Supplementary Methods & Results

Table of contents

1. Audio files demonstrating evolution of song culture

Attached audio file: song_culture.wav
Also available at
http://forum.sci.ccny.cuny.edu/Members/ofer/song-culture/cultural-evolution.wav

2. Tutoring lineages: imitation statistics and sonograms

	average	**CV**
Tutor: number of complex syllable types	1.85	0.58
Pupil: number of complex syllable types	1.92	0.54
% of complext syllables copied	92.31	0.17
% of complex syllables invented	5.08	2.44
Accurary (mean across syllables)	72.73	0.20
Tutor: number of short/long call types within motif	1.15	1.32
Pupil: number of short/long call types within motif	1.85	0.53
Tutor # of rare syllables	1.23	0.82
# of calls+rare copied	1.70	0.62
# of call types invented by pupil	0.38	1.32

Tutor	**19**	**19**	**19**	**19**	**1211**	**1211**	**1211**	**1238**	**1238**	**1247**	**1249**	**1249**	**1529**
Pupil	**1248**	**1340**	**1302**	**1661**	**1402**	**1566**	**1655**	**1342**	**1433**	**1315**	**1439**	**1530**	**1622**
Tut: complex	1	1	1	1	2	2	2	4	4	1	2	2	1
Pup: complex	1	1	1	1	2	3	3	4	3	1	2	2	1
% copied	100	100	100	50	100	100	100	100	75	75	100	100	100
% invented	0	0	0	0	0	33	33	0	0	0	0	0	0
Accuracy	81	63	72	39	83	76	85.5	85.5	81	49	80	75	75.5
Tut: call-syll	0	0	0	0	0	0	0	3	3	0	3	3	3
Pup: call-syll	1	1	1	3	2	3	0	2	3	1	3	2	2
Tut: rare syll	2	2	2	2	2	2	2	0	0	2	0	0	0
#calls copied	3	?	?	?	2	2	2	0	0	1	3	2	2
#calls invented	0	0	1	1	0	1	0	0	0	0	0	1	1

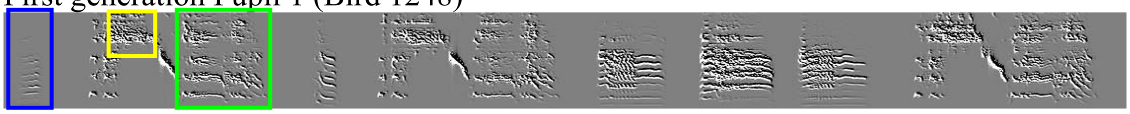
Supplementary Table 1 | Imitation of isolate tutors. Songs segmentation and imitation accuracy were done with Sound Analysis Pro 2, using symmetric comparisons and then averaged across syllables. The determination of syllable types and of their imitation was done subjectively, and confirmed syllable by syllable by accuracy measurements. In the isolate tutors, "within motif" refers to syllables that appears also in the middle of singing bouts (with no attempt to define motifs).

3. Sonograms of tutoring lineages

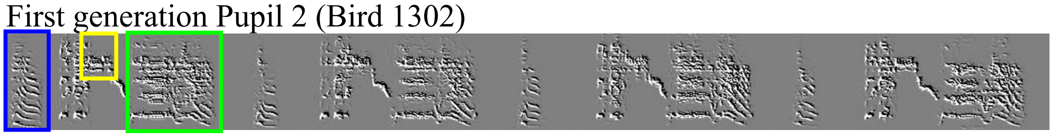
a. One-to-one tutoring – first generation

ISO Tutor 1 (Bird 19)

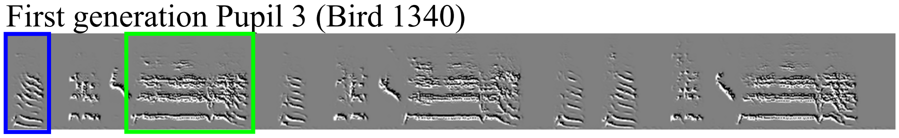
First generation Pupil 1 (Bird 1248)



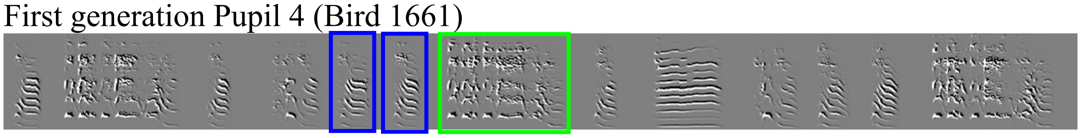
First generation Pupil 2 (Bird 1302)



First generation Pupil 3 (Bird 1340)

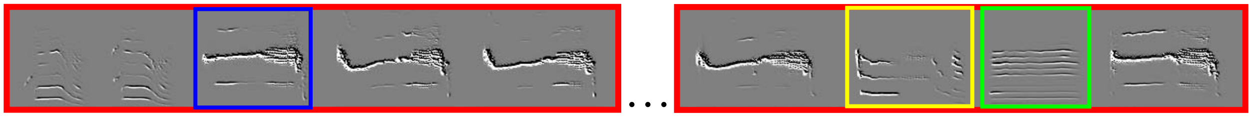


First generation Pupil 4 (Bird 1661)

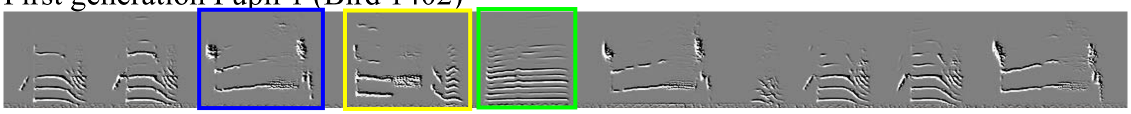


Tutor 1 was our oldest tutor bird and his song is highly stable, but his song has only one dominant syllable, which is atypical to wild type zebra finch song: it starts with a noisy note (in yellow) followed by a short, high-pitched note and a messy, scratchy harmonic. Occasionally, he produced scratchy calls (not shown) and those were copied by pupil 1. As seen above, all of the pupils of Tutor 1 imitated his song. Pupil 4 shows the least similarity. Here, the only recognizable elements are the introductory syllable (highlighted in blue) and the last, messy harmonic (green square), but instead of singing the latter continuously, as his tutor, the pupil tends to stutter this syllable, breaking it into short sections. The duration of the yellow note, as well as the green one, are decreased in the songs of Pupil 1 & 2. Pupil 3 omitted the yellow note altogether.

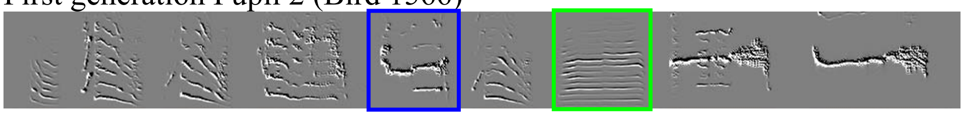
ISO Tutor 2 (Bird 1211)



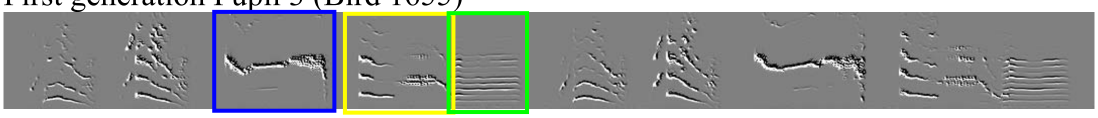
First generation Pupil 1 (Bird 1402)



First generation Pupil 2 (Bird 1566)

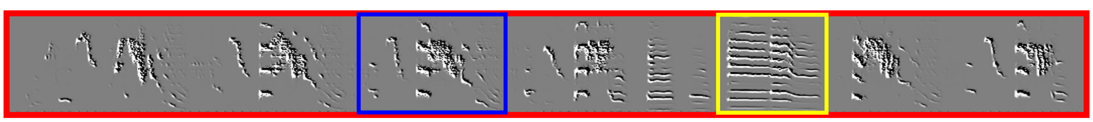


First generation Pupil 3 (Bird 1655)

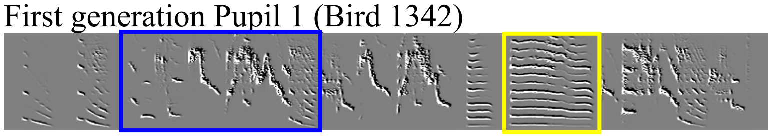


Tutor 2's song contains one long high-pitched note (blue) that is very untypical to wild type zebra finch song. The structure of that syllable was unstable with several rare variants, some of those were copied by his pupils (not shown). The other two syllables of his song were of similar duration and harmonic-like (yellow and green). Pupils of Tutor 2 imitated most of his song syllables. What is striking here is that none of the pupils produced back-to-back repetitions of the blue syllable – even though such repetitions dominated the tutor's song. On the other hand, the syllables that were rare in the tutor's song (yellow and green squares) were much more frequent in all of the pupils' imitations. Pupil 2 shortened the blue syllable and Pupil 3 fused the yellow and green syllables.

