Scientific Reports

https://doi.org/10.1038/s41598-026-44602-5

Article in Press

A species rules syntax model accurately organizes birdsong syllables into songs

Received: 30 July 2025

Accepted: 12 March 2026

Published online: 24 March 2026

Cite this article as: Edwards J.A. & Woolley S.M.N. A species rules syntax model accurately organizes birdsong syllables into songs. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-44602-5

Jacob A. Edwards & Sarah M. N. Woolley

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

A species rules syntax model accurately organizes birdsong syllables into songs

Jacob Aaron Edwards[1,2]; Sarah Margaret Nicolay Woolley*[1,2]

[1]Mortimer B. Zuckerman Mind, Brain, and Behavior Institute, Columbia University, New York, NY 10027, USA

[2]Department of Psychology, Columbia University, New York, NY 10027, USA

*Correspondence to S. M. N. Woolley; Zuckerman Institute at Columbia University, Jerome L. Greene Science Center, 3227 Broadway, L3.031, New York, NY 10027, USA

Email: sw2277@columbia.edu

Keywords: birdsong; syntax; sequence; model; rules; universals

Abstract

Birdsong is a rare example of learned vocal communication. Like speech, song consists of acoustic units (syllables) that are learned from adults and organized into temporal sequences (syntax) during development. Tutors' and pupils' syllables and syntax are often the same, suggesting that both song features are learned. Syntax is highly species-specific and consistent across individuals and populations of the same species, unlike syllables. Here, we hypothesized that song syntax can be accurately predicted by computational models based on the relationship between syllable acoustics and order that is consistently observed in the songs of conspecific individuals. We tested this hypothesis using techniques inspired by natural language processing to generate and test a species rules model for each of two species of estrildid finches whose songs are composed of stereotyped syllable sequences that differ in syntax. We modeled species rules as the significant correlations between a syllable's acoustic features and sequence position in a song, measured by analyzing song acoustics in many individuals of the same species. We used only a bird's syllable types and its species rules to predict syllable sequence in the birds' adult songs, without using tutors' songs to train or test the model. We then quantified the accuracy of the species rules model in predicting a bird's syllable sequence compared to the accuracy of the tutor's song in predicting a bird's syllable sequence. Results showed that species rules models predicted birds' actual sequences as well as did tutors' sequences, in both species and across different colonies. Results support the hypothesis that species-specific rules based on syllable acoustics and order can explain the species-specific syntax of birdsong. The modeling approach developed here has general utility for detecting, predicting and comparing sequential structure in complex audio signals.

Main Text

Introduction

Vocal communication signals often consist of acoustic units organized into temporal sequences. In speech, words are temporally organized into sentences using syntactic rules that determine the information conveyed by the speaker. Fundamental tasks in speech analysis and natural language processing are the modeling of syntactic rules to understand word sequences, and the use of those models to predict or generate new word sequences[1,2]. In birdsong, syllables are temporally organized into songs using syntax (syllable sequencing or order) that is similar to tutors' songs[3-6] and other conspecific songs[6-8]. Because syllable acoustics and syntax are often shared between tutor and pupil songs[4,6,9,10], both aspects of song structure appear to be learned. Song syntax is highly species-specific, however; syllable sequencing is different across species and consistent across individuals of a species[11-14]. This suggests that the species-specific syntax of birdsong may be accurately predicted by computational models that account for the statistical relationships between syllable acoustics and order found in the songs of conspecifics.

Tutoring manipulation studies suggest that species-specific statistical biases for syllable sequencing contribute to song organization. For example, the songs of untutored birds, those reared without hearing song, are composed of highly abnormal syllables that are organized into species-typical sequences[8,15-18]. Untutored zebra finches sing stereotyped repeats of syllable sequences (motifs; *e.g.* ABCABCABC), as do normal, tutored zebra finches[7,19-21]. Song sparrows and swamp sparrows show reliable species differences in song structure, such as the number of song types, song duration, and syllable rate, whether birds are tutored or untutored[22]. Cross-species tutoring experiments show that birds learn other species' syllables, but organize them into temporal sequences that are unlike those of their tutors' songs[7,23,24]. The songs of pupils tutored with artificial songs also support the hypothesis that species-specific biases contribute to song syntax[25-28]. For example, canaries tutored with artificial songs in which syllables are organized into glissandos rather than trills copy the artificial syllables and then produce them in trills, like normal canary song and unlike the artificial tutor song[25]. These findings suggest that song syntax may be predicted by models that capture species-specific statistical patterns in syllable order.

Within a species, correlations between syllable acoustics and sequence position have been demonstrated[11,26,29,30]. White-crowned sparrow song begins with a whistle[8,12,29], zebra finch song ends with long-duration syllables[19,26], and long-tailed finch song begins with short-duration, broadband syllables[7,31,32]. These species-specific syllable acoustic and syntax traits are consistent across conspecific populations of wild and captive birds[11-14,26], suggesting that computational models based on correlations between syllable acoustics and order may accurately predict song syntax. Studying statistical structure in birdsong offers unique opportunities for solving natural language processing tasks because no general lexicons describing relational structure between vocal units exist for birdsong[3]. "Vocabularies" of syllables are often unique to individuals and families, requiring an abstraction of syllables into types in order to analyze sequences[33]. Thus, computational modeling of birdsong syntax requires a generalizable strategy using correlations between acoustic structure and sequence position found in vocal signals[34].

Here, we tested the hypothesis that computational models of species syntax "rules" accurately predict syllable order in adult pupils' songs, without using tutors' songs to train or test the models. By quantifying correlations between the acoustic features and sequential positions of syllables in the songs of normal (reared and tutored by parents) birds, we generated a model of species rules for song syntax in each of two species, zebra finches (Z; *Taeniopygia guttata castanotis*) and long-tailed finches (L; *Poephila acuticauda*). The songs used were recorded from birds in three university colonies in different geographical regions and were not experimentally manipulated. Zebra finches and long-tailed finches have highly similar audiograms[35], developmental trajectories[20], song development stages[24,31,36,37] and ranges of acoustic frequencies in song[32,38]. In contrast to these similarities, syllable acoustics and syntax differ between the two species[24,31,32]. Below, we describe the generation of syntax rules models for each species, using each species' rules model to predict the syllable sequences of individuals' songs, and testing the accuracy of predicted songs by comparing them to each bird's actual song and tutor song.

Materials and Methods

Ethics statement

All animal handling and use procedures were conducted under the guidelines and approval of the Institutional Animal Care and Use Committee (IACUC) of Columbia University. The study reported here was conducted in compliance with recommendations in the ARRIVE guidelines.

Animals

We generated models of species rules and predictions of syllable sequences using adult songs of zebra finches (Z; *Taeniopygia guttata castanotis*) and long-tailed finches (L; *Poephila acuticauda*). Songs were recorded from 74 birds across 3 colonies that were reared and tutored by conspecific adults.

Birds were bred and reared in single-family enclosures. Breeding birds reared their young and served as tutors. Tutor birds were all normally-reared (tutored by conspecifics). All birds from which songs were recorded were ≥ 110 days old and < 4 years old, producing adult song. Birds of the two species were bred from parents of the same species, and received typical tutoring. Juveniles had unrestricted social access to both parents and siblings. Birds in the Columbia University colony could hear the vocalizations of other birds of both species as juveniles and adults. Song was recorded from two generations of normal birds so that the songs of tutors (Z, n = 13; L, n = 10) and pupils (Z, n = 13; L, n = 21) could be compared. We also obtained and analyzed the songs of 13 normally-reared Z birds (6 tutor-pupil pairs and one bird used for model training) reared and recorded at Tufts University, and 4 normally-reared Z birds (2 tutor-pupil pairs) reared and recorded at the University of California, San Francisco. The songs of birds from Columbia University constituted the Colony 1 dataset, and the songs of birds from Tufts University and UCSF constituted the Colonies 2 and 3 dataset. The family lineages of all tutor and pupil birds in the dataset are shown in **Supplementary Fig. S1**. All Z siblings were from different clutches (hatch dates > 5 months apart). Four pairs of L siblings were from the same clutch. For all analyses except those shown in **Figure 4c**, the Z song data were pooled across colonies, which resulted in a total sample of 22 tutors and 21 pupils.

Song recording and syllable extraction

To record singing behavior, we placed birds individually in sound-attenuating chambers equipped with microphones routed through an audio interface to a computer running Sound Analysis Pro[39]. Recorded song bouts (epochs of continuous singing) were digitized at 44100 Hz and stored as uncompressed wave files. Birds typically produced enough song for analysis over 2 to 4 days of recording. From each bird’s recorded song bouts, we randomly selected 25-30 seconds of song for analysis. Recordings were band-pass filtered (250-8000 Hz) and root-mean-square power matched (65 dB SPL).

To delineate syllables in song bouts, we used the amplitude of the sound pressure waveform[24]. Syllables were units of continuous energy that were at least 50% above the noise floor. For complex syllables that contained some amplitude fluctuations, we considered a syllable to be a single unit if any drops below threshold were shorter than 5 ms. We excluded introductory notes, which occur only at the beginning of a song bout[40,41]. This procedure resulted in a total of 12,187 syllables included for analysis (average number of syllables per bird: Z: 124.40, range: 65-234; L: 154.42, range: 67-342).

Syllable type labeling

We measured 26 acoustic features of each syllable using the *warbleR* package[42,43] for R (v4.1.2). Features measured for each syllable included duration, mean dominant frequency, spectral entropy, *etc*. These features capture a broad range of spectrotemporal acoustics and have been used to discriminate syllables of different species[38,44]. We labeled syllables using letters, according to their order in a bird's song (*e.g.* A, B, C, D) and confirmed each bird's syllable types from principal component analysis (PCA) of each syllable's acoustic features (**Fig. 1c**)[7,38]. The full list of features and PC loadings are provided in **Supplementary Table S1** and **Supplementary Table S2**. After performing PCA, syllables with similar acoustic features formed clusters. Syllables within a cluster were assigned the same letter label. All syllables labeled as the same type were verified by plotting and cross-correlation of syllable spectrograms[7,24].

