

Discussion

We have presented an experiment that allowed us to study cultural evolution in real-time in a controlled laboratory setting. We investigated what happens to isolate zebra finch song when it is imitated by generations of song learners. We studied this problem in different social settings. We used a recursive training paradigm in an impoverished social setting where young males were housed with adult ISO tutors one-to-one, and upon maturation, they became the tutors of the next generation of learners, and so on. However, since multi-generational changes in a complex social setting may be more representative of natural cultural evolutionary processes, we established a semi-natural colony with an adult ISO male and some females and let them reproduce for a few generations. In the colony young birds had unrestricted interactions with adult males, females, older siblings and clutch mates.

To quantify changes in the song, we developed a method free of qualitative judgment that takes into consideration three levels of song structure: ms-to-ms spectral features, the duration of acoustic state (which describes note-level changes) and song rhythm (which is a characterization of the song at the level of the song bout). We have found that in all song features, regardless of social setting, isolate zebra finch song progresses towards wild-type song over multiple generations of learners, and within 3-4 generations the ISO song features turn into WT features. The extent and time-line of the changes were similar in the two social settings, although spectral features became more structured in one-to-one training and rhythm improved much more in the colony. Since in the colony there was more raw material to imitate (including female calls and siblings' "errors"), there was more song diversity and more variations in song motifs. In the one-to-one setting, birds did not improvise, but instead, often differentiated simple syllables into complex ones.

In the Introduction, we proposed three scenarios that could be responsible for such changes: sexual selection of WT song, selective imitation and biased imitation. Based on our results, we can conclude that the birds used biased imitation to drive the evolution of

WT song from ISO. They imitated ISO-like and WT-like song syllables indiscriminately, but changed the ISO-like features so that in succeeding learning generations those syllables and song features became increasingly WT-like. Since biased imitation was the cause of the changes, females and a complex social environment were not necessary. According to Boyd and Richerson (2005), biased imitation produces nonrandom modifications that can accelerate the rate of evolution. In our experiment, we documented an extremely rapid change in song culture. In the impoverished condition, most of the changes occurred in the first learning generation. In the colony the process was somewhat slower and more gradual. However, in both settings there was a slowing of the progression towards WT in later generations, and the changes seemed to asymptote as the song features approached WT distributions. As noted in the Introduction, this could be because the imitation biases have a smaller effect when the deviation from the WT song culture is smaller, and this deviation was indeed smaller in later generations.

An experiment analogous to ours has recently been conducted in humans (Kirby et al. 2008). Kirby and colleagues trained participants on an “alien” language that included colored objects in motion and random strings of consonant-vowel pairs that were supposed to label these objects. Participants were trained on a set of these labeled moving shapes, which involved looking at the objects and the corresponding labels. After training, participants were tested on a different set that included the set that they were trained on but also objects they had never seen. During testing, participants were provided with a picture and they had to type in the “word” they thought the aliens would produce for that object. Then another group of participants were trained on the output of the last testing session of the previous group. They continued the experiment for 10 generations of learners. They found that transmission error decreased significantly, the largest decrease being in the first 5 generations. Moreover, the increased learnability of the “languages” corresponded to a significant rise in structure. The first presentation of the language was completely random, therefore, it contained no structure at all. However, learners gradually introduced structure by using the same string for the same motion, for example. On occasion, exceptions to rules were reduced by further learning generations.

The outcome of this experiment is analogous to ours in that learners produced increasing structure through an iterated learning process. In both cases, there are innate predispositions to learn something that resembles the WT culture more. Of course, it is fairly clear that in humans random labels that involve rote memorization are more difficult to learn than structured strings with semantic rules. In birds, in the absence of semantics, we cannot be sure that ISO song features are more difficult to learn and that this is what drives the transition to WT, but this could be a plausible hypothesis.

In a sense, our results represent a multigenerational song phenotype that is encoded in every zebra finch but takes multiple generations to appear. It is an extended developmental process. We have observed some similarities between the individual song development in juveniles and the multigenerational development of WT song. For example, in early development, some young males produce back-to-back repetitions of syllable types, but they abandon the repetitions at some point in their development. We saw that syllable repetitions were sometimes present in ISO song, but pupils did not imitate them. It would be interesting to find out if there are other parallels, such as neural and other behavioral changes, that take place during development as well as during the development of the multigenerational WT phenotype.

