Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants

Dina Lipkind[1,2], **Gary F. Marcus**[3], **Douglas Bemis**[3], **Kazutoshi Sasahara**[4], **Nori Jacoby**[5,6], **Miki Takahashi**[4], **Kenta Suzuki**[4], **Olga Feher**[1,2,7], **Primoz Ravbar**[2,7], **Kazuo Okanoya**[4], and **Ofer Tchernichovski**[1,2,7]

[1]Department of Psychology, Hunter College, City University of New York, New York, NY, USA

[2]Department of Biology, City College, City University of New York, New York, NY, USA

[3]Department of Psychology, New York University, New York, NY, USA

[4]Laboratory for Biolinguistics, RIKEN Brain Science Institute, Wako, Saitama, Japan

[5]The Edmond and Lily Safra Center for Brain Sciences, Hebrew University, Jerusalem, Israel

[6]The Music Department, Bar Ilan university, Ramat Gan, Israel

[7]The City University of New York Graduate Center, New York, NY, USA

Abstract

Human language, as well as birdsong, relies on the ability to arrange vocal elements in novel sequences. However, little is known about the ontogenetic origin of this capacity. We tracked the development of vocal combinatorial capacity in three species of vocal learners, combining an experimental approach in zebra finches with an analysis of natural development of vocal transitions in Bengalese finches and pre-lingual human infants and found a common, stepwise pattern of acquiring vocal transitions across species. In our first study, 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. All 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. Similarly, in the more complex songs of Bengalese finches, branching points and bidirectional transitions in song-syntax were acquired in a stepwise manner, starting from a more restrictive set of vocal transitions. The babbling of pre-lingual human infants revealed a similar developmental pattern: instead of a single developmental shift from reduplicated to variegated babbling (i.e., from repetitive to diverse sequences), we observed multiple shifts, where each novel syllable type slowly acquired a diversity of pair-wise transitions, asynchronously over development. Collectively, these results point to a common generative process that is conserved across species, suggesting that the long-

noted gap between perceptual versus motor combinatorial capabilities in human infants[1] may arise from the challenges in constructing new pair-wise transitions.

In the three species we studied, vocal behavior spans a broad range of combinatorial capabilities: zebra finches sing mostly linear sequences of syllables; Bengalese finch song includes branching sequences; pre-lingual human infants develop a capacity to transition between many syllables, eventually allowing flexible imitation of a potentially infinite array of words[2]. In zebra finches, we tested how the birds solve two combinatorial tasks: swapping syllable order, and inserting syllables into strings. In Bengalese finches, we examined the ontogenetic origin of combinatorial plasticity in specific vocal transitions, while in human infants we examined statistically, how diversification of many vocal transitions comes about. Across these levels of analysis we tested whether the capacity to flexibly rearrange vocal units is the starting point of vocal learning. Alternatively, the combinatorial machinery might develop slowly, via growth or learning, with individual vocal transitions introduced gradually. Such an early process could enable selective pruning later on. In the first case, we would observe simultaneous and parallel appearance of many syllable transitions during learning; in the latter case, we would observe a stepwise addition of particular transitions to the vocal repertoire.

We trained young zebra finches to imitate playbacks of one song (source), selected birds (17 out of 87) who imitated it fast enough (by day 63 post hatch), and switched their training to a variant song (target)[3,4] where syllable order was altered: ABC-ABC → ACB-ACB (Fig. 1a). We then examined the entire time course of the shift from source to target song (Suppl. 1; Suppl. Fig. 1)[5,6]. A bigram Markov model was found to account for the bulk of song sequence structure during the experimental period (Suppl. 2; Suppl. Fig. 2).

In the birds that completed the task (n=8; Fig. 1b–g), the target song appeared abruptly after 17 ± 4.4 days (mean $\pm$ s. e. m. here and thereafter; Fig. 1c; Suppl. Fig. 3a). Extinction of the source song occurred prior to, or concurrently with, the appearance of the target, with a time lag of 3 ± 2.8 days between source disappearance and target first appearance, indicating that the target song was generated via intermediate steps, but with no persistence of old singing habits once the entire target song was in place. To quantify intermediate steps, we tracked the appearance of the target pair-wise transitions (bigrams AC, CB and BA), the increase (adjustment) of their frequencies, and the extinction of no longer required source transitions (Fig. 1d–g).

New transitions appeared sparsely over development with a lag of 10.4 ± 1.91 days from training onset, and a 6.4 ± 3.5 days' gap between consecutive bigram appearances (Fig. 1d–e; Suppl. 3.1; Suppl. Fig 3b, d and e). Each gap included several thousand renditions of a single newly acquired bigram with no concurrent increase in the (zero or near zero) frequency of target bigrams that were not yet acquired (Suppl. 3.2-3.3; Suppl. Fig. 3f–h). Time gaps showed no developmental trend (no significant correlation between 1st– 2nd and 2nd–3rd transition gaps; $r^2=0.073$). In contrast, both adjustments and extinctions of transitions showed strong developmental trends (Fig 1g): the appearance of each new target bigram was followed by a fast adjustment to endpoint frequency (phase transition), the speed of which increased strongly with the order of bigram appearance: the time interval

from 25% to 75% of the endpoint frequency was 9.6 ± 2 days for the 1st bigram, 4.0 ± 0.9 days for the 2nd, and only 2.3 ± 0.2 days for the 3rd and final bigram (Fig. 1g, left; p=0.018, paired t-test 1st versus 3rd bigram). Extinction of source bigrams lagged behind the appearances of 1st and 2nd target bigrams (4.5 ± 3 and 6 ± 3 days respectively), but occurred almost simultaneously with the appearance of the 3rd target bigram (-0.3 ± 0.3 days) resulting in a prompt switch to exclusive target performance (Fig. 1g, right; Suppl. Fig. 3c). The prompt and rapid changes observed once the last bigram appeared are likely to mirror capabilities not fully expressed earlier (Suppl. 3.4), suggesting that bigram appearance was a rate limiting stage; namely, once a bigram was performed at all, or above some very low threshold, its frequency could change rapidly to match the target.