To compare syllables between tutors and pupils, we used a set of shared labels for each tutor-pupil pair. We first assigned syllable type labels for a tutor bird by using PCA to reduce the acoustic space of the tutor's syllables alone. Then, we assigned type labels to syllables within each cluster. After labeling each tutor syllable by type, we projected the pupil's syllables onto the tutor's PC coordinates[45] and assigned type labels to pupil syllables based on proximity to tutor clusters (**Fig. 1c**). For example, if a pupil produced a syllable cluster near or overlapping the tutor's type A cluster, those syllables were labeled A. In some cases, a pupil bird copied a syllable from his tutor but separated it into two distinct syllables. We annotated these syllables as modifications of a tutor's type (*e.g*. A- and A+), and considered them novel types produced by the pupil[4]. Letters labeled the syllable types in the songs of tutor-pupil pairs rather than across all birds of a species. For example, syllables labeled with the letter A were unique to each tutor-pupil pair. The syllable labeling procedure identified 124 syllable types across all Z songs (average number of types per song: 4.77; range: 2-7) and 67 syllable types across all L songs (average number of types per song: 4.71; range: 3-7).

To verify the accuracy of syllable type classification and pupils' copying of tutors' syllables, we quantified syllable similarity by cross-correlating syllable spectrograms within each bird and between pupils and tutors (**Fig. 1d-e**)[24]. For both self-comparisons and tutor-pupil comparisons, we randomly selected 20 (where possible, but no fewer than 5) examples of each syllable type for analysis. We then computed syllable spectrograms (log-transformed) with 2 ms temporal resolution and 100 Hz spectral resolution, and rescaled the amplitude of spectrograms to range from 0 to 1. Then, we cross-correlated each pair of syllables: the shorter spectrogram was convolved with the longer in 2 ms time increments, and the peak of the correlation function was taken as the similarity measure. Cross-correlation coefficients above 0.45 consistently identified syllables of the same type[7]. Visual inspection of every syllable classified as belonging to a type (*e.g.* A) was used to confirm the accuracy of type classification[7,26,30,46-49].

Quantification of differences in syllable acoustics across sequence positions

Song syntax transition diagrams were calculated and plotted as in[7] (**Fig. 1b**). Transition diagrams show the direction and probability of transitions between syllable types and from/to the start/end of a song bout. For each bird, the numbers of transitions between pairs of syllables were counted across all song bouts and converted to probabilities[30,48,50,51].

All measures of syllable sequences used to make and evaluate predictions were conducted in and across song motifs. A motif was defined as a stereotyped sequence of syllables that was repeated in and across a bird's song bouts. Each adult male Z bird and L bird has one song motif, which is sung once or repeated multiple times to form a song bout[14,32,52]. To quantify differences in syllable acoustics depending on syllable position, we first binned syllable positions. For this, we chose the number of position bins (5 bins for Z, 10 bins for L) to approximate the average number of syllables per motif in each species (Z: 4.77 ± 1.54; L: 13.66 ± 6.49 mean syllables per motif ± SD). We then quantified the acoustics of syllables at each sequence position. We averaged the values of each of the 26 acoustic features for the first syllable in a motif, across all conspecific songs, then the second syllable and each of the following syllables in a bird's motif. We did this for each bird's song and averaged the acoustic feature values for each position across all birds of a species (**Fig. 2a**). Statistical significance of differences by position for each feature was tested using Bonferroni-corrected one-way ANOVAs (see *Statistical Analysis*).

Generation of species rules

We defined "species rules" as a model that relates the specific sequence positions of syllables to the acoustic features of those syllables. Species rules models were calculated from Z and L tutor songs. First, we calculated a species rules matrix that was organized along the rows with only the acoustic features that significantly differed by position across the songs of each species. The matrix columns were organized as positions in the sequence, with the leftmost column being the first position, and the rightmost column being the last position. To calculate matrix values, we first min-max scaled each feature across all syllables within a motif for each tutor bird (min: -1, max: 1; **Fig. 2b**). We then computed the relative position (0-1; first-to-last) of each syllable within each motif. Then, we divided relative positions into equally-spaced bins (5 bins for Z, 10 bins for L), and averaged the scaled syllable feature values across all birds for each position bin. Finally, we smoothed the feature values across positions (along the matrix columns) using an interpolation kernel. The resulting species rules matrix was an M syllable feature × P sequence position matrix (**Fig. 2c** and **Fig. 3a**, middle). For each pupil sequence prediction, the data from the tutor and any related bird (relatedness > 0.125; first cousin) was excluded from the species' dataset.

Generation of predicted and random sequences

We used species rules matrices to predict the order of syllables in pupils' motifs by calculating the alignment of each syllable with each sequence position. To do so, we first organized each pupil's syllables into an N syllables × M syllable feature matrix, whose values were min-max scaled within each feature (**Fig. 3a**, left). Then, we computed the product of the pupil's syllable feature matrix ($N \times M$) and the species rules matrix ($M \times P$). The resulting product was an N syllables × P positions matrix, in which the values were the alignments of each syllable with each sequence position (**Fig. 3a**, right). Positive values in the alignment matrix resulted from feature-position correlations, whereas negative values were anti-correlations. We then obtained position alignment curves for each syllable type by averaging alignment values for each position (**Fig. 3b**). Our models of species rules related the sequential position of any syllable to a linear combination of that syllable's acoustic features.

From the position alignments curves, we generated sequence predictions using a priority-based positioning algorithm. First, for each motif produced by the pupil, we counted the number of syllable positions to be filled for the prediction (*e.g.* 5 total positions). Then, we found the syllable type (*e.g.* type A) with the highest alignment with any position (*e.g.* position 1) and assigned that syllable type to that position. After filling that position, we proceeded with the next-highest alignment syllable type for any open position (*e.g.* type B in position 2), and continued until all positions were filled. The only constraint we imposed on the prediction algorithm was that the motif had to include each syllable type. We used the same procedure to generate a random arrangement of pupils' syllables by randomizing the position alignment curves. This procedure resulted in motif sequences predicted from species rules (*e.g.* ABEDC) and random sequences (*e.g.* DAECB) that we then compared to the pupil's actual motif sequence (*e.g.* ABCDE).

Sequence match quantification

The match between two syllable sequences was quantified using the Levenshtein distance (LD) metric (**Fig. 3c**), which is widely used to quantify similarity between two strings[34,53,54] and has been previously used to compare syllable sequence similarity in birdsong[55,56]. The LD is a cost to convert a source sequence (*e.g.* pupil's predicted motif) into a target sequence (pupil's actual motif)[54], using the minimum number of operations. Operations include syllable insertions, deletions, and swaps (exchanging one syllable for another). We weighted swap operations using the acoustic similarity between the actual and misplaced syllables[53,57]. A swap cost between similar syllable types was <1, whereas a swap cost between maximally dissimilar types was 1. We then added all costs and divided by the total number of sequence positions to obtain a total cost between 0 (no errors; identical sequences) and 1 (all errors; completely dissimilar sequences)[58]. Sequence match was calculated as 1 - total cost, which ranged between 1 (perfect match) and 0 (complete mismatch).

Sequence match was used to assess similarity of predicted pupil sequences to actual pupil sequences (**Fig. 3c, Fig. 4**). The same metric was used to assess the match between pupil sequences and tutor sequences, and between pupil sequences and random arrangements of pupil syllables, which served as an expected lower bound for model performance[59]. We also used sequence match to quantify sequence similarity across renditions of motifs within individual birds (**Fig. 4c**). Within-bird sequence similarity was computed by comparing actual sequences (*e.g.* motif 1) to other actual sequences (*e.g.* motif 2) sung by the same bird.

We also computed, for each syllable type in a bird’s song, the repeat probability as the number of times the syllable type was repeated out of the total number of transitions for that syllable type. We then averaged, for each bird, all repeat probabilities for all syllable types in the bird’s song.

Statistical Analysis

We used α = 0.05 as a significance threshold for all statistical tests. We tested for syllable acoustic similarity differences using one-way ANOVAs with type comparison (same vs. copied vs. different) as factors. We tested for sequence match differences using nested one-way ANOVAs with sequence type (*e.g.* predicted vs. tutor vs. random) as factors, and bird nested within sequence type. We tested for significant differences in syllable acoustics by sequence position using one-way ANOVAs with sequence position as a factor. Because of the large number of syllable acoustic features measured (26 for each species), we used a Bonferroni correction to adjust the significance criterion for each test (corrected $\alpha = 0.0019$). We performed post-hoc multiple comparisons using the Tukey-Kramer method. We used paired *t*-tests to test for differences in syllable repeat probability between actual and predicted sequences.

Results

Syllable acoustics and sequences are similar in pupil and tutor song

We recorded, analyzed and compared the adult song syllables and syntax of tutors and their pupils in 43 zebra finches (Z; *Taeniopygia guttata*; 22 tutors, 21 pupils) from 3 colonies (26 from Columbia, 13 from Tufts, and 4 from UC San Francisco) and 31 long-tailed finches (L; *Poephila acuticauda*; 10 tutors, 21 pupils) from our colony at Columbia. **Figure 1a** shows representative spectrograms of a tutor song and a pupil song for each species (Z, top; L, bottom). As reported by others[14,26,60-63], Z songs were composed of broadband syllables with prominent harmonics organized into sequences of each syllable type to form motifs, which are repeated multiple times in a singing bout (*e.g.* ABCDABCDABCD). Also as previously reported[24,31,32], L songs began with short duration, broadband syllables, which were followed by more tonal mid-duration syllables and then long, frequency-modulated syllables, with sequences including repeats of the same syllable (e.g. AAABBBBBBCCCD). **Figure 1b** shows syllable transition diagrams for the songs shown in **Fig. 1a**. The Z transition diagrams show the same syllable sequences in tutor and pupil song. The L transition diagrams show the same syllable order but different repeats of the same syllable type; the tutor repeats syllable C and the pupil repeats syllable D.