In the imitation biases, we have identified innate components of culture. Without exposure to song tutoring the biases do not surface to drive the developing song towards WT-like features, but as soon as tutoring is present, the innate biases become activated. It is almost as if they are suppressed in the ISO birds. What is most interesting about our results is that tutoring acts as a “switch” of some sort that turns on the capacity to sing WT-like zebra finch song. We are very interested in the minimum requirement for this capacity. We already saw that simple social interactions are sufficient, but perhaps social interactions are not necessary at all. We are currently training young males with playbacks of ISO songs to investigate this possibility. Moreover, there is a possibility that WT song can be induced within a single bird in the absence of tutoring. When birds are raised in complete social and acoustic isolation, they rely on direct auditory feedback to shape their songs. As a result, they improvise ISO song. However, we know that we do

not hear ourselves the same way as others hear us, so if we recorded the developing songs of our isolate birds and played it back to them, we would provide a different level of auditory feedback that may allow the birds to modify their developing ISO songs to be more like WT songs. This experiment is also underway.

We hope that this work, which is currently in press (Fehér et al. 2009), has contributed to the understanding of the development of song cultures, and that it was successful in identifying a good model system to tackle further questions.

Appendix I. Excluding exposure to adult song in ISO birds

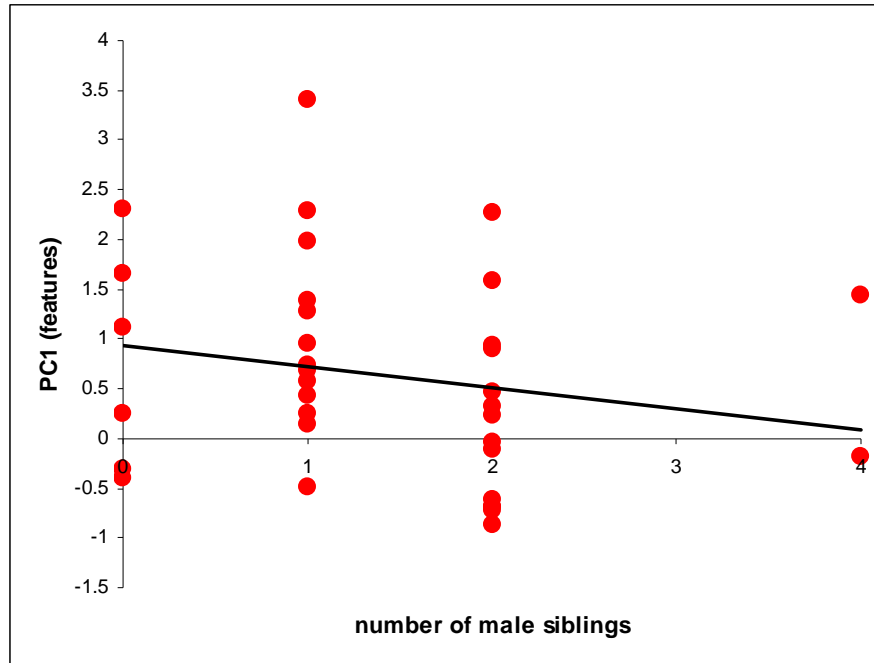
Excluding possible exposure to WT songs prior to day 7

In zebra finches there is no behavioral evidence to early influence of song exposure prior to day 25. Zebra finches are altricial birds: the chicks are helpless and tiny upon hatching, and it takes about 20 days for the auditory system to become fully functional.

Measurements obtained on day 10 after hatching show elevated auditory thresholds (about 20-30dB higher than in adults). Therefore, zebra finches do not hear much prior to day 7 post hatch. Zebra finches live in dense social groups and their song is soft and low amplitude, which further decrease the chance of affecting the young chicks. It is also interesting to note that during the first days after hatching the parents tend to be very quiet, and if the males sing at all, they tend to sing away from the nest, facing the other way (Tchernichovski et al. 1998). Of course, the considerations above should not apply to the island colony experiment since chicks hatched there were only exposed to the ISO song.

