Balanced imitation sustains song culture in zebra finches

Ofer Tchernichovski

Hunter College

Sophie Eisenberg-Edidin

Hunter College CUNY

Erich Jarvis (✉ ejarvis@rockefeller.edu)

Rockefeller University <https://orcid.org/0000-0001-8931-5049>

Article

Keywords: zebra finch, songs, vocal learning, song imitation

Posted Date: September 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-80497/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on May 7th, 2021.
See the published version at <https://doi.org/10.1038/s41467-021-22852-3>.

Abstract

Juvenile songbirds acquire songs by imitation, as humans do speech. Although imitation should drive convergence within a group and divergence through drift between groups, zebra finch songs sustain high diversity within a colony, but only mild variation across colonies. We investigated this phenomenon by analyzing vocal learning statistics in 160 tutor-pupil pairs from a large breeding colony. Song imitation was persistently accurate in some families, but poor in other families. This could not be attributed to genetic differences, as fostered pupils copied their tutors' vocal sounds as accurately or as poorly as the tutor's biological pupils. We discovered two effects that explained the finding: first, even in cases of accurate imitation, pupils often recombined imitated syllables to form new units, and, therefore, distributions of syllable types in pupils' songs were not well correlated with their tutors'; second, rare vocalizations in tutors' songs became more abundant in their pupils' songs, and vice versa. Consequently, cultural transmission of tutor songs that were high in acoustic diversity were stronger than those that were low in diversity. We suggest that a frequency dependent balanced imitation of vocal repertoires prevents the extinction of rare song elements and the overabundance of common ones. Within a group it promotes repertoire diversity, while across groups it constrains drift. Together with syllable recombination, balanced imitation sustains cross-generational homeostasis that prevents the collapse of vocal culture into either complete uniformity or chaos.

Background

Vocal culture is the cornerstone of spoken language, but is not unique to humans¹⁻⁵. Like humans, songbirds acquire their vocal repertoire via imitation (i.e. vocal learning)⁶⁻⁹, a process that can give rise to local dialects that persist over hundreds of generations^{10,11}. However, the repertoire of vocal learning birds also has a strong genetic component¹¹⁻¹³. Across populations, innate biases in song perception, production, and learning sustain species-specific song repertoires¹³⁻¹⁵. Canaries, for example, will faithfully imitate songs of abnormal combinatorial structure, but later, as they reach maturity, alter their songs to match a species typical song syntax to which they have not been exposed¹⁶. Similarly, zebra finch males (females do not sing) that are trained with random combinatorial transitions of syllable types will generate combinations that are biased toward the species typical^{17,18}. Innate biases may unfold at the scale of generations, too; the descendants of isolate zebra finch tutors, who produce aberrant songs, produce increasingly species typical species songs^{2,19}.

Theoretically, vocal imitation should drive song repertoire convergence within groups and divergence across groups²⁰⁻²². Meanwhile, innate biases in imitation might constrain drift. In reality, however, zebra finch songs remain highly diverse within groups and vary only mildly across them²². We do not know if this diversity serves any function in domesticated zebra finches, but high similarity between songs could potentially generate impoverished communication systems that convey little information about individual identity^{23,24}. In wild songbirds, across species, and even subspecies, the magnitude of individual song variability differs strongly, often for no apparent reason. For example, the songs of the wild Australian

zebra finch (*Taeniopygia guttata castanotis*) are much more variable across individuals than those of the closely related wild Timor zebra finch (*Taeniopygia guttata guttata*)²⁵. This variability persists despite the fact that they live in similar climates and have similar social organization.

Here we test how acoustic diversity of zebra finch songs is sustained during cultural transmission²⁶ by studying the statistics of song imitation in a large colony. We observed high variability in song imitation outcomes across families. However, song imitations were similarly accurate across biological and foster pupils, suggesting an environmental effect. Indeed, we found that pupil song outcome was influenced by the acoustic diversity of the tutor song. This finding led us to investigate the cultural transmission of song acoustic diversity. We first examined how song acoustic diversity is sustained in our colony, and then examined song diversity in other independent zebra finch colonies.

Results

We recorded the songs of 160 zebra finch tutor-pupil pairs (68 tutors & 160 pupils; 228 birds overall) at the Rockefeller University Field Research Center colony, which consisted of over 800 birds during the 1-year period of recording. Of the 160 pupils, 130 pupils were housed with their biological parents, and 30 pupils with foster parents. All birds were housed in individual breeding cages with parents (either biological or foster) and other offspring, and kept visually isolated from other breeding cages. With this social regimen, we found no evidence of song imitation across families. From each bird we recorded undirected songs (produced in isolation) for over a week to obtain a sample of at least 1000 song syllables per bird.

Imitation outcome varied across families

We first measured similarity²⁷ between tutor and pupil songs based on acoustic features²⁸. We observed considerable variability in imitation outcomes (mean song similarity = 69%, S.D.=18.9%; **Suppl. Fig. 1**). In order to evaluate imitation outcomes within and across families we identified 25 families that had multiple clutches including males. For each clutch, we calculated the mean song similarity between tutor and pupils. This allowed us to exclude the effect of song convergence between siblings²⁰, and compare the variance in imitation outcome within and across families. We found that imitation similarity was much more variable across families than within families (Kruskal-Wallis chi-squared = 44.727, df = 24, p-value = 0.006). In certain families, across clutches, song imitation tended to be almost exclusively accurate (top quartile), in some modest (middle quartile), and others generally poor (**Fig. 1a**).