To group syllables by type (*e.g.* A, B and C), we conducted PCA of 26 acoustic features measured from each syllable[42,43], described in detail in[7,38] (**Fig. 1c**). The full list of acoustic features and PC loadings are shown in **Supplementary Table S1** and **Supplementary Table S2**. **Figure 1c** shows the projected syllable type clusters for the songs in **Fig. 1a** and **b**. Syllables of different types formed separate clusters while the same types shared by tutors and pupils formed overlapping clusters. Syllable type ellipses showed no overlap for 79.4% of types, indicating clear separability of syllable types based on acoustic features. Tutor and pupil versions of the same syllable type overlapped completely or partially in 83.7% of syllable types, indicating highly similar acoustic features. Spectrogram cross-correlation confirmed the classification of syllables into types using PCA (**Fig. 1d-e**)[7,24,64-67]. **Figure 1d** shows spectrogram cross-correlation heatmaps for the songs shown in **Fig. 1a**, with warm colors showing greater similarity. **Figure 1e** boxplots show the spectrogram cross-correlation similarity scores for pupil syllables of the same type (same), pupil and tutor syllables of the same type (copy), and pupil syllables of different types, across all birds. The acoustic similarity of the same syllable type within an individual's song (same), between tutor and pupil songs (copy), and different syllable types (different) differed across Z birds (one-way ANOVA; $F_{2,75} = 297.57$; $p = 2\times10^{-36}$) and L birds ($F_{2,60} = 398.35$; $p = 2\times10^{-35}$). The acoustic similarity of same-type syllables was highest within a bird (Z: 0.79 ± 0.07; L: 0.75 ± 0.07; mean ± SD) and high between pupils and tutors (Z: 0.63 ± 0.07; L: 0.62 ± 0.08; mean ± SD), but significantly lower than within a bird's song in both species (Z same vs. copy: $p = 3\times10^{-11}$; L: $p = 1\times10^{-6}$). The acoustic similarity of different syllable types was low (Z: 0.33 ± 0.07; L: 0.18 ± 0.05; mean ± SD), and significantly different from that of the same syllable type within a bird (Z: $p = 5\times10^{-29}$; L: $p = 9\times10^{-29}$) and the same syllable type shared by tutors and pupils (Z: $p = 1\times10^{-20}$; L: $p = 9\times10^{-24}$). The results of acoustic analyses showed that all pupils reliably produced syllables of distinct types and accurately copied syllables from tutors (**Fig. 1e**).

Syllable acoustics differ with position in a song motif

We derived species rules for song syntax by correlating syllable acoustics and sequence position. **Figure 2a** shows the values of three out of twenty-six acoustic features (duration, spectral entropy, and mean dominant frequency) across syllables in each sequence position, in Z motifs (left) and L motifs (right). Each of the three features showed a significant difference across sequence positions using Bonferroni-corrected one-way ANOVAs (corrected $\alpha = 0.0019$), in Z motifs (duration: $F_{4,188} = 12.97$; $p = 2\times10^{-9}$; spectral entropy: $F_{4,188} = 5.34$; $p = 4\times10^{-4}$; mean dominant frequency: $F_{4,188} = 5.59$; $p = 3\times10^{-4}$) and L motifs (duration: $F_{9,295} = 27.73$; $p = 1\times10^{-34}$; spectral entropy: $F_{9,295} = 3.63$; $p = 3\times10^{-4}$; mean dominant frequency: $F_{9,295} = 19.90$; $p = 5\times10^{-26}$). We used Tukey-Kramer pairwise comparisons to assess feature differences between specific sequence positions. Across Z motif positions, duration differed between positions 1 and 2-5 (all $p < 5\times10^{-4}$), spectral entropy differed between positions 1 and 2 ($p = 0.033$) and between positions 2 and 4-5 (both $p < 0.012$), and mean dominant frequency differed between positions 1 and 2 ($p = 0.035$), between positions 1 and 5 ($p = 5\times10^{-5}$), and between positions 4 and 5 ($p = 0.037$). Across L motif positions, duration differed between positions 1-3 and 8-10 (all $p < 0.023$), spectral entropy differed between positions 1-3 and 8 (all $p < 0.039$) and between 8 and 10 ($p = 0.049$), and mean dominant frequency differed between positions 1-4 and 7-10 (all $p < 0.0039$). Of the total 26 features measured, 16 showed at least one significant difference across positions in Z song (61.5%), and 21 showed at least one significant difference across positions in L song (80.8%; Bonferroni-corrected one-way ANOVAs; detailed results are shown in **Supplementary Table S3** and **Supplementary Table S4**). Results indicated that in both species' songs, the measured syllable acoustics differed significantly over the course of syllable sequences.

Using the species-specific relationships between syllable acoustics and position in the sequence, we constructed a "species rules" model of song syntax for each species. **Figure 2b** shows a spectrogram of a Z bird's motif with five syllables in positions one through five and values of three min-max re-scaled acoustic features (duration, spectral entropy and mean dominant frequency) at each position. To create a syllable feature-position vector for every acoustic feature, we averaged the re-scaled values of each feature at each motif position across all motifs within bird, and then across all conspecific birds. We then organized the feature-position vectors into a matrix to get the "species rules." This analysis yielded position-weight matrices like those used to predict genetic sequences or word order in speech[68,69]. **Figure 2c** shows the species rules matrices for Z song (left) and L song (right). Consistent with previous reports, the Z species rules predicted long-duration syllables at motif ends, high-entropy syllables at the beginning positions, and high-frequency syllables in middle positions and end positions[19,26]. The L species rules predicted a similar trajectory of syllable features, but with gradual changes from short to long duration, high to low entropy, and low to high frequency across a sequence, as reported previously[7,24,31,32]. The procedure used to generate species rules models was identical for both species, requiring only the measurement of syllable features and the positions of syllables in sequences.

Prediction of syllable sequences from species rules and quantification of sequence match

To test the hypothesis that song syntax can be predicted by models of species-specific relationships between syllable acoustics and order, we used the species rules model for each species to predict the syllable sequences of unrelated birds' songs. Using the species rules, we calculated sequence position alignments for each syllable type in a pupil's song and used those alignments to predict the syllable sequence in that bird's song (**Fig. 3**). **Figure 3a** illustrates the result of multiplying a matrix of one sequence of one pupil's syllables (left) and their features with the species rules matrix (middle). The product is a syllable-by-position matrix (right) that quantifies the alignment between the acoustic feature vectors for each syllable and each position. We calculated a matrix of position alignments for each recorded syllable, for each bird. The mean ± SD for the number of times each syllable type was represented in syllable position alignment matrices was 30.2 ± 11.3 for Z birds and 32.4 ± 16.5 for L birds.

For each syllable type for each bird, we then generated an alignment curve that quantified the alignment of that syllable type in a bird's song with each possible sequence position, according to the species rules (**Fig. 3b**). Syllable sequence predictions resulted from alignment curves by the assignment of syllables to their highest open positions (red markers in **Fig. 3b**) in order starting with the syllable with the highest alignment with any position. In the example in **Fig. 3**, there are 5 positions in the sequence. Syllables A, B, and D were each assigned their highest-alignment positions (1, 2 and 4). Because syllable E's highest position was already taken (position 1), it was assigned its highest open position, 3. The same rule resulted in the placement of syllable C in position 5. We used this procedure for all syllable position assignments for each sequence from each bird until all positions were filled, resulting in a predicted syllable sequence (ABEDC; **Fig. 3c**).

We tested the accuracy of using species rules to predict syllable sequences by comparing the syllable sequences of predicted and actual motifs. For this, we used the Levenshtein distance metric[53-56,58], in which each syllable that did not match between predicted and actual sequences resulted in a cost (**Fig. 3c**). In the example shown in **Fig. 3c**, the bird sings ABCDE while species rules predicted the sequence ABEDC. Syllables A, B and D were correctly predicted. Syllables C and E were incorrectly swapped in the prediction, resulting in a total weighted cost of 1.22 (0.61 + 0.61). The sequence match between the actual and predicted sequences was therefore calculated to be 0.76. For each prediction of a pupil's syllable sequence, we excluded the tutor's song and the songs of related birds (relatedness > 0.125) from the species rules model.

Species rules accurately predict song syntax

The accuracy of syllable sequences predicted by species rules and by tutor song did not differ (**Fig 4**). **Figure 4a** shows spectrograms of actual motifs recorded from two Z pupils (top) and two L pupils (bottom). Below the actual motifs are motifs generated using species rules, tutors' motifs, and motifs constructed by randomly assigning pupils' syllables to sequence positions. Additional examples are shown in **Supplementary Fig. S2**.

Comparisons of sequence matches to pupils' actual motifs across birds showed that species rules were as accurate at predicting normal Z and L pupil sequences as were actual tutor sequences. **Figure 4b** shows quantification of sequence matches of species rules predictions, tutor motifs and random pupil syllable sequences, across all Z pupils (nested ANOVA $F_{2,20} = 19.55$; $p = 1\times10^{-6}$) and L pupils ($F_{2,20} = 12.41$; $p = 6\times10^{-5}$). Matches of Z species rules predictions and Z tutor motifs to pupil motifs were significantly better than matches of Z random sequences (species vs. random: $p = 1\times10^{-5}$; tutor vs. random: $p = 7\times10^{-6}$). Matches of Z species rules predictions and Z tutor motifs to pupil motifs did not differ ($p = 0.98$). Matches of L species rules predictions and L tutor motifs to pupil motifs were significantly better than those of L random sequences (species vs. random: $p = 3\times10^{-4}$; tutor vs. random: $p = 3\times10^{-4}$). Matches of L species rules predictions and L tutor motifs to pupil motifs did not differ ($p = 0.99$).