Excluding possible exposure to ISO subsong prior to day 30

We kept the siblings together until day 30 being well aware that the onset of early subsong is a bit ill-defined. Low-amplitude subsong-like vocalization can be sometimes recorded shortly after fledging using sensitive microphones (we are not sure how common it is, but those sounds are very faint and unstructured). As expected, our analysis show no evidence that the evolution of song culture is affected by the presence of male siblings before day 30. Fig. App. 1 presents the number of male siblings in a clutch, versus the first principal component of the WT/ISO features for all birds used in our study:



Appendix I Figure 1 | Number of male siblings vs. first Principal Component of WT/ISO features of all experimental birds.

As shown, there is no clear trend ($r^2=0.04$) and looking separately in each experimental groups also show no significant or consistent trend. Similar results were obtained in duration of acoustic state and in rhythm.

Level of isolation in the sound chambers

Our custom made sound chambers provide sufficient level of isolation for zebra finch songs, which rarely exceed 85dB. At 4000Hz (mean frequency of zebra finch songs), we played a 100 dB sound inside the box and measured the sound level outside the box. We measure a sound level of 63 dB, while the baseline noise was 61 dB. Therefore, our boxes contribute a 37 dB sound attenuation. We recorded the entire song development of each bird, and although we listened to recorded sound samples frequently, we heard no traces of songs from other boxes.

Appendix II. Sound chamber equipment for isolated semi-natural colony

Life-sustaining equipment

Jehm Co. linear piston air compressor pumped air through a ¼ inch diameter tube into the left-side chamber. A hole an inch in diameter was drilled in the back of the right-side chamber. This provided sufficient airflow.

The middle chamber was equipped with a 12 inch long 8 Watt light bulb and the 2 outside chambers with 12 inch LED lights. We did this, because the birds voluntarily appointed the middle chamber as a communal feeding area and the outside chambers as nesting areas. The LED lights provided a dimmer light, which we observed had a good effect on nesting.

Water was provided ad libitum in 13 water cups mounted onto the front grid, and food (ABBA seed dry seed and soaked seed mixed with chick food) was placed in small trays in the middle chamber fresh every day.

Monitoring equipment

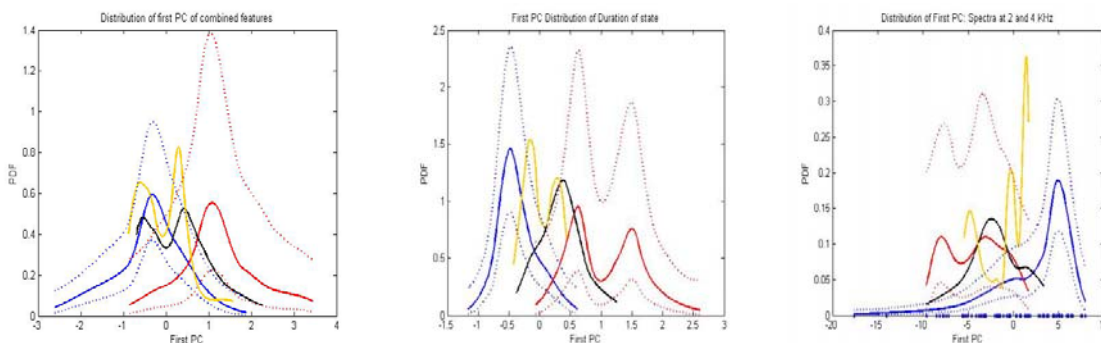
For video monitoring, we used 2 systems in parallel. Labtec webcam USB video cameras were connected to XtraSense surveillance software, and Watec (WAT-230A) and Genwac (GW-230A) color cameras connected to GeoVision surveillance software. The USB cameras gave high-definition colored pictures and simultaneous sound recordings, so for bird identification, this system was much better. The other system was used for reliable continuous observation of nests.

Three microphones were placed in the box, one into each chamber. These recorded sound continuously as long as there were juvenile males (<120 day old) in the colony. Sound recording and analysis was done by Sound Analysis Pro. Developmental song data was studied but not analyzed.

