishes this dimension from the previous one. Variation in the size of the learned repertoire is one of the most readily apparent dimensions of learning to vary among species (Table 1). For example, within the Passerellidae clade of New World sparrows repertoire sizes vary widely, from species in which males have just a single song in their repertoire, like the white-crowned sparrow, *Zonotrichia leucophrys,* (Marler and Tamura, 1962), to others in which males have a multi-song repertoire of 7–12 songs like the song sparrow, *Melospiza melodia* (Searcy et al., 2014), to the Bachman's sparrow, *Peucaea aestivalis*, in which a single male might have upwards of 50 distinct songs in its repertoire (Ali and Anderson, 2018). This pattern is replicated across the Passeriformes, with many clades including both species with single-song repertoires and species with multi-song repertoires, suggesting that the size of the learned song repertoire is an evolutionarily labile trait (MacDougall-Shackleton, 1997). More generally, there is evidence of variation in the size of the vocal repertoire across many species (Marler, 2004a), but the complexity of cataloging the entire vocal repertoire of a species across all contexts, seasons, sexes and life stages, coupled with the effort required to unambiguously determine which parts of it are learned (Box 1), means that outside of these species where only the song is learned and sampling the song repertoire is straightforward, the number of species for which the true extent of learned acoustic diversity is known is small.

One longstanding hypothesis for the presence of multi-song repertoires in males is that they are under directional selection via female preferences for larger repertoires. Although intuitively appealing, this hypothesis has received mixed support. Although some studies find evidence of female preferences for larger repertoires, others do not, and a meta-analysis found the overall effect size for the association between the two variables to be fairly small (Soma and Garamszegi, 2011). Furthermore, comparative studies indicate that single-and multi-song repertoires are approximately equally frequent across passerine species, and

phylogenetic reconstructions suggest that there have been frequent state reversals between single- and multi-song repertoires across passerine species, implying no general pattern of selection for larger song repertoires across passerines (MacDougall-Shackleton, 1997). Testing which of the many alternative hypotheses regarding the benefits of a given repertoire size will require comparative datasets on life history, social complexity and repertoire diversity, highlighting a need for further gathering of these data.

There are clearer patterns regarding the neural mechanisms underlying larger song repertoires. One of the first studies to examine this question found a relationship between the number of syllables in the song repertoire and the volume of the neural song production centers RA and HVC across individual male canaries, *Serinus canaria* (Nottebohm, 1981). Subsequent studies have found consistent support for a relationship between repertoire size and size of the HVC both within species (Garamszegi and Eens, 2004) and across different species (Devoogd et al., 1993). It remains unclear, though, whether the size of the vocal repertoire is determined by the amount of specialized tissue or whether the number of vocalizations learned affects the degree of development of these brain regions (Garamszegi and Eens, 2004); this question can only be answered by coupling new information from behavioral ecologists on what vocalizations are learned with neurogenomic assays performed by neurobiologists.

Dimension 3: When are vocalizations learned?

Those studying bird song have long made a distinction between closed-ended and open-ended vocal learners, with imitative learning in the former restricted to a discrete period, typically early in life, while learning in the latter may occur throughout life. Classic examples include the zebra finch, a closed-ended learner which learns its single song during a critical period that ends around 90 days post-hatch (Slater and Jones, 1995) and the budgerigar, *Melopsittacus undulatus*, which continually modifies its small repertoire of contact calls throughout its life (Farabaugh and Dooling, 1996). These two types of learners actually represent two ends of a continuum (Table 1) (Beecher and Brenowitz, 2005). Between these two extremes lie species with discrete learning periods that extend past the early juvenile phase into the end of the first year of life (Beecher, 2008), and others in which the critical period reopens seasonally (Nottebohm et al., 1986). A further consideration is that learning may occur in two distinct phases- the *sensory*, or memorization phase, during which an individual memorizes what it is to produce, and the *sensorimotor* or production phase, during which an individual starts to produce and refine what will eventually become its fully developed vocal signal (Soha, 2017). In some species, like the zebra finch, these two phases are largely overlapping, while in others they may be temporally separate, as in white-crowned sparrows which memorize their future songs in the fall of their first year but don't start producing these songs until the following spring (Brainard and Doupe, 2002; Hultsch and Todt, 2004). Furthermore, the timing of learning may be plastic to some degree and vary depending on tutor availability (Gobes et al., 2019).

Differences in the timing of song learning may have adaptive significance. Open-ended learning could allow for the continual expansion of the vocal repertoire, if such is favored by sexual selection, as in northern mockingbirds, *Mimus polyglottos* (Howard, 1974). It

could also favor continual turnover of calls used for recognition within dynamic social systems, such as found in budgerigars, *Melopsittacus undulatus* (Dahlin et al., 2014). Conversely, limiting the timing of learning to a discrete period may be beneficial if there are physiological costs to learning, as suggested by the relationship between brain size and vocal repertoire size discussed above. There also might be functional costs to open-ended learning if learning could occur at the wrong time, or from the wrong model (Table 1). The relative benefits and costs of learning at a given life stage may vary even within a species. For example, in the white-crowned sparrow, the critical learning period for the migratory subspecies *oriantha* occurs earlier and is more temporally constrained than that for the sedentary subspecies *nuttalli*; these differences are apparent even when the two subspecies are raised in a common laboratory environment, and are thought to have evolved due to the temporal constraints placed on learning by migration and to differences in the timing of settlement onto territories (Nelson et al., 1995).

One candidate neural mechanism governing the timing of vocal learning is the gene *FoxP2*. Expression levels of this transcription factor within the songbird neural learning center Area X increase relative to those in surrounding striatal tissue during song crystallization in zebra finches (Haesler et al., 2004). Conversely, *FoxP2* levels in Area X show distinct down-regulation relative to the surrounding striatum when adult males are singing an acoustically plastic "practice song" in isolation (Teramitsu and White, 2006). Based on these data, Miller, White and colleagues proposed that FoxP2 acts as a "plasticity gateway" such that when FoxP2 levels are high in Area X, the synaptic plasticity required for vocal learning is reduced, while when FoxP2 levels are low, synaptic plasticity is promoted and, if maintained for a sufficient amount of time, new learning can occur (Miller et al., 2010). Consistent with this view is the finding that in the budgerigar, which shows persistent plasticity and open-ended learning of contact calls, FoxP2 levels in MMST, the parrot analog of Area X, are consistently down-regulated relative to the surrounding striatum (Hara et al., 2015; Whitney et al., 2015). Further tests of whether FoxP2 expression is responsible for evolved differences among species in the timing of learning could come from collaborative surveys of FoxP2 activity in species with different critical periods by field and lab biologists, and from experimental manipulations of expression patterns at different life stages (Box 2).

Dimension 4: From whom are vocalizations learned?