The species rules model trained on the songs of birds from one colony accurately predicted the songs of birds from two other colonies, in different regions of the country. **Figure 4c** shows quantification of sequence matches of species rules predictions based only on Colony 1 data, tutor motifs and random pupil syllable sequences, for Z pupils from Colonies 2 and 3 (nested ANOVA $F_{2,7} = 11.05$; $p = 0.0013$). Matches of Colony 1 species rules predictions and Colonies 2 and 3 tutor motifs to Colonies 2 and 3 pupil motifs were significantly better than matches of pupil's random sequences (Colony 1 vs. random: $p = 0.015$; tutor vs. random: $p = 0.0012$). Matches of Colony 1 species rules predictions and Colonies 2 and 3 tutor motifs to pupil motifs did not differ ($p = 0.41$). Results demonstrate that the model captured the syntax rules of the species and not colony-specific dialects[70].

Because sequence matches were worse overall for L songs compared to Z songs, we considered the possibility that greater motif-to-motif variability in L songs may have lowered prediction accuracy for that species. We tested this hypothesis by quantifying within-bird motif-to-motif similarity for tutors and pupils using the same sequence match measure (**Fig. 4d**). Within-individual motif similarity differed between Z and L song (ANOVA $F_{3,71} = 27.76$; $p = 5\times10^{-12}$). In both tutors and pupils, Z song motifs were more consistent within an individual than were L song motifs (all $p < 1\times10^{-4}$). Motifs were equally consistent between tutors and pupils, in either species ($p > 0.51$). Results showed that L songs had greater motif-to-motif variability than Z songs.

Actual motifs and those predicted by species rules rarely included repeats of the same syllable type in Z song, but often included repeats of the same syllable type in L song. Because syllable repeat probability was not used in calculating species rules, we hypothesized that syllable repetition was a sequence feature that emerged from syntax species rules. To test this, we compared syllable repeat probabilities in actual syllable sequences and those predicted by species rules (**Fig. 4e**). In Z song, syllable repeat probabilities were low in actual (mean: 0.0025) and predicted (mean: 0.00023) sequences, and did not differ (paired *t*-test: $t_{20} = 1.00$; $p = 0.33$). In L song, syllable repeat probabilities were higher in actual (mean: 0.65) than in predicted (mean: 0.56) sequences ($t_{20} = 4.78$; $p = 1\times10^{-4}$). Therefore, the species rules model predicted syllable repeats in L song, but with a lower probability than in actual songs (**Fig. 4e**). The large species differences and within-species accuracy of species rules predictions suggest that repeating the same syllable type is a trait that emerges from species rules of song syntax.

Discussion

By relating song syllable acoustics to sequential position across many individuals, we generated a computational model of “species rules” that predicts a bird’s song syntax, without using the tutor’s song to train or test the model. Syllable sequences predicted by species rules models matched pupils’ actual sequences as well as did tutors’ sequences, across birds of two species with different song syntax, and across birds of the same species from different colonies. Our findings show that a model of syntax rules based on a species’ unique correlations between syllable acoustics and sequence position can accurately predict an individual’s song.

Birdsong presents unique challenges and opportunities for the computational study of sequencing processes, including natural language processing tasks. Language processing tasks are aided by knowledge of lexicons that describe the shared vocabulary of same-language users and the statistical relations between words, such as the probability of one word following another[1,2]. One challenge presented by birdsong is that syllables are learned by pupils from tutors and therefore syllable types are often unique to families[71,72]. Our solution to this issue was to use a wide array of measured acoustic features for each syllable for training the species rules models, instead of using symbolic representations (*e.g.* ABCD)[33,38]. Symbolic representations of syllables were used only when generating and testing the accuracy of predicted sequences. A second, related challenge is that statistical relations between syllable types (*e.g.* the probability of B following A) are individual-specific because syllables A and B may be different across individuals[3,4,10,73]. Our species rules models overcame this challenge by capturing correlations between syllable acoustics and sequence positions across the songs of many individuals.

Because our computational strategy relied only on the identification of vocal units and sequence positions, the approach reported here has general utility for the detection and analysis of sequential structure in complex audio signals, including speech or other animal vocalizations[34,53,55,56]. Understanding sequential structure in vocalizations is of interest to behavioral neuroscientists, ecologists and evolutionary biologists, machine learning researchers and quantitative linguists[34]. For example, comparisons of the vocalizations of juvenile or experimentally manipulated songbirds to species-typical vocalizations have been used for decades to understand mechanisms of auditory-vocal plasticity[56,74]. Studies of changes in individual humpback whales’ song sequences have revealed rapid, population-level changes in social exchanges akin to cultural “revolutions”[75]. The advancement of large-language models that predict word sequences are transforming how we handle social information as humans, in large part because they convincingly emulate our language rules[2]. One benefit of a “species rules”-like model developed here is that it provides an “expected value” for which to compare observed sequences[59,76]. As shown here, by quantifying the match to an expected sequence (species rules), researchers can quantify the extent to which an observed sequence is similar to or deviates from a hypothesized sequence. Additionally, we used the same procedure to quantify the variability of sequences produced by individuals, finding that differences in species rules prediction accuracy correlated with species-level (Z versus L) within-individual variability. Such metrics could be used to quantify and compare complexity in vocal sequences measured from animals of any species[34], provided that vocal acoustics can be measured and sequence positions distinguished.

The use of species “rules” or “universals” is supported by the finding of species-typical syntax in the songs of pupils even when tutor syntax information is missing or abnormal. Zebra finches tutored with random arrangements of syllables ABCDE produced adult songs with the same syllables arranged non-randomly in sequences that followed rules similar to those that we found; long-duration syllables were statistically more likely to be produced at the end of motifs, for example[26]. Canaries tutored with glissandos of syllables eventually arranged a subset of those syllables into phrases of trills, the syntax typical of their species but that which they had never heard[25]. The phrased songs of white-crowned sparrows always start with a whistle, whether birds are tutored, untutored, or tutored with phrases containing no whistles[28,77]. The species-specific whistle is such a strong acoustic-syntactic cue that two of eight birds studied by Soha and Marler [29] incorporated squirrel calls into their songs when tutored by those calls attached to whistles. Our study adds to this work by developing, testing and validating a generalizable computational model of syntax using statistical relationships between syllable acoustics and sequence position.

Our approach to modeling song syntax differs from previous studies, such as those that have modeled the complex song syntax of Bengalese finches[73,78]. Bengalese finch song consists of frequent repeats of the same syllable type and probabilistic transitions from one syllable type to another[50,51,79]. Despite their complex syntax, the use of Markov models with hidden states has shown that Bengalese finch song can be described with comparatively simple, first- or second-order transition models, such that the next element in a sequence can be predicted by one or two elements in the past[73,78]. Similar to us, the authors trained computational models of song syntax using relationships between syllable acoustics and order. In contrast to us, their models of song syntax were specific to the syllable transition patterns found in the songs of each individual, though common sequence structures were found across individuals[73]. Our work complements these studies in that we first modeled the species-specific song syntax found across the songs of many conspecifics, and then used those models to predict the specific syllable sequences of unrelated individuals.

Short- and long-range dependencies may be one set of mechanisms by which species rules are used by neural processes for syllable sequence production. Short-range dependencies may result from moment-to-moment auditory feedback while the bird is singing. For example, hearing a low-frequency syllable A may trigger the production of a high-frequency syllable B. Evidence for this type of process comes from the rapid degradation of song syntax following the loss of auditory feedback by deafening in Bengalese finches[74]. Sensitivity to temporal order emerges in the songbird auditory cortex[80] and neurons in the higher-order auditory cortex (caudomedial nidopallium; NCM) are sensitive to syllable order and temporal organization on the order of milliseconds[81]. However, it has been shown in zebra finches, white-crowned sparrows and canaries that auditory input to the premotor nucleus HVC, where song sequence is encoded[82], is inhibited while a bird is singing[83]. Long-range dependencies in song syntax have been shown in other species that sing long, phrased songs. In song sparrows that sing multiple song types, the observed cycling of song types is consistent with a dependency on song types that were sung up to 30 minutes prior[84]. In canaries, HVC neurons encode information about which phrases occurred up to four seconds and up to forty syllables in the past[85]. Short- or long-range sensorimotor dependencies may be a general mechanism by which animals actively guide behaviors through sequences[59,86].

Neural processes that set “macrostructural” parameters of behavior may also influence how species rules contribute to vocal sequences. For example, species-specific limits on the length of song bouts[84] or motifs[7] could interact with syntax-generating mechanisms. Neurons of the songbird anterior forebrain pathway (AFP) are involved in song learning[87-89] and song initiation and termination[90]. Lesions and gene knockdowns in an AFP region in adult zebra and Bengalese finches do not affect syllable structure but disrupt syntax by altering song initiation and termination, and dramatically increase syllable repetition[90-92]. One possibility is that the AFP sets a parameter like “the number of syllables in a sequence”. Our modeling and prediction procedure had two constraints: 1) each syllable in a bird’s song must be used once, and 2) the number of positions in each sequence was known. In our data, Z birds often had the same number of syllable types as sequence positions, so syllables did not repeat. By contrast, L birds had fewer syllable types than sequence positions, forcing syllable types to appear more than once. Our models predicted that L syllables would be repeated (AAABBBCCC) and not arranged randomly or cyclically (ABCABCABC). Species-specific macrostructural rules, in addition to those that govern syllable-to-syllable transitions, are supported by experiments in which cross-species tutored birds' songs were composed of heterospecific syllables with the motif structure of conspecifics[7,23,93].

