Vocal learning beyond imitation: mechanisms of adaptive vocal development in songbirds and human infants

Ofer Tchernichovski[1] and **Gary Marcus**[2]

[1]Dept. of Psychology, Hunter College, City University of New York

[2]Dept. of Psychology, New York University

Abstract

Studies of vocal learning in songbirds typically focus on the acquisition of sensory templates for song imitation and on the consequent process of matching song production to templates. However, functional vocal development also requires the capacity to adaptively diverge from sensory templates, and to flexibly assemble vocal units. Examples of adaptive divergence include the corrective imitation of abnormal songs, and the decreased tendency to copy overabundant syllables. Such frequency-dependent effects might mirror tradeoffs between the assimilation of group identity (culture) while establishing individual and flexibly expressive songs. Intriguingly, although the requirements for vocal plasticity vary across songbirds, and more so between birdsong and language, the capacity to flexibly assemble vocal sounds develops in a similar, stepwise manner across species. Therefore, universal features of vocal learning go well beyond the capacity to imitate.

Vocal imitation is among the most mysterious forms of developmental learning. In humans vocal play (babbling) begins shortly after birth, and it develops gradually over the first year of life [1]. Early babbling vocalizations are composed of simple and unstructured sounds, but within several months one can identify vocal elements that are clearly derived from the native language, and those structured sounds then further evolve into words [2]. However, little is known about the developmental trajectories leading from primitive vocalizations into specific words, and about the brain mechanisms involved in developmental vocal learning in humans. Much more is known about vocal development in songbirds. Song development goes through similar stages to those of infant babbling [3]. Juvenile songbirds can copy song syllables of adult birds ('tutors') with striking accuracy [4][5]. It is now possible to track trajectories of vocal changes leading to imitation continuously over development [6], and to identify the emergence of syllable types and the differentiation of prototype syllables [6,7]. Song learning and production is controlled by a set of distinct brain nuclei collectively called 'the song-system' [8]9]. The role of specific song nuclei in song production and learning is relatively well understood, including mechanisms of producing rhythm [10]

[11,12], spectral features [13][14], and gestures [15]. Patterns of gene expressions in song nuclei [16–18] were identified and associated with the song learning process. Even the role of transitory behavioral states, such as sleep, on song learning can be examined at behavioral and brain levels [19–21] over development. We understand fairly well the role of reinforcement learning [22] in correcting vocal errors [23] and in matching the fine temporal structure of the sensory template [24],[25]. Of particular interest is how vocal exploration (variability) is actively regulated by specific song nuclei during the natural time course of song learning [26][27].

However, the success of birdsong neuroscience in discovering mechanisms of vocal imitation detracts attention from the fact that *exact* vocal imitation is not the typical developmental endpoint in songbirds [28] or in human infants [29]. For example, over-regularizations during early speech development are not direct imitations of the caregivers [30,31]. Further, language capacities of infants may surpass that of their parents [31] or otherwise diverge from the spoken language they perceive leading to a rapid language changes across generations [30]. Similar effects were observed in songbirds. When a single juvenile zebra finch is raised together with a single adult bird (tutor) a near-exact match to tutor song typically emerges [28]. However, if instead of one-to-one vocal tutoring, five juvenile males are raised together with a single tutor, some birds will diverge from the tutor song by copying only parts of it [28,32]. Juvenile birds often copy song elements from each other [33,34], or improvise and remix syllables across tutors to create new song types [35,36]. Further, in both songbirds and human infants social feedback can strongly affect vocal development [37]. Overall, imitation per se is only one component of a complex vocal development, and we suggest that song learning can be used for studying vocal learning beyond imitation, as a model for language development.