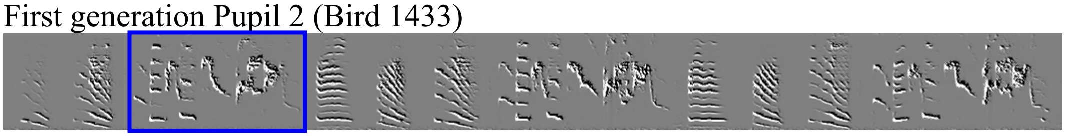
ISO Tutor 3 (Bird 1238)



First generation Pupil 1 (Bird 1342)



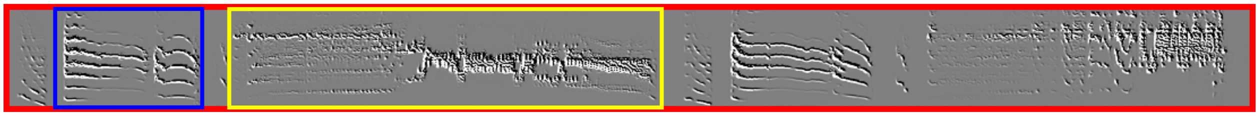
First generation Pupil 2 (Bird 1433)



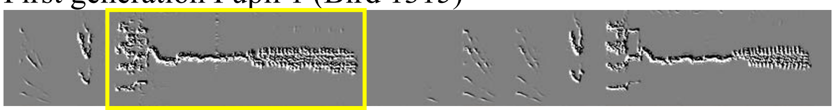
Tutor 3 produced a song dominated by a single syllable, which appears similar to WT syllables except for its high pitch. For instance, it has fast transitions between different notes that are quite complex acoustically. However, it does contain some features that are typical only to isolate song, such as the repetition of the same syllable over and over (blue syllable). Also, there are three harmonic elements that follow the repeated syllable, and the last one of these is long and has a very short stop in the middle (in yellow). Both

of these isolate-like features were changed by the pupils. We can observe a reduction, as in Pupil 1, or an omission, as in Pupil 2, of the long harmonic syllable, and a decrease in the number of syllable repetitions to 2 and 1 in Pupil 1 and 2, respectively.

ISO Tutor 4 (Bird 1247)



First generation Pupil 1 (Bird 1315)

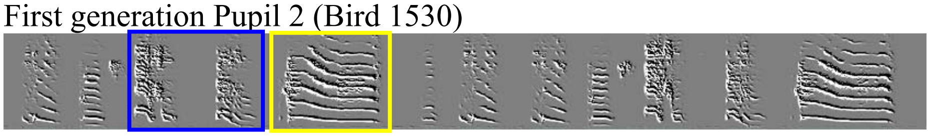


Tutor 4 had the most abnormal song among our tutors. There is a harmonic element (blue square) that is long and call-like (although it was not included in every single bout) followed by an unstructured, broadband, scratchy syllable (in yellow) that can be more than a second long, which is longer than the typical zebra finch song motif. This syllable is highly variable in the pupil in both duration and bandwidth and internal structure. The pupil of this tutor did not imitate the call-like harmonic, and greatly reduced the length and the bandwidth of the scratchy syllable. He has created a stable motif and bout structure out of a highly unstable song.

ISO Tutor 5 (Bird 1249)

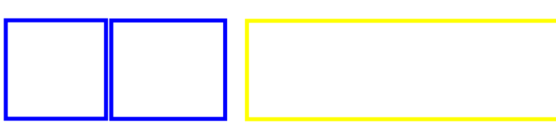
First generation Pupil 1 (Bird 1439)

First generation Pupil 2 (Bird 1530)

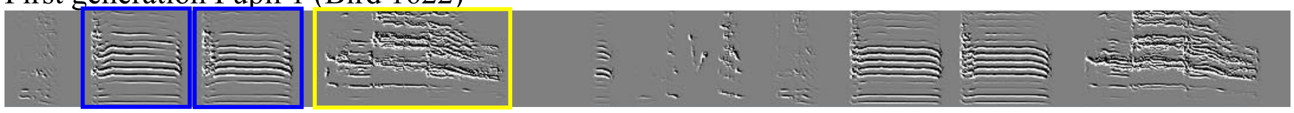


Tutor 5 has a song with a few short syllables (in blue square) that are within the normal, wild type range in duration as well as other acoustic features, but the last syllable (in yellow) is an extremely long call-like harmonic which is too long to be a part of a normal zebra finch motif. Both of the pupils of this bird shortened the long call into a medium-length harmonic (a typical element of wild type song) but copied all or some of the other elements with high accuracy.

ISO Tutor 6 (Bird 1529)



First generation Pupil 1 (Bird 1622)

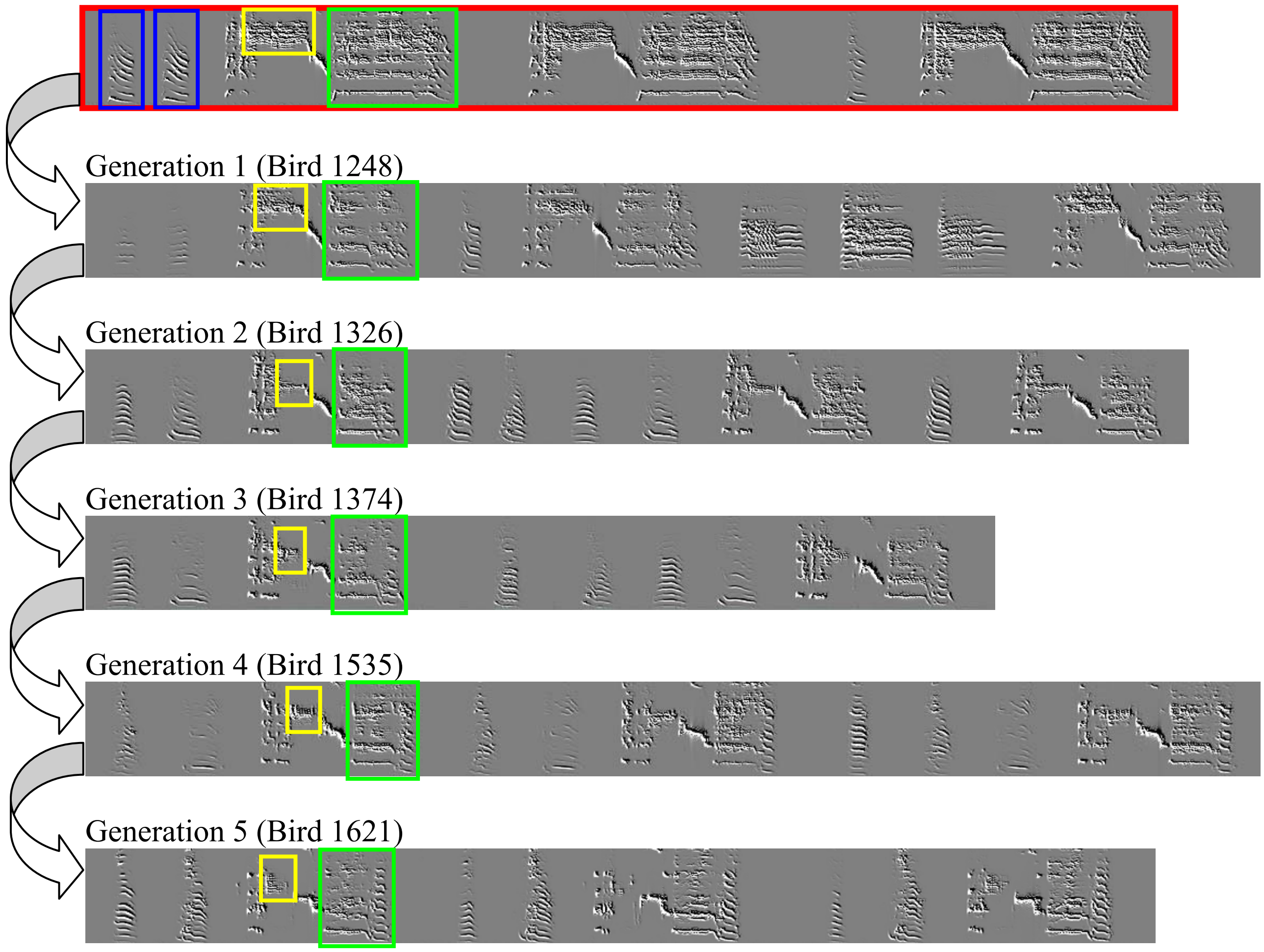


Tutor 6 produced a very simple and highly abnormal, though quite stable, song. There are two long harmonics (blue) and an extremely long high-pitched harmonic (yellow). The pupil imitated the harmonics but shortened the long syllable. He not only shortened it, but also started to differentiate it into notes.