Our finding that computational models of older birds' song sequences accurately predicted song syntax in younger birds of two species suggests that neural systems subserving song production and perception are adapted to the species-specific organization of conspecific song[8]. Across animal taxa, communication signals evolve under the constraints and biases of species-specific sensory and motor capacities[94], including the temporal and sequential patterning of acoustic elements in vocalizations[31,34,95,96]. In many species studied, including frogs, bats, and songbirds, temporal tuning of auditory neurons matches the temporal properties of behaviorally relevant sounds[80,97-99]. In songbirds, manipulations of neural structures suggest that syllables and syllable order are separately processed[91,92,100-104]. Neural systems for complex sequenced behaviors may adaptively balance flexibly-learned elements with species-specific sequencing rules.

Data Availability Statement

The analysis code and datasets used here are available from the corresponding author on request.

References

1. Cambria, E. & White, B. Jumping NLP curves: a review of natural language processing research. *IEEE Comput. Intell. Mag.* **9**, 48-57 (2014).
2. Min, B. *et al.* Recent advances in natural language processing via large pre-trained language models: a survey. *ACM Comput. Surv.* **56**, 1-40 (2023).
3. Berwick, R. C., Okanoya, K., Beckers, G. J. L. & Bolhuis, J. J. Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* **15**, 113-121 (2011).
4. James, L. S., Sun, H., Wada, K. & Sakata, J. T. Statistical learning for vocal sequence acquisition in a songbird. *Sci. Rep.* **10**, 2248 (2020).

5 Lewis, R. N., Soma, M., de Kort, S. R. & Gilman, R. T. Like father like son: cultural and genetic contributions to song inheritance in an estrildid finch. *Front. Psychol.* **12**, 654198 (2021).

6 Menyhart, O., Kolodny, O., Goldstein, M. H., DeVoogd, T. J. & Edelman, S. Juvenile zebra finches learn the underlying structural regularities of their fathers' song. *Front. Psychol.* **6**, 571 (2015).

7 Edwards, J. A., Rivera, M. & Woolley, S. M. N. The temporal organization of learned vocal behavior is predicted by species rather than experience. *J. Neurosci.* **45**, e0576242025 (2025).

8 Marler, P. A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.* **71**, 1-25 (1970).

9 Liu, W. C., Gardner, T. J. & Nottebohm, F. Juvenile zebra finches can use multiple strategies to learn the same song. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 18177-18182 (2004).

10 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).

11 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).

12 Marler, P. & Tamura, M. Culturally transmitted patterns of vocal behavior in sparrows. *Science* **146**, 1483-1486 (1964).

13 Sturdy, C. B., Phillmore, L. S. & Weisman, R. G. Note types, harmonic structure, and note order in the songs of zebra finches (*Taeniopygia guttata*). *J. Comp. Psychol.* **113**, 194-203 (1999).

14 Zann, R. Structure, sequence and evolution of song elements in wild Australian zebra finches. *Auk* **110**, 702-715 (1993).

15 Hinde, R. A. *Bird vocalizations: their relation to current problems in biology and psychology*. (Cambridge University Press, 1969).

16 Love, J., Hoepfner, A. & Goller, F. Song feature specific analysis of isolate song reveals interspecific variation in learned components. *Dev. Neurobiol.* **79**, 350-369 (2019).

17 Searcy, W. A., Soha, J., Peters, S. & Nowicki, S. Variation in vocal production learning across songbirds. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200257 (2021).

18 Thorpe, W. H. The learning of song patterns by birds, with special reference to the song of the chaffinch, *Fringilla coelebs*. *Ibis* **100**, 535-570 (1958).

19 James, L. S., Davies, J., R., Mori, C., Wada, K. & Sakata, J. T. Manipulations of sensory experiences during development reveal mechanisms underlying vocal learning biases in zebra finches. *Dev. Neurobiol.* **80**, 132-146 (2020).

20 Price, P. H. Developmental determinants of structure in zebra finch song. *J. Comp. Physiol. Psychol.* **93**, 260-277 (1979).

21 Williams, H., Kilander, K. & Sotanski, M. L. Untutored song, reproductive success and song learning. *Anim. Behav.* **45**, 695-705 (1993).

22 Marler, P. & Sherman, V. Innate differences in singing behaviour of sparrows reared in isolation from adult conspecific song. *Anim. Behav.* **33**, 57-71 (1985).

23 Clayton, N. S. The effects of cross-fostering on selective song learning in estrildid finches. *Behaviour* **109**, 163-175 (1989).

24 Moore, J. M. & Woolley, S. M. N. Emergent tuning for learned vocalizations in auditory cortex. *Nat. Neurosci.* **22**, 1469-1476 (2019).

25 Gardner, T. J., Naef, F. & Nottebohm, F. Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science* **308**, 1046-1049 (2005).

26 James, L. S. & Sakata, J. T. Learning biases underlie "universals" in avian vocal sequencing. *Curr. Biol.* **27**, 3676-3682 (2017).

27 Peters, S., Soha, J., Searcy, W. A. & Nowicki, S. Are song sequencing rules learned by song sparrows? *Anim. Behav.* **192**, 75-84 (2022).

28 Rose, G. J. *et al.* Species-typical songs in white-crowned sparrows tutored with only phrase pairs. *Nature* **432**, 753-758 (2004).

29 Soha, J. A. & Marler, P. A species-specific acoustic cue for selective song learning in the white-crowned sparrow. *Anim. Behav.* **60**, 297-306 (2000).

30 Wohlgemuth, M. J., Sober, S. J. & Brainard, M. S. Linked control of syllable sequence and phonology in birdsong. *J. Neurosci.* **30**, 12936-12949 (2010).

31 Woolley, S. M. N. & Moore, J. M. Coevolution in communication senders and receivers: vocal behavior and auditory processing in multiple songbird species. *Ann. N. Y. Acad. Sci.* **1225**, 155-165 (2011).

32 Zann, R. A. Variation in the songs of three species of estrildine grassfinches. *Emu* **76**, 97-108 (1976).

33 Sainburg, T. & Gentner, T. Q. Toward a computational neuroethology of vocal communication: from bioacoustics to neurophysiology, emerging tools and future directions. *Front. Behav. Neurosci.* **15**, 811737 (2021).

34 Kershenbaum, A. *et al.* Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biol. Rev. Camb. Philos. Soc.* **91**, 13-52 (2016).

35 Yeh, Y., Rivera, M. & Woolley, S. M. N. Auditory sensitivity and vocal acoustics in five species of estrildid songbirds. *Anim. Behav.* **195**, 107-116 (2023).

36 Immelmann, K. in *Bird vocalisations* (ed R. A. Hinde) 61-74 (Cambridge University Press, 1969).

37 Slater, P. J. B., Eales, L. A. & Clayton, N. S. Song learning in zebra finches (*Taeniopygia guttata*): progress and prospects. *Adv. Study Behav.* **18**, 1-34 (1988).

38 Rivera, M., Edwards, J. A., Hauber, M. E. & Woolley, S. M. N. Machine learning and statistical classification of birdsong link vocal acoustic features with phylogeny. *Sci. Rep.* **13**, 7076 (2023).

39 Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B. & Mitra, P. P. A procedure for an automated measurement of song similarity. *Anim. Behav.* **59**, 1167-1176 (2000).

40 Kao, M. H. & Brainard, M. S. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *J. Neurophysiol.* **96**, 1441-1455 (2006).

41 Zevin, J. D., Seidenberg, M. S. & Bottjer, S. W. Limits on reacquisition of song in adult zebra finches exposed to white noise. *J. Neurosci.* **24**, 5849-5862 (2004).

42 Araya-Salas, M. & Smith-Vidaurre, G. warbleR: an R package to streamline analysis of animal acoustic signals. *Methods Ecol. Evol.* **8**, 184-191 (2016).

43 Sueur, J., Aubin, T. & Simonis, C. Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics* **18**, 213-226 (2008).

44 Keen, S. C. *et al.* A machine learning approach for classifying and quantifying acoustic diversity. *Methods Ecol. Evol.* **12**, 1213-1225 (2021).

45 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**, 564-568 (2009).

46 Matheson, A. M. M. & Sakata, J. T. Relationship between the sequencing and timing of vocal motor elements in birdsong. *PLoS One* **10**, e0143203 (2015).

47 Mets, D. G. & Brainard, M. S. An automated approach to the quantitation of vocalizations and vocal learning in the songbird. *PLoS Comput. Biol.* **14**, e1006437 (2018).

48 Sakata, J. T. & Brainard, M. S. Real-time contributions of auditory feedback to avian vocal motor control. *J. Neurosci.* **26**, 9619-9628 (2006).

49 Warren, T. L., Charlesworth, J. D., Tumer, E. C. & Brainard, M. S. Variable sequencing is actively maintained in a well learned motor skill. *J. Neurosci.* **32**, 15414-15425 (2012).

50 Okanoya, K. Song syntax in Bengalese finches: proximate and ultimate analyses. *Adv. Study Behav.* **34**, 297-346 (2004).

51 Takahasi, M., Yamada, H. & Okanoya, K. Statistical and prosodic cues for song segmentation learning by Bengalese finches (*Lonchura striata* var. *domestica*). *Ethology* **116**, 481-489 (2010).

52 Sossinka, R. & Böhner, J. Song types in the zebra finch *Poephila guttata castanotis*. *Z. Tierpsychol.* **53**, 123-132 (1980).

53 Garland, E. C. *et al.* Improved versions of the Levenshtein distance method for comparing sequence information in animals' vocalisations: tests using humpback whale song. *Behaviour* **149**, 1413-1441 (2012).

54 Kruskal, J. B. An overview of sequence comparison: time warps, string edits, and macromolecules. *SIAM Rev. Soc. Ind. Appl. Math.* **25**, 201-237 (1983).