Vocal learning capacities beyond imitation

Figure 1 illustrates that vocal learning is an iterative process with a strong inertia, such that vocal changes transform already-established vocal sounds. Quite often, the scope of those vocal changes is highly localized to a specific syllable type [26]. Vocal changes are subject to strong developmental constraints. For example, HVC is the principal song nucleus that drives song production in adult birds. However, HVC has no apparent role in the production of subsong (babbling) in juvenile birds. Instead, the early subsong is driven by a set of forebrain song nuclei, collectively called the Anterior Forebrain Pathway (AFP) [38]. During song learning, there is a gradual transition of control from AFP, which produces noisy and exploratory song patterns, to HVC, which produces stereotyped song patterns. Song learning takes place within the framework of this brain development [39].

How do those internal constraints and external social and auditory inputs shape vocal changes that eventually add up to an adaptive developmental outcome? Accurate imitation is an important capacity that allows birds to generate songs that attract mates, deter intruders, and carry information about group identity (song dialects) via cultural transmission [40]. But songs also carry information about the individual identity of the bird, and -- perhaps most importantly -- they should provide information about the qualities of the singing bird. From

this functional perspective, the cases where the outcome of song learning is *not* exact imitation are most interesting.

Range limited and frequency dependent vocal imitation

Spoken language can change remarkably across generations [30], and so does birdsong [41]. As noted earlier, in one-to-one tutoring setups zebra finches usually imitate their tutor very accurately. But there is an important exception to this rule, which is the biased imitation of 'abnormal' songs [42]. For example, canaries' songs include several back-to-back renditions (trills, rolls & tours) of the same note type (e.g., AAAA,BBB,…). Training canaries with species-atypical playbacks such as ABCDE… that lack repetition, initially results in imitation of that atypical song. But when the bird approaches sexual maturity (or after an injection of testosterone), the bird abruptly transforms its song into the typical reduplicative AAA,BBB… pattern – which it never heard [43].

Another example is the imitation of isolates' songs in zebra finches [41]. Zebra finches raised in complete social and acoustic isolation develop abnormal song, called isolate song. Abnormalities include long and monotonous syllable types, and deviation from the species-typical song syntax, wherein each syllable type normally appears only once (non-reduplicatively) within each song motif. Intriguingly, when juvenile zebra finches imitate isolates' songs they copy them in a biased manner, transforming acoustic features, rhythms and song-syntax to approximate the wild-type song – which they never heard. Similarly, human children can "regularize" inconsistent input in sign languages [31]. In songbirds, such biases can be studied experimentally to promote mechanistic understanding:

As shown in Figure 1, vocal learning might be thought of as an iterative function that transforms an input (say a song syllable heard), through vocal changes, onto a specific endpoint vocalization (a mature syllable produced). Within some range of variation, that mapping appears as if it were a simple identity function, in which learners acquire near perfect copies of the input. Outside that normal range, however, learners no longer directly imitate their models [42]; instead, they often "regularize" aberrant input towards the central tendencies of the wild type. Tracking the imitation of isolate zebra finch songs provides an opportunity to characterize the regularization function in a manner that can uncover mechanisms. For example, the imitation of isolate syllables of abnormally long duration often results in shorter copies (Fig. 2A). To quantify this effect, we plot the duration of isolate-song syllables versus the duration of the imitation of those syllables, across the entire range of syllable durations present in isolate songs of several birds (Fig. 2B)[41]. As shown, for up to about 270ms, imitation of syllable duration is faithful; however, beyond that range, imitation of duration becomes systematically shorter than that of the sensory template. The trivial implication of this effect is that across generations of learners, iterations of learning must eventually result in a distribution of syllable duration where 270ms is the upper bound –which is indeed the case in wild type zebra finch songs. That is, the wild type range of syllable durations can be fully explained by recursions of this multigenerational transform, as opposed to an alternative Darwinian processes such as sexual selection, where the distribution of durations could have been determined by females mating preferences, followed by imitation of father's song in the next generation.