We then compared song imitation between biological and foster pupils. Foster pupils imitated their tutor as well as biological ones (biological similarity: $68.2 \pm 1.7\%$, $n=130$; foster similarity: $70.0 \pm 3.6\%$, $n=30$, mean $\pm$ S.E.M. hereafter). Therefore, the variability we observed in imitation outcome across families cannot be explained by genetic variability. Instead, we found that variability in imitation among pupils was associated with tutor song structure. For example, tutor Aq12 had a very simple song and all of his pupils imitated it poorly, sometimes introducing apparently novel syllable types in developing their own

songs (**Fig. 1b**). In contrast, tutor DG1 had a more complex song, and all of his pupils imitated it much more accurately, with little to no introduction of novel syllables (**Fig. 1c**). In both kinds of cases, the birds still produced their syllables in the typical repeated song motif of 2-6 syllable types (**Fig. 1b,c**). This suggested to us that pupils might more accurately imitate tutor songs that are rich in acoustic structure (i.e., acoustically diverse), while improvising upon impoverished tutor songs. To test this hypothesis quantitatively, we compared the diversity of song syllable types in tutors and their pupils.

Syllable type diversity is not correlated between tutor and pupil songs

We selected a random group of 80 adult tutor-pupil pairs, and segmented their songs into syllable units using an amplitude threshold²⁷. Song syllables were automatically clustered into types based on their acoustic features (**Fig. 2a-b**)²⁷. We then calculated the relative frequency (abundance) of each syllable type and used Shannon information entropy²⁷ to estimate the diversity of song syllable types produced by each bird. Specifically, for each bird's song, we calculated the proportion of syllables produced for each syllable type i , and computed H , an entropy estimate of syllable type diversity. We then tested if the syllable type diversity was correlated in tutors and their pupils. Surprisingly, there was no statistically significant correlation between tutor and pupil syllable diversity (**Fig. 2c**, overall $R^2=0.079$, NS). Further, there was no apparent interaction between song similarity and syllable diversity (**Fig. 2c**, red vs. green markers). Next, we used the same Shannon information measure to estimate the entropy of sequence (bigram) transitions between syllable types. Again, we found no correlation between tutors and pupils (**Fig. 2d**, **Supp. Fig 2**).

Half of pupils recombine syllables

Puzzled by the lack of correlation between tutors' and pupils' syllable type and sequence diversity, we examined cases of most accurate imitation. We found frequent inconsistencies in the boundaries of corresponding syllables in the songs of tutors and their pupils. This was not primarily due to measurement (segmentation) errors, but because pupils often modified or recombined the units they imitated (**Fig. 2e**). We assessed a lower bound estimate of similarity in the syllable boundaries of tutor and pupil songs, restricting analysis to those syllables whose acoustic structure was clearly and fully imitated by the pupil (examples in **Fig. 2e-i**). With this strict criterion, analysis of syllable imitations in 33 randomly selected tutor-pupil pairs revealed modification of syllable boundaries in 47 cases (22%) of the copied syllables. Overall, 54% (18/33) of the pupils showed at least one case of altering syllables units. Interestingly, all 47 cases were of merging tutor syllables, rather than splitting.

Vocal state measures capture balanced imitation

Given the extent of syllable recombination, we next sought a quantitative measure that captures acoustic diversity at the sub-syllabic level, which would be, by design, insensitive to syllable recombination. For each of the 160 tutor-pupil family pairs, we calculated continuously (in 10ms FFT windows excluding silences, but without segmentation) three acoustic feature vectors: pitch, Wiener entropy (width of power

spectrum), and frequency modulation (**Fig. 3a**)²⁸. A 3D scatterplot of these features for all birdsongs reveals several concentrations, and we used the contours of these concentrations to partition the entire acoustic space of the songs into 10 regions (**Fig. 3b**). To better visualize these concentrations, we present 2D slices of the feature space according to four peaks in the distribution of pitch, that we labelled very low, low, medium, and high (**Fig. 3c**). These four slices show distinct concentrations of the 10 regions, that we will call vocal states (**Fig. 3d**). The two concentrations in the highest and lowest pitch regions consisted of down-modulated and up-modulated sounds, respectively (vocal states 1 & 2, for lowest pitch, and 9 & 10 for highest pitch). The two central pitch regions (low and medium) consisted of similar types vocal states, and two additional states (4 and 7) centered at zero frequency modulation of non-modulated harmonic sounds. With the vocal states of the population categorized, we next calculated the relative abundances of sounds within each vocal state for each bird.

Similar to syllable and transition diversity, for each bird's song, we calculated acoustic diversity over the 10 vocal states using Shannon information entropy²⁷, , but here i is the proportion of sounds within each vocal state i . Tutor and pupil vocal state diversities were positively and significantly correlated (**Fig. 3e**, $R^2 = 0.21$, linear fixed-effect model: $n=160$ pairs, estimate = 0.41, $t=6.53$, $p=8.2e-10$), unlike the syllable diversity measures. However, vocal state diversity of the tutor's song explains only a fraction (21%) of the variance observed in pupil songs. In cases of poor imitation (below median, red markers), the correlation was driven primarily by a few tutors of extremely low song diversity, all of which belong to a single branch in our colony family tree. This may call into question the validity of the correlation observed in lower quartiles. However, at least, in the top two quartiles the coefficient of determination seems robust ($R^2 = 0.23$, **Fig. 3e**) with no apparent interaction with specific family branches. We next investigated if song diversity at the level of vocal states can explain the variability in song imitation across families.