Appendix III. Statistical tests

Distributions of first principal components

In order to assess the effects of recursive tutoring, we considered song feature values of the following four bird groups: WT (52 birds), ISO (17 birds), 1st-generation (13 birds, henceforth denoted as F1), and higher generations (8 birds, denoted as F2+). The distributions of first PC of the four groups are shown in Appendix IV Fig. 1.



Appendix III Figure 1 | Distributions of first PC of combined song features, acoustic state duration, and rhythm spectra. The distributions of ISO, F1, F2+, and WT are shown in red, black, orange, and blue, respectively. The dotted lines are the 95% confidence intervals of WT and ISO. The distinctness between WT and ISO, as well as the shift toward WT due to tutoring, is demonstrated in all three timescales.

The distinctiveness of WT and ISO are very significant. A gradual, yet steady shift of distributions towards WT demonstrates the multi-generational phenotype. The significance of this trend can be subjected to statistical tests.

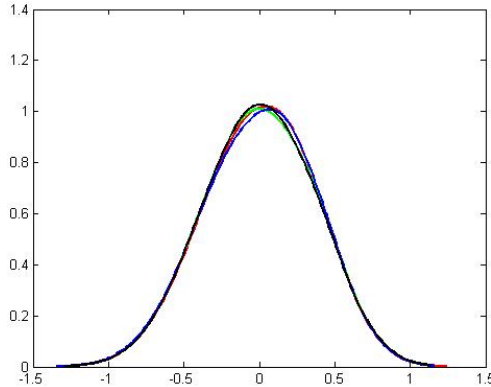
Statistical test methods

The significance of the shift towards WT can be characterized by statistical tests of the difference between the mean values. A two-sided independent *t*-test may be used for this purpose. Since some of the distributions cannot be approximated by Gaussian, we also perform a non-parametric alternative (permutation test) for additional validation. The results from the two methods approximately agree with each other.

Consider two groups of a song trait value (e.g., the first PC of state duration), which have means μ_1 and μ_2 , respectively. If the two means are equal under null hypothesis, the division of the whole population into the two groups is arbitrary, as far as mean is concerned. It follows that we can interchange the members of the two groups, without (significantly) changing the following relationship

$$\mu_1 - \mu_2 = 0$$

Interchanging of group members can be realized by permutation. For each (random) permutation, we calculate the statistic $\mu_1 - \mu_2$. We resample through random permutation thousands of times, and obtain a *permutation distribution* of $\mu_1 - \mu_2$. If the null hypothesis holds, the difference between the original group means should not significantly deviate from 0. App. III Fig. 2 shows examples for permutation distributions.



Appendix III Figure 2 | Four examples of permutation distribution for group mean differences of F1 and F2+. They are generated using 10000 permutations and show reasonable stability.

In summary, we test the following null hypothesis:

$$H_0 : \mu_1 - \mu_2 = 0$$

The CDF of the statistic $\mu_1 - \mu_2$ can be obtained from the permutation distribution, and the p-value of the two-sided test is given by

$$p = 1 - [F(|\mu_1 - \mu_2|) - F(-|\mu_1 - \mu_2|)].$$

Six hypotheses are subject to test among the four group means at each time scale. There are 18 hypotheses altogether to be tested on the four birdsong data sets, corresponding to the four bird groups. We apply the *false discovery rate* (FDR) control to correct for multiple comparisons.

Statistical results of group means

i. Spectral frame features

The mean of first PC of the four bird groups are

WT	Higher generations (F2+)	1 st -generation (F1)	ISO
-0.4235	0.0266	0.2992	1.2956

The group means clearly demonstrate progression from ISO to WT. The significances of the progressions are statistically tested below.

The p-values using *t*-test are shown below:

	WT	Higher generations	1 st -generation	ISO
WT		0.1736	0.0180	< 0.0001
Higher generations			0.4742	0.0045
1 st -generation				0.0126
ISO				

The p-values obtained from permutation test are:

	WT	Higher generations	1 st -generation	ISO
WT		0.2126	0.0155	< 0.0001
Higher generations			0.4984	0.0104
1 st -generation				0.0158
ISO				

The p-value obtained by permutation test may have small fluctuation with different set of 10000 permutations. Hypotheses marked in red can be rejected at the significance level of $\alpha = 0.05$.