In theory, transmission from model to learner may occur vertically from parents, horizontally from siblings, obliquely from unrelated territorial neighbors, or in a distributed fashion across a network of social group members (Lynch and Baker, 1993). There is ample evidence from experiments in natural conditions or ones that control potential tutors that there is variation among species in who serves as a primary model for vocal learning (Table 1). In two species of Darwin's finches, the medium ground finch, *Geospiza fortis*, and the cactus finch, *G. scandens*, young males learn their songs primarily from their fathers (Grant and Grant, 1996). In contrast, young male song sparrows acquire their multi-song repertoire from unrelated males who hold territories in the area where the young males are attempting to settle (Beecher, 2017). Young budgerigars develop their initial contact call from their individually distinctive but presumably unlearned begging calls, but then progressively develop a repertoire of multiple contact call types shared first with their

siblings and later with a broader array of social associates (Brittan-Powell et al., 1997). In zebra finches, early studies suggested males learned their songs primarily from their fathers (Immelmann, 1969); later work, though, has suggested a more complex picture in which learning may occur preferentially from fathers (Mann and Slater, 1995), obliquely from unrelated males (Williams, 1990), or even horizontally from siblings (Deregnaucourt and Gahr, 2013). Interactions between social factors and neural mechanisms lead to wide variation across species in whom individuals learn from. For example, juvenile zebra finches preferentially learn songs from males a) that they are housed with prior to their sensitive phase, b) that are paired with their mother, c) that are paired rather than single, d) direct more aggression towards them, and e) direct more songs to them (Chen et al., 2016; Jones and Slater, 1996; Mann and Slater, 1994). Furthermore, recent work has shown that juvenile male learning is guided by non-vocal feedback from adult females; juvenile males who received contingent visual reinforcement of a female "fluff-up" display after singing learned faster and more accurately than did males with non-contingent reinforcement (Carouso-Peck and Goldstein, 2019). In aggregate, these preferences should typically lead to juveniles learning songs from their fathers. In contrast, juvenile song sparrows preferentially learn songs a) from territorial males who survive the winter, b) from males who share songs with other males, and c) that they overhear being used in singing interactions between males (Beecher, 2017). They do not, however, learn more from males who are aggressive towards them, or of higher overall quality (Beecher, 2017). These preferences, coupled with a learning period that extends to the end of first year, leads to song sparrows learning their multi-song repertoire primarily from unrelated males who hold territories in the area where the juvenile settles. While the outcome is different for these two species, in both it is apparent that whom an individual learns from is governed by multiple factors. Importantly, it is evident that species vary not just in who they learn from, but also in the extent to which a given signal is learned from a single versus multiple models (see Dimension 6 below).

There is some evidence that the choice of a particular model type can be adaptive for the individual. Female Darwin's finches, Genus *Geospiza*, prefer males who sing conspecific song but avoid those who sing the same song as their fathers; thus, males who accurately learn their father's song can avoid both non-adaptive inbreeding and hybridization (Grant and Grant, 1996). In song sparrows, *Melospiza melodia*, young males who share more song types with their neighbors upon settlement are able to hold onto that territory for longer than those with fewer shared songs (Beecher et al., 2000), and females prefer local song types to ones from more distant populations (Searcy et al., 2002). In other species the benefits of learning from a particular model are less clear; for example, in budgerigars, females who share more contact call types within a particular social group engaged in more aggressive interactions with each other, but there is no apparent relationship between call sharing and affiliative interactions (Dahlin et al., 2014).

The mechanisms that govern choice of model remain relatively unexplored. One clear determinant is which models are available to an individual when it is learning. This availability depends on such factors as dispersal patterns, social organization, the timing of learning (Dimension 3) and the permissiveness of any internal template guiding learning (Dimension 5, below). In addition, mechanisms that govern social attention (i.e. the intensity and direction of social interaction) could be important determinants of whom an

individual learns from and how well it learns. Recent work in zebra finches has shown that catecholaminergic neurons in two midbrain centers, locus coeruleus and ventral tegmental area, that are generally implicated in social attention and learning, showed significantly higher expression of the immediate early gene EGR-1 in birds that were socially tutored for song than those who were passively tutored (Chen et al., 2016). This result is especially significant because a) socially tutored birds learned their songs with a higher degree of accuracy than passively tutored birds, and b) the two midbrain centers are known to have projections to the song learning circuits in the forebrain (Chen et al., 2016). Little is known about the importance of these two mid-brain centers in species beyond zebra finches, which invites further integrative collaborations.

Dimension 5: What is the extent of the internal template?

Dimension 5 describes the extent and composition of latent or innate information that is available to an individual to guide its vocal learning. As with Dimension 2 (see also Box 1), it is typically measured through experiments that control the auditory input available to an individual during vocal learning and examine the resulting vocalizations. For this dimension, however, the question is not whether or not a particular vocalization can develop at all in the absence of social input, but rather which aspects of the resulting vocalization are universal to a species and which are subject to modification via individual experience. Foundational work by Peter Marler and colleagues used this approach extensively to examine the learning of specific song traits in a suite of New World sparrow species (reviewed in Soha, 2017). One example from this body of work is a study of juvenile swamp sparrows, *Melospiza georgiana*, tutored with songs composed of either swamp sparrow or song sparrow syllables that were arranged with temporal patterns characteristic of one or the other species (Marler and Peters, 1977). The juveniles learned only those songs composed of swamp sparrow syllables, regardless of their temporal patterning, suggesting that a preference for species-specific syllable types was innate while a preference for species-typical temporal patterning was not. This study, and others similar to it, led Marler to propose the *auditory template hypothesis*, which held that young birds possessed an internal representation of species-specific song that guided song development in isolation, promoted selective attention to the song of conspecific adults, and honed successive renditions of bird's song(s) during practice (Marler, 1984; Soha, 2017). Subsequent work provided compelling evidence that the extent and specificity of this auditory template varies across different species (Marler, 2004b; Soha, 2017).

An alternative approach to understanding the nature of genetic contributions to vocal signals is to estimate the heritability of quantitative vocal traits. This can be done using pedigrees, selected lines or crosses between populations or species with different song traits. A large-scale study comparing heritability of acoustic traits in zebra finches found lower heritability measures for call traits in males, who learn their calls, than for calls traits in females, who do not (Forstmeier et al., 2009). Heritability was lower still for male song traits, with traits that were linked to the physical mechanisms of producing songs, such as mean frequency and timbre showing higher heritability than traits such as repertoire size and mean song length that may lack the same physiological constraints (Forstmeier et al., 2009).

While it is clear that species vary in the extent of the template they possess about their species-typical song, it is less clear what the adaptive value of this variation might be. At the most basic level, the ability to focus learning on appropriate conspecific models undoubtedly helps avoid maladaptive hybridization when these learned signals are used for mate choice. But why species should show such extensive variation in the nature of both the template and the cues used to identify species-specific song remain poorly understood. This gap may arise in part because, until recently, most work focused on the neural substrate of the birds' own song or tutor song (Bulhuis & Moorman 2014). We had limited understanding, however, of the mechanisms underlying the neural representation of species-specific song. This situation is starting to change. Work by Yazaki-Sugiyama and colleagues has demonstrated that, in zebra finches, species specificity is encoded in the timing of the gaps between syllables in song (Araki et al., 2016). Juveniles tutored with heterospecific song from Bengalese finches, *Lonchura striata domestica*, produced song composed of Bengalese finch syllables sung with temporal phrasing typical of zebra finches. This species-specific temporal patterning appeared to be encoded in a specific population of neurons in Field L, part of the ascending auditory pathway, that fired strongly in response to both natural and synthetic songs with syllable gaps typical of zebra finch song, and were less responsive to songs with conspecific frequency information but heterospecific timing (Araki et al., 2016). This works sets the stage for collaborative and comparative work examining the responsiveness of Field L to conspecific and heterospecific song features across a range of species.

Dimension 6: How is the template integrated with social learning and innovation?

Dimension 6 captures variation among species in how genetically inherited information is combined with information from social learning and with individual innovation to produce a learned vocal signal (Table 1). Social learning refers not only to external acoustic information but also to the influences of social interactions, such as those mentioned in Dimension 4 with potential tutors and even non-tutors in social groups.