As can be seen above, there was a general tendency among pupils of isolate birds to shorten certain syllables that were of long duration. Repetitions of syllables disappeared in the songs of the pupils and some syllables became more complex or more stereotyped. In all cases, the pupils' motifs were highly stable: they always repeated the syllables in the same sequence with only minor acoustic variations, even in cases where their tutors sang extremely variable songs.

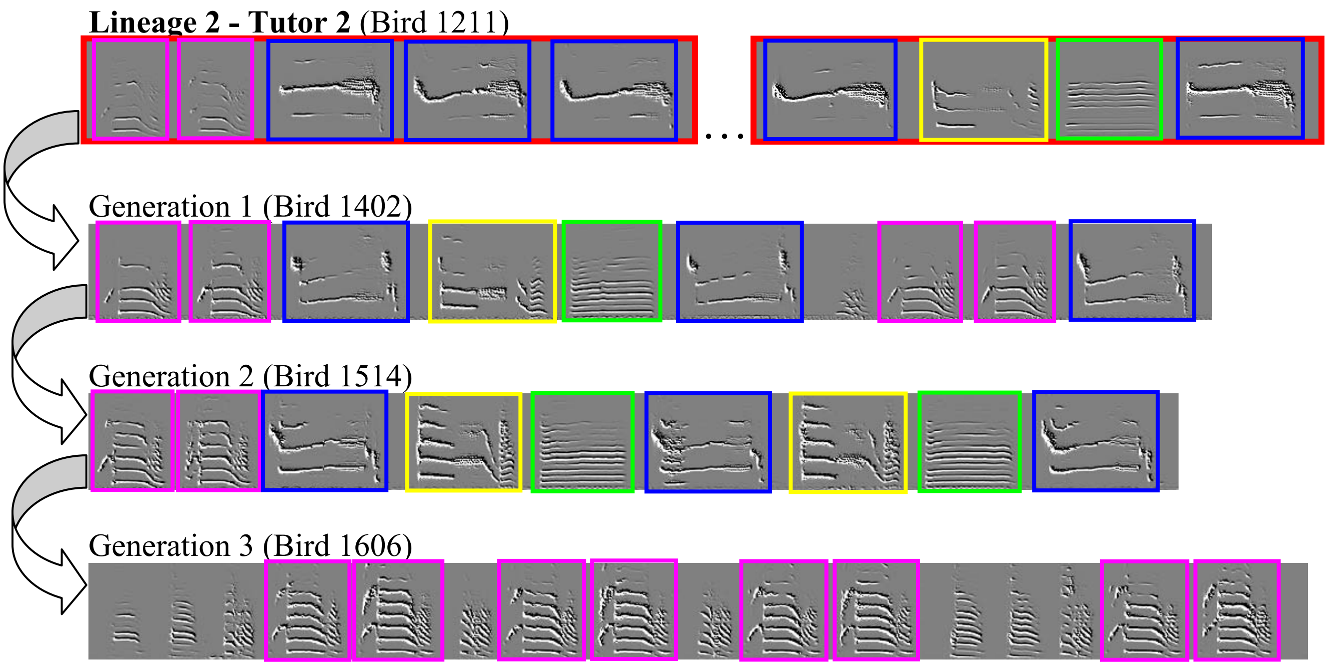
b. One-to-one tutoring – multiple lineages

Lineage 1 - Tutor 1 (Bird 19)



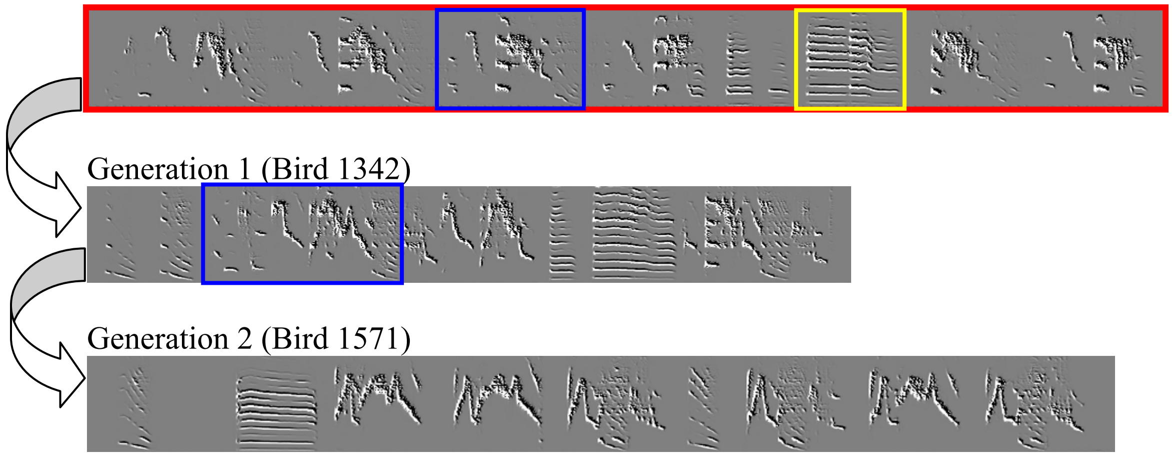
In the case of **tutor 1**, following the song of his pupils over generations reveals that much of the structure of his song is preserved by the pupils – but with modifications. Interestingly, the changes that the first generation pupil made to the song were repeated and accumulated in the succeeding generations. For example, the song duration of the green syllable was decreased further by the second generation pupil and its internal structure was modified and differentiated even by the fifth generation pupil who sang a short but very clear harmonic at the end of this syllable. Similarly, the yellow syllable became much reduced over generations of learners, and by the fifth generation, there was only a trace of the original syllable.

There were bout-level changes that were made to this song, as well. The bout got stretched out by the introduction of other, short (introductory-like) syllables between the renditions of the long syllable. The second generation pupil sang some medium duration modulated call-like syllables between the motifs. These syllables were not improvised, as they can be found in the isolate tutor's repertoire, although at much lower frequencies. His pupil sang them in nearly every bout. These syllables didn't become prevalent in the pupils' songs.



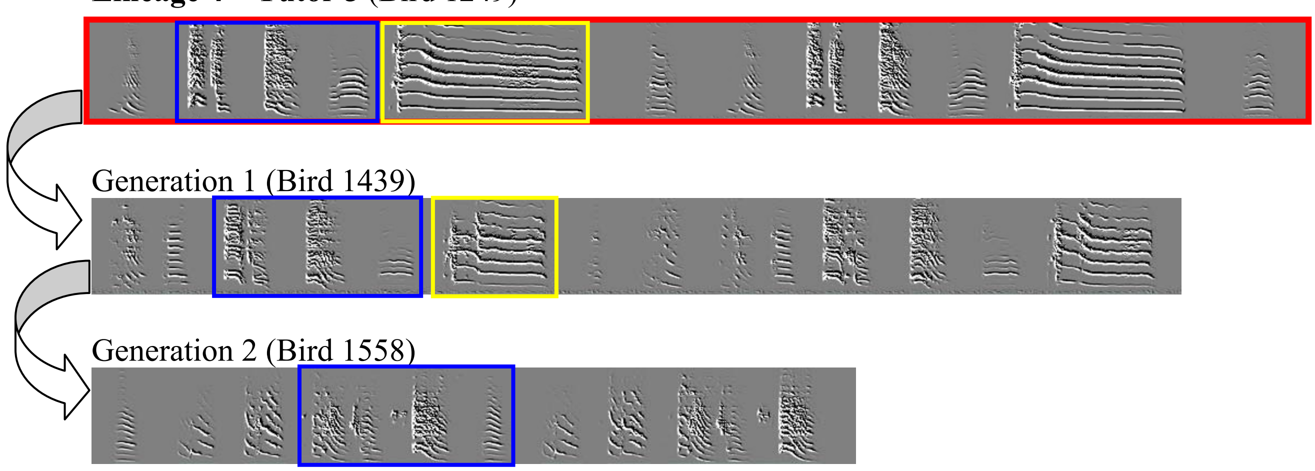
In this lineage, the isolate **tutor 2** sang a very long, high-pitched syllable as the dominant element in his motif (blue rectangle). This song was extremely unusual due not only to the structure of the main syllable but to the length of the motif: the bird repeated the main syllable multiple times. Two other syllable types of similar duration (in yellow and green), sung only once, were nested among these repetitions. Interestingly, the first generation pupil did not imitate the repetitions and constructed his motif out of the three long syllable types, singing them serially once and ending with the first one. Another syntactical reorganization happened in the second generation, when the song went from one rendition introduced by two renditions of the pink syllable (ppABCA) to ppABCABCA. This motif repetition and bout lengthening is reminiscent of the syntax changes that take place during development in an individual bird.