Although direct comparison between first generation pupils (F1) and higher generation pupils (F2+) cannot reject the null hypothesis, their relationships to WT and ISO show multi-generational effects:

- F1 is different from WT, while F2+ is not significantly different from WT.
- F1 and F2+ are rather similar, while F1 is very different from ISO. This suggests asymptotic behavior of the recursive tutoring.
- We group higher generations together because there are not enough data of 2nd-generation alone.

ii. Duration of state

The mean of first PC of duration of state are

WT	Higher generations	1 st -generation	ISO
-0.3472	0.0195	0.3549	1.0619

Similar observations as spectral frame feature apply.

The p-values obtained from t-test are:

	WT	Higher generations	1 st -generation	ISO
WT		0.0089	< 0.0001	< 0.0001
Higher generations			0.0467	< 0.0001
1 st -generation				0.0013
ISO				

The p-values obtained from permutation test are

	WT	Higher generations	1 st -generation	ISO
WT		0.0069	< 0.0001	< 0.0001
Higher generations			0.0633	< 0.0001
1 st -generation				0.0017
ISO				

At 0.05 significance level, the null hypothesis can be rejected for all pairs (t-test). It follows that the four groups are indeed different from each other.

iii. Rhythm

The mean of first PC of different bird groups are

WT	Higher generations	1 st -generation (F1)	ISO
-1.98	1.36	2.16	4.09

The p-values obtained from t-test are

	WT	Higher generations	1 st -generation	ISO
WT		0.017	0.0033	< 0.0001
Higher generations			0.5872	0.0566
1 st -generation				0.1693
ISO				

The p-values obtained from permutation test are

	WT	Higher generations	1 st -generation	ISO
WT		0.0775	0.0099	< 0.0001
Higher generations			0.6093	0.075
1 st -generation				0.1657
ISO				

At 0.05 significance level, the null hypotheses that can be rejected are marked in red.

iv Multiple comparisons

The 18 hypotheses of comparisons among four groups are tested simultaneously. We use the Benjamini-Hochberg procedure to control the false discovery rate. At FDR level of 0.05, all the hypotheses marked red can be rejected.

Distance metric from WT (tutor-pupil closeness to WT)

In order to study the effects of tutoring, we want to see if the pupil's songs become more similar to the WT songs than the tutor's. Since each bird is represented by a point in an n -

dimensional space (e.g., the CDF space of state duration), we can measure the difference between two birds (A and B) by the Euclidean distance between the two vectors:

$$d_{AB}^2 = \sum_{i=1}^n [F_A(x_i) - F_B(x_i)]^2 .$$

The pupil's song is said to be more similar to the WT songs than the tutor's, if, overall, the distances between the pupil's song and WT songs become smaller than those between the tutor's and WT.

Let us define the tutoring-induced distance change as

$$\Delta_i(P, T) = d_{P,i} - d_{T,i},$$

where $d_{P,i}$ is the distance between pupil's song and the i th WT song, and bird T is the tutor. The pupil's song is said to become more similar to a given WT song when

$$\Delta_i(P, T) < 0.$$

The probability for the pupil's song to become closer to a given WT song is then given by

$$p_i = P\{\Delta_i < 0\}$$

Since the tutor and its pupil are genetically randomized and never exposed to any WT song, it is reasonable to assume that all the probabilities, p_i , are equal,

$$p_1 = \dots = p_i = \dots = p_k = p.$$

We define a random variable X to be the number of WT songs for which $\Delta_i < 0$,

$$X = N\{\Delta_i < 0\},$$

which thus follows binomial distribution, $B(n, p)$: each of the Δ_i independently has the same probability p to be less than 0. Here, n is the total number of WT birds.

Hence, we test the following hypothesis:

$$H_0 : p \leq 0.5,$$

$$H_1 : p > 0.5.$$

If the null hypothesis holds, the pupil's songs are no more similar to WT than its tutor's.

The p-values are given in the following table.