55 Gil, D. & Slater, P. J. B. Song organisation and singing patterns of the willow warbler, *Phylloscopus trochilus*. *Behaviour* **137**, 759-782 (2000).

56 Margoliash, D., Staicer, C. A. & Inoue, S. A. Stereotyped and plastic song in adult indigo buntings, *Passerina cyanea*. *Anim. Behav.* **42**, 367-388 (1991).

57 Haldar, R. & Mukhopadhyay, D. Levenshtein distance technique in dictionary lookup methods: an improved approach. *ArXiv* **1101.1232** (2011).

58 Yujian, L. & Bo, L. A normalized Levenshtein distance metric. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 1091-1095 (2007).

59 Slater, P. J. B. in *Perspectives in Ethology* (eds P. P. G. Bateson & P. H. Klopfer) Ch. 5, 131-153 (Plenum Press, 1973).

60 Goffinet, J., Brudner, S., Mooney, R. & Pearson, J. Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. *eLife* **10**, e67855 (2021).

61 Mandelblat-Cerf, Y. & Fee, M. S. An automated procedure for evaluating song imitation. *PLoS One* **9**, e96484 (2014).

62 Theunissen, F. E., Sen, K. & Doupe, A. J. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.* **20**, 2315-2331 (2000).

63 Wang, H. *et al.* Transcriptional regulatory divergence underpinning species-specific learned vocalization in songbirds. *PLoS Biol.* **17**, e3000476 (2019).

64 Aronov, D., Andalman, A. S. & Fee, M. S. A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* **320**, 630-634 (2008).

65 Lahti, D. C., Moseley, D. L. & Podos, J. A tradeoff between performance and accuracy in bird song learning. *Ethology* **117**, 802-811 (2011).

66 Nelson, D. A. & Marler, P. Selection-based learning in bird song development. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10498-10501 (1994).

67 Peters, S., Derryberry, E. P. & Nowicki, S. Songbirds learn songs least degraded by environmental transmission. *Biol. Lett.* **8**, 736-739 (2012).

68 Bin Raies, A., Mansour, H., Incitti, R. & Bajic, V. B. Combining position weight matrices and document-term matrix for efficient extraction of associations of methylated genes and diseases from free text. *PLoS One* **8**, e77848 (2013).

69 Siddharthan, R. Dinucleotide weight matrices for predicting transcription factor binding sites: generalizing the position weight matrix. *PLoS One* **5**, e9722 (2010).

70 Wang, D. *et al.* Machine learning reveals cryptic dialects that explain mate choice in a songbird. *Nat. Commun.* **13**, 1630 (2022).

71 Catchpole, C. K. & Slater, P. J. B. *Bird Song: Biological Themes and Variations*. (Cambridge University Press, 2008).

72 Lipkind, D. *et al.* Songbirds work around computational complexity by learning song vocabulary independently of sequence. *Nat. Commun.* **8**, 1247 (2017).

73 Katahira, K., Suzuki, K., Okanoya, K. & Okada, M. Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLoS One* **6**, e24516 (2011).

74 Woolley, S. M. N. & Rubel, E. W. Bengalese finches *Lonchura striata domestica* depend upon auditory feedback for the maintenance of adult song. *J. Neurosci.* **17**, 6380-6390 (1997).

75 Garland, E. C. *et al.* Quantifying humpback whale song sequences to understand the dynamics of song exchange at the ocean basin scale. *Journal of the Acoustical Society of America* **133**, 560-569 (2013).

76 Lawrence, C. E. & Reilly, A. A. An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences. *Proteins: Structure, Function, and Genetics* **7**, 41-51 (1990).

77 Plamondon, S. L., Rose, G. J. & Goller, F. Roles of syntax information in directing song development in white-crowned sparrows (*Zonotrichia leucophrys*). *J. Comp. Psychol.* **124**, 117-132 (2010).

78 Jin, D. Z. & Kozhevnikov, A. A. A compact statistical model of the song syntax in Bengalese finch. *PLoS Comput. Biol.* **7**, e1001108 (2011).

79 Honda, E. & Okanoya, K. Acoustical and syntactical comparisons between songs of the white-backed munia (*Lonchura striata*) and its domesticated strain, the Bengalese finch (*Lonchura striata* var. *domestica*). *Zoolog. Sci.* **16**, 319-326 (1999).

80 Lim, Y., Lagoy, R., Shinn-Cunningham, B. G. & Gardner, T. J. Transformation of temporal sequences in the zebra finch auditory system. *eLife* **5**, e18205 (2016).

81 Schneider, D. M. & Woolley, S. M. N. Sparse and background-invariant coding of vocalizations in auditory scenes. *Neuron* **79**, 141-152 (2013).

82 Yu, A. C. & Margoliash, D. Temporal hierarchical control of singing in birds. *Science* **273**, 1871-1875 (1996).

83 McCasland, J. S. & Konishi, M. Interaction between auditory and motor activities in an avian song control nucleus. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7815-7819 (1981).

84 Searcy, W. A., Soha, J., Peters, S. & Nowicki, S. Long-distance dependencies in birdsong syntax. *Proc. R. Soc. Lond. B Biol. Sci.* **289**, 20212473 (2022).

85 Cohen, Y. *et al.* Hidden neural states underlie canary song syntax. *Nature* **582**, 539-544 (2020).

86 McCullough, M. H. & Goodhill, G. J. Unsupervised quantification of naturalistic animal behaviors for gaining insight into the brain. *Curr. Opin. Neurobiol.* **70**, 89-100 (2021).

87 Bottjer, S. W., Miesner, E. A. & Arnold, A. P. Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901-903 (1984).

88 Doya, K. & Sejnowski, T. J. A novel reinforcement model of birdsong vocalization learning. *Adv. Neural Inf. Process. Syst.* **7**, 101-108 (1995).

89 Doya, K. & Sejnowski, T. J. in *Central Auditory Processing and Neural Modeling* (eds P. F. Poon & J. F. Brugge) 77-88 (Plenum Press, 1998).

90 Xiao, L. *et al.* Expression of FoxP2 in the basal ganglia regulates vocal motor sequences in the adult songbird. *Nat. Commun.* **12**, 2617 (2021).

91 Kobayashi, K., Uno, H. & Okanoya, K. Partial lesions in the anterior forebrain pathway affect song production in adult Bengalese finches. *Neuroreport* **12**, 353-358 (2001).

92 Kubikova, L. *et al.* Basal ganglia function, stuttering, sequencing and repair in adult songbirds. *Sci. Rep.* **4**, 6590 (2014).

93 Eales, L. A. Do zebra finch males that have been raised by another species still tend to select a conspecific song tutor? *Anim. Behav.* **35**, 1347-1355 (1987).

94 Bradbury, J. W. & Vehrencamp, S. L. *Principles of Animal Communication*. (Sinauer Associates, 2011).

95 Brumm, H. & Slabbekoorn, H. Acoustic communication in noise. *Adv. Study Behav.* **35**, 151-209 (2005).

96 Gerhardt, H. C. The evolution of vocalization in frogs and toads. *Annu. Rev. Ecol. Syst.* **25**, 293-324 (1994).

97 Gooler, D. M. & Feng, A. S. Temporal coding in the frog auditory midbrain: the influence of duration and rise-fall time on the processing of complex amplitude-modulated stimuli. *J. Neurophysiol.* **67**, 1-22 (1992).

98 Suga, N., O'Neill, W. E. & Manabe, T. Cortical neurons sensitive to combinations of information-bearing elements of biosonar signals in the mustache bat. *Science* **200**, 778-781 (1978).

99 Woolley, S. M. N., Fremouw, T. E., Hsu, A. & Theunissen, F. E. Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.* **8**, 1371-1379 (2005).

100 Basista, M. J. *et al.* Independent premotor encoding of the sequence and structure of birdsong in avian cortex. *J. Neurosci.* **34**, 16821-16834 (2014).

101 Foster, E. F. & Bottjer, S. W. Lesions of a telencephalic nucleus in male zebra finches: influences on vocal behavior in juveniles and adults. *J. Neurobiol.* **46**, 142-165 (2001).

102 Hosino, T. & Okanoya, K. Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *Neuroreport* **11**, 2091-2095 (2000).

103 Koparkar, A. *et al.* Lesions in a songbird vocal circuit increase variability in song syntax. *eLife* **13**, RP93272 (2024).

104 Williams, H. & Vicario, D. S. Temporal patterning of song production: participation of nucleus uvaeformis of the thalamus. *J. Neurobiol.* **24**, 903-912 (1993).

Funding Declaration

This study was supported by a National Research Service Award (F31DC020904) to J.A.E. and grants from the US National Science Foundation (IOS-1656825) and National Institutes of Health (DC009810) to S.M.N.W.

Acknowledgements

We thank Mimi Kao for generously providing song recordings of zebra finches from the Tufts University and University of California, San Francisco colonies.

Author Contributions

J.A.E. and S.M.N.W. conceived and designed this study. J.A.E. and S.M.N.W. conducted the experiments and analyses. J.A.E. and S.M.N.W. wrote the manuscript text. J.A.E. and S.M.N.W. reviewed and edited the manuscript.

Competing Interest Statement

The authors have declared that no competing interests exist.