Figure 2C shows another example of the imitation of isolate zebra finch song. Here there are only two syllable types in the song, and one of them (B) is about 80% abundant, which is highly abnormal. The imitation of this song already appears fairly 'normalized' (similar to wild-type songs). A vocal learning graph across the entire range of syllable abundance (relative frequency) present in isolate songs (Fig. 2D) shows that as long as the abundance of a syllable type is 30% or less, acquired song faithfully reflects the model, but above that range, the relative frequency is no longer copied; here the mapping asymptotes at about 25%, again closely matching the upper-bound of wild type range of syllable abundance in a wild-type zebra finch song. The interpretation of this effect is very interesting: it is a special case of a *negative frequency dependent selection* [44] -- a mechanism that can maintain polymorphism. For example, if a predator develops preferences to the most abundant morph of its pray, this morph is selected against and becomes rare in the next generation. In this way, the frequency of rare morphs increases, preventing extinction by genetic drift. Here, a negative frequency dependent imitation might serve to preserve polymorphism in syllable-type diversity in the population, which is likely to promote both individual and group identity in the song repertoire of individuals.

In sum, studying the mappings from auditory input to acquired songs systematically over a broad range of acoustic features can help us understand mechanisms that shape the development of species-specific vocal repertoire, and the establishment of vocal culture across generations. Whether such effects can explain language evolution is, however, an open question: frequency dependent effects were observed in speech development. For example, the chances that a child will over-regularize a verb (e.g. sing->singed) is inversely proportional to the frequency of that verb in the spoken language [45]. More generally, infants are sensitive to statistical distributions of synthetic words [46]. Studying such effects in the context of adaptive divergence and language change could be complemented by mechanistic investigations in songbirds.

The development of combinatorial abilities

The ability to flexibly rearrange vocal elements is the cornerstone of human language. Although human languages differ greatly from songbird systems of vocalization, in recent work we found a surprising commonality in how vocal combinatorial capacities were acquired across species [47]. Juvenile zebra finches were trained to perform one song and then the training target was altered, prompting the birds to swap syllable order, or insert a new syllable into a string. Birds solved these permutation tasks in a series of steps, gradually approximating the target sequence by acquiring novel pair-wise syllable transitions (sometimes too slowly to fully accomplish the task). In the more complex songs of Bengalese finches, branching points and bidirectional transitions in song-syntax were acquired *in a similar stepwise manner*, starting from a more restrictive set of vocal transitions. For example, when a Bengalese finch acquires two syllable types, say A and B, the transitions AB and BA emerge one by one, often weeks apart.

Strikingly, the babbling of pre-lingual human infants revealed a similar developmental pattern: instead of observing single developmental shift from reduplicated to variegated babbling (i.e., from repetitive to diverse sequences), as the literature seemed to imply [1], we

observed multiple shifts, where each novel syllable type slowly acquired a diversity of pair-wise transitions, asynchronously over development. Thus, for example, after a syllable type (say 'A') emerges, transitions with that syllables are acquired slowly (say AB, AC, AD…,) over about 20 weeks until the newly acquired syllable type appears to be 'connected' to the entire repertoire of syllables. Collectively, these results point to the possibility of a common generative process that is conserved across species. They also suggest that a long-noted gap between perceptual versus motor combinatorial capabilities in human infants may arise in part from the challenges in constructing new pair-wise transitions.

Thus, the ability to flexibly rearrange vocal sounds is not present in young infants or songbirds. Instead, this capacity is acquired over development. In zebra finches, transitions are acquired 'as needed' based on auditory input (namely, for the purpose of imitation). However, humans eventually acquire the capability to promptly perform vocal sequences they never heard or performed before. Therefore, the capacity to rearrange syllables might be crucial for both imitation and creative expression, although currently there is no direct evidence to confirm this interpretation of the results.