Tutoring pair	Feature	State Duration	Rhythm
19 → 1248	< 0.0001	< 0.0001	< 0.0001
19 → 1302	< 0.0001	< 0.0001	< 0.0001
19 → 1340	< 0.0001	< 0.0001	< 0.0001
19 → 1661	< 0.0001	< 0.0001	< 0.0001
1247 → 1315	0.2442	< 0.0001	> 0.9999
1238 → 1342	0.06317	> 0.9999	0.0182
1238 → 1433	< 0.0001	0.0002	< 0.0001
1211 → 1402	< 0.0001	< 0.0001	< 0.0001
1211 → 1566	< 0.0001	0.0039	< 0.0001
1211 → 1655	< 0.0001	< 0.0001	< 0.0001
1249 → 1439	< 0.0001	< 0.0001	< 0.0001
1249 → 1530	< 0.0001	< 0.0001	< 0.0001
1529 → 1622	0.6611	< 0.0001	> 0.9999

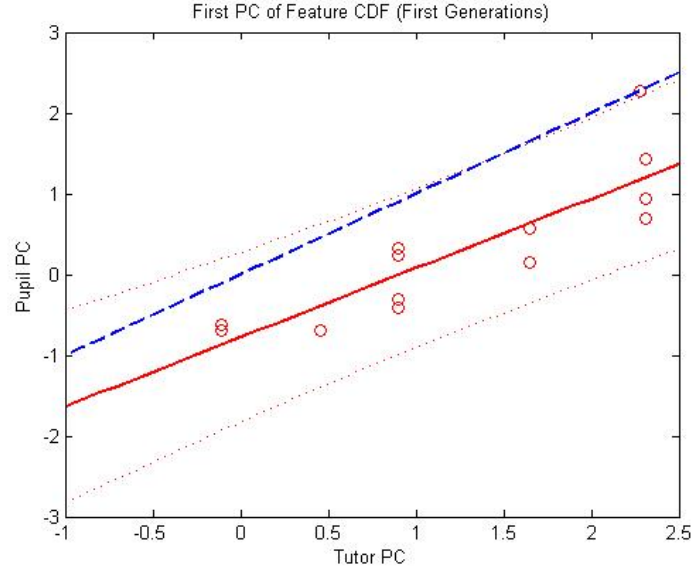
There are 39 hypotheses. Using the Benjamini-Hochberg procedure, those null hypotheses that *cannot* be rejected at a FDR level of 0.01 are marked blue in the table above. We conclude that the pupil's songs become more similar to WT as a result of tutoring.

Tutor-Pupil Correlation: Using First PC of Feature CDF

We now turn to the question how the song trait values of tutor and pupil are related, using the first PC of feature CDF as an example. The relationship at the first generation of tutoring may be approximated by linear regression:

$$y = -0.78 + 0.86x$$

where y is the pupils' PC1, and x represents the tutors' PC. This relationship is shown as the solid red line in Appendix III Fig. 3.



Appendix III Figure 3 | The relationship between tutor’s and pupil’s trait values. The red solid line is obtained by linear regression, while the dotted lines are the 95% confidence interval. The blue dashed line, $y = x$, represents faithful copying of tutor’s song (phenotypic value) by the pupil.

We test if the copying is faithful or partial. Hence the hypotheses are

$$H_0 : y = x$$

$$H_1 : y = \beta_0 + \beta_1 x, \quad \beta_1 < 1$$

Using the likelihood ratio test and define

$$\lambda = \frac{\max[L(H_0)]}{\max[L(H_1)]}$$

where $L(H_0)$ is the likelihood function under null hypothesis. Under regular conditions, $-2\log(\lambda)$ follows chi-square distribution with 2 degree of freedom. With first PC of feature CDF, we have

$$-2\log(\lambda) = 23.14$$

$$p = 9.4 \times 10^{-6}$$

where p is the p-value. Hence, we shall reject the null hypothesis and conclude that the copying by the pupil is partial.

Increased stability in pupil's song compared to ISO tutor

Pupils’ songs were more stable than those of their tutors. Here we examine stability in the duration of syllable renditions (within-type variance). The coefficient of variance (CV) of syllable durations was significantly higher in the isolates’ syllables compared to the imitation of those syllables (ISO: CV=15%, range 2-57%; pupils: CV=4%, range 2-9%; $p < 0.05$, Wilcoxon sign test).