From the nine birds that failed to complete the task, five learned it partially (Fig 1h–j; Suppl. 4; Suppl. Fig. 4). Their learning process was similar to that of successful birds, except for failing to perform a single (and in one case two) target transition. Consequently, the end point of unsuccessful birds resembled intermediate stages of learning in successful birds (Fig 1f and j), including a higher transition entropy which merely mirrored the co-existence of source and target bigrams (Suppl. 5; Suppl. Fig. 5). Thus, despite performing millions of syllable renditions, unsuccessful birds had no measurable capability to produce the entire set of target transitions.

To test if newly acquired syllables can form transitions more easily, we constructed a task that elicited combinatorial changes in newly formed syllables, training birds to incorporate newly formed B syllables into strings of A^n syllables, (AAAA→ABAB, namely A^n→$(AB)^n$, Fig. 2a–b; Suppl. 6; Suppl. Fig. 6). Note that syllable B can be inserted into the string even as an unstructured precursor of B. This task was indeed easier: 15 out of 28 birds learned the source song and ten of them also imitated the target song (Fig. 2c–d; Suppl. 6; Suppl. Fig 7a). However, birds did not directly insert syllable B into A^n strings, and instead, acquired two novel transitions, AB and BA (Fig. 2e, Suppl. Fig. 7b), with an initial delay of 9.7 ± 1.9 days, and time gaps of 4.9 ± 1.52 days between their appearances (Fig. 2h), comparable to appearance gaps in the sequence rearrangement task. As in the ABC→ACB task, adjustment durations tended to decrease with bigram order (7.8 ± 1.6 and 5.1 ± 0.9 days for the 1st and 2nd bigram; Fig. 2i), and extinction of the source bigram (AA) usually occurred simultaneously with the appearance of the last target bigram (-1 ± 1.5 days; Fig. 2j; Suppl. 7).

In this task, the newly formed syllable type should initially appear exclusively at the song's edge until it can be 'connected' by two distinct bigrams. For example (Fig. 2f–g), if AB is learned first, the bird must stop after singing B, confining B to appear at the end of A^n strings until the second bigram (BA) is learned. To test for such 'edge effect', we calculated daily frequencies of the occurrence of B at the start of the song (BA^n), at its end (A^nB) and in its middle (ABA). As expected, B was initially performed exclusively at one edge of the song, (BA^n in 5 birds and A^nB in 3 birds, Fig 2f; Suppl. 7; Suppl. Fig 7c). In all cases, syllable B appeared in the middle of song bouts immediately once the second bigram was learned. Namely, we did not observe cases of a $BA^nB \mid n>1$ stage prior to $(AB)^n$, indicating that the only obstacle for incorporating B into the bout center was inability to perform both AB and BA transitions, as opposed to difficulties in breaking AA transitions. We observed a similar 'edge effect' also in naturally occurring syntax development (Suppl. 8; Suppl. Fig.

8). Therefore, stepwise acquisition of bigrams generalizes to earlier stages of vocal development, and to a different learning task, where we juxtaposed the formation of a novel syllable type with a sequence rearrangement task.

By selecting for fast-learning birds and training them unnaturally, we might have underestimated their combinatorial capabilities. To address this, we studied Bengalese finches, raised in a semi-natural aviary (n=8; Suppl. 9). While altered-target training was necessary to induce sequence rearrangement in zebra finches, Bengalese finches naturally rearrange syllables as adults[2,7]. We examined the ontogentic origin of song-syntax plasticity by tracing the development both of fixed and of variable parts of the adult song. Consider a case of a bi-directional transition in the mature song A B (Fig. 3a): this plasticity might be a residual of an early stochastic performance of transitions, including both AB and BA. Alternatively, transitions are acquired sparsely, say AB and later BA. We identified seven bidirectional transitions in the endpoint song of five of our experimental birds, and tracked the frequencies of both bigrams (AB and BA) from the earliest time point when both syllable types A and B could be recognized (days 65–83 post hatch) to the end of development (Fig. 3a; Suppl. Fig. 9a). We found long gaps between bigram appearances (17.7 ± 8.7 days; Fig. 3a–c; Suppl. Fig. 9b), and adjustment durations were shorter (8.9 ± 2.2 days; Fig. 3c).

Next, we traced the ontogenetic origin of unidirectional transitions (AB): in 15 out of 16 cases (Suppl. 9), as early as the clusters corresponding to A and B could be identified, significant frequency of AB transitions could be identified, but the frequency of the reverse transitions (BA) was zero or near zero ($20 \pm 2\%$ versus $1 \pm 0.9\%$ for AB and BA respectively, $p<0.001$, paired t-test). Therefore, both unidirectional (Fig. 3d, top) and bidirectional (Fig. 3d, bottom) transitions tended to originate from unilateral transitions, which is inconsistent with the notion of highly stochastic transitions early on.