Lineage 3 – Tutor 3 (Bird 1238)



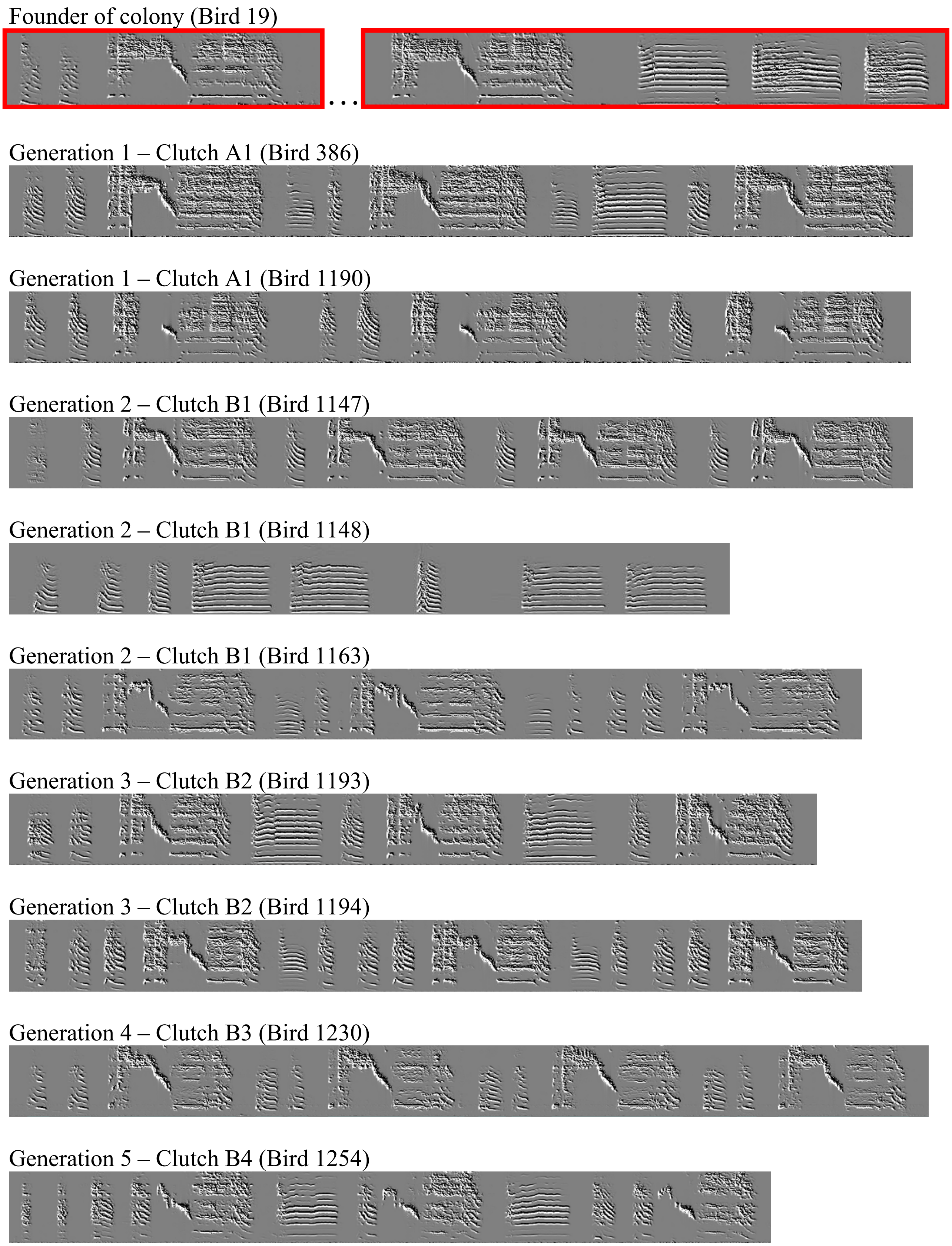
Tutor 3 sang a complex syllable (blue rectangle) that, based on its acoustic features and fast transitions, could be classified as a wild type song. However, the syllable repetition is highly unusual and the stuttered harmonics (longest in yellow) that follow these repetitions are also abnormal. The number of repetitions decreases in the song of the first generation pupil and the blue syllable becomes even more complex and differentiated so that it is not repeated exactly the same way each time. The second generation pupil breaks up this long syllable into two independent types, and though he repeats the first which is still typical to isolates, we see the stability of elements and a clear syntax emerge in this bird's song.

Lineage 4 – Tutor 5 (Bird 1249)



Tutor 5 sang a simple song with some short, introductory-like notes (blue rectangle) followed by a short and a very long harmonic (yellow rectangle). In this case, that syllable type was not copied by the second generation pupil.

a. Imitations of ISO song in a semi-natural colony



3. **Experimental Methods**

a. Animal care

All experiments were performed in accordance with guidelines of the National Institutes of Health and have been reviewed and approved by the Institutional Animal Care and Use Committee of the City College of New York.

b. Experimental design

We used zebra finches (*Taenyopygia guttata*) from the City College of New York breeding colony. Colony management and isolation procedures have been described previously[1]. Zebra finches do not imitate songs heard prior to day 20 post hatch[2], and we kept our ISO acoustically isolated from songs after day 7. Birds were raised by their mothers from day 7-29 post hatch with no adult males present. Except for the colony experiment, all birds were kept either singly (isolates) or pair-wise (live tutored) in sound attenuating chambers from days 30 to 120 after hatching, or longer.

c. Experimental groups

<u>Wild type (WT) songs (n=52):</u> To obtain a baseline for wild type songs produced by domesticated zebra finches in typical aviaries, we used recordings from birds raised in two well-established colonies: The Rockefeller University Field Research Center colony and Hunter College zebra finch colony. Both colonies have existed for over 20 years. Birds were kept in breeding rooms including family cages (with 6-12 birds) or in larger semi-natural aviaries. All birds were raised in a social environment including mixed company of males and females. We used 52 songs altogether. We thank Fernando Nottebohm, Noam Leader and Cheryl Harding for making the wild-type songs accessible to us.

<u>Early isolation from songs:</u> all birds (except for the isolated colony birds) were raised by their parents in a dedicated cage until day 7 post hatch (we start counting from the hatching of the first egg). The father was then removed, and the cage (with the nest box) was taken to a nursery area housing mothers (who do not sing) and chicks only. Birds were raised by their mother, and at day 30, when the young can already feed themselves (and just at the onset of subsong), birds were placed in sound attenuation chambers as described below.

<u>Isolates (n=17):</u> Birds kept visually and acoustically separated from other birds during the sensitive period for song learning are called isolates. 17 birds were isolated from songs from day 7-29 post-hatch. On day 30, male birds were placed singly in sound attenuation chambers. They were kept in complete isolation from day 30 until day 120 post-hatch or later. Six of these birds were used as tutors. We used 4 of them as tutors 2-4 times, and confirmed in each case that their song did not change (based on visual inspection & feature distributions). Tutors' age ranged between 140-1571 days (median age at beginning of first tutoring = 316 days). One tutor was first used as a founder of our "island colony", and then as a one-to-one tutor. In this case as well, no measurable changes in the isolate song structure (namely, no apparent change in syllabic structure and feature distribution, or addition of syllable types) could be detected.

<u>Pupils of isolate tutors (n=13)</u>:
Early isolation from song: Birds were isolated from songs from day 7 to 29. On day 30, birds were placed in sound attenuation chambers together with an adult isolate (one of

our 6 isolate tutors). The isolate tutors were removed when the pupils were 120 days old. At that time we recorded the song of the pupil, and also obtained an additional recording from the isolate tutor (in a separate box) to test if his song remained unchanged.

One-to-one tutoring: We randomly selected hatchlings from 40 breeding pairs, and paired them with one of 6 isolate tutors on day 30. The isolate tutor and his pupil were kept together for 90 days in a sound isolated chamber. Subsequently, we recorded the pupil's song and compared it to that of his tutor. To evaluate the effect of individual tutors, four of the isolate tutors were used 2-4 times to train unrelated pupils. Additional training was performed serially (one-to-one) after confirming that the isolate tutor song remained stable over the tutoring period.