The distribution of song diversity remained stable over the lifespan of our colony (**Suppl. Fig. 3**). The highest theoretically possible diversity, with a uniform distribution of the 10 vocal states, is 10 times $-0.1(\log_2(0.1) = 3.32$ bits. Interestingly, the median song diversity of the population was 3.14 bits (**Fig. 3f**), fairly close to the upper theoretical limit, suggesting a trend to develop acoustically "balanced" songs with respect to the 10 vocal states. The distribution of song diversity was asymmetric with a longer left tail (**Fig. 3f**) going down to about 2.3 bits, which is equivalent to about 50% vacant vocal states. Plotting song diversity of each tutor against the mean song diversity of all of his pupils showed that the mean song diversity in pupils of low diversity (below median) tutors is often higher than that of their tutors, and vice versa (**Fig. 3g**). That is, despite the positive correlation between tutor and pupil song diversity, we see frequent reversals such that a large proportion of pupils with low song diversity had tutors with high song diversity, and vice versa.

We further explored these reversals across multiple generations, and analyzed 14 family branches, where we had song imitation data across two generations of pupils. We found that in the families where the first-generation pupils imitated poorly, there was often some recovery in the second-generation, the grand-pupils (**Fig. 4a**). For example, in the two lineages (HP10 and DG4) with the greatest number of first-

generation pupils that imitated poorly, all of the grand-pupils imitated more accurately. Sonograms revealed that, in both lineages, the grand-tutor songs were atypical. Tutor HP10 had a very high-pitched song (**Fig. 4b**), whereas tutor DG4's song included numerous harmonic stacks (**Fig. 4c**). In both cases, their pupils developed songs that appear to be more acoustically "balanced," and ones that the grand-pupils imitated accurately (**Fig. 4b-c**). In other cases, however, low similarity was simply due to poor imitation rather than balancing, e.g., in lineage (LB12), where the song imitation became progressively worse because a grand-pupil dropped a syllable during imitation (**Fig. 4d**).

To better estimate how tutor song vocal state diversity may affect cultural transmission, we calculated song acoustic similarity in reverse, from pupil to tutor. We call this a measure of "influence" because it tells us how much of the pupil's song is influenced by the tutor. For example, an imitation ABC->ABCDE will give us 100% imitation similarity because all of the tutor's sounds are present in pupil's song, but only 50% influence because half of the pupil's song is improvised. We found that tutors with low song diversity had a weaker influence on the songs of their pupils compared to high diversity tutors (**Fig. 4e**, $R^2=0.25$, $t=4.8$, $p=4.2e-6$). Interestingly, however, imitation similarity of syllables was not significantly affected by tutor's song diversity (**Fig. 4f**, $R^2=0.08$, NS). That is, pupils of tutors with low song diversity imitated them, but also made additions, and were therefore less influenced by them. We suspect that these additions are improvisations, as they did not resemble syllables of neighboring birds' songs. Assuming a natural trend to develop low diversity songs either via imitation or improvisation, we asked why several birds still developed low diversity songs. We found that pupils that imitated poorly, regardless of tutor song diversity, tended to have low diversity songs ($R^2=0.20$). This may suggest that low song diversity is an outcome of poor imitation. To directly test for interaction between imitation accuracy and song diversity, we ran a linear mixed-effect model to explain pupil song diversity with two fixed effects: the diversity of the tutor song, and the acoustic similarity to the tutor song (how much of it was copied). Results confirmed that both factors contribute about equally to pupil song diversity (imitation similarity: $t=5.0$, $p=1.4e-6$; tutor song diversity: $t=4.6$, $p=7.9e-6$).

Our results so far are consistent with the following interpretation of cross-generational dynamics: Pupils of tutors with high song diversity who imitate well, produce songs of comparably high diversity. Pupils who imitate less accurately, though, tend to develop songs that are lower in diversity and, therefore, lower in influence as well. Grand-pupils imitate the deficient songs of these tutors, but they also improvise, thus increasing the diversity of their songs.

Balanced imitation of vocal state abundances

Our measures up to now summarize the distribution of vocal states within a song. To test our hypothesis further, we look at each vocal state separately and measure how raw frequencies (abundances) of vocal states are imitated. In prior studies, we noted that vocal imitation in zebra finches is inversely related to model abundance. That is, too much exposure to a tutored song could reduce learning²⁹. Here we test if this is the case also for abundances of vocal states within a song.

We calculated the abundance of each vocal state in tutor vs. pupil songs. We then partitioned the vocal state data into quartiles based on the overall acoustic similarity between tutor and pupil songs. For each tutor-pupil pair, in each quartile, we then plotted the relative abundances of all 10 corresponding vocal states in the tutor's song versus the pupil's song (**Fig. 5a-d**). We found that relative abundances of all 10 states were correlated, for each quartile. As expected, tutor-pupil vocal state abundances were more strongly correlated with higher syllable imitation; for example, the residual coefficient of determination was much higher in the top similarity quartile, explaining about 35% of the variance in cases of highest song similarity (**Fig. 5a**), and about 9% of the variance in the bottom quartile (**Fig. 5d**).