Figures

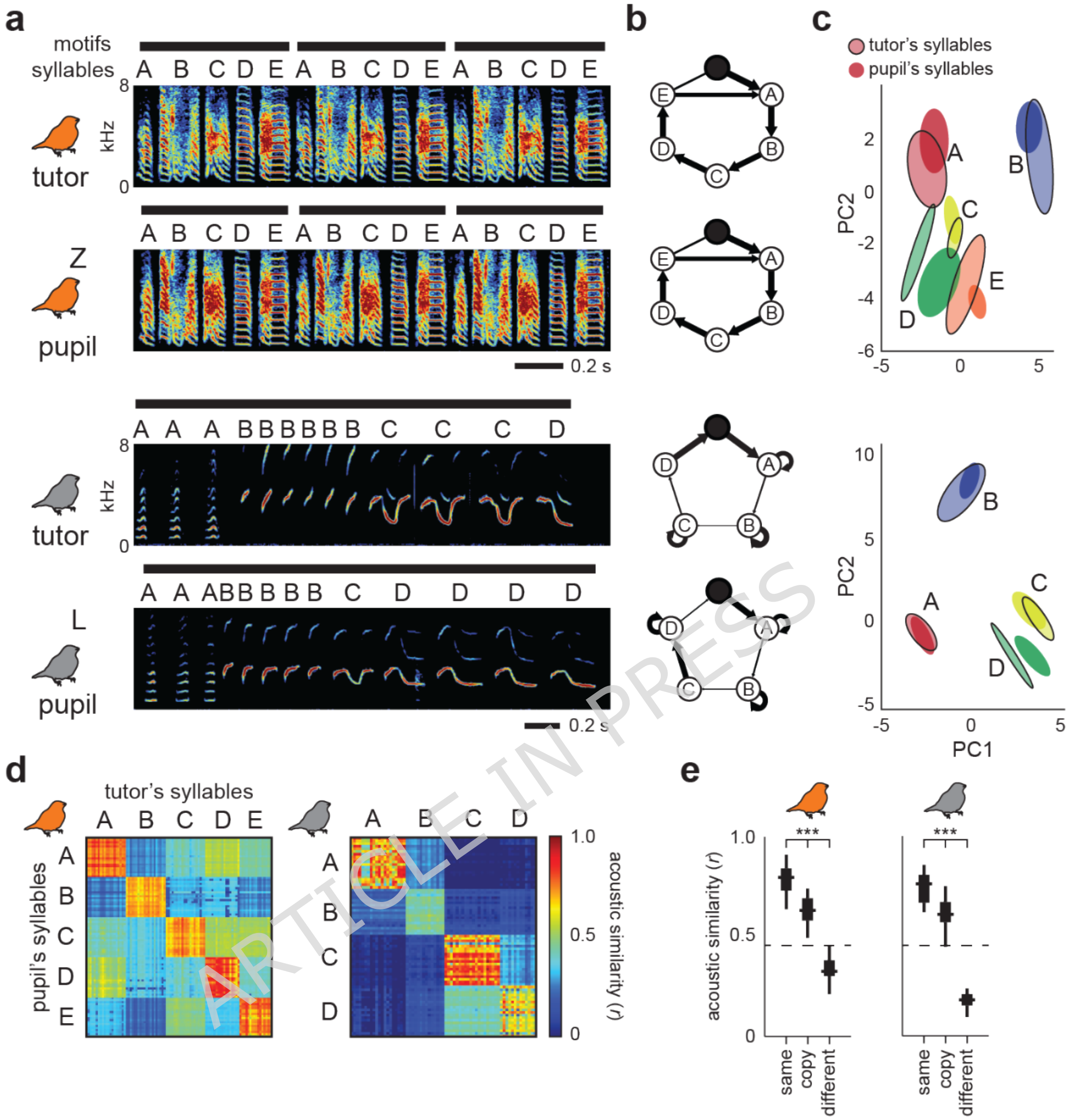

Figure 1. Pupil songs resemble tutor songs in syllable acoustics and sequence. **(a)** Spectrograms of the songs of a zebra finch tutor (orange; top) and pupil (orange; bottom), and the songs of a long-tailed finch tutor (gray; top) and pupil (gray; bottom). Syllables are labeled by letter and motifs are indicated by black bars above spectrograms. **(b)** Transition diagrams for the songs in (a). Syllable types are labeled by letter, transitions between two types are indicated by arrows, with the thickness of the arrow representing the probability of a transition. Syllable repeats are represented by an arrow looping back to the same type. The start/end of a song bout

is represented by a black circle. **(c)** Projections of tutors' (light ellipses) and pupils' (dark ellipses) syllables onto the first two principal components (PC1 & PC2) derived from 26 measured syllable features. Radii of ellipses extend ±1 SD from each syllable cluster's center. Syllable types are the same as shown in (a) and are indicated by letter and color. **(d)** Syllable similarity matrices showing spectrogram cross-correlation values (r) for pairs of each pupil's syllables (rows) and tutor's syllables (columns). Syllables are grouped and labeled by type. **(e)** Boxplots showing distributions of average similarity values for all pupil syllables labeled as the same type (same), all pupil syllables copied from tutors (copy), and all pupil syllables labeled as different types (different), across all birds of each species (Z $n = 22$; L $n = 21$). Dashed lines show the similarity cut-off value of 0.45. Center lines show medians, and boxes extend to the upper and lower quartiles. Whiskers extend to ±1.5×IQR. (***: $p < 0.001$)

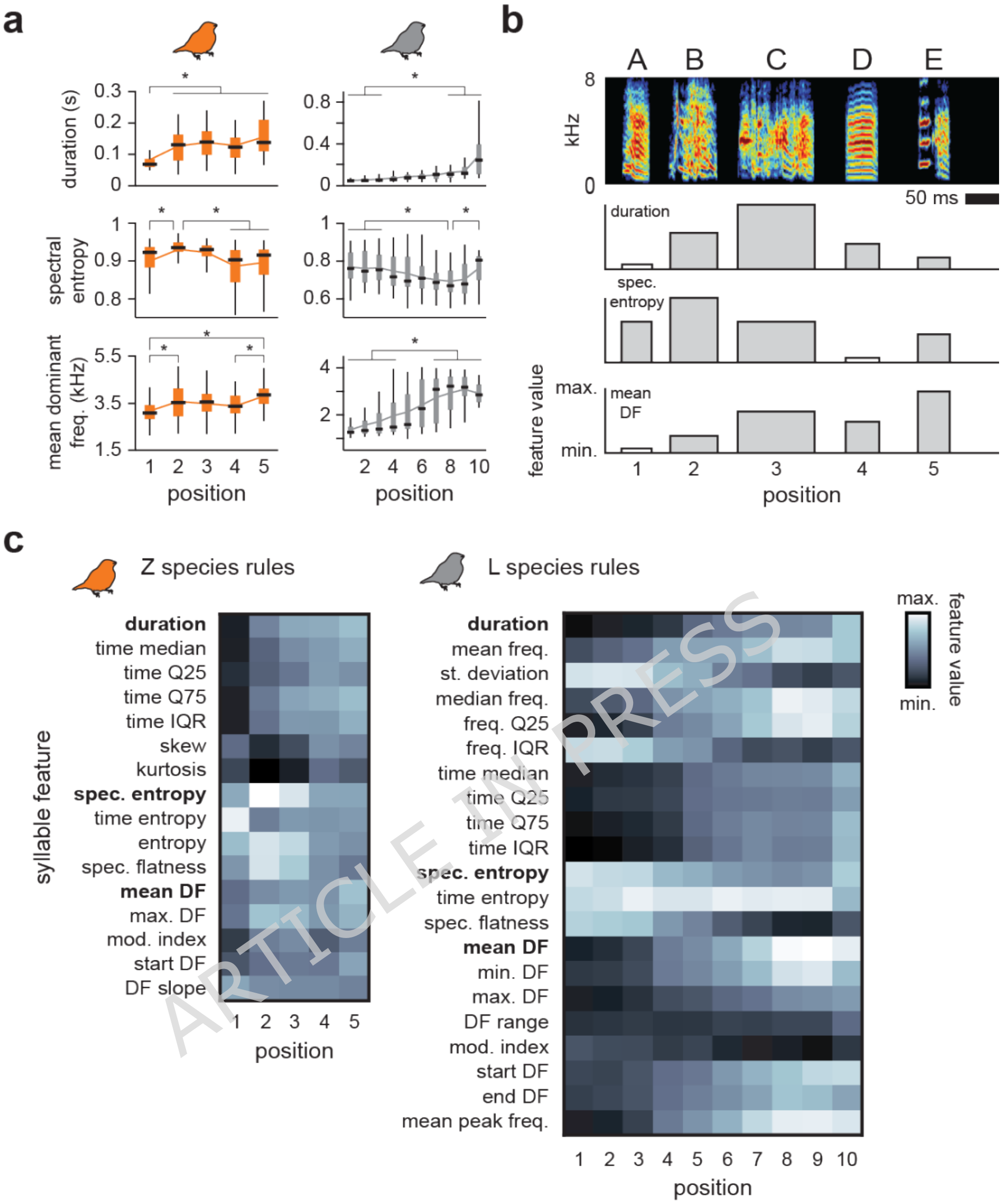

Figure 2. Generation of species rules based on the correlation between syllable acoustics and sequence position. **(a)** Box plots show acoustic differences by position for 3 of the 26 syllable features: duration, spectral entropy, and mean dominant frequency, in Z song over 5 positions (orange; left; $n = 43$) and L song over 10 positions (gray; right; $n = 31$). Lines track feature means across positions. Center lines show medians, and boxes extend to the upper and lower quartiles. Whiskers extend to ±1.5×IQR. (*: $p < 0.05$; n.s.: not significant; significance refers to Tukey-

Kramer pairwise comparisons between positions) **(b)** Spectrogram of a Z pupil’s motif, with min-max normalized values of syllable duration, spectral entropy, and mean dominant frequency below. **(c)** Species rules matrices derived from all tutor zebra finches (left) and long-tailed finches (right) songs. Only syllable features that significantly differed by position were included in each matrix. The names of features shown in examples (duration, spectral entropy, mean dominant frequency) are highlighted in bold. (Q25: 25th quartile; Q75: 75th quartile; IQR: inter-quartile range; DF: dominant frequency)