<u>Tutoring lineages</u>: For 4 out of the 6 isolate tutor songs, we established a line of learners, where the first generation pupils tutored another generation which, upon reaching adulthood (between day 120-140) tutored another generation. This allowed us to track the same song as it was passed down over a few generations.

d. Isolate colony setting

We constructed a large isolation chamber from an old 20 cubic ft refrigerator (Suppl. Fig 1). The chamber contained three separate compartments, each equipped with a nest box, microphones and video cameras (Labtec webcam USB cameras). The birds had free movement between the chambers.

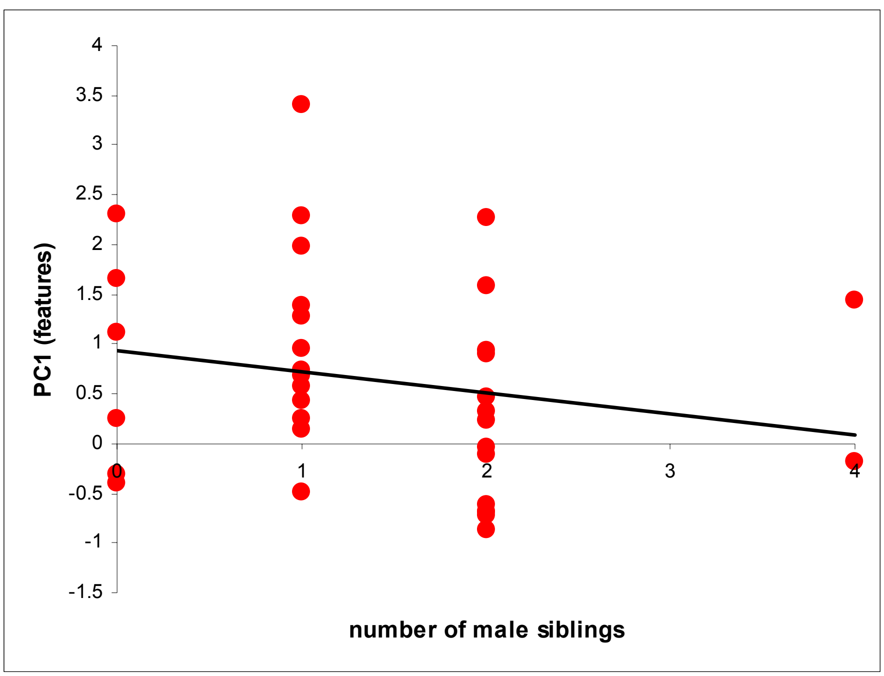
Supplementary Figure 1 | Island Colony setting.

We used one of our isolate tutors to establish the "island colony". Three females and this isolate male were placed into the chamber and kept completely isolated, acoustically as well as socially from other birds over a period of 2 years. Once a pair-bond was established, we followed (by audio and video recording) the evolution of this colony. All birds in the colony (except for the 3 female founders) were the descendants of the founder male. However, he only fathered one clutch, after which one of his sons from this first clutch paired up with another of the original female founders and produced all the successive clutches. Based on partial video observations we suspect that the rest of the colony birds were all descendents of this pair. Although we didn't establish certain genetic relatedness, this pair tended to all the succeeding chicks, hatching them and feeding them until fledging. We allowed the colony to grow with the occasional removal of female offspring over five generations of learners. The colony founder was removed just prior to the hatching of the 5th generation learner.

e. Isolation methods & validation

Possible exposure to WT songs prior to day 7: In zebra finches there are 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 quite, and if the males sing at all, they tend to sing away from the nest, facing the other way (Tchernichovski et al, Animal Behav. 1998). Of course, the considerations above should not apply to the island colony experiment since chick hatched there were only exposed to the ISO song.

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. Supplementary Fig. 2 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:



Supplementary Figure 2 | Number of male siblings vs. first Principal Component of WT/ISO features of all experimental birds.

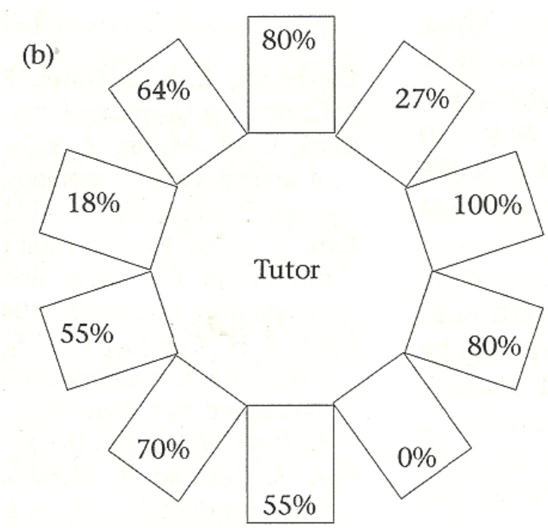
As shown, there is no clear trend (r^2=0.04) and looking separately in each experimental groups also show no significant or consistent trend. Similar results were obtained in duration of acoustic state and in rhythm.

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.

f. Social isolation or lack of tutoring

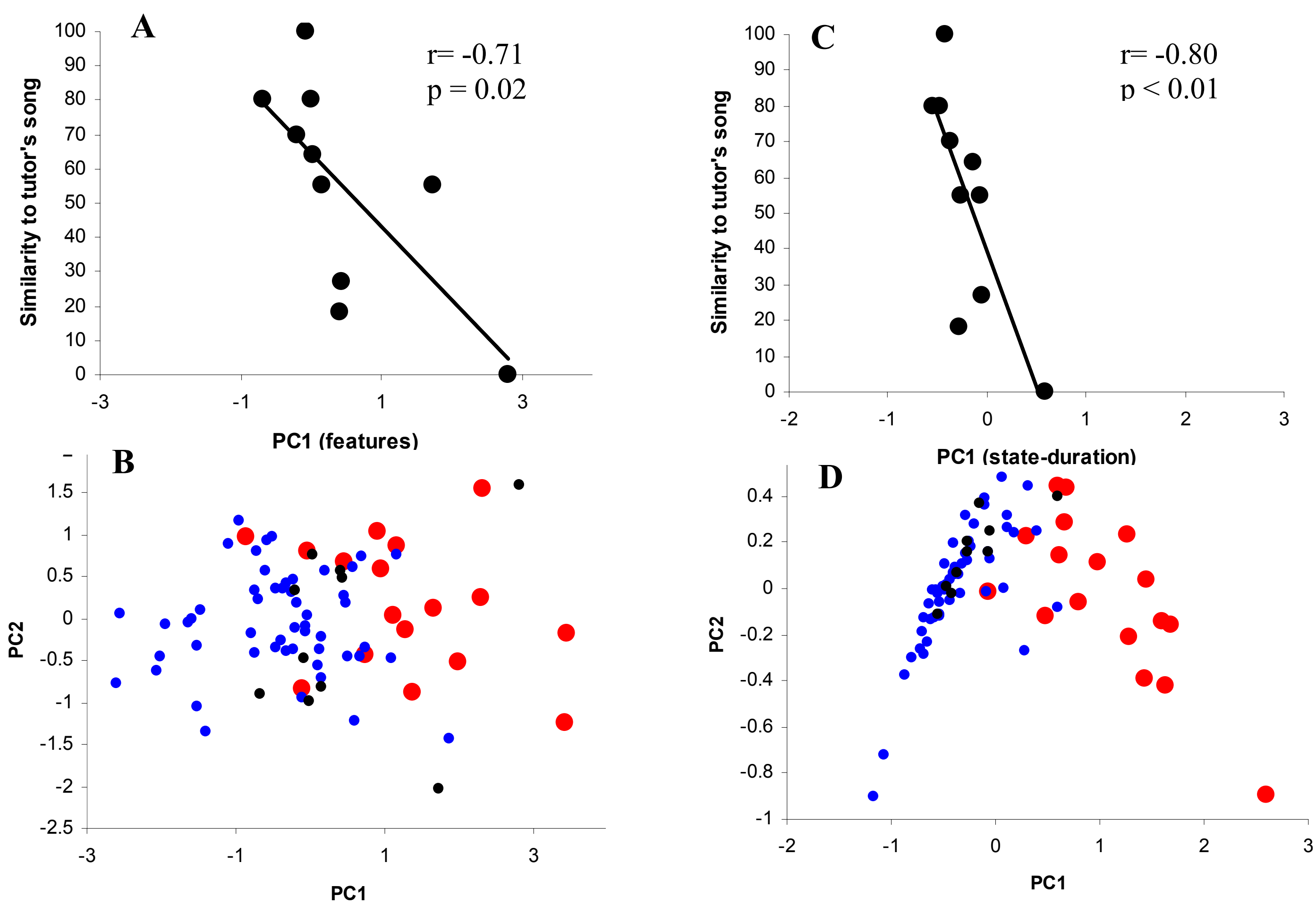
The difference between WT and ISO songs we observed (figures 1, 2) might be caused by the lack of tutoring, but also by social factors unrelated to acoustic experience. For instance, in sedge warblers, deprivation of song tutoring does not result in less structured songs[3]. We examined if social inhibition of song imitation[4] can lead to ISO-like songs, by allowing 10 pupils to interact with a single tutor simultaneously. The analysis below demonstrates that the songs were WT-like in pupils who imitated accurately, but more ISO-like the less the pupil copied from the tutor, confirming that zebra finch ISO songs are, to a large extent, an outcome of tutoring deprivation.