We noted that in all quartiles, the slope of the correlation was less than one (**Fig. 5a-d**), meaning that data tend to be above the diagonal when the abundance of a vocal state is low in a tutor's song, and below the diagonal when the abundance is high. To visualize this result across the entire data set, we graphed the difference in all 10 vocal state abundances between all tutor and pupil pairs, which showed a clear regression to the mean (**Fig. 5e**). We would expect this regression towards the mean to be strong in cases of poor imitation (i.e., indicating randomness), and weaker in cases of good imitation. To evaluate if this was the case, we calculated the median abundance of all 10 vocal states in small abundance bins for tutors and pupils. For each bin, we plotted the abundance ratio between tutors and pupils. A gain value of 1 (y-axis in **Fig. 5f**) represents identical abundance of all 10 vocal states in pupil and tutor. A gain value of 2 indicates a doubling of abundances in the pupil (amplification), and a value of 0.5 halving (attenuation). Interestingly, the gain-loss curves have similar shapes and magnitude across all four quartile groups (**Fig. 5f**). In all cases, the zero crossing, a gain of 1 (where abundance tends to be identical across pupils and their tutors), was at 11-12% abundance, which is fairly close to the center of the distribution (= 10%, since we have 10 vocal states). These findings suggest that the regression we noted is not an entirely random effect. For example, in Q1, where mean similarity is 93%, we see that when tutor state abundance is above 0.2, the corresponding pupil abundance is lower in 10 out of 11 cases (**Fig. 5a**). In all these cases, the corresponding vocal sounds were imitated, but produced either less often, or with biased features, by the pupil.

To visualize and quantify relative vocal state abundances in tutors vs their pupils, we reduced the ten vocal states into four color codes, and graphed them along the sonograms of each bird (**Fig. 5g-i**). In cases where the tutors' songs included many high-pitched vocalizations (vocal states 9 and 10), their pupils imitated, but lowered the pitch, thereby decreasing the abundance of those states (**Fig. 5g,h**). In another example, where the tutor's song had high abundance of harmonic stacks (states 4 and 7), their pupil imitated only some of them (**Fig. 5i**). In turn, in the following generation, the pupil's pupil further differentiated his song to include more balanced vocal states (**Fig. 5i**). Taken together, song imitation appears to be highly sensitive to the relative abundances of vocal states, suggesting a balancing mechanism that prevents song diversity from becoming too low, perhaps independently of imitation.

Finally, as we did earlier for individual syllables, we asked whether foster pupils imitate their tutor's song vocal states as accurately as biological pupils. Analysis at the level of vocal states allowed us to compare how abundances of vocal states are influenced by foster vs. biological fathers. For reference,

imitation of vocal state abundances between the 130 biological pupils and their fathers had an $R^2=0.16$ (**Fig. 5j**; $t=5.9$, $p=3.9e-09$). The 30 foster pupils relative to their foster fathers had a similar $R^2=0.19$ (**Fig. 5k**; $t=2.5$, $p=0.01$). However, there was near zero correlation between fostered pupils and their biological fathers (**Fig. 5l**, residual $R^2=0.01$, $t=0.46$, NS). Therefore, the similarities we observed in vocal state abundances between tutors and their pupils reflect learning with no detectable genetic effect at this level of analysis.

How balanced imitation constrains distributions of song features

High diversity songs are, by definition, balanced, and cannot be extreme in their mean feature values. Low diversity songs can, in principle, have average features that are close to the population mean, but are more likely to have extreme mean feature values. For example, a song containing mostly high-pitched sounds is both low diversity and extreme in its mean pitch. Can we predict imitation outcomes based on the mean features of a tutor song? To show this, we plotted the mean pitch of tutor songs against the pitch of their pupil's songs. Indeed, we see that the distribution of mean song pitch is tighter for the top quartile of tutor-pupil song imitation. Note, for example, that all songs with mean pitch above 2000Hz were imitated poorly (these extreme songs are also of low diversity, **Fig. 6a**, histograms). As expected, in the top quartile of best imitations, the pitch values were highly correlated between tutors and pupils, whereas in the bottom quartile they were not correlated (**Fig. 6a**, regression graphs). A similar effect can be seen in Wiener entropy (**Fig. 6b**) and frequency modulation (**Fig. 6c**). Green brackets above the green histograms outline 90% interval of mean feature values of the top quartile pupil imitations: mean pitches 1000-2000 Hz, mean Wiener entropy between -2.7 to -2.0, and mean frequency modulation between 30-50. We next superimposed these intervals on ranges of mean song features obtained in other colonies. Figure 6d-f show the distribution of mean song features (pitch, Wiener entropy, and frequency modulation) in a database of four zebra finch colonies (including the current one). Across the three other colonies, the distribution of mean song features was to a large extent confined within the range of good imitations in our colony. Therefore, the range of mean feature values of best imitations, but not of worst imitations, seems consistent across zebra finch colonies. This range, in turn, can be explained by balanced imitation.