Focusing on bidirectional transitions was necessary to overcome biases in the detection level of syllables during early development: because of symmetry, such biases should not affect the relative frequencies of AB and BA. However, once all syllable types were in place we were able to examine all transition types (Fig. 3e–f). During that period, 5 out of 8 birds kept adding and removing transitions. As in early song development, this process was biased so as to increase connectivity across syllable types (15 additions versus 6 deletions) and decrease repetitive sequences (zero additions versus 4 deletions; Fig 3e). Further, looking at branching points, the mean number of variegated transitions (excluding reduplications) per syllable increased over time (3.28 ± 0.24 and 3.88 ± 0.19 for start point and end point respectively; $p=0.04$, paired t-test; Fig 3f;). Thus, combinatorial plasticity observed in the adult bird developed from a more restricted syntax, in a stepwise manner.

Finally, we examined the development of phonetic syntax of infant babbling. Classical studies identified a transition from predominantly reduplicated utterances (e.g., ba ba ba) to variegated utterances (e.g., ba gu ge)[8,9], which could perhaps mirror a stepwise acquisition of variable transition types. However, later studies failed to reliably replicate this effect[10,11], and instead found variegated utterances throughout babbling development (Suppl. 10). This could suggest that, unlike songbirds, human infants can rearrange syllables early on with

relative ease. However, infants' large repertoire of syllable types is acquired gradually, so that at any time point the infant is producing a mixture of newly acquired and old syllable types: if for each syllable type the number of available transitions increases gradually, then we would expect less variegation in newly acquired syllables and more in old syllables. The mixture of old and new syllable types at any developmental time could mask a syllable specific increase in variegation. We therefore tested the existence of a developmental trend in variegation, in reference to the development of specific syllables.

We analyzed databases of phonologically transcribed babbling sessions (CHILDES[12,13]) from nine US toddlers recorded bi-weekly at ages of 9–28 months, which we segmented into syllables (Methods, Section 6; Suppl. 11.1 and 11.2). We pooled all measures across syllable types in each child, and adjusted our measures through a bootstrapped normalization in each session to control for effects due to developmental changes in the number of syllable types and in utterance length (Suppl. 11.3).

We first tracked the frequencies of reduplicated transitions over infants' age, aligning the data in reference to the age where speech/babbling ratio reached 50%. Throughout development, reduplicated utterances were performed $15 \pm 5.7\%$ above chance ($p < 0.001$). However, we did not observe any changes in the tendency to reduplicate syllables over development (Fig. 4a, adjusted $r^2 = 0.01$; $p = 0.32$).

Next, we calculated the same measure again, aligning the data in reference to the appearance time of each syllable type (Fig. 4b). Strikingly, a clear shift from high to low reduplication frequency was now observed (adjusted $r^2 = 0.26$; $p < 0.001$), occurring very slowly, over 20–30 weeks from the time of appearance. Namely, syllables tended to be repeated (reduplicated) when first acquired, and this was followed by a gradual acquisition of transitions to other syllables (variegation). Therefore, previous failures to find a developmental shift from reduplicated to variegated babbling[10,11] may be explained by a masking effect due to asynchronous appearance of new syllable types.

Our findings in songbirds predict that in infants, new syllable types should first appear at an utterance edge, and indeed, newly generated syllable types appeared more frequently at utterance edges ($8.6 \pm 4\%$ above chance, $p < 0.001$), and this tendency decreased slowly over 20 weeks or so (adjusted $r^2 = 0.13$; $p < 0.01$;Fig. 4c). Finally, we found that the rate of acquiring transitions was lower than expected by chance, taking into account that new syllable types are continuously added to the child's vocabulary, necessarily enlarging the pool of potential syllable transitions ($20 \pm 8\%$ below chance at first session; $p < 0.001$; Fig. 4d). Namely, the increase in the number of bigrams lagged behind the increase in syllable vocabulary.

Our results across species suggest that, in contrast to predictions of previous theories[14,15] novel vocal transitions are not acquired rapidly during early stages of development. Further, studies of movement sequence learning in adult monkeys showed, in contrast to our results, that target sequences appeared very rapidly while frequency adjustments took weeks[16,17]. Are these differences due to age, experience, or learning modules (movement versus vocal)? The latter seems less likely, since a gradual, generative process was observed in the

development of non-learned movement sequences[18,19]. The role of experience may be resolved by comparing the development of learned and non-learned movement sequences.

Our findings point to a prolonged developmental stage of stepwise acquisition of vocal combinatorial capacity, which may be accompanied or followed by pruning of unnecessary transitions[20,21]. Dynamics of a trial-and-error learning alone are unlikely to explain the zero or near zero frequencies of many transitions for prolonged developmental epochs (Suppl. 12). Instead, we propose that stepwise development of combinatorial diversity might stem from the dynamics of constructing new links between representations of vocal gestures in the motor system: In songbirds, vocal production gradually differentiates into distinct syllable types, represented by chains of neuronal activity[22] in the motor song system, which are thought to code sequences of vocal gestures[23]. During singing, neuronal activity must propagate from the tail of one chain of gestures to the head of the other[24]. The construction of such connections might be initially sparse, limiting vocal sequences to a small set of transitions, and reduplications. Adding and removing tail-to-head connections, should allow additions and deletions of vocal transitions (AB↔ABC) but not swapping (ABC→ACB) or insertions (AA→ABA). If the process is dominated by additions, we should see more and more branching sequences (as in Bengalese finches), and eventually (perhaps in human infants), an all-to-all network with a single connected component might emerge, allowing free access to any element, which is later pruned to produce speech[20,21]. According to this model, vocal babbling is shaped by a slowly evolving inter-syllabic network, where freedom gain due to acquiring new transitions is counterbalanced by the acquisition of new syllable types that are not yet connected and tend to reduplicate or break the sequence. Such process could explain the mismatch between infants' precocious ability to perceive complex grammars, and their initially limited ability to produce vocal sequences[1]. A similar gap may also exist in songbirds, whose perceptual capacity for syntax learning is a debated question[25–29]. However, there is also a fascinating parallel between the perceptual ability of songbirds to assemble memories of phrase pairs into a complete multi-phrase song template[30], and the phenomenon shown here, of birds and infants using pair-wise syllable transitions to transform one multi-syllable string into another.