While conceptually related to Dimension 5, Dimension 6 differs in focusing not so much on the nature or extent of genetically coded information, but on how different species integrate this information with external influences and innovation. One approach to addressing this question is to examine variation in a learned signal across individuals and populations of a given species. A number of foundational studies in songbirds used this approach to identify some aspects of song that are shared by all individuals in a species. These "species universals" are inferred to represent innate, or unlearned, contributions to the vocal output that are then elaborated or recombined during the learning process (Marler, 2004b). One early example is work by Emlen, Payne and others identifying species universals in the songs of indigo buntings, *Passerina cyanea* (Marler, 2004b; Payne, 2006). These songbirds sing a single song that is composed of multiple elements. Surveys across the wide range of this species found that all individuals produced a song composed of elements drawn from the same catalog of only 100 or so elements, suggesting elements are at least partially innate. This species also showed evidence of social learning of song in the form of local song neighborhoods, in which clusters of neighboring males sing similar songs, and of innovation (or possibly learning errors) in the smaller-scale variation from individual to individual in song (Payne, 2006). This pattern of species universals that are elaborated upon by social

learning and innovation to form regional- and individual-specific repertoires appears to be commonplace across vocal learning species (Marler, 2004b; Wright and Dahlin, 2018).

Variation in how different forms of information are combined is both widespread among species and evolutionary labile. At the grossest level this can be assessed by the propensity for a species to produce heterospecific vocalizations (e.g. mimicry), for which there is unlikely to be genetically coded information available, versus solely conspecific vocalizations, which may have at least some degree of genetically coded information. Goller and Shizuki examined the phylogenetic distribution of heterospecific mimicry across the oscine songbirds (339 species in 43 families) and when mimicry was mapped as a character state on the oscine phylogeny they estimated that mimicry was not ancestral in the oscines, but was gained at least 237 times and lost at least 52 times (Table 1) (Goller and Shizuka, 2018). They proposed that mimicry evolved repeatedly either through a relaxation of constraints on conspecific learning, or through active selection for mimicry as a means of increasing repertoire size or complexity (Goller and Shizuka, 2018). In either case, the transition to heterospecific mimicry would require either a broadening of, or a reduction in reliance on an internal template for conspecific song, and an increase in learning from the external environment.

We know very little about how the brains of mimics differ from those of non-mimics, or more broadly, how the brain integrates information acquired through individual or social learning with internally represented information. The recent discovery of a putative duplication of the vocal learning circuit in parrots may offer some insight into the vocal flexibility and heterospecific mimicry that is common in this group (Chakraborty et al., 2015). Chakraborty and colleagues found that many of the previously identified vocal learning regions in the brain of budgerigars had anatomically distinct core and shell sub-regions (Chakraborty et al., 2015) These sub-regions were defined by different patterns of expression of learning-related genes, and that core sub-regions primarily projected to other cores, and shells to other shells. When these regions were examined across a small suite of parrot species, the authors noted a general tendency for there to be a higher ratio of shell to core in species with more developed mimicry abilities (Chakraborty et al., 2015), suggesting that the duplication and subsequent elaboration of the shell system from an ancestral core system contributed to the advanced mimicry abilities seen in parrots (Chakraborty et al., 2015). Broader scale comparisons of mimicry patterns across parrot species as well as comparisons among songbirds, which lack the core-shell structure, could provide key insight into how the learning brain integrates information acquired from different sources.

Tools for further advances

Comparative studies.

The comparative approach is a powerful method to test the adaptive significance and mechanistic underpinnings of different dimensions of vocal learning. Marler and Peters made extensive use of this approach to test the hypothesis of selective learning in songbirds. As mentioned above, this now classic work trained song and swamp sparrows on the other species' songs to reveal species-level differences in the extent and type of latent information guiding the content and structure of songs (Marler and Peters, 1977; Marler and

Peters, 1988). Since that time, the comparative approach has proved fruitful for examining a number of different dimensions of the learning phenotype, including understanding the adaptiveness of variation in the timing of song learning (Nelson et al., 1995) and variation in repertoire size (Kroodsma and Canady, 1985). Many of these comparative studies are limited to a comparison of two to three populations or species. However, broader taxonomic comparisons are becoming increasingly feasible. The first 'all birds' phylogeny (9,993 species) was published in 2012 (Jetz et al., 2012), and there are a number of recent well-supported phylogenies for lineages of birds that learn their song (Burns et al., 2014; Gardner et al., 2010; McGuire et al., 2014; Provost et al., 2018). Accompanying these improvements in phylogenetic tools are advances in the comparative analysis tools (Jombart et al., 2010; O'Meara et al., 2006; Rabosky et al., 2014; Revell, 2012; Slater et al., 2012) needed to examine evolution of complex traits such as learned songs and calls (Mason et al., 2017; Medina-García et al., 2015). One critical component that remains scarce, though, is data on the vocal dimensions themselves. Outside of a few species, little is known about when vocalizations are learned by each sex (D1), from whom individuals learn (D4) or the extent of the internal template (D5). Collecting these data is no simple task, but it may be possible to use machine learning approaches to leverage acoustic data found in curated collections (e.g., Borror Laboratory of Bioacoustics and the Macaulay Library of Sound) and online (e.g., xeno-canto) to quantify variation along some dimensions of learning. For example, the extent of the internal template (D5) could be approximated indirectly by comparing geographically separated populations of the same species to extract information about the universal (and most likely genetically-determined) features of species' songs (Lachlan et al., 2010). Collecting these types of data broadly across avian and mammalian lineages with vocal learning would facilitate progress in understanding the evolutionary history of genetically determined constraints on learned acoustic signals.

Selected lines.

Another potentially powerful approach that, to date, has been rarely used to explore vocal learning is laboratory lines that have been specifically selected for differences in a particular dimension of vocal learning. In some cases, the time-consuming work of creating these lines has already been done by aviculturists selectively breeding for particular song traits. For example, Mundinger and Lahti (2014) made use of the fact that many strains of canary have been bred for specific acoustic characteristics of their learned song (Guttinger, 1985). They bred two such strains, rollers and borders, to each other and performed backcrosses to create lineages that differed in the degree of genetic complement inherited from each strain (Mundinger and Lahti, 2014). These groups were then tutored with either low frequency syllables characteristic of rollers or high frequency syllables characteristic of borders. They found a strong correlation between the degree of genetic complement from one strain and the propensity to learn the syllables characteristic of that strain, with an outsized effect of the Z sex chromosome (Mundinger and Lahti, 2014). They also found a strong bias for individuals with the same genetic complements to learn the same syllables, suggesting that their internal template biased learning towards particular acoustic characteristics (Mundinger and Lahti, 2014). A complementary study of hearing in similar backcrosses suggests that some of these latent biases may arise from differences in peripheral sensory abilities and not solely higher-order cognitive processes (Wright et al., 2004). While the creation of

lines selected for differences in specific learning dimensions would be time-consuming and costly, this approach should not be overlooked as a powerful means to testing both mechanistic and adaptive explanations for variation in these dimensions.

Candidate genes.