Discussion

We analyzed song learning statistics in a large zebra finch colony. We observed high variability in song imitation outcomes across families. It did not stem from genetic variability, but rather was explained by an environmental effect: the acoustic structure of the tutor's song. Pupil song imitation was more influenced by tutors who produced songs of high acoustic diversity. In order to identify this relationship, it was necessary to develop a measure of song diversity at the sub-syllabic level, where we detected 10 vocal states common across zebra finch songs. We found that pupils copy the abundances of vocal states in their songs from the songs of their tutors, but that they do so in a balanced manner, such that

highly abundant vocal states in a tutor's song become less abundant in their pupils' song and vice versa. We discovered that extreme mean song features, which are associated with low acoustic diversity, are also associated with poor imitation. The converse is associated with good imitation. Similar moderate mean song features were more highly present in three independent colonies of zebra finches, suggesting a species-specific mechanism that can be explained by an innate bias to produce acoustically balanced songs. Our findings suggest that this bias is highly sensitive to even mild fluctuations in vocal state abundance in the tutor's song and is independent from imitation outcomes. We call the process, where vocal imitation of sounds is balanced by vocal states abundance, "balanced imitation." We suggest that balanced imitation prevents vocal cultural learning from converging too much into complete uniformity or diverging too much into chaos.

Given that the mean vocal state diversity of birdsongs in our colony was close to the theoretical upper limit of diversity, one might wonder why tutors with low diversity songs are not rarer in our colony. We observed that low diversity songs are often corrected toward high diversity songs in the grand-pupils of low diversity song tutors. This correction via cross generational dynamics may sustain homeostasis of high song diversity. But it is not unlikely that other social forces drive songs toward low diversity. This study, as well as previous ones, reported strong variability in song imitation success across birds. For example, we previously reported social inhibition of song learning in clutches that contain more than two male siblings³⁰. This partial imitation of the tutor's song due to social inhibition may lead to impoverished songs. The current study confirms that a proportion of pupils of high diversity tutors acquired very low diversity songs, perhaps due to social inhibition of song imitation. Finally, we note anecdotally that some of these poor song learners were aggressive, dominant birds that were among the best breeders in our colony, in terms of number of offspring sired. This is counterintuitive, given that zebra finch females are less likely to ovulate in response to males who imitated poorly³⁰. Further studies should look into fitness consequences of such phenotypes.

The current study generalizes upon previous studies^{16,31,32} that documented the emergence of song culture from the impoverished song of isolated founders^{2,19}. Songs of isolate zebra finches often contain abnormally high abundance of some song elements. Pupils of isolate tutors normalized these high abundances in their imitations, reducing them to the species typical range. Our statistical analysis suggests that in those studies², song normalization across generations was driven, at least in part, by the balanced imitation discovered here: song syllables that were abnormally high in abundance in the isolate tutors were copied at lower frequencies, and vice versa, syllable types that were rare in the isolate song were copied at higher frequency. Interestingly, this effect is not driven by the overall abnormality of the isolate songs. Instead, it occurs when tutor songs are well within the normal range, at the micro level of balancing vocal state abundances. Further, biases in the statistical learning of vocal state abundance can explain variability in imitation outcome across families, and allow us to distinguish between low song similarity due to putative imitation failure, as opposed to "corrective" deviations, leading to a cross generational homeostasis of songs with high acoustic diversity.

It would be interesting to test if balanced imitation parameters are different across species. Variation in the intensity of the trend to sustain high song diversity and that of the trend to imitate songs accurately could lead to equilibriums that differ according to species and possibly even the ecological conditions in which a species lives. Perhaps species with songs that are similar across individuals engage in weak balanced imitation and vice versa. For example, to explain why the songs of the Timor zebra finch are much more similar across individuals compared to the Australian zebra finch²⁵, our findings would predict a weaker balanced imitation gain in the Timor zebra finches compared to zebra finches.

Regardless of possible prevalence across species, accounting for balanced imitation in zebra finches might be necessary in order to properly interpret vocal learning outcomes. This is particularly important because mechanisms of vocal learning are studied extensively in Estrildid finches, among which song learning outcomes vary considerably across individuals. In part, this variability is associated with factors like genetics and with tutoring mismatches^{12,33}. Our results indicate that, in addition, deviations from tutor song through reorganization and transformation of copied vocal sounds may be driven by an inclination to maximize diversity. This can be regarded as a discrete form of error correction during song learning. That is, balanced imitation involves correcting errors from states of minimal (and perhaps also maximal) vocal diversity. In the framework of error correction³⁴, the developmental question is when and how the vocal learning bird balances between error correction exclusively in reference to tutor sounds to error correction in reference to a state of minimal or maximal sound diversity. Better understanding of this balance and possible transition could reveal the mechanism through which a species specific level of cultural song diversity is determined²³. Another observation that requires further study is the recombination of syllable units. In the cases we observed, pupils combined tutored syllables into new and more complex units, but splitting appears to be rare. This could suggest a tendency to compress the tutored song. It is possible that improvised syllable types tend to be simple and are transformed across generations into complex types. This hypothesis suggests that perhaps we should consider not only the overall diversity of a tutored song, but also the diversity per unit time. Here too, variation across species can be explained: as opposed to zebra finches, in Bengalese finches, syllable level analysis show correlation in song and transition diversity across tutors and pupils³⁵. We do not know if syllable recombination is common in Bengalese finches, but we do know that they usually produce less complex syllable types compared to zebra finches, which could suggest that Bengalese finches are less inclined to compress their songs.