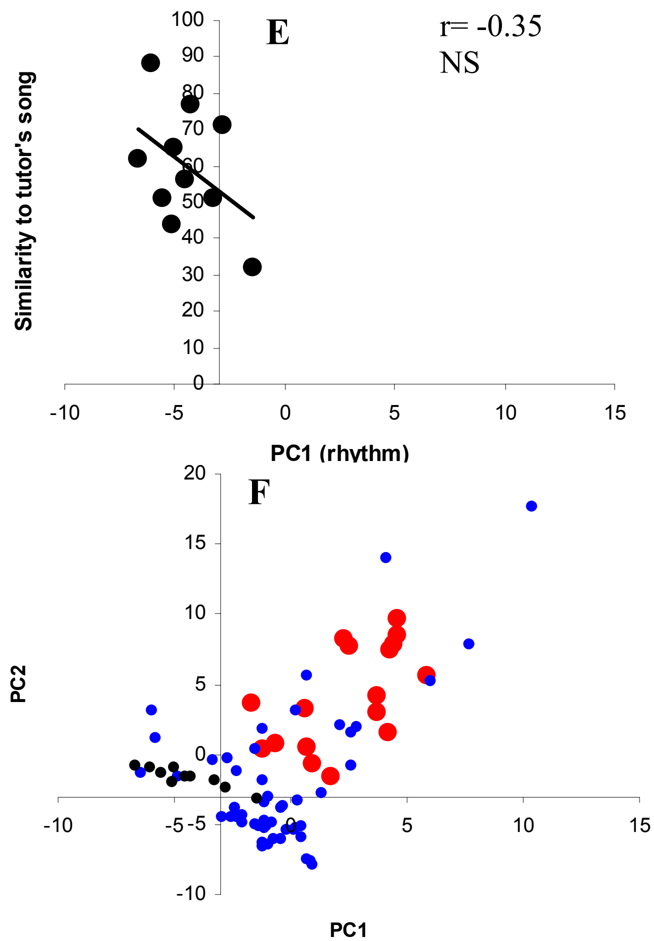
The interpretation of our results would be quite different if the difference between WT and ISO songs using our three song traits (features, duration of acoustic state, rhythm) are mostly due to social isolation, as opposed to the lack of tutoring. To judge this, we used data from birds raised in an arena as shown on Supplementary Fig. 3 (data are from Tchernichovski at al. 2000, Animal Behavior[5]).



Supplementary Figure 3 | Arena of cages with 10 pupils organized around a single tutor. Such one to many arrangements causes inhibition of song imitation, so that one or two pupils imitate well, while other pupils diverge by copying less from the model, and in the extreme cases (e.g., zero similarity, songs composed of invented syllables). Similarity scores show broad range, with no apparent correlation between scores of adjacent cages.

The tutor was placed at the center, and the 10 pupils around him in a circle. The similarity scores between each pupil and the tutor (manual score) are shown for each pupil. In such one-to-many arrangements, songs do not converge but diverge (we suspect that this is how song polymorphism is maintained in zebra finch colonies). Regardless of the underlying cause, imitation is inhibited in some pupils but not in others. So given that some birds are more “tutored” than others, we can test if songs that were less influenced by the tutor are more similar to isolate songs. Supplementary Fig. 4 presents the similarity to the tutor song against the first PC of the PCAs of the three song traits. The correlation is statistically significant for features ($p=0.02$) and for duration of acoustic state ($p<0.01$) but not for rhythm. For the first two traits, the PC values of birds that imitated well project on the WT distribution, and vice versa (doted lines indicates the median PC). For rhythm, however, all birds appear to be within the WT range even though the trend is similar.





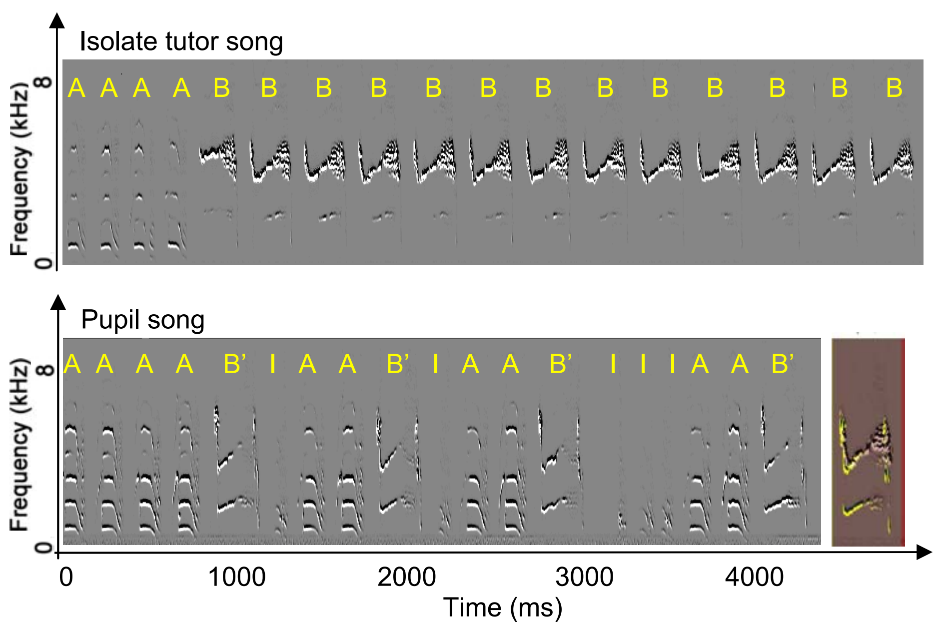
Supplementary Figure 4 | Correlation between imitation outcome in the arena configuration (Suppl. Figure 2) and the first principal component of WT/ISO song traits. For features trait (A-B) higher similarity correlated significantly with correlates with lower PC1 (p=0.02). Similar trend (C-D) was observed for the duration of acoustic state (r=-0.8, p<0.01). However, for rhythm we see only a mild (and non-significant) correlation with all data projection on the WT range of the rhythm PCA (E-F).

These results support the notion that the isolate song is, to a large extent, an outcome of the lack of tutoring. There might be an additional effect of isolation stress on zebra finch song development, and such an effect can be tested by raising birds by mute parents.

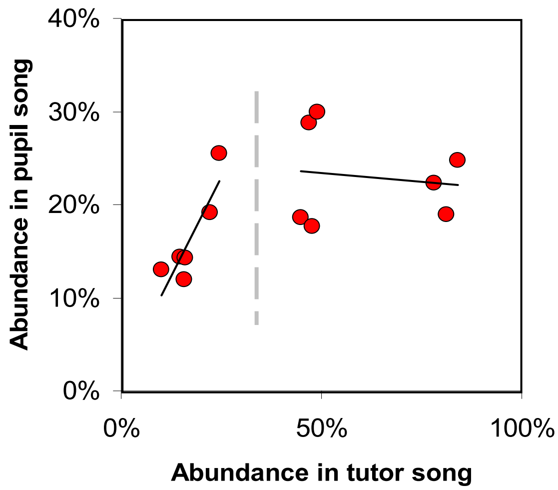
g. Range-limited copying of syllable abundance

Supplementary Fig. 5a presents an example of global reorganization of temporal structure in a pupil's song compared to the tutor song. The isolate tutor sang back-to-back renditions of two syllable types (denoted as *A* and *B*). In the tutor, the abundance (relative frequency) of syllable *B* was 81%. The pupil imitated both syllables but syllable B was altered (B') and its relative frequency decreased to 19%. Across birds we found that isolate tutors varied markedly in the relative frequency of their most abundant syllable (mean=41%, range=10-84%). These "most abundant syllables" were copied by all 13 pupils, but the relative frequencies of the same syllables in pupil's songs were significantly lower and less variable (mean= 20%, range= 12-30%, $p<0.01$, Wilcoxon sign test). Interestingly, when the relative frequency of the most abundant tutor syllable was 30% or lower, the relative frequency of the same syllable in the pupil song followed the tutors' values (Supplementary Fig. 5b**,** $r^2=0.77$, slope=0.85, $p=0.02$, n=6 birds). However, for syllables in the tutor song with relative frequencies higher than 30%, there was no correlation ($r^2=-0.02$, slope=0.04, NS, n=7 birds), and relative frequencies in pupil song decreased to 20-30%. Overall, the ranges of relative frequencies in pupil songs corresponded to that of WT songs, where they rarely reach 30%.

a



b



Supplementary Figure 5 | Range-limited copying of syllable abundance. a, Song bouts of an isolate tutor and his pupil. The pupil rearranged the syllables, altered syllable B (B'), and reduced stuttering. An overlay of syllable B and its imitation, B' (yellow) is shown on the bottom right. **b**, Relative frequencies (abundance) of copied syllables in pupils vs. the relative frequencies in tutor song. When abundance of ISO syllables is 30% or lower, the relative frequency in pupil song matches that of the tutor's.