Candidate genes are another powerful approach for testing both proximate and ultimate hypotheses explaining variation in a particular learning dimension. In the past such genes were generally first identified in genetic model systems like *Drosophila*, *Mus* or *Caenorhabditis* and then examined in other species for roles in the vocal learning process (Mello et al., 1992). This approach is limited by the fact that none of these systems are capable of vocal production learning. In a few cases candidate genes have been identified by association studies in humans and then subsequently tested for effects in vocal learning birds. Such was the case for the gene Foxp2, whose role in vocal learning was first identified in a human family with an inherited speech learning disability (Lai et al., 2001). More recent candidates have been identified using large-scale screens employing microarrays, RNA sequencing, or other means of identifying transcriptionally active genes, followed by bioinformatic approaches that compare patterns of expression (Burkett et al., 2018). As the sensitivity and specificity of these approaches has increased, these experiments have become increasingly sophisticated in targeting their comparisons to particular time points, behavioral treatments, brain regions, and even, with the advent of single cell nucleus sequencing (Wang et al., 2019), to specific cell types.

Identification of candidate genes is merely a first step, however, in determining whether they play a role in mediating variation in a particular learning dimension. Ideally such studies are followed by experimental manipulation of gene expression levels within specific neural regions and measurement of the resulting learning phenotype (See Box 2). Such manipulations provide strong tests of mechanistic hypotheses at the molecular level. For example, both experimental knockdowns and virally-mediated overexpression of FoxP2 targeted to the Area X of juvenile zebra finches interferes with song crystallization and reduces the accuracy of song learning (Haesler et al., 2007; Heston and White, 2015; Murugan et al., 2013) (Box 2). Such studies can also be effective ways of testing whether the proposed dimensions are truly independent or share the same underlying mechanisms; changes in FoxP2 expression affect both how quickly individuals learn and the degree to which their song matches their tutor's song (Haesler et al., 2007; Heston and White, 2015; Murugan et al., 2013). What is sometimes overlooked is that these molecular manipulations can also be a powerful tool for testing functional hypotheses for the selective value of variation. If a manipulation does result in changes in phenotype along a particular learning dimension, then the relative fitness of resulting variants can be examined under a variety of conditions (Box 2, Figure 1).

Conclusions and research prospectus

By articulating the different dimensions of vocal production learning, we highlight a framework for studying the extensive diversity found in imitative vocal learning programs. Our examples are drawn primarily from birds, but these dimensions should apply equally

to mammalian vocal learners. Our intent is to stimulate new avenues of research that better integrate evolutionary and mechanistic approaches. We see potential for a virtuous cycle whereby improved understanding of the mechanistic underpinning of specific vocal dimensions will help refine and test theories for the adaptive significance of variation, which in turn will give new insight into potential mechanisms. The process of defining vocal dimensions also raises a number of new empirical questions, including whether dimensions are independent or impose constraints on one another. For example, the timing of learning affects who is available for a young bird to learn from as well as the potential for social reinforcement of learning, yet we know little about how these three dimensions interact at the mechanistic level. Such interactions among dimensions may limit the phenotypic landscape of vocal production learning or result in unknown, emergent properties of this complex behavior (Bradbury and Vehrencamp, 2014). We propose that further advances will require collecting behavioral and neurogenetic data from a wider range of species and lineages, including those bird species in which females learn as well as among mammalian vocal learners of both sexes, to take advantage of the wide range of state space already explored by evolution. New tools should aid this advancement, particularly new molecular tools for manipulating behavior and new comparative tools for modeling complex trait evolution. In particular, close collaborations between behavioral ecologists and behavioral neurobiologists will generate novel discoveries. In one such case, work on categorization of vocal notes led to the discovery of mirror neurons in birds (Prather et al., 2009). Future collaborations using a shared multi-dimensional framework for defining the vocal learning phenotype should produce further advances in our understanding of this complex cognitive trait.

Acknowledgements

Appendices

Box 1:

Is a vocal signal learned?

A fundamental issue for all these dimensions is determining whether the particular vocal signal is socially learned or, alternatively, develops innately without substantial social input from conspecifics. Vocal learning is typically demonstrated in one of three ways. The first approach is through controlled exposure experiments. Isolation, deafening, live or recorded tutoring, and cross-fostering have all been used by investigators to control the social and acoustic input available to individuals; the fully developed signal is compared to non-treated controls or across treatments to ask whether altering the stimuli alters the course of development (e.g. Marler and Peters, 1977; Podos et al., 2004). The

second is by examining naturally occurring patterns of acoustic variation in a particular vocalization and associating these to patterns of social interactions. Such studies might focus on detecting statistical associations between the songs of parents and offspring, or they might look for geographic patterns, such as vocal dialects, that suggest learning of local call types (Marler and Tamura, 1962). Although such studies are more naturalistic, they typically lack the same degree of experimenter control as exposure studies, and extra care must be taken to rule out the influence of alternative explanations for acoustic similarity, such as genetic relatedness or shared environmental factors (e.g. Forstmeier et al., 2009; Wright and Wilkinson, 2001). A notable exception is a recent study that elegantly controlled auditory exposure of free-living birds to experimentally demonstrate both learning and cultural transmission of novel song types in an island population of savannah sparrows, *Passerculus sandwichensis* (Mennill et al., 2018). The third approach to testing whether a vocalization is learned is altering the underlying neural substrate responsible for learning. Investigators have used lesions, genetic tools, electrophysiology, and pharmacological manipulations to interfere with specific regions of the brain and examine the results on developed vocalizations (e.g. Burkett et al., 2018; Plummer and Striedter, 2002, see also Box 2). While such experiments rely on a theoretical framework for the regulation of vocal learning that is still under development, they can be effective in determining whether currently known mechanisms of learning are involved in the development of specific vocalizations in a given species.

Box 2.

Molecular tools for manipulating learning

Newly developed molecular tools offer a promising approach to understanding both the mechanisms and adaptive significance of variation in specific dimensions of vocal learning. Tools such as RNA-interference knockdowns, virally-mediated overexpression, or CRISPR-CAS gene editing can be used to alter the expression of targeted genes within specific neural regions to test for effects on specific vocal learning dimensions. Manipulations that alter learning behavior can also be used to test the adaptive significance of this behavior. One example of this approach comes from the work of White and colleagues on the role of FoxP2 in vocal learning in the zebra finch. They tested the plasticity gateway hypothesis discussed in Dimension 3 by increasing the expression of FoxP2 in the Area X of both juvenile and adult male zebra finches using an adeno-associated virus (Burkett et al., 2018; Day et al., 2019; Heston and White, 2015). As predicted, this manipulation reduced the accuracy of vocal learning in juvenile birds (Box 2 Figure 1) (Burkett et al., 2018; Heston and White, 2015). Effects were more subtle in adults, who had already ended their learning by the time of manipulation. Preference trials showed that females preferred the songs of control males to those that had experienced FoxP2 overexpression (Day et al., 2019). Similar experiments are underway in adult budgerigars, which do learn new contact calls and show chronic FoxP2 under-expression in their Area X analog, MMST; as predicted, birds with FoxP2 overexpression in MMST show reduced call learning compared to controls (G. Kohn, T. Wright et al, in prep). There is also increasing interest in other genes in the FoxP family,