In conclusion, by recording and analyzing families of vocal learning birds in a large colony, we have gained a deeper understanding into mechanisms that constrain learned vocal repertoires of a species. This mechanism may regulate the level of convergence or divergence in long time scales across generations, while sustaining certain level of acoustic diversity within a population. While our study is far from exhaustive, publishing our song imitation library at the Linguistic Data Consortium catalog (<https://www ldc.upenn.edu>) should allow others to further test the generality of our findings using different approaches at different degrees of granularity.

Methods

Animals

All experiments were approved by the Rockefeller University IACUC. Each bird was raised by his parents in a flight cage until day 90. Partitions kept each family visually isolated from neighboring flight cages. Cages were distributed across three rooms (~250-500 sq. ft each).

Audio recording

Birds were placed singly in sound attenuation chambers³⁶ and their vocal activity was recorded continuously over one week using Sound Analysis Pro²⁸. All songs analyzed in this study are undirected songs.

Song Library

All song recording data generated at the Rockefeller University Field Center Colony, including songs from 160 tutor-pupil pairs analyzed here, were deposited at the Linguistic Data Consortium (LDC) and will be made available at the time of publication at <https://www ldc.upenn.edu/>.

Data Analysis & Statistics

All data were analyzed using MATLAB and R, including the Sound Analysis Tool package for MATLAB, which was used to extract song features. We deposited our MATLAB code for computing song diversity together with the entire song library raw data at the Linguistic Data Consortium (URL and access # pending).

Similarity measurement: For each tutor-pupil pair we calculated similarity measurements with Sound Analysis Pro 2011²⁸ using the default settings. For each song, we outlined a motif and calculated %similarity of tutor vs. pupil song (asymmetric song similarity). We repeated this calculation five times and used the median %similarity as our estimate for each tutor-pupil pair.

Classification of vocal states: For each bird, we analyzed singing bouts of 6-8 seconds each. We used the Sound Analysis for Matlab tool box (<http://soundanalysispro.com/matlab-sat>) to calculate song features using the default settings of FFT window size = 10ms in steps of 1ms. We set a 50dB threshold, below which data were regarded as silences and excluded from further analysis. For each 1ms window, we calculated vocal states as following: We first detected the pitch category (see boundaries in **Supplementary Fig. 1**). We then identified clusters in each slice according to the boundaries of the heat maps outlined in **Fig. 1b**. We chose to use this simple method in order to cover the entire vocal space of the bird. That is, we classify each FFT window as belonging to one of the ten states, without residuals.

Calculation of song diversity: For each bird's song (including bouts of 6-8 seconds), we calculated vocal states as shown above. We then calculated the proportion of vocal sounds within each vocal state and

calculated information entropy²⁷ over the 10 clusters, , where p = the proportion of sounds within each vocal state i .

Fixed effects statistical models: for statistically independent measures such as syllable diversity, vocal state diversity, similarity and influence, we used the Matlab fixedEffects(lme) function. Where lme is a statistical model such as “PupilSongDiversity ~ TutorSongDiversity”.

Linear mixed effect statistical models for vocal state abundance: We used the Matlab linear mixed-effects model fitlme function for statistical analysis. Because abundances of vocal states are not independent, with some vocal states being more frequent than others in the population, we first need to account for this global trend. We refer to R^2 s as “residual” to indicate the removal of this global trend. Prior to statistical analysis we removed the overall trends in vocal state abundances, so as to make each vocal state equally abundant in the population, which guarantees a zero correlation when tutor and pupil identities are shuffled (which we confirmed by shuffling). Because of repeated observations (10 states per tutor-pupil pair), we accounted for random effects of both tutor and pupil identities using mixed effect models. See Supplementary Material for complete description of the models.

Statistical treatment of multiple tests: p-values that we call ‘statistically significant’ are all < 0.01 after Bonferroni adjustment.

Acknowledgements

We thank Michael Yartsev and Lomax Boyd for useful comments and suggestions, and the Rockefeller University Field Research Center staff Gillian Permuy and Lotem Tchernichovski for helping with audio recordings. Supported by NIH grant NIH DC04722 to OT and by HHMI and Rockefeller University funding to EDJ.

References

1. Doupe, A. J. & Kuhl, P. K. Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999).
2. Fehér, O., Wang, H., Saar, S., Mitra, P. P. & Tchernichovski, O. De novo establishment of wild-type song culture in the zebra finch. *Nature* **459**, 564–8 (2009).
3. Lipkind, D. & Tchernichovski, O. Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl**, 15572–9 (2011).
4. Bolhuis, J. J., Okanoya, K. & Scharff, C. Twitter evolution: converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11**, 747–759 (2010).
5. Jarvis, E. D. Evolution of vocal learning and spoken language. *Science* **366**, 50–54 (2019).
6. Tamura, P. M. and M. Song ‘Dialects’ in Three Populations of White-Crowned Sparrows. *Condor* **Vol. 64**, 368–377 (1962).