4. Analysis Methods

All the analysis presented below was performed using *Matlab 7*, except for feature calculations, which were done using *Sound Analysis Pro 2*.

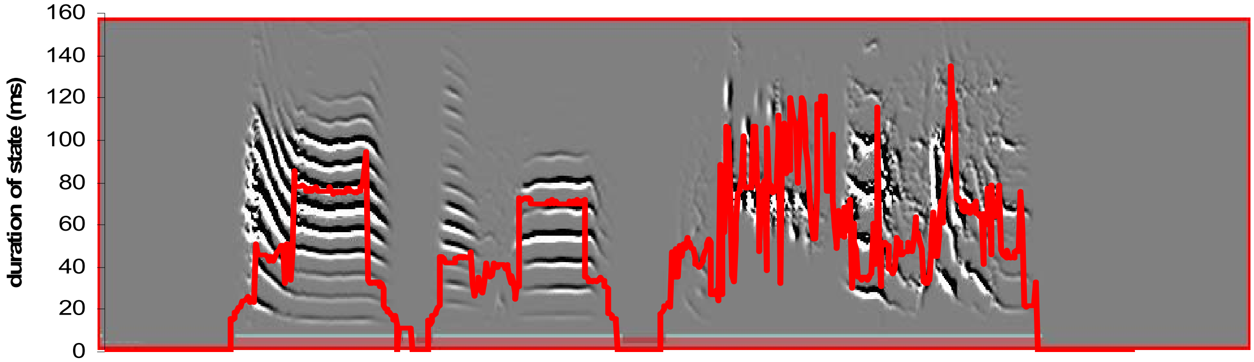
a. Spectral frame (10 ms) features

We used *Sound Analysis Pro 2* to calculate song features: pitch, frequency modulation (FM), amplitude modulation (AM), Wiener entropy & goodness of pitch. Features were calculated in 10 ms windows, every 1ms of singing. We analyzed 20 sec of singing bouts for each bird. We computed distribution histograms as well as cumulative histograms (Cumulative Density Figures, CDF) and compared CDFs of ISO birds to those of WT birds. To construct the Principal Component Analysis shown in Fig. 2a**,** we used the three features that showed the best separation between ISO and WT in the CDFs (Supplementary Fig. 5): FM, AM and Goodness of pitch.

b. Duration of acoustic state

Isolate song syllables and notes are often prolonged and monotonic. To quantify this notion, we estimated correlation time, namely, the interval where acoustic features remain highly correlated. Song correlation time can be calculated by computing the spectral auto-correlation of the song bout, and measuring the intervals, starting from the diagonal of the auto-correlation matrix, where the correlation coefficient (*r)* is higher than a certain threshold. However, during noisy sounds (e.g., a long monotonic buzz), spectral correlation time is short, even though articulatory state probably remains unchanged during a monotonic noisy sound. Using features that correlate with the articulatory state can provide better estimates of correlation time. We therefore calculated the duration of acoustic state based on pitch, FM, Wiener entropy & Goodness of pitch.. We used an algorithm developed earlier[6] to calculate the period of repetition in songs. Briefly, we scaled the features to units of statistical distances based on WT zebra finch feature distribution. We then constructed a similarity matrix based on Euclidian distance between features calculated every 1 ms. Thereafter, computation of correlation time is similar to that of spectral auto-correlation matrix. We used a threshold of 2.5 Median

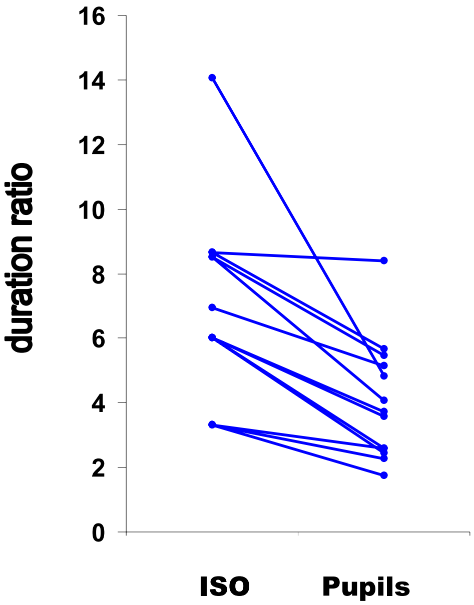
Absolute Deviation (Supplementary Fig. 6). This procedure is implemented in *Sound Analysis Pro 2*.



Supplementary Figure 6 | Duration of acoustic state. The red line represents the duration of acoustic state values at each time point of the sonogram.

c. Note duration ratio

Since *duration of acoustic state* is used here for the first time, we wanted to test if a similar outcome (wider range of durations in ISO) can be seen based on the distribution of note durations. We sampled 10 song pairs of ISO tutor and his pupil. A sample of 10 song motifs for each bird was subjectively segmented to notes based on visual inspection of the spectral derivatives and detecting intervals of continuous sounds. We only considered song syllables and excluded isolated long calls. We then calculated the average duration of the longest and shortest syllables and the ratio between them (Supplementary Fig. 7).This ratio was higher in all ISO tutors compared to their pupil.



Supplementary Figure 7 | Note duration ratios decrease in pupils of isolate tutors (n=13 ISO and pupil pairs). The left column shows the duration ratios of isolate tutors, the right column those of their pupils. In every case, the ratio decreased, showing that pupils had less variability in syllable

d. Rhythm spectrum

Rhythm spectrum[7] was used to detect periodicity (rhythm) in song features over the time scales of the song bout. Rhythm frequencies can capture patterns of repetitions at the syllabic level and at the song-motif level. We used a nested spectral analysis method. First, the song feature time series were estimated. Although feature values at a given time point depend on the fine temporal structure of the waveform with millisecond resolution, the features themselves change with a slower timescale of 10–100ms. The continuous (not segmented) feature time series were subjected to a second spectral analysis, and the result was a "rhythm" spectrum. Here, rhythm spectrum was calculated across song features over 20s of song bouts in ISO and WT birds. We used a low frequency cutoff at 0.5 HZ, to examine time scales up to 1 sec.

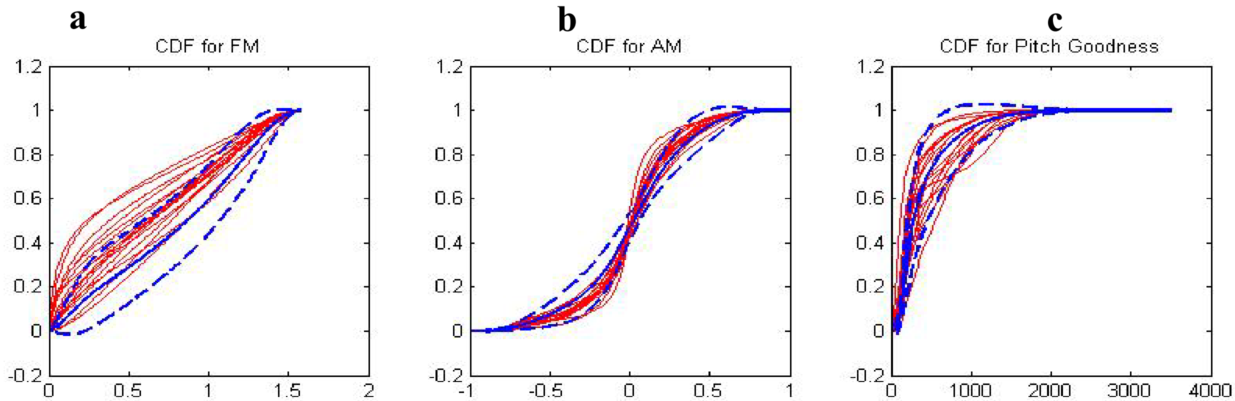
e. Constructing the song features PCA

As shown above, the song of each individual bird is described by a set of feature vectors spanning multiple time scales. For example, when analyzing 20 sec of singing, we obtained several spectral frame features (pitch, FM, etc), with 20,000 time-series values for each feature. We first computed cumulative frequency distributions (CDF) for each feature time series (Supplementary Fig. 8). The CDFs, which summarize the distribution of each feature in a song, are the input vectors of the Principal Component Analysis figure shown in Fig. 2a. Note that each red dot in Fig. 2a is a two-dimensional projection of CDFs (red lines) presented in Supplementary Fig. 8 for each bird (combining FM, AM and Goodness of pitch).