A comparative approach for studying vocal learning beyond imitation

We reviewed evidence showing that birdsong is an appropriate model system for studying vocal learning beyond imitation focusing on two types of mechanisms: those that promote adaptive deviation from templates, and those that allow flexible production of combinatorial sequences. These mechanisms can bridge between birdsong learning, language development, and language evolution research. For example, Neyogi [30] and Kirby [48] attempted to resolve the tension between language learning and language evolution by studying the multi-agent (i.e., learning from parents, learning from peers) and multi-generational language learning process. Those studies attempt to explain the changes that occur in spoken language across generations by either modeling, or experimenting with the complex dynamics of recursive learning. We suggest that studying mechanisms of vocal learning beyond imitation in songbirds may provide detailed mechanistic explanation to some of those phenomena. For example, frequency-dependent vocal learning observed in songbirds [27] might have an equivalent effect in language learning. This would predict that if some features of spoken language become too abundant, infants would then tend to diverge, leading to language change across generation. If this is true, than discovering brain mechanisms that might be shared across song development and language development could become feasible. Such lines of research might bridge between brain mechanisms of learning at the individual levels and at time scales of cultural evolution [48].

In sum, adaptive vocal development, both in humans and songbirds, extends significantly beyond mere imitation and template-matching. Studying such processes across songbirds and human infants, and across levels of brain function, is an exciting new direction for future research.

References

1. Oller, DK. The Emergence of the Speech Capacity (Google eBook) [Internet]. Psychology Press; 2000.

2. Oller DK, Wieman LA, Doyle WJ, Ross C. Infant babbling and speech [Internet]. J. Child Lang. 2008; 3:1–11.

3. Doupe AJ, Kuhl PK. Birdsong and human speech: common themes and mechanisms. [Internet]. Annu. Rev. Neurosci. 1999; 22:567–631. [PubMed: 10202549]

4. Thorpe WH. Learning and instinct in animals. [date unknown].

5. Marler P. A comparative approach to vocal learning: Song development in white-crowned sparrows. [date unknown], [no volume].

6. Tchernichovski O, Mitra PP, Lints T, Nottebohm F. Dynamics of the vocal imitation process: how a zebra finch learns its song. [Internet]. Science. 2001; 291:2564–2569. [PubMed: 11283361] This is the first detailed study about how trajectories of vocal changes are leading to song imitation. It contain description of how the diversity of vocal state increases in the onset of learning, tracking of non-monotonic vocal changes, and documentation of how prototype syllables differentiate to create a range of syllable types.

7. Aronov D, Veit L, Goldberg JH, Fee MS. Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. [Internet]. J. Neurosci. 2011; 31:16353–16368. [PubMed: 22072687]

8. Nottebohm F, Arnold A. Sexual dimorphism in vocal control areas of the songbird brain [Internet]. Science (80-.). 1976; 194:211–213.

9. Roberts TF, Gobes SMH, Murugan M, Olveczky BP, Mooney R. Motor circuits are required to encode a sensory model for imitative learning. [Internet]. Nat. Neurosci. 2012; 15:1454–1459. [PubMed: 22983208] This study shows, using optogenetics, that song nucleus HVC is involved in the acquisition of song template in real-time. This is the first documentation of direct involvement of sensory-motor system in the very initial stage of acquiring a memory template for vocal learning.

10. Hahnloser RHR, Kozhevnikov AA, Fee MS. An ultra-sparsecode underlies the generation of neural sequences in a songbird. [Internet]. Nature. 2002; 419:65–70. [PubMed: 12214232]

11. Saar S, Mitra PP. A technique for characterizing the development of rhythms in bird song. [Internet]. PLoS One. 2008; 3:e1461. [PubMed: 18213370]

12. Rajan R, Doupe AJ. Behavioral and Neural Signatures of Readiness to Initiate a Learned Motor Sequence [Internet]. 2013

13. Amador A, Perl YS, Mindlin GB, Margoliash D. Elemental gesture dynamics are encoded by song premotor cortical neurons [Internet]. Nature. 2013; 495:59–64. [PubMed: 23446354]

14. Leonardo A, Fee MS. Ensemble coding of vocal control in birdsong. [Internet]. J. Neurosci. 2005; 25:652–661. [PubMed: 15659602]