7. Stoddard, P. K., Beecher, M. D., Campbell, S. E. & Horning, C. L. Song-type matching in the song sparrow. *Can. J. Zool.* **70**, 1440–1444 (1992).
8. Marler, P. & Pickert, R. Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Anim. Behav.* **32**, 673–689 (1984).
9. Okanoya, K. & Yamaguchi, A. Adult bengalese finches (*Lonchura striata* var. domestica) require real-time auditory feedback to produce normal song syntax. *J. Neurobiol.* **33**, 343–356 (1997).
10. García, N. C., Arrieta, R. S., Kopuchian, C. & Tubaro, P. L. Stability and change through time in the dialects of a Neotropical songbird, the Rufous-collared Sparrow. *Emu* **115**, 309 (2015).
11. MacDougall-Shackleton, E. A. & MacDougall-Shackleton, S. A. CULTURAL AND GENETIC EVOLUTION IN MOUNTAIN WHITE-CROWNED SPARROWS: SONG DIALECTS ARE ASSOCIATED WITH POPULATION STRUCTURE. *Evolution (N. Y.)*. **55**, 2568–2575 (2001).
12. Mets, D. G. & Brainard, M. S. Genetic variation interacts with experience to determine interindividual differences in learned song. *Proc. Natl. Acad. Sci.* **115**, 421–426 (2018).
13. Robinson, C. M., Snyder, K. T. & Creanza, N. Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning. *Elife* **8**, (2019).
14. Soha, J. & Marler, P. A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Anim. Behav.* **60**, 297–306 (2000).
15. Fehér, O., Ljubičić, I., Suzuki, K., Okanoya, K. & Tchernichovski, O. Statistical learning in songbirds: from self-tutoring to song culture. *Phil. Trans. R. Soc. B*.
16. Gardner, T. J., Naef, F. & Nottebohm, F. Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science* **308**, 1046–9 (2005).
17. Lachlan, R., Verhagen, L. & Peters, S. Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taeniopygia guttata*),. *J. Comp.* (2010).
18. James, L. & Sakata, J. Learning biases underlie 'universals' in vocal sequencing. *Curr. Biol.* (2017).
19. Fehér, O., Ljubičić, I., Suzuki, K., Okanoya, K. & Tchernichovski, O. Statistical learning in songbirds: From selftutoring to song culture. *Philos. Trans. R. Soc. B Biol. Sci.* **372**, (2017).
20. Derégnaucourt, S. & Gahr, M. Horizontal transmission of the father's song in the zebra finch (*Taeniopygia guttata*). *Biol. Lett.* **9**, 20130247 (2013).
21. Weiss, M., Hultsch, H., Adam, I., Scharff, C. & Kipper, S. The use of network analysis to study complex animal communication systems: a study on nightingale song. *Proc. Biol. Sci.* **281**, 20140460 (2014).
22. Lachlan, R. F., van Heijningen, C. A. A., ter Haar, S. M. & ten Cate, C. Zebra Finch Song Phonology and Syntactical Structure across Populations and Continents—A Computational Comparison. *Front. Psychol.* **7**, 980 (2016).
23. Tchernichovski, O., Feher, O., Fimiarez, D. & Conley, D. How social learning adds up to a culture: from birdsong to human public opinion. *J. Exp. Biol.* **220**, 124–132 (2017).

24. Akçay, Ç., Tom, M. E., Campbell, S. E. & Beecher, M. D. Song type matching is an honest early threat signal in a hierarchical animal communication system. *Proc. Biol. Sci.* **280**, 20122517 (2013).
25. Lansverk, A. L. *et al.* The variability of song variability in zebra finch (*Taeniopygia guttata*) populations. *R. Soc. Open Sci.* **6**, 190273 (2019).
26. Tchernichovski, O. & Marcus, G. Vocal learning beyond imitation: Mechanisms of adaptive vocal development in songbirds and human infants. *Curr. Opin. Neurobiol.* **28**, 42–47 (2014).
27. Shannon, C. E. Prediction and Entropy of Printed English. *Bell Syst. Tech. J.* **30**, 50–64 (1951).
28. Tchernichovski, O., Nottebohm, F., Ho, C., Pesaran, B. & Mitra, P. A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167–1176 (2000).
29. Tchernichovski, O., Lints, T., Mitra, P. P. & Nottebohm, F. Vocal imitation in zebra finches is inversely related to model abundance. *Proc. Natl. Acad. Sci.* **96**, 12901–12904 (1999).
30. Tchernichovski, O. & Nottebohm, F. Social inhibition of song imitation among sibling male zebra finches. *Proc. Natl. Acad. Sci.* **95**, 8951–8956 (1998).
31. Janney, E. *et al.* Temporal regularity increases with repertoire complexity in the Australian pied butcherbird's song. *Royal Society Open Science* **3**, (2016).
32. Fee, M. S. & Scharff, C. The songbird as a model for the generation and learning of complex sequential behaviors. *ILAR J.* **51**, 362–77 (2010).
33. Sober, S. J. & Brainard, M. S. Vocal learning is constrained by the statistics of sensorimotor experience. *Proc. Natl. Acad. Sci.* 1–5 (2012). doi:10.1073/pnas.1213622109
34. Gadagkar, V. *et al.* Dopamine neurons encode performance error in singing birds. *Science* **354**, 1278–1282 (2016).
35. James, L. S., Sun, H., Wada, K. & Sakata, J. T. Statistical learning for vocal sequence acquisition in a songbird. *Sci. Rep.* **10**, 1–18 (2020).
36. Tchernichovski, O., Mitra, P. P., Lints, T. & Nottebohm, F. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* **291**, 2564–9 (2001).

Figures

Figure 1

Imitation outcome varies across families. a, 25 song tutoring lineages. All tutors had pupils in more than a single clutch. Tutor nodes are presented on the bottom and pupil nodes on the top. Similarity scores are presented as quartiles (green for best imitations & red for poorest). Lineages are sorted according to mean similarity between tutor and pupils from highest (top) to lowest (bottom). b, Example of song imitations from tutor (Aq12) with a low similarity family. c, Example of song imitations from tutor (DG1) with a high similarity family. Red bars outline the repeated song motifs of the tutors.