i. Constructing *PCA of 10 ms song features*

For each ISO and WT bird, we used 20s of song bouts to calculate CDFs for pitch, frequency modulation (FM), amplitude modulation (AM), Wiener entropy and Goodness of pitch using Sound Analysis Pro version 2. We found that three of those features: FM, AM, and Goodness of pitch showed different distributions across WT and ISO

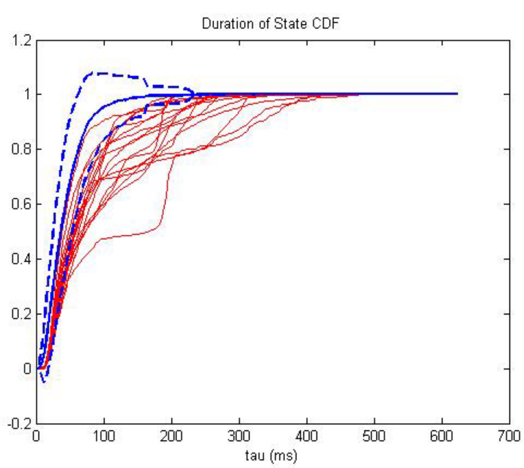
(Supplementary Fig. 8). We therefore used these three feature CDFs as inputs to the song features PCA.



Supplementary Figure 8 | Cumulative Frequency Distributions (CDFs) of frequency modulation (FM), amplitude modulation (AM) and goodness of pitch. Each red line represents an ISO bird song. The blue line represents the mean CDF values of 52 WT birds, and the dashed lines show 95% confidence intervals.

ii. Constructing *PCA of duration of acoustic state*

For each ISO and WT bird, we used 20s of song bout to calculate vectors of duration of acoustic state (see section 3b). Supplementary Fig. 9 presents the CDFs of the vectors for ISO and WT birds.



Supplementary Figure 9 | Cumulative Frequency Distributions (CDFs) acoustic state durations. Each red line represents the song of one ISO bird. The blue line represents the mean CDFs values of 52 WT birds, and the dashed line show LocFit 95%

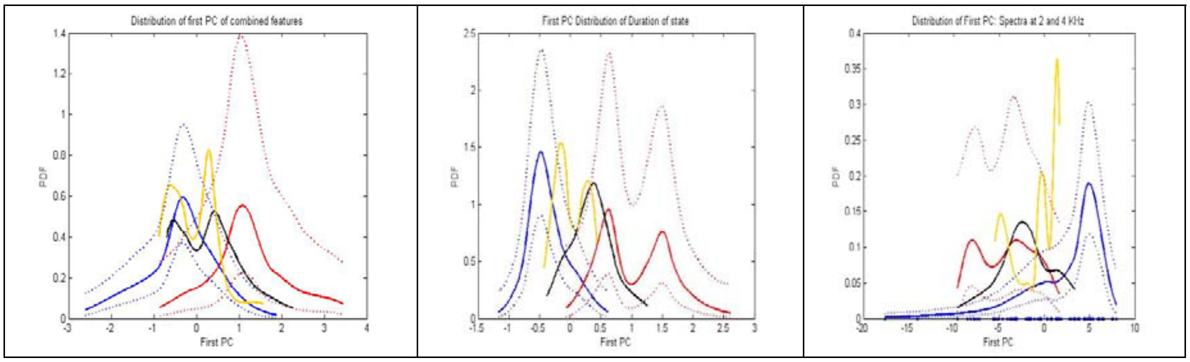
iii. Constructing *PCA of rhythm spectra*

Rhythm spectrum was calculated over 20s of singing as described in Section 4d above. The rhythm spectrum was then treated as a vector input to the PCA, as with the CDF vector for acoustic state duration.

5. Statistics

a. 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 Supplementary Fig. 10.



Supplementary Figure 10 | 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 time scales.

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.

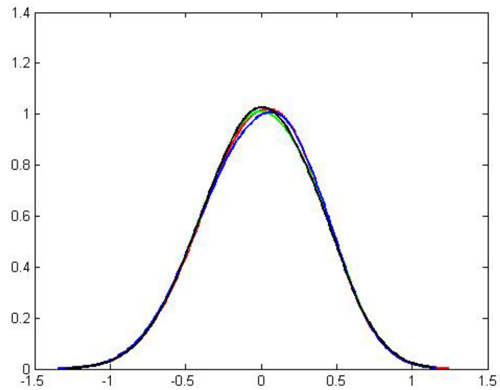
b. Statistical test methods

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. Supplementary Fig. 11 shows example permutation distributions.



Supplementary Figure 11 | Four examples 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.

c. Statistical results of group means

i. Spectral frame feature

The mean of first PC of the four bird groups are

WT	Higher generations (F2+)	1st-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	1st-generation	ISO
WT		0.1736	0.0180	< 0.0001
Higher generations			0.4742	0.0045
1st-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. So the later generations of tutoring indeed move the birds closer to 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 2^{nd}-generation alone.

<u>ii. Duration of state</u>

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

The p-values obtained from t-test are

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

The p-values obtained from permutation test are

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

d. 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,

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

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:

$$H_0 : p \le 0.5,$$

$$H_1 : p > 0.5.$$

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

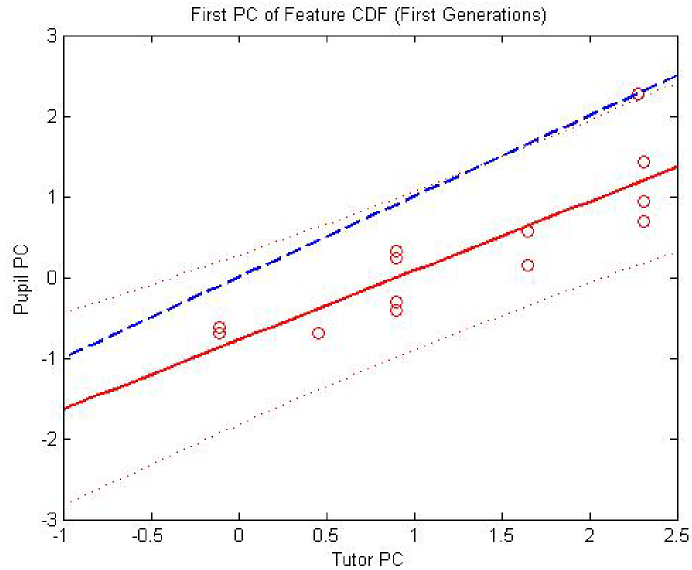
There are 39 hypotheses. Using the Benjamini-Hochberg procedure, those null hypotheses that *cannot* be rejected at a FDR level of 0.01 are marked blue in the table above. We conclude that the pupil's songs become more similar to WT as a result of tutoring.

e. 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 Supplementary Fig. 12.



Supplementary Figure 12 | The relationship between tutor and pupil's trait value. 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

$$H_0: \quad y = x$$

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

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.

f. 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). We , the coefficient of variance (CV) of syllable durations was significantly higher in the isolates' syllables compared to the imitation of those syllables (isolate: CV=15%, range 2-57%; pupils: CV=4%, range 2-9%; p<0.05, Wilcoxon sign test).

Supplementary references

1. Deregnaucourt, S., Mitra, P. P., Feher, O., Pytte, C. & Tchernichovski, O. How sleep affects the developmental learning of bird song. *Nature*, **433**, 710-716 (2005).

2. Roper, A. & Zann, R. The Onset of Song Learning and Song Tutor Selection in Fledgling Zebra Finches. *Ethology* **112**, 458-470 (2006).

3. Leitner, S., Nicholson, J., Leisler, B., DeVoogd, T. J., Catchpole, C. K. Song and the song control pathway in the brain can develop independently of exposure to song in the sedge warbler. *Proc. Biol. Sci.* **269**, 2519-2524 (2002).

4. Tchernichovski, O., Nottebohm, F. Social inhibition of song imitation among sibling male zebra finches. *Proceedings of the National Academy of Sciences, U. S. A.* **95**, 8951-8956 (1998).

5. Tchernichovski, O., Nottebohm, F., Ho, C.E., Bijan, P., Mitra, P.P. A procedure for an automated measurement of song similarity. *Animal Behaviour* **59**, 1167-1176(2000).

6. Tchernichovski, O., Mitra, P.P., Lints, T., Nottebohm, F. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* **291,** 2564-2569 (2001).

7. Saar, S. & Mitra, P. P. A technique for characterizing the development of rhythms in bird song. *PLoS ONE* **3**, 1461 (2008).