Methods

1. Animal care

Experiments were conducted in accordance with the guidelines of the US National Institutes of Health and have been reviewed and approved by the Institutional Animal Care and Use Committees of Hunter College & City College, City University of New York and of RIKEN Brain Science Institute.

2. Experimental design

Male zebra finches were bred at Hunter College and City College of The City University of New York, and reared in the absence of adult males between days 7–30 post hatch. Afterwards, birds were kept singly in sound attenuation chambers, and recorded continuously. Twelve out of seventeen birds were passively exposed to 30 playbacks per day of the source, occurring at random with a probability of 0.01 per second, from day 33–39

until day 43, in an attempt to increase success rate. On day 43, each bird was trained to press a key to hear song playbacks, with a daily quota of 20. Once birds learned the source, we switched to playbacks of the target. Only birds that learned the source before day 63 (n=17, 20% of the total birds trained with the source) were used. Recording and training were done using Sound Analysis Pro[6].

Source and target song models were synthetically composed of natural syllables. To balance the design of the sequence rearrangement task, we trained some birds (n=7) with ABC-ABC → ACB-ACB as source and target, and others (n=10) with ACB-ACB → ABC-ABC. The two groups were pooled, and for simplicity we refer to all as ABC-ABC → ACB-ACB.

Bengalese finches were reared in communal aviaries in RIKEN Brain Science Institute, Japan. From the age of 40–50 days, they were transferred singly into a sound proof cage at 3–4 days intervals and recorded for ~24 hours.

3. Data analysis (songbirds)

Song feature calculation and cluster analysis were performed using Sound Analysis Pro[6]. Further analysis was done using MATLAB 7. Cluster information was used to elucidate the order of syllable types sung by a bird on each developmental day and to test whether syllable types are reused in the learning of new syntax (Suppl. 1).

In zebra finches, the percent of clustered syllables in bouts (assessed by manual inspection of a sample of 10 song bouts per bird) was $96 \pm 3\%$ at the endpoint, and $90 \pm 2\%$ during the transition from source to target. Clustering Bengalese finch songs was more difficult, with $91 \pm 1\%$ at the endpoint, and $80 \pm 3\%$ at the starting point. Unidentified syllable types were regarded as missing data.

Song bouts were defined as sequences of identified syllable types with stop durations of less than a threshold that was determined by the typical stop duration in the endpoint song (150ms for ABC-ABC→ ACB-ACB and Bengalese finches; 100ms for AAAA→ABAB). The threshold for bigram occurrences was similarly defined by the typical stop duration within each bigram type at the endpoint (60–150ms). Daily frequencies of sequences (source and target songs and bigrams) were calculated as the proportion of syllables comprising a given sequence out of the total number of syllables sung on that day. Due to unavoidable misclassifications in the clustering process, we had to determine a margin of error to decide when an observed transition frequency is real. We empirically estimated our error level as about 2% (2.2 ± 0.5 %) by measuring the baseline levels of target bigrams on day 0, and set our threshold to detect the moment of appearance of a bigram transitions at 3% above noise (i.e., 5%). In an effort to assess the real performance rate of target transitions below noise level, we visually examined a sample of positively identified bigrams on days where their frequency was close to zero (Suppl. 2.2), and found that actual performance rates ranged from very low (0.01) to absolute zero.

Bengalese finches' songs contained more syllable types than those of the zebra finches (6–10 versus 2–3), resulting in a higher level of clustering errors. We therefore took a semi automatic approach for determining the bigram detection threshold. In the first stage, we

used 5%, as in zebra finches; in the second stage, we excluded from our analysis transitions that were clearly an outcome of clustering errors, based on visually examining 20 random instances of each transition type.

6. Analysis of babbling data

Data was obtained from 9 children in the Davis corpus[14] of the CHILDES database[13] On average, children were 9 months, 28.3 days old (2 months, 1.3 days std.) at the first session, and data was collected for an average of 1 year, 7 months (7 months, 12.8 days std.). Data consisted of 38.8 sessions on average (10.2 std.) per child, recorded an average of 16.07 days apart (6.4 days std.).

Only babbling utterances (i.e., utterances for which no lexical items were assigned in the CHILDES transcriptions) were analyzed. Utterances were parsed into syllables using a semiautomated method, described below. Only utterances that received a complete syllabic parse were analyzed (2135 utterances per child (924 std.) and 62.0 utterances per session (37.5 std.)).

6.1. Parsing algorithm—We used an iterative parsing process. An utterance was considered parsed if every phoneme in it was successfully assigned to a syllable by the algorithm, such that each phoneme was used exactly one time in a syllable. On each iteration, we first manually assigned complete syllabic parses to several unparsed utterances. We then added new syllable types to the set of possible syllables that could be used for parsing. Next, we automatically checked if all utterances could be exhaustively parsed using the current store of syllables. For example, an utterance *badaja* would be manually assigned the syllabic parse of *ba*, *da*, and *ja*. On the following iteration, an utterance *baja* could be parsed into the syllables *ba* and *ja*.

Utterances that could not be fully parsed using the set of defined syllables were manually parsed, adding to the set of acceptable syllables. Thus, every syllable used to parse the data was manually verified as a valid syllable within the data. If an utterance could be assigned two different parses, then we employed a heuristic such that we chose the parse with the greater number of two phoneme syllables (CV or VC). If multiple parses for an utterance were equal in this measure, we would manually assign a parse to the utterance or leave it as ambiguous, and exclude it from the analysis. Iterations were performed until a sizeable amount of the data had been parsed (58.2% of babbling utterances [19.0% std.]).

6.2. Initial Tabulation—From this set of parsed utterances, we tabulated the frequency of each syllable within each session and its placement within utterance. Analysis was restricted to syllables that reached a frequency threshold of 1% of the total number of syllables in the session, thus focusing on syllables that the child produced at a non-negligible rate. We also calculated the frequency of all transitions between the syllables. A transition was defined as any two sequential syllables within an utterance. On average, each child used 128 distinct syllables (8.12 std.) and constructed 763 distinct transitions (95.6 std.).

6.3. Bootstrap Normalization—Measures of the development of transition variability over time are affected by the growth in the number of syllable types and in utterance length.

To control for this, we used a bootstrapped normalization procedure for all measures. To establish a baseline value for each measure that reflected a random placement of syllables but held vocabulary size and utterance length constant, we shuffled syllables randomly within each recording session while maintaining the length of each utterance. Each measure was then recalculated over these bootstrapped randomizations to establish a baseline value, to which the observed data were compared.

6.4. Measures—All measures were calculated for each syllable type in each session. Sessions were then aligned on syllable types' first appearance, and a mean over syllable types was calculated for each session. For each measure, we evaluated trends over sessions by fitting a simple linear regression model to subject averages using R (R Development Core Team, 2007, freely available at http://cran.r-project.org). Separate models were fit for each measure with session number as a fixed factor. Only syllable types that appeared in the course of the experimental period (namely, that were not present at the first session) were analyzed.

Reduplication: the frequency of occurrences in reduplicated transitions per syllable type. This measure was calculated twice using two different alignments: by developmental stage (the first session where speech/babbling ratio reached 50%), and by syllable type's first appearance. Note that the sample size for the developmental alignment analysis was smaller (n=7) than for the syllables-specific alignment (n=9), due to insufficient number of sessions with more than 50% babbling utterances in two children.

Occurrence of new syllables at edges: the frequency of occurrences of a syllable type at the edge of an utterance compared to occurrences its middle. For this measure, we did not count reduplications as being in middle of utterances.

Addition of new transitions: the number of new transition types per syllable type in each session. This measure indicates how likely each syllable occurrence was to participate in a previously unseen transition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Marcus GF, Vijayan S, Bandi Rao S, Vishton PM. Rule learning by seven-month-old infants. Science (New York, NY). 1999; 283:77–80.

2. Berwick RC, Okanoya K, Beckers GJL, Bolhuis JJ. Songs to syntax: the linguistics of birdsong. Trends in Cognitive Sciences. 2011; 15:113–121. [PubMed: 21296608]

3. Eales L. Song learning in zebra finches: some effects of song model availability on what is learnt and when. Animal Behaviour. 1985; 33:1293–1300.

4. Plamondon SL, Rose GJ, Goller F. Roles of Syntax Information in Directing Song Development in White-Crowned Sparrows (Zonotrichia leucophrys). J Comp Psychol. 2010; 124:117–132. [PubMed: 20476811]

5. Derégnaucourt S, Mitra PP, Fehér O, Pytte C, Tchernichovski O. How sleep affects the developmental learning of bird song. Nature. 2005; 433:710–6. [PubMed: 15716944]

6. Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, Mitra PP. A procedure for an automated measurement of song similarity. Animal Behaviour. 2000; 59:1167–1176. [PubMed: 10877896]

7. Yamashita Y, et al. Developmental learning of complex syntactical song in the Bengalese finch: a neural network model. Neural Networks. 2008; 21:1224–1231. [PubMed: 18460417]

8. Oller DK. The emergence of the sounds of speech in infancy. Child phonology 1 Production. 1980; 1:93–112. Product.

9. Stark R. Stages of speech development in the first year of life. Child phonology 1 Production. 1980; 1:73–92.

10. Mitchell PR, Kent RD. Phonetic variation in multisyllable babbling. Journal of Child Language. 1990; 17:247–265. [PubMed: 2380268]

11. Smith BL, Brown-Sweeney S, Stoel-Gammon C. A quantitative analysis of reduplicated and variegated babbling. First Language. 1989; 9:175–190. ST – A quantitative analysis of reduplica.

12. MacWhinney B. The CHILDES project: Tools for analyzing talk. Child Language Teaching and Therapy. 2000; 8:217.

13. Davis BL, MacNeilage PF. The articulatory basis of babbling. Journal Of Speech And Hearing Research. 1995; 38:1199–1211. [PubMed: 8747814]

14. Edelman, G. Neural Darwinism. The Theory of Neuronal Group Selection. Basic Books; New York: 1987.

15. Hanuschkin A, Diesmann M, Morrison A. A reafferent and feed-forward model of song syntax generation in the Bengalese finch. Journal of computational neuroscience. 2011; 31:509–32. [PubMed: 21404048]

16. Hikosaka O, Rand MK, Miyachi S, Miyashita K. Learning of sequential movements in the monkey: process of learning and retention of memory. Journal of Neurophysiology. 1995; 74:1652–1661. [PubMed: 8989401]

17. Rand MK, et al. Characteristics of sequential movements during early learning period in monkeys. Experimental Brain Research. 2000; 131:293–304. [PubMed: 10789945]

18. Golani I. A Mobility Gradient in the Organization of Vertebrate Movement3: The Perception of Movement Through Symbolic Language. Behavioral and Brain Sciences. 1992; 15:249–308.

19. Dominici N, et al. Locomotor Primitives in Newborn Babies and Their Development. Science. 2012:997.

20. De Boysson-Bardies B, Vihman MM. Adaptation to language: Evidence from babbling and first words in four languages. Language. 1991; 67:297–319.

21. Vihman MM, Macken MA, Miller R, Simmons H, Miller J. From babbling to speech: A re-assessment of the continuity issue. Language. 1985; 61:397–445.

22. Jin DZ, Ramazano lu FM, Seung HS. Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. Journal of Computational Neuroscience. 2007; 23:283–299. [PubMed: 17440800]

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

24. Jin DZ. Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. Physical review. E, Statistical, nonlinear, and soft matter physics. 2009; 80:051902.

25. Abe K, Watanabe D. Songbirds possess the spontaneous ability to discriminate syntactic rules. Nature Neuroscience. 201110.1038/nn.2869

26. Beckers GJL, Bolhuis JJ, Okanoya K, Berwick RC. Birdsong neurolinguistics: songbird context-free grammars claim is premature. NeuroReport. 201210.1097/WNR.0b013e32834f1765

27. Gentner TQ, Fenn KM, Margoliash D, Nusbaum HC. Recursive syntactic pattern learning by songbirds. Nature. 2006; 440:1204–7. [PubMed: 16641998]

28. Katahira K, Suzuki K, Okanoya K, Okada M. Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. PloS one. 2011; 6:e24516. [PubMed: 21915345]

29. Van Heijningen, Caa; De Visser, J.; Zuidema, W.; Ten Cate, C. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:20538–43. [PubMed: 19918074]

30. Rose GJ, et al. Species-typical songs in white-crowned sparrows tutored with only phrase pairs. Nature. 2004; 432:753–8. [PubMed: 15592413]
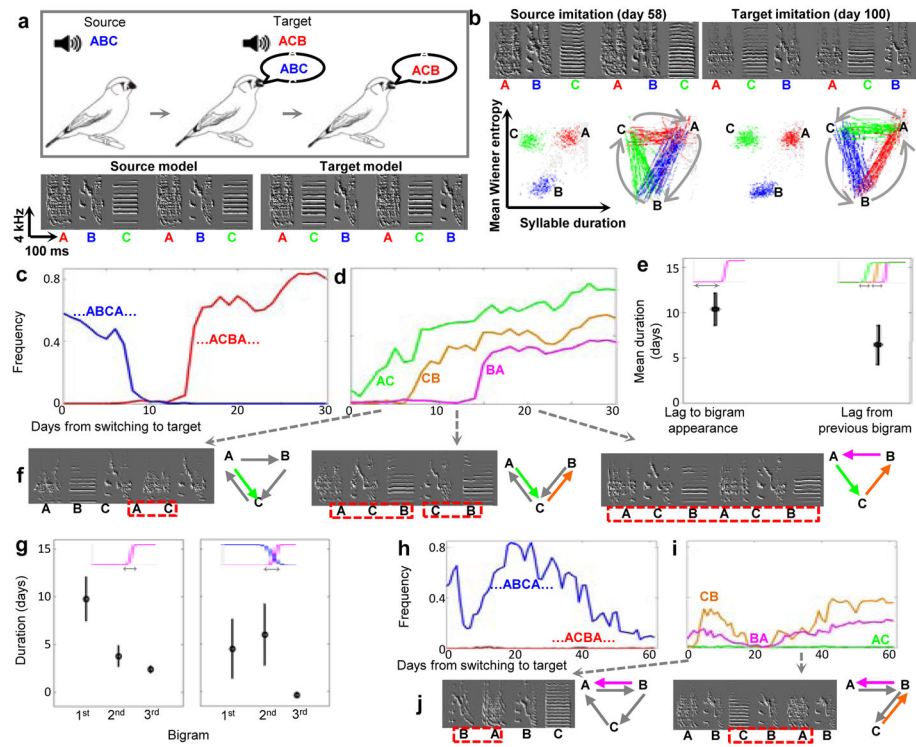
Figure 1. Syllable rearrangement task

a, Top, sequential training with two songs; bottom, training models. **b,** Song examples (top) and scatter plots of syllable features (bottom) after source and after learning in one bird. Clusters represent syllable types and lines represent transitions (colors represent transition end syllable). **c–d,** Daily frequencies (in one bird) of **c,** source and target songs; **d,** target bigrams; **e,** Learning phases in successful birds (means ± s.e.m.; n=8). **f,** Songs and syntax diagrams during learning (same bird). **g,** duration of adjustment and extinction according to bigram appearance order. **h–j,** Same as **c, d** and **f** in an unsuccessful bird.
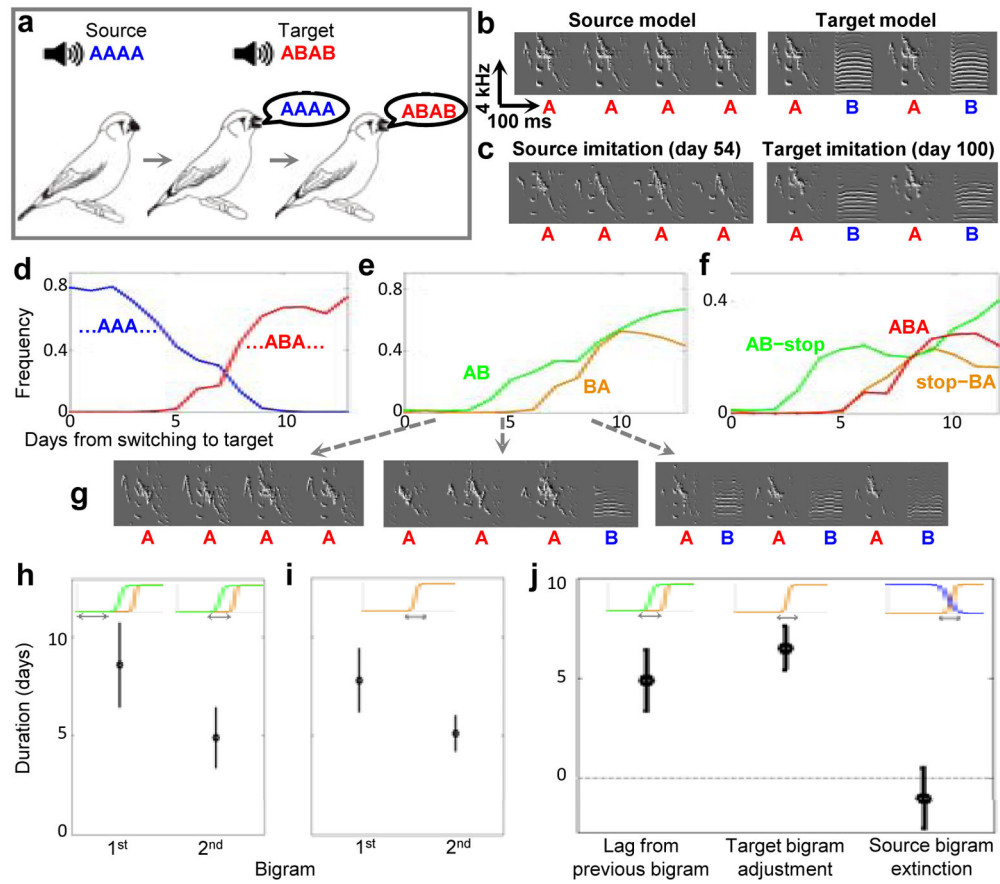
Figure 2. Syllable insertion task

a–b,: Training regime and models. **c,** Learning outcome in one bird. **d–f,** Daily frequencies of syllable sequences in one bird: **d,** Source and target songs; **e,** target bigrams; **f,** occurrences of syllable B at bouts' end (green), start (orange) and middle (red). **g,** Song examples during learning (same bird). **h–j,** Means ± s. e. m. (n=10) of **h,** appearance lags of target bigrams; **i,** adjustment durations; **j,** lags between target bigrams' appearance, adjustment of the 2nd target bigram, and extinction duration of the source bigram (AA).
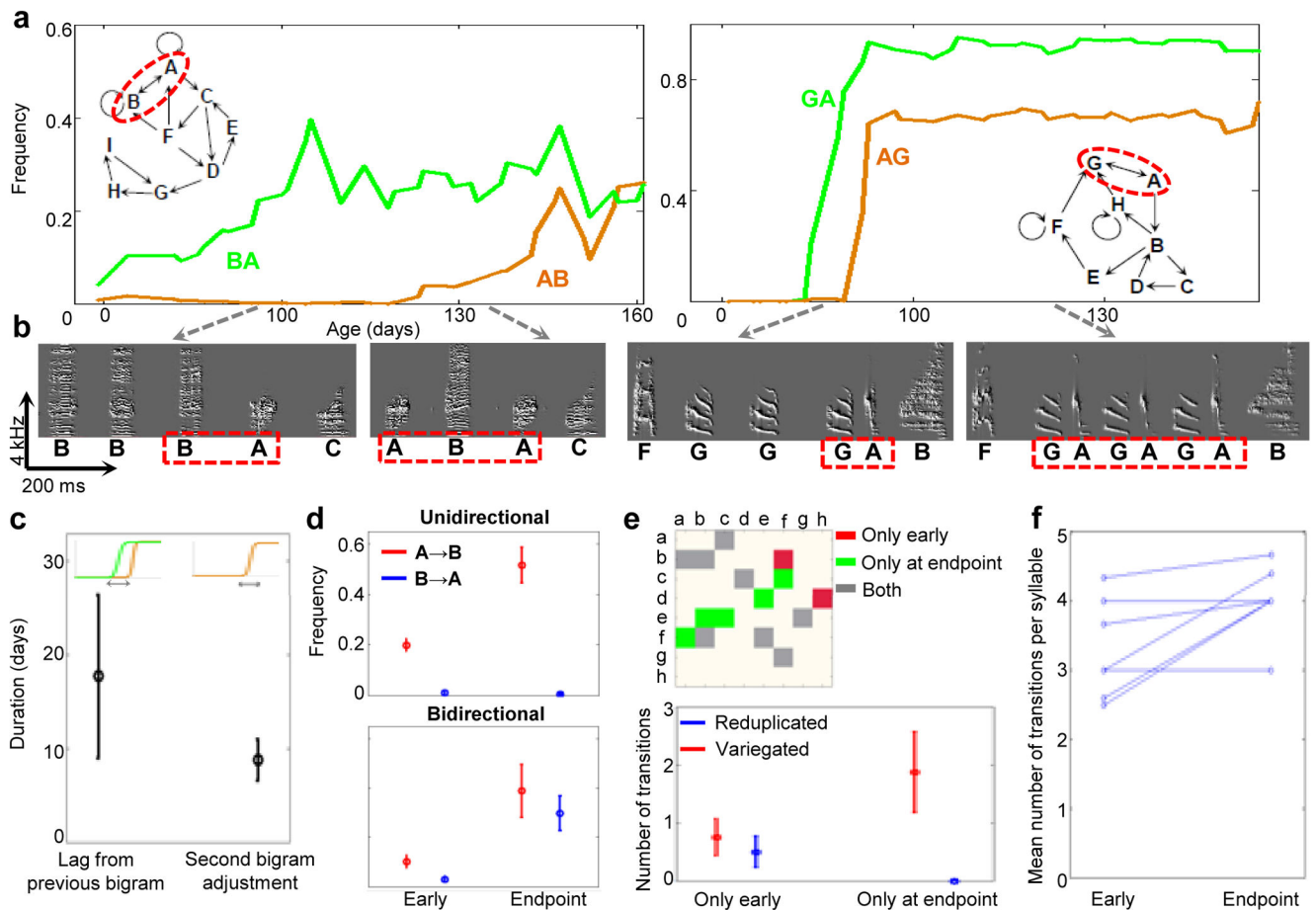
Figure 3. Combinatorial learning in Bengalese finches

a–b, Development of bidirectional transitions in two birds. Insets show endpoint syntax. **c,** Mean ± s.e.m. of appearance lag & adjustment duration in bidirectional transitions (n=7 transitions). **d,** Frequencies of unidirectional (top, n=16) and bidirectional (bottom, n=7) transitions in early development and at endpoint. **e,** Top, binary transition matrix (one bird), showing transitions present only early (green), only at endpoint (red), and in both (gray). Bottom, means ± s.e.m. across birds (n=8) of the number of transitions present only early or only at endpoint; **f,** Developmental changes in the mean number of transitions per syllable (in cases of variable transitions).
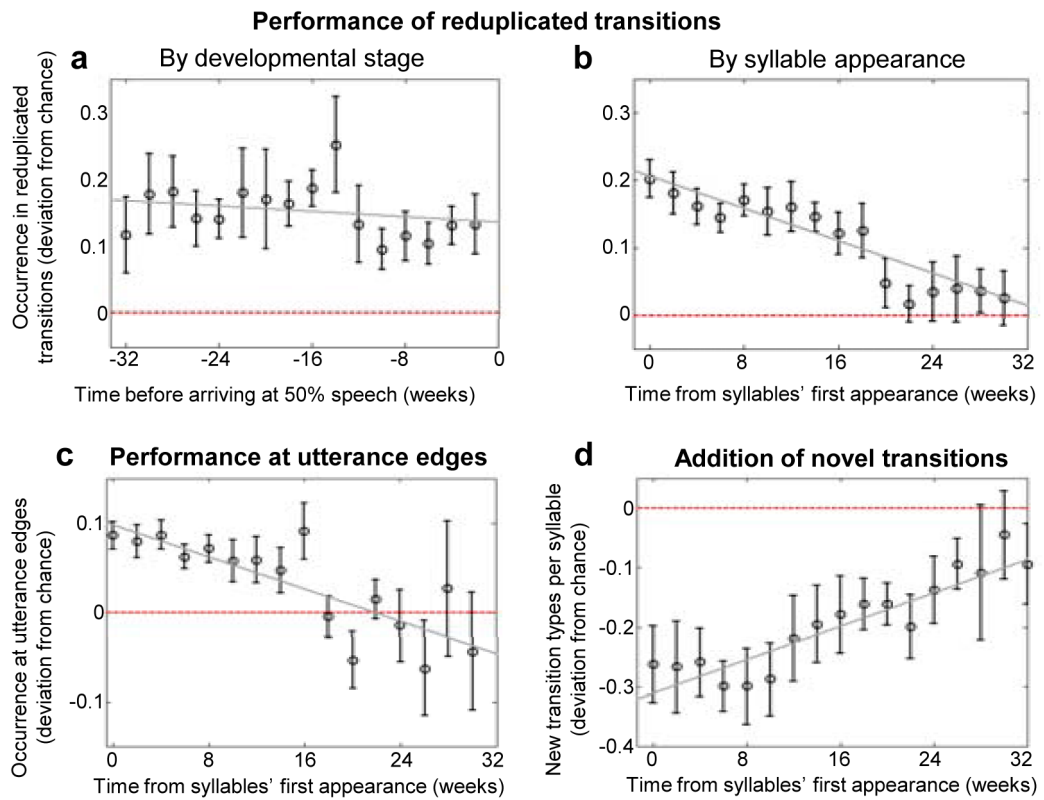
Figure 4. Incorporation of new syllables into infants' babbling utterances
a–b, Frequency of syllable occurrence in reduplicated transitions. **a,** Data aligned by developmental stage (time zero is the first session with >50% of speech utterances); **b,** Data aligned by each syllable type's first appearance. **c,** Frequency of syllable occurrence at utterance edges. **d,** New transition types added per syllable type. **a–d,** Means ± s. e. m. across children (n=9) are deviations from chance level (zero, red dashed line, assessed by bootstrap analysis). Grey lines, fitted linear model.