Figure 2

Syllable type diversity. a, Example sonograms of a tutor-pupil song pair. Syllables types are color-coded by lines above them. Smoothed amplitude curve indicates segmentation. Color line above each syllable indicate clusters computed separately for tutor and pupil in (b), which do not always correspond to the same color for the same syllable in tutor and pupil. Highlighted box syllable indicates novel syllable in pupil not found in tutor. b, 2D scatter plots of syllable acoustic features. The color of each marker indicates its computed syllable type (type = cluster in feature space). Colors of clusters correspond to syllable type colors shown in a. c, Regression analysis between tutor and pupil syllable type diversity, showing no correlation for high or low diversity tutors ($R^2=0.079$, Linear mixed-effects model, $t=1.96$, $p =$

0.2; n= 80 tutor-pupil pairs). d, Regression analysis between tutor and pupil syllable sequence transition diversity, also showing no correlation. e-i, Examples of five tutor-pupil pairs with syllable recombination and merging in pupil songs.

Figure 3

Vocal states and diversity in zebra finch songs. a, continuously measured song feature vectors plotted on top of a sonogram, including pitch (red), FM (aqua) and Wiener entropy (yellow). b, 3D scatter plot of song features pooled over the songs of 147 birds. Dotted rectangles show the partitioning of this space into ten vocal states. c, Pitch histogram of songs of the same birds, calculated in 10ms windows, the acoustic unit of time proposed to be controlled by the song production brain pathway³². Shadings show partitioning into four regions according to contours of the pitch distribution. d, Two-dimensional heatmaps of frequency modulation and Wiener entropy for each of the four pitch regions. Red circles outline 10 clusters around which vocal states are defined. e. Song diversity in tutor songs versus high and low imitation pupil songs. Colors show R^2 separately for high and low similarity birds (combined data $R^2 = 0.21$, $t=5.6$, $p= 9.5e-08$). f, Histogram of song diversity for all male birds recorded. g, Song diversity in

families, comparing the songs of tutors versus the mean diversity across all of their pupils' songs. The upper dotted line represents the upper theoretical bound where all vocal states are equally abundant. The lower dotted line indicates the median. Blue lines show decreases, and red show increases in pupil vocal state diversity. Note, below the median is dominated by low diversity tutors whose pupils increase, and vice versa.

Figure 4

Song diversity across generations. a, Song similarity across two generations of pupils (colors represent quartiles, as in Fig. 1a) in 14 family lineages. b, an example from lineage HP10 show cultural transmission from poor imitation in a first-generation pupil to accurate imitation in a grand pupil. c, Same as (b) for lineage DG4. d, A counter example in lineage LB12, where the grand pupil imitated poorly. e, Tutor song diversities vs. the influence of tutor song on his pupils ($R^2=0.25$, Linear mixed-effects model $t=4.8$, $p=4.2e-6$). Vertically aligned values are often birds from the same lineage. f, Tutor song diversities vs. similarity with pupil songs ($R^2=0.08$, $t=1.9$, NS). g, Pupil song diversity vs. song similarity to tutors.

Figure 5

imitation of vocal state abundances. a-d, raw distribution of tutor vs. corresponding pupil vocal state abundances according to quartiles of song similarities. Note that each bird is represented by 10 markers, which are not statistically independent. The residual correlations were computed after removing trends with bird identities included as random factors. Dashed lines are identity, slope = 1. Blue lines are regression of the data. e, Same data as in (a-d) combined, with y axis showing differences between tutor and pupil vocal states abundances. Positive diffs indicate higher abundance in pupil and vice versa. f, Median imitation gains for all state abundances, according to imitation quartiles. Gain of 1 indicates no bias, gain of 2 indicates doubling of abundance, and gain of 0.5 halving. Y-axis is log-scale. g-i, examples of vocal state diversity balancing. We simplified the 10 vocal states into 4 groups: yellow for high pitch states 9-10; mustard for medium pitch, high entropy states 6 & 8; light blue for non-modulated states 4 & 7; and dark blue for the rest 1, 2, 3, and 5. In (i), we present two generations of pupils. j, Imitation of vocal state abundances in biological tutors vs. pupils' songs. k, Same for fostered pupils. l, Lack of such imitation for fostered pupils vs. their biological father, which did not raise them. Dashed lines are identity, slope = 1. Blue lines are regression of the data.

Figure 6

Song diversity versus imitation a-c, Mean tutor's song features versus pupil's song features for pitch (a), Wiener entropy (b), and frequency modulation (c), for the top imitations (green markers, top quartile) and for poorest imitations (red markers, bottom quartile). Plotted at the bottom are histograms of tutor features for top and bottom quartiles. d-f, Box plot distribution of mean song features in four colonies for pitch (d), Wiener entropy (e) and frequency modulation (f). Each marker represents the mean value for one bird. Light blue shaded areas correspond to feature ranges in colony RU2019 that are outside the

90% range of top imitations. Descendants of the RU2002 colony make up about 20% of the RU2019 colony. The remainder of the 2019 colony originated from Duke University.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.png](#)
- [FigS2.png](#)
- [FigS3.png](#)