15. Ashmore RC, Renk JA, Schmidt MF. Bottom-up activation of the vocal motor forebrain by the respiratory brainstem. [Internet]. J. Neurosci. 2008; 28:2613–2623. [PubMed: 18322104]

16. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, et al. The genome of a songbird. [Internet]. Nature. 2010; 464:757–762. [PubMed: 20360741]

17. Hilliard AT, Miller JE, Fraley ER, Horvath S, White SA. Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. [Internet]. Neuron. 2012; 73:537–552. [PubMed: 22325205]

18. Fee MS, Scharff C. The songbird as a model for the generation and learning of complex sequential behaviors. [Internet]. ILAR J. 2010; 51:362–377. [PubMed: 21131713]

19. Dave AS. Behavioral State Modulation of Auditory Activity in a Vocal Motor System [Internet]. Science (80-.). 1998; 282:2250–2254.

20. Dave AS. Song Replay During Sleep and Computational Rules for Sensorimotor Vocal Learning [Internet]. Science (80-.). 2000; 290:812–816.

21. Derégnaucourt S, Mitra PP, Fehér O, Pytte C, Tchernichovski O. How sleep affects the developmental learning of bird song. [Internet]. Nature. 2005; 433:710–716. [PubMed: 15716944]

22. Tumer EC, Brainard MS. Performance variability enables adaptive plasticity of "crystallized" adult birdsong. [Internet]. Nature. 2007; 450:1240–1244. [PubMed: 18097411] This study shows how vocal exploration is regulated in a manner that enables adaptive vocal plasticity leading to imitation.

23. Tschida K, Mooney R. The role of auditory feedback in vocal learning and maintenance. [Internet]. Curr. Opin. Neurobiol. 2012; 22:320–327. [PubMed: 22137567]

24. Glaze CM, Troyer TW. The Development of Temporal Structure in Zebra Finch Song. [Internet]. J. Neurophysiol. 2012

25. Ali F, Otchy TM, Pehlevan C, Fantana AL, Burak Y, Ölveczky BP. The Basal Ganglia Is Necessary for Learning Spectral, but Not Temporal, Features of Birdsong [Internet]. Neuron. 2013; 80:494–506. [PubMed: 24075977]

26. Ravbar P, Lipkind D, Parra LC, Tchernichovski O. Vocal exploration is locally regulated during song learning. [Internet]. J. Neurosci. 2012; 32:3422–3432. [PubMed: 22399765] ** shows that variability in vocal patterns is regulated at time scales of milliseconds, so that variability is high only in song elements that need more learning. This result suggest that the song is learned in a piecewise manner, where each part of the song is learned independently by gating variability (exploration) in an adaptive manner.

27. Murugan M, Harward S, Scharff C, Mooney R. Diminished FoxP2 Levels Affect Dopaminergic Modulation of Corticostriatal Signaling Important to Song Variability [Internet]. Neuron. 2013; 80:1464–1476. [PubMed: 24268418]

28. Tchernichovski O, Nottebohm F. Social inhibition of song imitation among sibling male zebra finches [Internet]. Proc. Natl. Acad. Sci. 1998; 95:8951–8956. [PubMed: 9671785]

29. Byun TM. Bidirectional perception–production relations in phonological development: evidence from positional neutralization [Internet]. Clin. Linguist. Phon. 2012; 26:397–413. [PubMed: 22489733]

30. Niyogi, P. The Computational Nature of Language Learning and Evolution (Current Studies in Linguistics) [Internet]. The MIT Press; 2009.

31. Singleton JL, Newport EL. When learners surpass their models: The acquisition of American Sign Language from inconsistent input [Internet]. Cogn. Psychol. 2004; 49:370–407. [PubMed: 15342259]

32. Tchernichovski O, Lints T, Mitra PP, Nottebohm F. Vocal imitation in zebra finches is inversely related to model abundance [Internet]. Proc. Natl. Acad. Sci. 1999; 96:12901–12904. [PubMed: 10536020]

33. Derégnaucourt S, Gahr M. Horizontal transmission of the father's song in the zebra finch (Taeniopygia guttata). [Internet]. Biol. Lett. 2013; 9:20130247. [PubMed: 23760166]

34. Volman SF, Khanna H. Convergence of untutored song in group-reared zebra finches (Taeniopygia guttata). [date unknown], [no volume].

