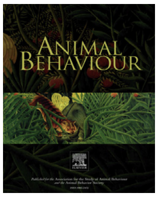
Animal Behaviour 210 (2024) 331–345

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav



Detecting cultural evolution in a songbird species using community science data and computational modelling

Yakov Pichkar [a, b, 1], Abigail M. Searfoss [a, c, 1], Nicole Creanza [a, b, *]

[a] Department of Biological Sciences, Vanderbilt University, Nashville, TN, U.S.A.
[b] Evolutionary Studies Initiative, Vanderbilt University, Nashville, TN, U.S.A.
[c] Program for Chemical and Physical Biology, Vanderbilt University, Nashville, TN, U.S.A.

ARTICLE INFO

Article history:
Received 26 January 2023
Initial acceptance 22 March 2023
Final acceptance 2 October 2023
Available online 16 February 2024
MS. number: A23-00047R

Keywords:
birdsong
community science
cultural evolution
learned behaviour
selective pressure

Song in oscine birds is learned across generations, and aspects of the song-learning process parallel genetic transmission: variation can be introduced into both cultural and genetic traits via copy error, and both types of traits are subject to drift and selective pressure. Similarly to allele frequencies in population genetics, observing frequencies of birdsong features can improve our understanding of cultural transmission and evolution. Uniquely, community science databases of birdsong provide rich spatiotemporal data with untapped potential to evaluate cultural evolution in songbirds. Here we use both community science and field study recordings of chipping sparrows, *Spizella passerina*, to examine trends across nearly seven decades of song. We found that some syllable types tended to persist in the population for much longer than others. Persistent songs tended to contain more syllables of shorter duration than songs that were observed across fewer years. To draw inferences about the effects of learning biases on chipping sparrow syllables, we constructed a spatially explicit agent-based model of song learning. By comparing our empirical analysis to simulated song distributions using three different song-learning strategies (neutral transmission, conformity bias and directional selection), we suggest that chipping sparrows are unlikely to select tutors neutrally or with a conformity bias and that they learn their songs with a remarkably low copy error rate.

For oscine songbirds, song has many important functions, including territory defence, species identification and mate attraction (Catchpole & Slater, 2003; Searcy & Andersson, 1986). In contrast to a closely related outgroup (suboscines), oscines must learn their songs, making the process of song learning critical to the reproductive success of individuals across this diverse clade (Kroodsma & Miller, 1996; Kroodsma et al., 1982; Mason et al., 2017; Thorpe, 1958). The evolutionary dynamics of learned song exhibit parallels to those of human cultural evolution, where long-lasting traditions can coexist (Aplin, 2016; Hoppitt & Laland, 2013; Kandler & Laland, 2009; Tomasello et al., 1993; Whiten, 2017). By studying song learning, we can better understand which aspects of human learning and cultural evolution are shared with other species and which properties are unique.

The transmission of information between individuals underpins both human and avian cultural evolution. Laboratory and field studies have shed light on how song is transmitted in avian populations. Some of these studies have measured the properties of cultural transmission: the similarity between learned song and tutor song, error rates in song matching, the invention of new songs and the frequency of songs in a population, among other factors (Cardoso & Atwell, 2016; Marler & Peters, 1982; Marler & Tamura, 1962; Slater, 1986; Thorpe, 1958). Others have used field site data to address questions of song change over time. For example, some studies have tracked cultural evolution using recordings taken in one population over multiple decades (Ju et al., 2019; Williams et al., 2013), and other studies have demonstrated that temporal changes in song are discerned by the current population by showing that birds react more strongly to modern recordings than to historical ones (Derryberry, 2007, 2011). Field study recordings can ensure coverage of local song repertoires, facilitate direct observation of song tutors and provide samples from the entire

---

[*] Corresponding author.
E-mail address: nicole.creanza@vanderbilt.edu (N. Creanza).
[1] Contributed equally.

https://doi.org/10.1016/j.anbehav.2024.01.009
0003-3472/

site's population. Due to the limits on the time period and geographical range they can cover, field studies are snapshots of the cultural evolution of syllables, and larger-scale studies can help bridge the gap between local behaviours and cultural evolution.

By tracking songs and reproductive success over time, as well as by determining which song features correspond to stronger responses in current populations, researchers have gained insight into the types of selective pressures that operate on song (Derryberry, 2007, 2011; Williams et al., 2013). In parallel, evolutionary biologists and population geneticists, generally without access to time series data, have synthesized evolutionary models with evidence from existing distributions of allele frequencies to understand whether regions of the genome have undergone selection (Bamshad & Wooding, 2003; Bustamante et al., 2001; Ford, 2002; Gutenkunst et al., 2009; Nielsen, 2005; Williamson et al., 2005). A genetic variant can become more frequent in a population because it is associated with a fitness advantage (selection) or due to random chance (genetic drift). A genetic region under selection will tend to have a different distribution of allele frequencies than those regions not under selection (Nielsen, 2005). Thus, one approach in population genetics is to simulate the evolution of a trait under different selection pressures and population histories. By comparing data from real populations to predictions from evolutionary models, researchers have identified which of these models best explains the data (Akashi & Schaeffer, 1997; Gutenkunst et al., 2009; Kryazhimskiy & Plotkin, 2008; Williamson et al., 2005). Some researchers apply this theoretical approach to the cultural evolution of song by examining the distribution of song within populations (Lynch & Baker, 1993, 1994; Lynch et al., 1989; Mcgregor & Krebs, 1982; Parker et al., 2012) and by developing individual-based or agent-based simulations of song learning that are compared to field site data (Crozier, 2010; Ellers & Slabbekoorn, 2003; Lachlan et al., 2018; