LETTER

doi:10.1038/nature12173

Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants

Dina Lipkind[1,2], Gary F. Marcus[3], Douglas K. Bemis[3], Kazutoshi Sasahara[4], Nori Jacoby[5,6], Miki Takahasi[4], Kenta Suzuki[4,7], Olga Feher[1,2,8], Primoz Ravbar[2,8], Kazuo Okanoya[4,7] & Ofer Tchernichovski[1,2,8]

Human language, as well as birdsong, relies on the ability to arrange vocal elements in new sequences. However, little is known about the ontogenetic origin of this capacity. Here we track the development of vocal combinatorial capacity in three species of vocal learners, combining an experimental approach in zebra finches (*Taeniopygia guttata*) with an analysis of natural development of vocal transitions in Bengalese finches (*Lonchura striata domestica*) and pre-lingual human infants. We find 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 new pairwise syllable transitions, sometimes too slowly to accomplish the task fully. Similarly, in the more complex songs of Bengalese finches, branching points and bidirectional transitions in song syntax were acquired in a stepwise fashion, starting from a more restrictive set of vocal transitions. The babbling of pre-lingual human infants showed a similar pattern: instead of a single developmental shift from reduplicated to variegated babbling (that is, from repetitive to diverse sequences), we observed multiple shifts, where each new syllable type slowly acquired a diversity of pairwise 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 partly from the challenges in constructing new pairwise vocal transitions.

In the three species we studied, vocal behaviour 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. 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 rearrange vocal units flexibly is the starting point of vocal learning. Alternatively, the combinatorial machinery might develop slowly, through 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 after hatching), 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 (Supplementary Information section 1 and Supplementary Fig. 1)[5,6]. A bigram Markov model was found to account for the bulk of song sequence structure during the experimental period (Supplementary Information section 2 and Supplementary 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., and hereafter; Fig. 1c and Supplementary Fig. 3a). Extinction of the source song occurred before, 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 by 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 pairwise transitions (bigrams AC, CB and BA), the increase (adjustment) of their frequencies and the extinction of source transitions that were no longer required (Fig. 1d–g).

New transitions appeared sparsely over development, with a lag of 10.4 ± 1.91 days from training onset, and a gap of 6.4 ± 3.5 days between consecutive bigram appearances (Fig. 1d, e, Supplementary Information section 3.1 and Supplementary Fig. 3b, d, 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 (Supplementary Information sections 3.2–3.3 and Supplementary Fig. 3f–h). Time gaps showed no developmental trend (no significant correlation between first–second and second–third 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 end-point frequency (phase transition), the speed of which increased strongly with the order of bigram appearance: the time interval from 25% to 75% of the end-point frequency was 9.6 ± 2 days for the first bigram, 4.0 ± 0.9 days for the second and only 2.3 ± 0.2 days for the third and final bigram (Fig. 1g, left; $P = 0.018$, paired t-test first versus third bigram). Extinction of source bigrams lagged behind the appearances of first and second target bigrams (4.5 ± 3 and 6 ± 3 days, respectively), but occurred almost simultaneously with the appearance of the third target bigram (-0.3 ± 0.3 days), resulting in a prompt switch to exclusive target performance (Fig. 1g, right, and Supplementary Fig. 3c). The prompt and rapid changes observed once the last bigram appeared probably mirror capabilities not fully expressed earlier (Supplementary Information section 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.

[1]Department of Psychology, Hunter College, City University of New York, New York, New York 10065, USA. [2]Department of Biology, City College, City University of New York, New York, New York 10031, USA. [3]Department of Psychology, New York University, New York, New York 10003, USA. [4]Laboratory for Biolinguistics, RIKEN Brain Science Institute, Wako, Saitama 351-0198, Japan. [5]The Interdisciplinary Center for Neural Computation, Hebrew University, 91904 Jerusalem, Israel. [6]The Department of Music, Bar Ilan University, Ramat Gan 5290002, Israel. [7]JST ERATO Okanoya Emotional Information Project, Wako, Saitama 351-0198, Japan. [8]The City University of New York Graduate Center, New York, New York 10016, USA.

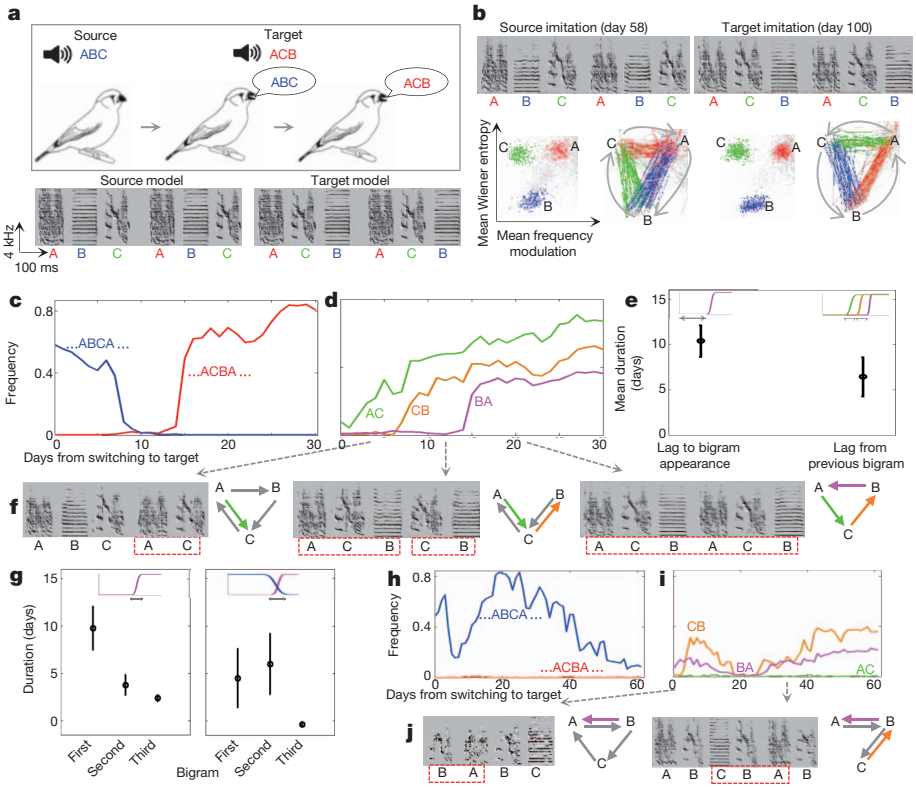

Figure 1 | Syllable rearrangement task. **a**, Top, sequential training with two songs; bottom, training models. **b**, Song examples (top) and scatter plots of syllable features (bottom) after source and after target learning in one bird. Clusters represent syllable types and lines represent transitions (colours represent transition end syllable). **c**, **d**, Daily frequencies (in one bird) of (**c**) source and target songs and (**d**) target bigrams. **e**, Learning phases in successful birds (means ± s.e.m.; $n = 8$). **f**, Songs and syntax diagrams during learning (same bird as in **c**, **d**). **g**, Duration of adjustment (left) and extinction (right) according to bigram appearance order. **h**–**j**, Same as **c**, **d** and **f** in an unsuccessful bird.

From the nine birds that failed to complete the task, five partly learned it (Fig. 1h–j, Supplementary Information section 4 and Supplementary 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, j), including a higher transition entropy which merely mirrored the coexistence of source and target bigrams (Supplementary Information section 5 and Supplementary Fig. 5). Thus, despite performing millions of syllable renditions, unsuccessful birds had no measurable capability of producing 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 $A^n \rightarrow (AB)^n$ (Fig. 2a, b, Supplementary Information section 6 and Supplementary 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, Supplementary Information section 6 and Supplementary Fig. 7a). However, birds did not directly insert syllable B into A^n strings; instead, they acquired two new transitions, AB and BA (Fig. 2e and Supplementary 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 first and second bigrams; 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 and Supplementary Information section 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

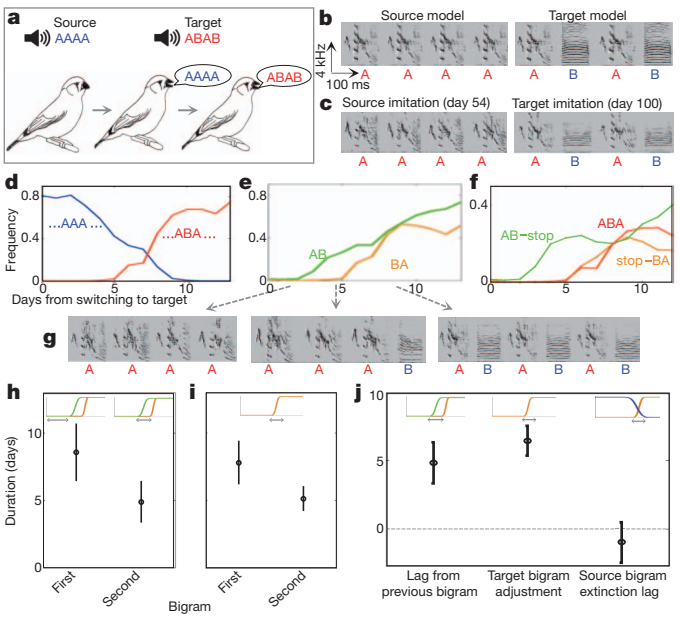

Figure 2 | Syllable insertion task. **a**, **b**, Training regime and models. **c**, Learning outcome in one bird. **d**–**f**, Daily frequencies of syllable sequences in one bird: **d**, source and target songs; **e**, target bigrams; **f**, occurrences of syllable B at bouts' end (green), start (orange) 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 and (**j**) lags between target bigrams' appearance, adjustment of the second target bigram and extinction of the source bigram (AA).

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 an '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 five birds and A^nB in three birds; Fig. 2f, Supplementary Information section 7 and Supplementary 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 before $(AB)^n$, indicating that the only obstacle for incorporating B into the bout centre 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 (Supplementary Information section 8 and Supplementary 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 new syllable type with a sequence rearrangement task.

By selecting for fast-learning birds and training them unnaturally, we might have underestimated the full range of combinatorial capabilities that birds might express under more natural conditions. To address this, we studied Bengalese finches, raised in a semi-natural aviary ($n = 8$; Supplementary Information section 9). Although 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 bidirectional 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 end-point 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 after hatching) to the end of development (Fig. 3a and Supplementary Fig. 9a). We found long gaps between bigram appearances (17.7 ± 8.7 days; Fig. 3a–c and Supplementary 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 (Supplementary Information section 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 ± 2% versus 1 ± 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, five out of eight birds kept adding and removing transitions. As in early song development, this process was biased to increase connectivity across syllable types (Fig. 3e; 15 additions versus six deletions across birds) and decrease repetitive sequences (zero additions versus four deletions). 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 the start and end points, 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 (for example, 'ba ba ba') to variegated utterances (for example, 'ba gu ge')[8,9], which could perhaps mirror a stepwise acquisition of variable transition types. However, later studies failed to replicate this effect reliably[10,11], and instead reported variegated utterances throughout babbling development (Supplementary Information section 10). This could suggest that, unlike songbirds, human infants can rearrange syllables early on with relative ease. However,

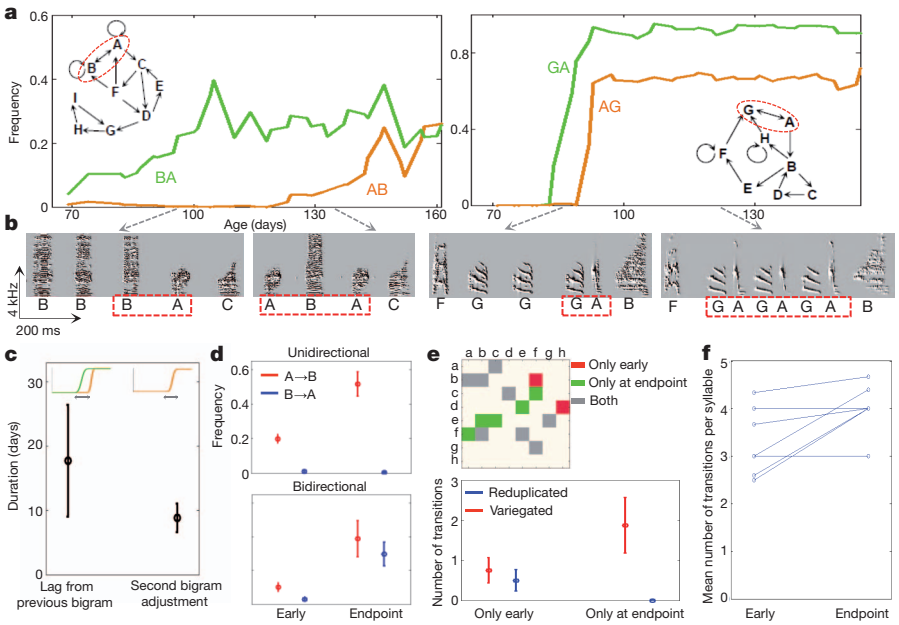

Figure 3 | Combinatorial learning in Bengalese finches. **a**, **b**, Development of bidirectional transitions in two birds. Insets show end-point syntax. **c**, Mean ± s.e.m. of appearance lag and adjustment duration in bidirectional transitions ($n = 7$ transitions). **d**, Means ± s.e.m. of the frequencies of unidirectional (top, $n = 16$) and bidirectional (bottom, $n = 7$) transitions in early development and at end point. **e**, Top, binary transition matrix (one bird), showing transitions present only early (green), only at end point (red) and in both (grey). Bottom, means ± s.e.m. across birds ($n = 8$) of the number of transitions present only early or only at end point. **f**, Developmental changes in the mean number of transitions per syllable (in cases of variable transitions).

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 analysed databases of phonologically transcribed babbling sessions (CHILDES[12,13]) from nine US infants recorded once every 2 weeks at ages 9–28 months, which we segmented into syllables (Methods, see section on 'Analysis of babbling data', and Supplementary Information sections 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 (Supplementary Information section 11.3).

We first tracked the frequencies of reduplicated transitions over infants' ages, aligning the data in reference to the age where the 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. 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 ($28 \pm 8\%$ 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.

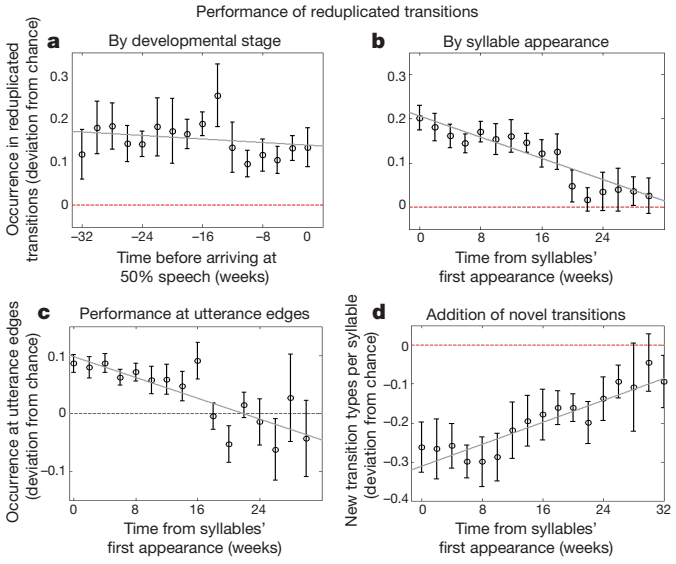

Figure 4 | Incorporation of new syllables into infants' babbling utterances. **a**, **b**, Frequency of syllable occurrence in reduplicated transitions: (**a**) data aligned by developmental stage (time zero is the first session with more than 50% of speech utterances); (**b**) data aligned by each syllable 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.

Our results across species suggest that, in contrast to predictions of previous theories[14,15], new vocal transitions are acquired slowly during early stages of development. A similar, gradual, generative process was observed in the development of non-learned movement sequences[16,17], although intriguingly, not in studies of movement sequence learning in adult monkeys, where target sequences appeared very rapidly but frequency adjustments took weeks[18,19]. Whether these differences are due to age, experience or learning modules (movement versus vocal) should be investigated in future work.

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 (Supplementary Information section 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 gained due to acquiring new transitions is counterbalanced by the acquisition of new syllable types that are not yet connected and that tend to reduplicate or break the sequence. Such a 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 pairwise syllable transitions to transform one multi-syllable string into another.

METHODS SUMMARY

Animal care. All experiments were conducted in accordance with the guidelines of the US National Institutes of Health and were reviewed and approved by the Institutional Animal Care and Use Committees of Hunter College and City College of the City University of New York, and of RIKEN Brain Science Institute.
Experimental design. Male zebra finches were reared and housed singly, in sound attenuation chambers as previously described[5]. Audio-recording and training with song playbacks used Sound Analysis Pro[6]. Bengalese finches were reared in communal aviaries in RIKEN Brain Science Institute, Japan. Starting from 40 to 50 days of age, they were transferred singly to soundproof cages at 3- to 4-day intervals and recorded for approximately 24 h.

Data analysis. We used Sound Analysis Pro[6] for song feature calculation and cluster analysis. We used Matlab for the rest of the analysis. Cluster information was used to elucidate the order of syllable types sung. Daily frequencies of sequences (source and target songs and bigrams) were calculated as the proportion of syllables constituting the sequence out of the total number of syllables sung on that day.

Infant babbling data were obtained from the Davis corpus[13] of the CHILDES database[12]. Only babbling utterances were analysed. Utterances were semi-athomatically parsed into syllables (Methods, see section on 'Analysis of babbling data'). Bootstrap normalization was used to establish a value for each measure reflecting a random placement of syllables with vocabulary size and utterance length held constant (Methods, see section on 'Analysis of babbling data'). Measures were calculated for each syllable type in each session, and a mean over syllable types was calculated for each time point.

Full Methods and any associated references are available in the online version of the paper.

Received 5 April 2012; accepted 8 April 2013.
Published online 29 May 2013.

1. Marcus, G. F., Vijayan, S., Bandi Rao, S. & Vishton, P. M. Rule learning by seven-month-old infants. *Science* **283,** 77–80 (1999).
2. Berwick, R. C., Okanoya, K., Beckers, G. J. L. & Bolhuis, J. J. Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* **15,** 113–121 (2011).
3. Eales, L. Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Anim. Behav.* **33,** 1293–1300 (1985).
4. Plamondon, S. L., Rose, G. J. & Goller, F. Roles of syntax information in directing song development in white-crowned sparrows (*Zonotrichia leucophrys*). *J. Comp. Psychol.* **124,** 117–132 (2010).
5. Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C. & Tchernichovski, O. How sleep affects the developmental learning of bird song. *Nature* **433,** 710–716 (2005).
6. Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. & Mitra, P. P. A procedure for an automated measurement of song similarity. *Anim. Behav.* **59,** 1167–1176 (2000).
7. Yamashita, Y. *et al.* Developmental learning of complex syntactical song in the Bengalese finch: a neural network model. *Neural Netw.* **21,** 1224–1231 (2008).
8. Oller, D. K. in *Child Phonology* Vol. 1 (eds Yeni-Komshian, G., J. Kavanagh, J. & Ferguson, C.) 93–112 (Academic, 1980).
9. Stark, R. in *Child Phonology* Vol. 1 (eds Yeni-Komshian, G., J. Kavanagh, J. & Ferguson, C.) 73–92 (Academic, 1980).
10. Mitchell, P. R. & Kent, R. D. Phonetic variation in multisyllable babbling. *J. Child Lang.* **17,** 247–265 (1990).
11. Smith, B. L., Brown-Sweeney, S. & Stoel-Gammon, C. A quantitative analysis of reduplicated and variegated babbling. *First Lang.* **9,** 175–189 (1989).
12. MacWhinney, B. The CHILDES project: tools for analyzing talk. *Child Lang. Teach. Ther.* **8,** 217–218 (1992).
13. Davis, B. L. & MacNeilage, P. F. The articulatory basis of babbling. *J. Speech Hear. Res.* **38,** 1199–1211 (1995).
14. Edelman, G. *Neural Darwinism. The Theory of Neuronal Group Selection* (Basic Books, 1987).
15. Hanuschkin, A., Diesmann, M. & Morrison, A. A reafferent and feed-forward model of song syntax generation in the Bengalese finch. *J. Comput. Neurosci.* **31,** 509–532 (2011).
16. Golani, I. A mobility gradient in the organization of vertebrate movement?: the perception of movement through symbolic language. *Behav. Brain Sci.* **15,** 249–308 (1992).
17. Dominici, N. *et al.* Locomotor primitives in newborn babies and their development. *Science* **334,** 997–999 (2012).
18. Hikosaka, O., Rand, M. K., Miyachi, S. & Miyashita, K. Learning of sequential movements in the monkey: process of learning and retention of memory. *J. Neurophysiol.* **74,** 1652–1661 (1995).
19. Rand, M. K. *et al.* Characteristics of sequential movements during early learning period in monkeys. *Exp. Brain Res.* **131,** 293–304 (2000).
20. De Boysson-Bardies, B. & Vihman, M. M. Adaptation to language: evidence from babbling and first words in four languages. *Language* **67,** 297–319 (1991).
21. Vihman, M. M., Macken, M. A., Miller, R., Simmons, H. & Miller, J. From babbling to speech: a re-assessment of the continuity issue. *Language* **61,** 397–445 (1985).
22. Jin, D. Z., Ramazanoğlu, F. M. & Seung, H. S. Intrinsic bursting enhances the robustness of a neural network model of sequence generation by avian brain area HVC. *J. Comput. Neurosci.* **23,** 283–299 (2007).
23. Amador, A., Perl, Y. S., Mindlin, G. B. & Margoliash, D. Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* **495,** 59–64 (2013).
24. Jin, D. Z. Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. *Phys. Rev. E* **80,** 051902 (2009).
25. Abe, K. & Watanabe, D. Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neurosci.* **14,** 1067–1074 (2011).
26. Beckers, G. J. L., Bolhuis, J. J., Okanoya, K. & Berwick, R. C. Birdsong neurolinguistics: songbird context-free grammars claim is premature. *NeuroReport* **23,** 139–145 (2012).
27. Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. Recursive syntactic pattern learning by songbirds. *Nature* **440,** 1204–1207 (2006).
28. Katahira, K., Suzuki, K., Okanoya, K. & Okada, M. Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLoS ONE* **6,** e24516 (2011).
29. Van Heijningen, C. A. A., de Visser, J., Zuidema, W. & ten Cate, C. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl Acad. Sci. USA* **106,** 20538–20543 (2009).
30. Rose, G. J. *et al.* Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* **432,** 753–758 (2004).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Benichov, J. Hyland Bruno, I. Ljubičić and C. Roeske for help with data analysis. We also thank A. Vouloumanos, M. Hauber, L. Parra and V. Valian for reading the manuscript. The study was supported by a US Public Health Service grant to O.T., by a Grant in Aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan to K.O., and by a Grant in Aid for Japan Society for the Promotion of Science Fellows to M.T.