35. Eales LA. Song learning in zebra finches: some effects of song model availability on what is learnt and when [Internet]. Anim. Behav. 1985; 33:1293–1300.

36. Jones AE, ten Cate C, Slater PJB. Early experience and plasticity of song in adult male zebra finches (Taeniopygia guttata). [date unknown], [no volume].

37. Goldstein MH, Schwade JA. Social feedback to infants' babbling facilitates rapid phonological learning. [Internet]. Psychol. Sci. 2008; 19:515–523. [PubMed: 18466414]

38. Aronov D, Andalman AS, Fee MS. A specialized forebrain circuit for vocal babbling in the juvenile songbird. [Internet]. Science. 2008; 320:630–634. [PubMed: 18451295]

39. Ölveczky BP, Otchy TM, Goldberg JH, Aronov D, Fee MS. Changes in the neural control of a complex motor sequence during learning. [Internet]. J. Neurophysiol. 2011; 106:386–397. [PubMed: 21543758]

40. Tamura PM, M. Song "Dialects" in Three Populations of White-Crowned Sparrows. Condor. 1962; 64:368–377.

41. Fehér O, Wang H, Saar S, Mitra PP, Tchernichovski O. De novo establishment of wild-type song culture in the zebra finch. [Internet]. Nature. 2009; 459:564–568. [PubMed: 19412161] This study shows how vocal culture emerges in songbirds under controlled conditions. Starting from an isolate founder of a colony, each generation of pupils performed imitation in a manner that was biased more and more closely toward the wild type song of zebra finches. Results suggest that the most basic features of zebra finch songs are encoded in each bird, but take several generations of tutoring to become fully expressed.

42. Podos J, Peters S, Nowicki S. Calibration of song learning targets during vocal ontogeny in swamp sparrows, Melospiza georgiana [Internet]. Anim. Behav. 2004; 68:929–940.

43. Gardner TJ, Naef F, Nottebohm F. Freedom and rules: the acquisition and reprogramming of a bird's learned song. [Internet]. Science. 2005; 308:1046–1049. [PubMed: 15890887] This study shows how canaries 'reprogram' song patterns that diverge from the species specific template over development. It demonstrated that canaries can copy 'random walk' song-syntax, but later on, without any external stimulation, they spontaneously transform the sequence to a 'canary like' song-syntax (which they never heard).

44. Allen JA, Clarke BC. Frequency dependent selection: homage to E.B. Poulton [Internet]. Biol. J. Linn. Soc. 1984; 23:15–18.

45. Marcus GF, Pinker S, Ullman M, Hollander M, Rosen TJ, Xu F, Clahsen H. Overregularization in language acquisition. [date unknown], [no volume].

46. Saffran JR, Aslin RN, Newport EL. Statistical learning by 8-month-old infants. Science (80-.). 1996; 274:1926–1928.

47. Lipkind D, Marcus GF, Bemis DK, Sasahara K, Jacoby N, Takahasi M, Suzuki K, Feher O, Ravbar P, Okanoya K, et al. Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. [Internet]. Nature. 2013; 498:104–108. [PubMed: 23719373] This study shows how songbirds and human infants acquire vocal transitions one by one during development. Birds were trained with vocal 'tasks' that required them to alter their syntax: either to swap syllable order or to insert syllables into strings. The results showed that each transition is acquired independently, sparely over development. Similar effect were observed in human infants.