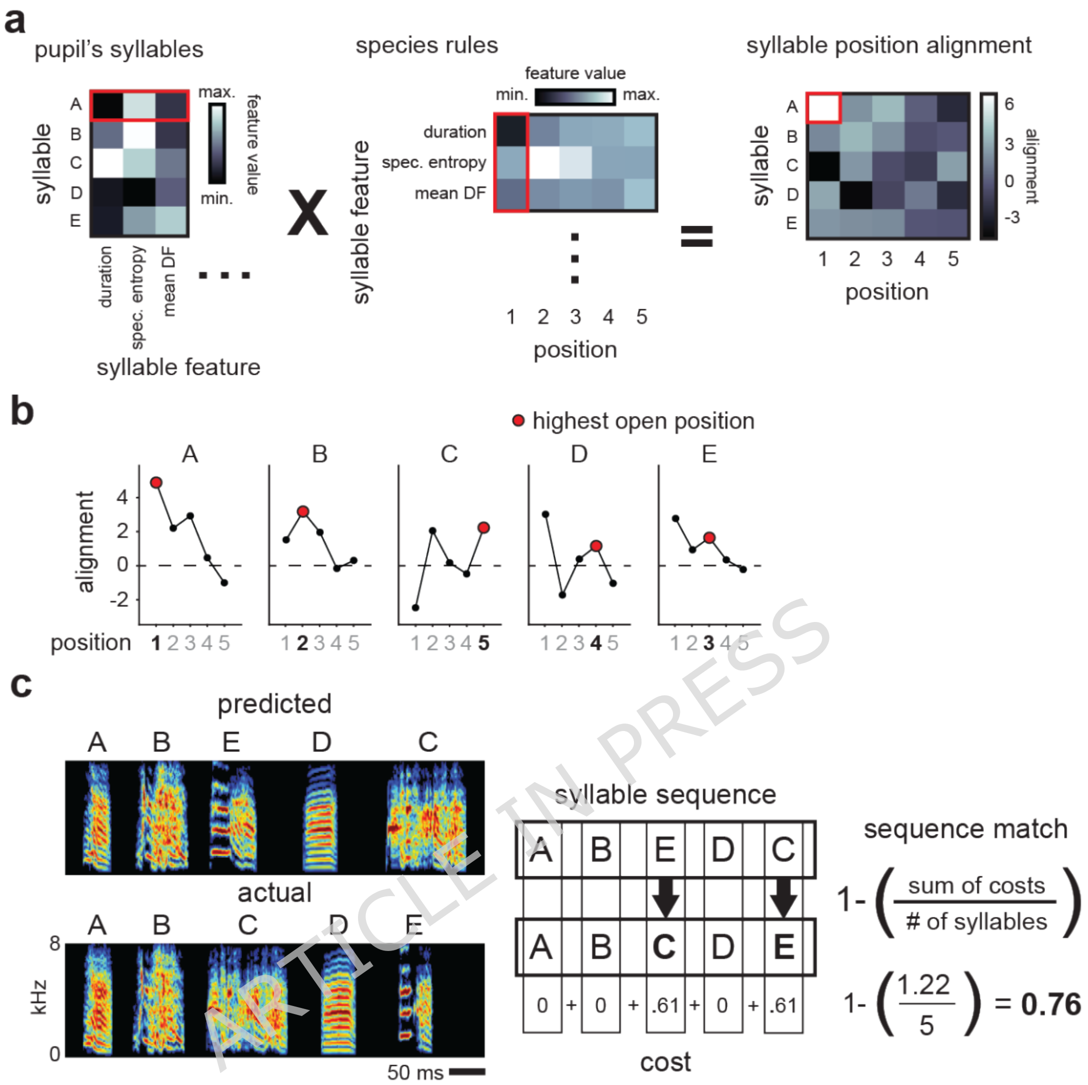

Figure 3. Pupil syllable sequences are predicted by species rules. **(a)** The features of a pupil's syllables (left matrix) multiplied by the species rules (middle matrix) yields the position alignment of each syllable in a pupil's motif (right matrix). Red outlines indicate the row × column multiplication that yields position alignment value for one rendition of syllable A. **(b)** Syllable position alignment for each syllable type shown in (a), averaged across all renditions of each syllable type (*e.g.* all renditions of syllable A). Syllables were assigned to open positions in order from highest to lowest alignment. **(c)** Spectrogram of a pupil's predicted motif showing the sequence of syllables predicted by species rules (left, top). Spectrogram of the pupil's actual motif

(left, bottom). The positions of syllables C and E do not match between the predicted and actual motifs. Next to the spectrograms are letter representations of the predicted and actual syllable sequences, the cost for each operation and the calculated sequence match.

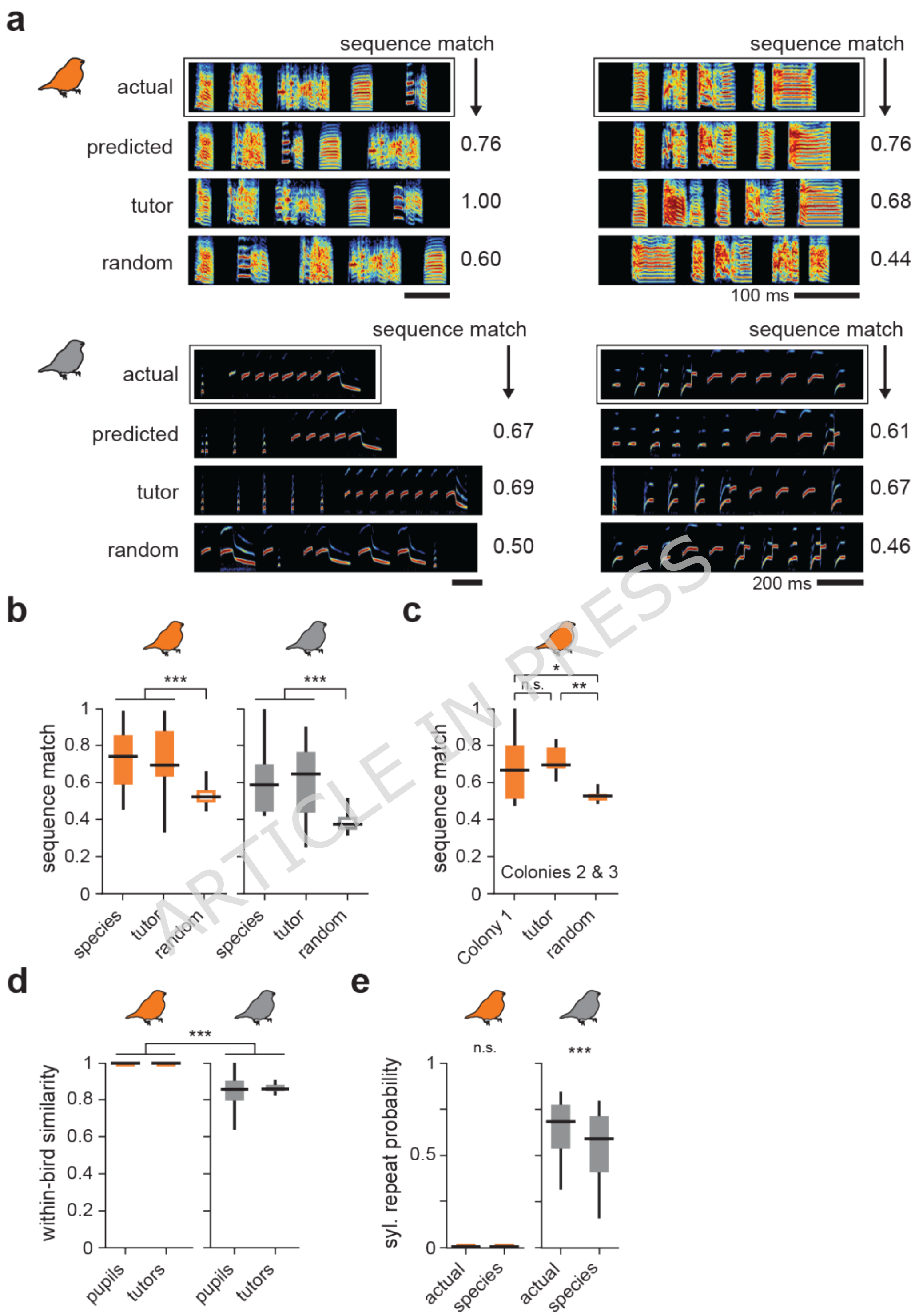

Figure 4. Species rules predict the order of syllables in pupils' motifs. **(a)** Spectrograms of two exemplar Z pupils' motifs (top) and L pupils' motifs (bottom), with spectrograms of pupils' predicted motifs based on species rules, tutors' motifs,

and motifs with pupil syllables in random order below. Actual motifs are surrounded by black boxes, and sequence match scores are to the right of each spectrogram. **(b)** Sequence matches between pupils' actual syllable sequences and sequences predicted using species rules, tutors' syllable sequences, and random arrangements of pupils' syllables for zebra finches (Z; left; $n = 21$) and long-tailed finches (L; right; $n = 21$). **(c)** Sequence matches between Colonies 2 and 3 pupils' actual syllable sequences and sequences predicted using species rules, tutors' syllable sequences, and random arrangements of pupils' syllables (Colonies 2 and 3 pupils; $n = 8$). Here, the species rules model was generated using only data from birds of Colony 1 ($n = 13$). **(d)** Within-bird motif-to-motif sequence similarity for Z and L pupils (Z $n = 21$; L $n = 21$) and tutors (Z $n = 22$; L $n = 10$), using the same sequence match metric as in (b-c). **(e)** Syllable repeat probabilities in actual and predicted pupil sequences (Z $n = 21$; L $n = 21$). Center lines show medians, and boxes extend to the upper and lower quartiles. Whiskers extend to ±1.5×IQR. Points show outliers. (***: $p < 0.001$; **: $p < 0.01$; *: $p < 0.05$; n.s.: not significant)