Author Contributions D.L., O.T., G.F.M., D.K.B., Ka.S. and K.O. designed the research. D.L., O.F. and O.T. performed experiments on zebra finches. D.L., G.F.M., O.F., P.R., N.J. and O.T. analysed data of zebra finches. Ka.S., M.T., Ke.S. and K.O. designed and conducted experiments on Bengalese finches. D.L., Ka.S. and O.T. analysed data of Bengalese finches. D.K.B. analysed infant babbling data, with contributions from G.F.M., O.T. and D.L. D.L., G.F.M., D.K.B., K.O. and O.T. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.L. (dina.lipkind@gmail.com).

METHODS

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

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 and 30 after hatching. 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 days 33 to 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 used 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 soundproof cage at 3- to 4-day intervals and recorded for approximately 24 h.

Data analysis (songbirds). Song feature calculation and cluster analysis were performed using Sound Analysis Pro[6]. We used MATLAB 7 for further analysis. 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 were reused in the learning of new syntax (Supplementary Information section 1).

In zebra finches, the percentage of clustered syllables in bouts (assessed by manual inspection of a sample of ten song bouts per bird) was 96 ± 3% at the end point, and 90 ± 2% during the transition from source to target. Clustering Bengalese finch songs was more difficult, with 91 ± 1% at the end point and 80 ± 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 end-point song (150 ms for ABC–ABC → ACB–ACB and Bengalese finches; 100 ms for AAAA → ABAB). The threshold for bigram occurrences was similarly defined by the typical stop duration in each bigram type at the end point (60–150 ms). Daily frequencies of sequences (source and target songs and bigrams) were calculated as the proportion of syllables constituting a given sequence out of the total number of syllables sung on that day. Because of unavoidable misclassifications in the clustering process, we had to determine a margin of error to decide when an observed transition frequency was 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 transition at 3% above noise (that is, 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 (Supplementary Information section 3.2), 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 to determine 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, on the basis of visually examining 20 random instances of each transition type.

Analysis of babbling data. Data were obtained from nine children in the Davis corpus[13] of the CHILDES database[12]. On average, children were 9 months 28.3 days old (s.d. 2 months 1.3 days) at the first session, and data were collected for an average of 1 year 7 months (s.d. 7 months 12.8 days). Data consisted of 38.8 sessions on average per child (s.d. 10.2), recorded an average of 16.07 days apart (s.d. 6.4).

Only babbling utterances (that is, utterances for which no lexical items were assigned in the CHILDES transcriptions) were analysed. Utterances were parsed into syllables using a semi-automated method, described below. Only utterances that received a complete syllabic parse were analysed (2135 utterances per child (s.d. 924) and 62.0 utterances per session (s.d. 37.5)).

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 in the data. If an utterance could be assigned two different parses, we used a heuristic such that we chose the parse with the greater number of two phoneme syllables (CV or VC). If several 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 (s.d. 19.0)).

From this set of parsed utterances, we tabulated the frequency of each syllable in each session and its placement in an utterance. We restricted analysis 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 in an utterance. On average, each child used 128 distinct syllables (s.d. 8.12) and constructed 763 distinct transitions (s.d. 95.6).

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 in 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.

All measures were calculated for each syllable type in each session. Sessions were then aligned on the first appearance of syllable types, 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, available at http://cran.r-project.org). Separate models were fitted 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 analysed.

Reduplication was 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 the first appearance of the syllable type. Note that the sample size for the developmental alignment analysis was smaller ($n = 7$) than for the syllable-specific alignment ($n = 9$), because of an insufficient number of sessions with more than 50% babbling utterances in two children.

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

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