48. Scott-Phillips TC, Kirby S. Language evolution in the laboratory [Internet]. Trends Cogn. Sci. 2010; 14:411–417. [PubMed: 20675183]

Highlights

Adaptive vocal development extends beyond imitation and template-matching

Songbirds and humans share the capacity to converge towards species-typical ? vocalization

The capacity to flexibly assemble vocal sounds develops in a similar, stepwise manner ? across species.

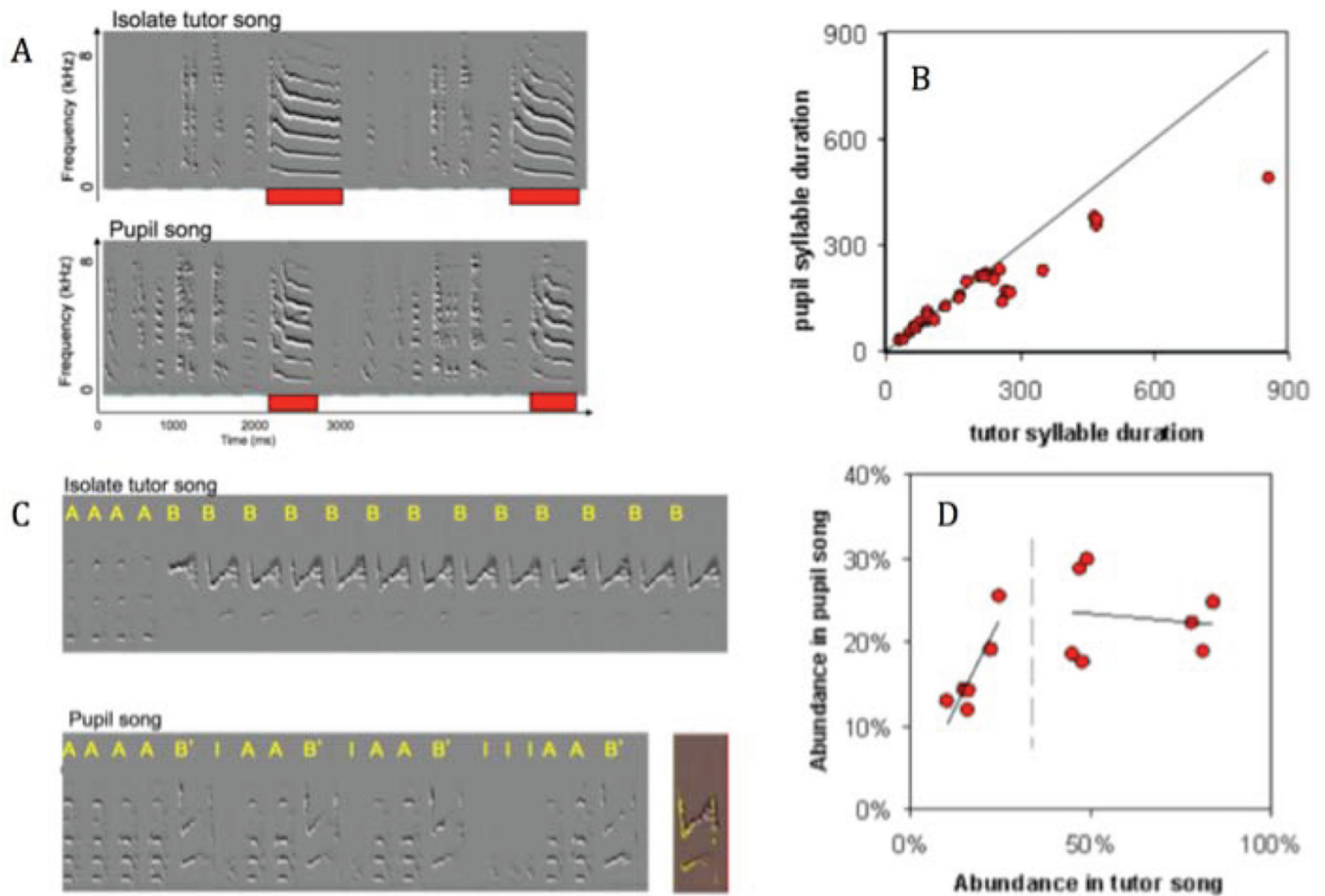Those capacities can explain multi-generational phenomena such as vocal culture