including FoxP1 and FoxP4, both of which appear to affect the learning phenotype differently than FoxP2 (Norton et al., 2019). Such work is technically demanding and requires preliminary knowledge of the vocal learning circuits of a species. But we are rapidly approaching the point at which these techniques can be used more widely across species to test comparative hypotheses for the underlying mechanisms and functional significance of variation in specific dimensions of vocal learning.
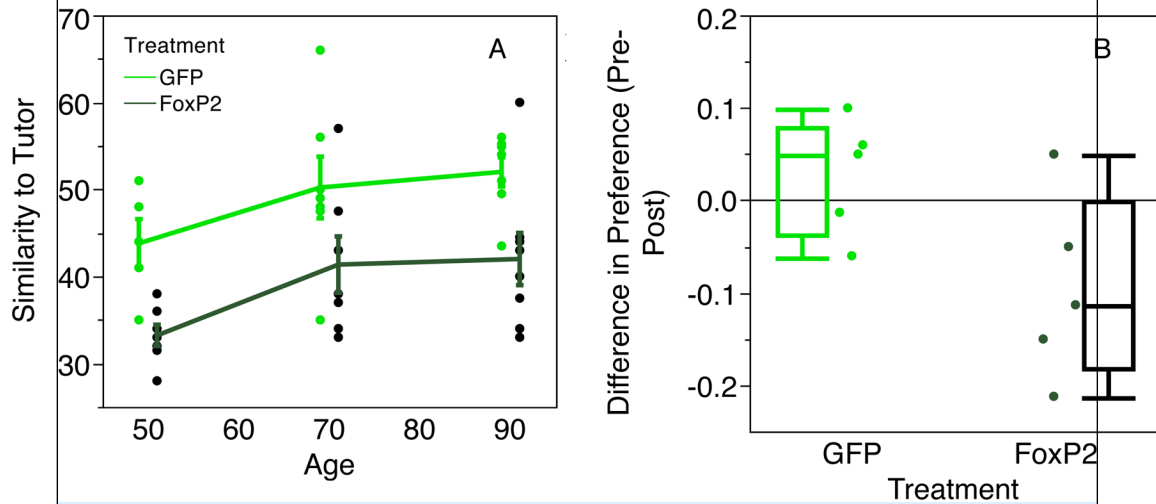
Box 2 Figure 1.

a) Virally-mediated overexpression of the full-length FoxP2 protein (FoxP2.FL) in the songbird vocal learning region Area X reduces the ability of juvenile zebra finches to match the song of their adult tutors relative to that seen in controls with green fluorescent protein (GFP) expression. B) In adult males, this same manipulation of FoxP2 expression causes males to produce song that is less preferred by females than is song from control GFP males. These targeted manipulations of expression levels reveal the importance of this gene in controlling vocal learning, and illustrate how molecular manipulations can provide a powerful tool for testing hypotheses about the functional significance of variation in different vocal learning dimensions. Panel a) redrawn with permission of authors from (Burkett et al., 2018), panel b) redrawn with permission of authors from (Day et al., 2019).

References

Ali S, Anderson R, 2018. Song and aggressive signaling in Bachman's sparrow. Auk 135, 521–533.

Araki M, Bandi MM, Yazaki-Sugiyama Y, 2016. Mind the gap: neural coding of species identity in birdsong prosody. Science 354, 1282–1287. [PubMed: 27940872]

Baptista LF, 1990. Dialectical variation in the raincall of the chaffinch. Die Vogelwarte 35, 249–256.

Beecher MD, 2008. Function and mechanisms of song learning in song sparrows, in: Brockmann HJ, Roper TJ, Naguib M, Wynne-Edwards KE, Barnard C, Mitani JC (Eds.), Advances in the Study of Behavior, Vol 38. Elsevier Academic Press Inc, San Diego, pp. 167–225.

Beecher MD, 2017. Birdsong learning as a social process. Animal Behaviour 124, 233–246.

Beecher MD, Brenowitz EA, 2005. Functional aspects of song learning in songbirds. Trends in Ecology & Evolution 20, 143–149. [PubMed: 16701358]

Beecher MD, Campbell SE, Nordby JC, 2000. Territory tenure in song sparrows is related to song sharing with neighbours, but not to repertoire size. Animal Behaviour 59, 29–37. [PubMed: 10640364]

Bradbury JW, Vehrencamp SL, 2014. Complexity and behavioral ecology. Behavioral Ecology 25, 435–442.

Brainard MS, Doupe AJ, 2002. What songbirds teach us about learning. Nature 417, 351–358. [PubMed: 12015616]

Brenowitz EA, Beecher MD, 2005. Song learning in birds: diversity and plasticity, opportunities and challenges. Trends in Ecology and Evolution 20, 127–132.

Brittan-Powell EF, Dooling RJ, Farabaugh SM, 1997. Vocal development in budgerigars (*Melopsittacus undulatus*): Contact calls. Journal of Comparative Psychology 111, 226–241. [PubMed: 9286093]

Burkett ZD, Day NF, Kimball TH, Aamodt CM, Heston JB, Hilliard AT, Xiao X, White SA, 2018. FoxP2 isoforms delineate spatiotemporal transcriptional networks for vocal learning in the zebra finch. eLife 7.

Burns KJ, Shultz AJ, Title PO, Mason NA, Barker FK, Klicka J, Lanyon SM, Lovette IJ, 2014. Phylogenetics and diversification of tanagers (Passeriformes: Thraupidae), the largest radiation of Neotropical songbirds. Molecular Phylogenetics and Evolution 75, 41–77. [PubMed: 24583021]

Burt JM, Campell SE, Beecher MD, 2001. Song type matching as threat: A test using interactive playback. Animal Behaviour 62, 1163–1170.

Byers BE, Kroodsma DE, 2009. Female mate choice and songbird song repertoires. Animal Behaviour 77, 13–22.

Carouso-Peck S, Goldstein MH, 2019. Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. Curr Biol 29, 631–636.e633. [PubMed: 30713105]

Catchpole CK, Slater PJB, 2008. Bird Song: Biological Themes and Variations, 2 ed. Cambridge University Press, Cambridge.

Chakraborty M, Jarvis ED, 2015. Brain evolution by brain pathway duplication. Philosophical Transactions of the Royal Society B-Biological Sciences 370, 12.

Chakraborty M, Walløe S, Nedergaard S, Fridel EE, Dabelsteen T, Pakkenberg B, Bertelsen MF, Dorrestein GM, Brauth SE, Durand SE, Jarvis ED, 2015. Core and shell song systems unique to the parrot brain. PLoS ONE 10, e0118496. [PubMed: 26107173]

Chen YN, Matheson LE, Sakata JT, 2016. Mechanisms underlying the social enhancement of vocal learning in songbirds. Proceedings of the National Academy of Sciences of the United States of America 113, 6641–6646. [PubMed: 27247385]

Dahlin CR, Young A, Halford D, Cordier B, Wright TF, 2014. The function of vocal convergence in female budgerigars, *Melopsittacus undulatus*. Behavioral Ecology and Sociobiology 68.

Day NF, Hobbs TG, Heston JB, White SA, 2019. Beyond critical period learning: striatal FoxP2 affects the active maintenance of learned vocalizations in adulthood. eneuro 6, ENEURO.0071–0019.2019.

Deregnaucourt S, Gahr M, 2013. Horizontal transmission of the father's song in the zebra finch (*Taeniopygia guttata*). Biol. Lett 9, 4.

Devoogd TJ, Krebs JR, Healy SD, Purvis A, 1993. Relations betwen song repertoire size and the volume of brain nuclei related to song - comparative evolutonary analysis amongst oscine birds. Proceedings of the Royal Society B-Biological Sciences 254, 75–82.

Farabaugh SM, Dooling RJ, 1996. Acoustic communication in parrots: laboratory and field studies of budgerigars, *Melopsittacus undulatus*, in: Kroodsma DE, Miller EH (Eds.), Ecology and Evolution of Acoustic Communication in Birds. Cornell University Press, Ithaca, New York, pp. 97–117.

Fisher SE, Marcus GF, 2006. The eloquent ape: genes, brains and the evolution of language. Nature Reviews Genetics 7, 9.

Forstmeier W, Burger C, Temnow K, Deregnaucourt S, 2009. The genetic basis of zebra finch vocalizations. Evolution 63, 2114–2130. [PubMed: 19453380]

Garamszegi LZ, Eens M, 2004. Brain space for a learned task: strong intraspecific evidence for neural correlates of singing behavior in songbirds. Brain Research Reviews 44, 187–193. [PubMed: 15003393]

Gardner JL, Trueman JW, Ebert D, Joseph L, Magrath RD, 2010. Phylogeny and evolution of the Meliphagoidea, the largest radiation of Australasian songbirds. Molecular Phylogenetics and Evolution 55, 1087–1102. [PubMed: 20152917]

Gobes SMH, Jennings RB, Maeda RK, 2019. The sensitive period for auditory-vocal learning in the zebra finch: Consequences of limited-model availability and multiple-tutor paradigms on song imitation. Behav Processes 163, 5–12. [PubMed: 28743517]

Goller M, Shizuka D, 2018. Evolutionary origins of vocal mimicry in songbirds. Evolution Letters 2, 417–426. [PubMed: 30283692]

Grant BR, Grant PR, 1996. Cultural inheritance of song and its role in the evolution of Darwin's finches. Evolution 50, 2471–2487. [PubMed: 28565664]

Guttinger HR, 1985. Consequences of domestication on the song structures in the canary. Behaviour 94, 254–278.

Haesler S, Rochefort C, Georgi B, Licznerski P, Osten P, Scharff C, 2007. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. PLoS Biology 5, 2885–2897.

Haesler S, Wada K, Nshdejan A, Morrisey EE, Lints T, Jarvis ED, Scharff C, 2004. FoxP2 expression in avian vocal learners and non-learners. J Neurosci 24, 3164–3175. [PubMed: 15056696]

Hara E, Perez JM, Whitney O, Chen Q, White SA, Wright TF, 2015. Neural FoxP2 and FoxP1 expression in the budgerigar, an avian species with adult vocal learning. Behavioural brain research 283, 22–29. [PubMed: 25601574]

Heston JB, White SA, 2015. Behavior-linked FoxP2 regulation enables zebra finch vocal learning. J. Neurosci 35, 2885–2894. [PubMed: 25698728]

Howard RD, 1974. Influence of sexual selection and interspecific competition on mockingbird song (Mimulus-polyglottos). Evolution 28, 428–438. [PubMed: 28564842]

Hultsch H, Todt D, 2004. Learning to sing, in: Marler P, Slabbekoorn H (Eds.), Natures' Music: The Science of Birdsong. Elsevier, San Diego, pp. 80–107.

Immelmann K, 1969. Song development in the zebra finch and other estrelid finches, in: Hinde RA (Ed.), Bird Vocalizations. Cambridge University Press, Cambridge, pp. 64–74.

Janik VM, 2014. Cetacean vocal learning and communication. Curr. Opin. Neurobiol 28, 60–65. [PubMed: 25057816]

Jarvis ED, 2019. Evolution of vocal learning and spoken language. Science 366, 50–54. [PubMed: 31604300]

Jetz W, Thomas G, Joy J, Hartmann K, Mooers A, 2012. The global diversity of birds in space and time. Nature 491, 444. [PubMed: 23123857]

Jombart T, Balloux F, Dray S, 2010. Adephylo: new tools for investigating the phylogenetic signal in biological traits. Bioinformatics 26, 1907–1909. [PubMed: 20525823]

Jones AE, Slater PJB, 1996. The role of aggression in song tutor choice in the zebra finch: Cause or effect? Behaviour 133, 103–115.

Krebs J, Ashcroft R, Webber M, 1978. Song repertoires and territory defence in the great tit. Nature 271, 539.

Kroodsma DE, Canady RA, 1985. Differences in repertoire size, singing behavior, and associated neuroanatomy among marsh wren populations have a genetic basis. The Auk 102, 439–446.

Lachlan RF, Slater PJB, 2003. Song learning by chaffinches: how accurate, and from where? Animal Behaviour 65, 957–969.

Lachlan RF, Verhagen L, Peters S, ten Cate C, 2010. Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (Fringilla coelebs), zebra finches (Taenopygia guttata), and swamp sparrows (Melospiza georgiana). Journal of Comparative Psychology 124, 92. [PubMed: 20175600]

Lai CSL, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP, 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 413, 519–523. [PubMed: 11586359]

Lattenkamp EZ, Vernes SC, 2018. Vocal learning: a language-relevant trait in need of a broad cross-species approach. Current opinion in behavioral sciences 21, 209–215.

Lynch A, Baker AJ, 1993. A population memetics approach to cultural evolution in chaffinch song: meme diversity within populations. The American Naturalist 141, 597–620.

MacDougall-Shackleton SA, 1997. Sexual selection and the evolution of song repertoires. Current Ornithology 14, 81–124.

Mann NI, Slater PJB, 1994. What causes young male zebra finches, Taeniopygia guttata, to choose their father as song tutor? Animal Behaviour 47, 671–677.

Mann NI, Slater PJB, 1995. Song tutor choice by zebra finches in aviaries. Animal Behaviour 49, 811–820.

Marler P, 1984. Song learning: Innate species differences in the learning process, in: Marler P, Terrace HS (Eds.), The Biology of Learning Springer-Verlag, Berlin, pp. 289–309.

Marler P, 2004a. Bird calls: a cornucopia for communication, in: Marler P, Slabbekoorn H (Eds.), Nature's Music: The Science of Birdsong. Elsevier, San Diego, pp. 132–177.

Marler P, 2004b. Science and birdsong: the good old days, in: Marler P, Slabbekoorn H (Eds.), Natures' Music: The Science of Birdsong. Elsevier, San Diego, pp. 1–38.

Marler P, Peters S, 1977. Selective vocal learning in a sparrow. Science 198, 519–521. [PubMed: 17842140]

Marler P, Peters S, 1988. The role of song phonology and syntax in vocal learning preferences in the song sparrow, Melospiza melodia. Ethology 77, 125–149.

Marler P, Pickert R, 1984. Species-universal microstructure in the learned song of the swamp sparrow (Melospiza georgiana). Animal Behaviour 32, 673–689.

Marler P, Tamura M, 1962. Song 'dialects' in three populations in white-crowned sparrows. The Condor 64, 368–377.

Mason NA, Burns KJ, Tobias JA, Claramunt S, Seddon N, Derryberry EP, 2017. Song evolution, speciation, and vocal learning in passerine birds. Evolution 71, 786–796. [PubMed: 28012174]

McGregor PK, Krebs JR, 1989. Song learning in adult great tits (Parus major): effects of neighbours. Behaviour, 139–159.

McGuire JA, Witt CC, Remsen JV, Corl A, Rabosky DL, Altshuler DL, Dudley R, 2014. Molecular phylogenetics and the diversification of hummingbirds. Current Biology 24, 910–916. [PubMed: 24704078]

Medina-García A, Araya-Salas M, Wright TF, 2015. Does vocal learning accelerate acoustic diversification? Evolution of contact calls in Neotropical parrots. Journal of Evolutionary Biology 28, 1782–1792. [PubMed: 26189657]

Mello C, Vicario D, Clayton D, 1992. Song presentation induces gene expression in the songbird forebrain. Proceedings of the National Academy of Sciences of the United States of America 89, 6818–6822. [PubMed: 1495970]

Mennill DJ, Doucet SM, Newman AEM, Williams H, Moran IG, Thomas IP, Woodworth BK, Norris DR, 2018. Wild birds learn songs from experimental vocal tutors. Current Biology 28, 3273–3278.e3274. [PubMed: 30293720]

Miller JE, Hilliard AT, White SA, 2010. Song practice promotes acute vocal variability at a key stage of sensorimotor learning. PloS one 5, e8592–e8592. [PubMed: 20066039]

Mischler SK, Karlin EJ, MacDougall-Shackleton SA, 2020. Call production induces motor-driven ZENK response in the song control system of black-capped chickadees. Animal Behaviour 163, 145–153.