Figure 1. A scheme of developmental vocal learning
Vocal learning is a slow, iterative developmental process of progressively shaping vocal change that eventually leads to an integration in a complex communication system that endows an animal with individual identity (e.g., a unique song), a group identity (integrated in a vocal culture) and a means for coordinating behavior with other individuals (including affiliation, aggression and courtship behaviors). The capacity to imitate is only one factor that shapes the outcome of vocal learning over development.
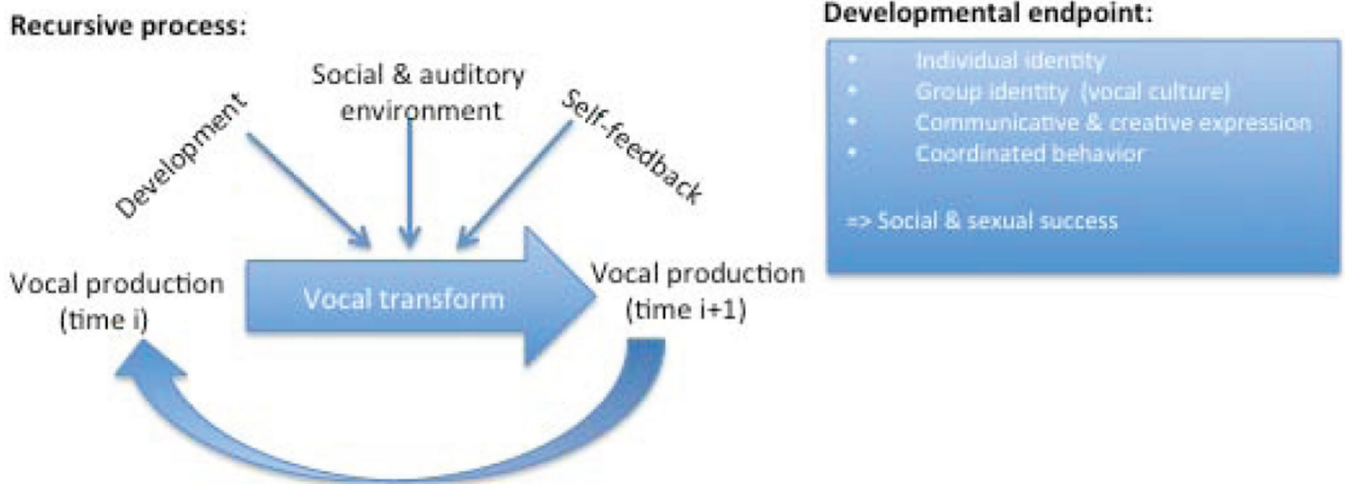
Figure 2. Vocal learning as a functional transform
A–B quantitative examination of the input/output relationship with respect to syllable
duration. A, sonograms of isolate song (top) and its imitation (bottom). Red bars outline an
isolate syllable of long duration, and its shorter version in the imitation of that song by a
juvenile zebra finch. B, a plot summarizing the imitation of isolate syllables with respect to
syllable-duration across eight birds who imitated isolate songs. As shown, syllable durations
are copied faithfully within an input range of up to about 270ms. Above this range, imitation
is invariably shorter than the sensory input. **C–D**, frequency dependent imitation. C,
sonogram of isolate song showing a rare syllable type (A) and an abundant syllable type (B).
In the imitation of this song, the rare syllable becomes more abundant and the abundant
syllable becomes more rare (note also that its spectral features are not accurately copied). D,
a plot, same as in B, but for relative frequency (abundance). Note that the imitation of
relative frequency is faithful in the range of 10–25%, but for higher abundance, the imitation
asymptotes on about 25%. Altered from Feher et al. [41]