Mundinger PC, Lahti DC, 2014. Quantitative integration of genetic factors in the learning and production of canary song. Proceedings of the Royal Society B-Biological Sciences 281.

Murugan M, Harward S, Scharff C, Mooney R, 2013. Diminished FoxP2 levels affect dopaminergic modulation of corticostriatal signaling important to song variability. Neuron 80, 1464–1476. [PubMed: 24268418]

Nelson DA, Marler P, Palleroni A, 1995. A comparative approach to vocal learning: Intraspecific variation in the learning process. Animal Behaviour 50, 83–97.

Norton P, Barschke P, Scharff C, Mendoza E, 2019. Differential song deficits after lentivirus-mediated knockdown of FoxP1, FoxP2, or FoxP4 in Area X of juvenile zebra finches. The Journal of Neuroscience 39, 9782–9796. [PubMed: 31641053]

Nottebohm F, 1981. A brain for all seasons - cyclical anatomic changes in song control Science 214, 1368–1370.

Nottebohm F, Nottebohm ME, Crane L, 1986. Developmental and seasonal-changes in canary song and their relation to changes in the anatomy of song-control nuclei. Behav. Neural Biol 46, 445–471. [PubMed: 3814048]

Nowicki S, Searcy WA, 2014. The evolution of vocal learning. Curr. Opin. Neurobiol 28, 48–53. [PubMed: 25033109]

O'Meara BC, Ané C, Sanderson MJ, Wainwright PC, 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution 60, 922–933. [PubMed: 16817533]

Odom KJ, Hall ML, Riebel K, Omland KE, Langmore NE, 2014. Female song is widespread and ancestral in songbirds. Nature Communications 5.

Payne RB, 2006. Indigo Bunting (*Passerina cyanea*), version 2.0., in: Poole AF (Ed.), The Birds of North America. Cornell Lab of Ornithology,, Ithaca, NY.

Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, Roulhac PL, Howard JT, Wirthlin M, Lovell PV, Ganapathy G, Mouncastle J, Moseley MA, Thompson JW, Soderblom EJ, Iriki A, Kato M, Gilbert MTP, Zhang G, Bakken T, Bongaarts A, Bernard A, Lein E, Mello CV, Hartemink AJ, Jarvis ED, 2014. Convergent transcriptional specializations in the brains of humans and song-learning birds. Science 346, 1333-+.

Plummer TK, Striedter GF, 2002. Brain lesions that impair vocal imitation in adult budgerigars. Journal of Neurobiology 53, 413–428. [PubMed: 12382268]

Podos J, Peters S, Nowicki S, 2004. Calibration of song learning targets during vocal ontogeny in swamp sparrows, Melospiza georgiana. Animal Behaviour 68, 929–940.

Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R, 2009. Neural correlates of categorical perception in learned vocal communication. Nature neuroscience 12, 221. [PubMed: 19136972]

Provost KL, Joseph L, Smith BT, 2018. Resolving a phylogenetic hypothesis for parrots: implications from systematics to conservation. Emu-Austral Ornithology 118, 7–21.

Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, Huang H, Larson JG, 2014. BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution 5, 701–707.

Revell LJ, 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217–223.

Riebel K, Odom KJ, Langmore NE, Hall ML, 2019. New insights from female bird song: towards an integrated approach to studying male and female communication roles. Biol. Lett 15.

Searcy WA, Akcay C, Nowicki S, Beecher MD, 2014. Aggressive signaling in song sparrows and other songbirds, in: Naguib M, Barrett L, Brockmann HJ, Healy S, Mitani JC, Roper TJ, Simmons LW (Eds.), Advances in the Study of Behavior, Vol 46. Elsevier Academic Press Inc, San Diego, pp. 89–125.

Searcy WA, Nowicki S, Hughes M, Peters S, 2002. Geographic song discrimination in relation to dispersal distances in song sparrows. Am. Nat 159, 221–230. [PubMed: 18707375]

Sewall KB, 2009. Limited adult vocal learning maintains call dialects but permits pair-distinctive calls in red crossbills. Anim Behav 77, 1303–1311.

Sewall KB, Young AM, Wright TF, 2016. Social calls provide novel insights into the evolution of vocal learning. Animal Behaviour 120, 163–172. [PubMed: 28163325]

Simpson HB, Vicario DS, 1990. Brain pathways for learned and unlearned vocalizations differ in zebra finches Journal of Neuroscience 10, 1541–1556. [PubMed: 2332796]

Slater GJ, Harmon LJ, Wegmann D, Joyce P, Revell LJ, Alfaro ME, 2012. Fitting models of continuous trait evolution to incompletely sampled comparative data using approximate Bayesian computation. Evolution 66, 752–762. [PubMed: 22380438]

Slater PJB, Jones AE, 1995. The timing of song and distance call learning in zebra finches. Animal Behaviour 49, 548–550.

Soha J, 2017. The auditory template hypothesis: a review and comparative perspective. Animal Behaviour 124, 247–254.

Soma M, Garamszegi LZ, 2011. Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. Behavioral Ecology 22, 363–371.

Spector DA, 1994. Definition in Biology: The Case of "Bird Song". Journal of Theoretical Biology 168, 373–381.

Ter Maat A, Trost L, Sagunsky H, Seltmann S, Gahr M, 2014. Zebra finch mates use their forebrain song system in unlearned call communication. Plos One 9.

Teramitsu I, White SA, 2006. FoxP2 regulation during undirected singing in adult songbirds. J Neurosci 26, 7390–7394. [PubMed: 16837586]

Todt D, 1971. Aquivalente und konvalente gesangliche Reaktion einer extrem regelmaBig singenden Nachtigall. Zeitschrift für vergelichende physiologie 71, 262–285.

Todt D, Böhner J, 1994. Former experience can modify social selectivity during song learning in the nightingale (Luscinia megarhynchos). Ethology 97, 169–176.

Wang X, Park J, Susztak K, Zhang NR, Li M, 2019. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. Nature Communications 10, 380.

Whitney O, Pfenning AR, Howard JT, Blatti CA, Liu F, Ward JM, Wang R, Audet JN, Kellis M, Mukherjee S, Sinha S, Hartemink AJ, West AE, Jarvis ED, 2014. Core and region-enriched networks of behaviorally regulated genes and the singing genome. Science 346, 1334-+.

Whitney O, Voyles T, Hara E, Chen Q, White SA, Wright TF, 2015. Differential FoxP2 and FoxP1 expression in a vocal learning nucleus of the developing budgerigar. Dev Neurobiol 75, 778–790. [PubMed: 25407828]

Williams H, 1990. Models for song learning in the zebra finch - fathers or others. Animal Behaviour 39, 745–757.

Wirthlin M, Chang EF, Knörnschild M, Krubitzer LA, Mello CV, Miller CT, Pfenning A, Vernes SC, Tchernichovski O, Yartsev MM, 2019. A modular approach to vocal learning: disentangling the diversity of a complex behavioral trait. Neuron 104, 87–99. [PubMed: 31600518]

Wright TF, Brittan-Powell EF, Dooling RJ, Mundinger PC, 2004. Sex-linked inheritance of hearing and song in the Belgian Waterslager canary. Proceedings of the Royal Society B-Biological Sciences 271, S409–S412.

Wright TF, Dahlin CR, 2018. Vocal dialects in parrots: patterns and processes of cultural evolution. Emu 118, 50–66. [PubMed: 29962561]

Wright TF, Wilkinson GS, 2001. Population genetic structure and vocal dialects in an amazon parrot. Proceedings of the Royal Society of London, series B 268, 609–616. [PubMed: 11297178]
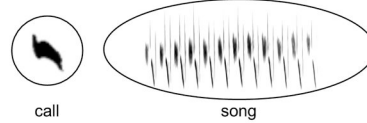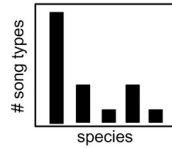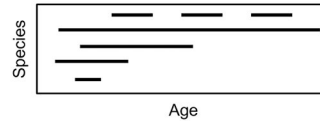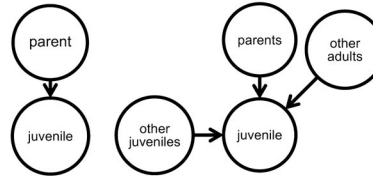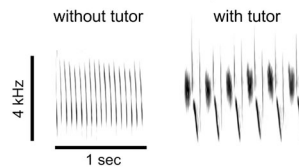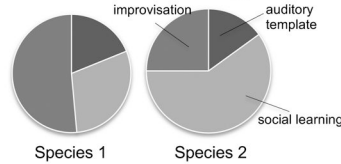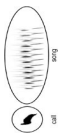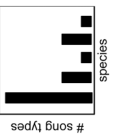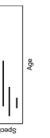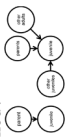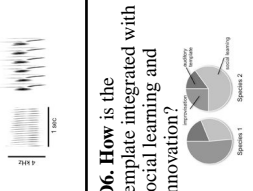
Figure 1.
Visualization of six dimensions of vocal production learning. Dimension 1 (D1) is illustrated by a red crossbill, *Loxia curvirostra* Type 2 call (left) and a swamp sparrow, *Melospiza georgiana* song (right). Dimension 2 is illustrated by species-level variation in the number of different song types. Dimension 3 is illustrated by variation among species in the timing of learning. Dimension 4 is illustrated by a contrast between a species that learns primarily from one individual, as in the medium ground finch, *Geospiza fortis* or from multiple sources, such as in the zebra finch, *Taeniopygia guttata*. Dimension 5 is illustrated by an approach to characterizing the internal template by comparing vocalizations of individuals with and without tutors, as illustrated by these spectrograms from a song learning experiment with swamp sparrows. Dimension 6 is illustrated by a contrast between two species that differ in the degree to which their vocalizations are shaped by their template, social learning, and improvisation.

Table 1.

Overview of six dimensions of vocal production learning, including evidence for variation and functional and mechanistic hypotheses.

Vocal Dimensions	Evidence for Variation	Functional Hypotheses	Mechanistic Hypotheses
D1. Which vocalizations are learned?	• Species can have both learned and innate vocalizations, e.g. the zebra finch (Slater and Jones, 1995; Ter Maat et al., 2014) • Vocal categories that are innate in many species can be learned in some e.g., learned contact calls in the red crossbill, *Loxia curvirostris* (Sewall, 2009) and many parrots (Wright and Dahlin, 2018)	• Sexual selection via mate preference for more complex signals (Nowicki and Searcy, 2014) • Flexible labeling of relationships within dynamic social systems (Sewall et al., 2016)	• Same neural centers for learning songs and calls (Simpson and Vicario, 1990; Ter Maat et al., 2014) • Different neural substrates for learned and unlearned vocalizations (Chakraborty and Jarvis, 2015)
D2. How many are learned?	• Closely related species can vary widely in # of song types, e.g., within the Passerellidae clade, single song repertoires of white-crowned sparrow, *Zonotrichia leuophrys* (Marler and Tamura, 1962) to > 50 distinct songs in repertoire of Bachman's sparrow, *Peucaea aestivalis* (Ali and Anderson, 2018)	• Sexual selection via mate preferences for larger repertoires (mixed support) (Byers and Kroodsma, 2009) • Selection for communicative complexity (Byers and Kroodsma, 2009)	• Volume of neural song production centers positively correlates with repertoire size (Devoogd et al., 1993; Garamszegi and Eens, 2004)
D3. When it is learned?	• Continuum of close-ended to open-ended learners, e.g., the song sparrow critical period ends early in life (Beecher, 2008), canaries reopen a critical period annually (Nottebohm et al., 1986), while the budgerigar learns continuously throughout life (Dahlin et al., 2014). • Learning phases also vary in duration and degree of overlap, e.g., the sensory and sensorimotor phases overlap in zebra finches but not in white-crowned sparrows (Hultsch and Todt, 2004)	• Mate preferences for open-ended learning for continual expansion of vocal repertoire, e.g., in northern mockingbirds, *Mimus polyglottos* (Howard, 1974) • Selection for close-ended learning to reduce physiological or social costs to learning, e.g., critical learning period of white-crowned sparrow subspecies varies with migratory behavior (Nelson et al., 1995)	• Plasticity gateway hypothesis: expression levels of the *FoxP2* transcription factor modify synaptic plasticity required for vocal learning (Miller et al., 2010) (See Box 1)
D4. From whom is it learned?	• Social influences may be vertical, oblique, horizontal or a social network (Lynch and Baker, 1993). The medium ground finch, *Geospiza fortis* learns primarily from the father (Grant and Grant, 1996) whereas budgerigars' contact calls morph from unlearned begging calls to calls shared with siblings and then a social network (Brittan-Powell et al., 1997)	• Selection to avoid non-adaptive inbreeding and hybridization (Grant and Grant, 1996) • Song sharing increases resource holding potential (Beecher et al., 2000) • Call sharing increases affiliative interactions (no current support) (Dahlin et al., 2014)	• Availability of vocal models during critical learning periods, depends on dispersal, social organization, dimensions 3 and 5 • Mechanisms that govern social attention, e.g., gene expression in catecholaminergic neurons with projections to song learning circuits (Chen et al., 2016)
D5. What is the extent of the internal template?	• Variation in the extent and specificity of auditory templates, e.g., swamp sparrows, *Melospiza georgiana* show innate preferences for species specific syllables but not species typical temporal	• Adaptive value of the degree and form of the internal template is largely not articulated beyond	• Species specific temporal patterning encoded in Field L, part

Vocal Dimensions	Evidence for Variation	Functional Hypotheses	Mechanistic Hypotheses
	patterning whereas song sparrows show the reverse (Marler, 1984)	facilitating learning of conspecific song (Marler 1984).	of the ascending auditory pathway (Araki et al., 2016)
D6. How is the template integrated with social learning and innovation?	• Variation in species universals, e.g., 100 shared elements in indigo buntings (Payne, 2006) compared to 6 shared elements in swamp sparrows, *Melospiza georgiana* (Marler and Pickert, 1984) • Mimicry is evolutionarily labile, e.g., in oscine songbirds gained at least 237 times and lost at least 52 times (Goller and Shizuka, 2018)	• Mate preferences for individually distinctive or more novel repertoires promote innovation • Mate preferences for increased repertoire size or complexity lead to mimicry via reduced internal template	• Duplication and elaboration of core and shell sub-regions in vocal learning circuit in parrots may facilitate mimicry (learning heterospecific vocalizations)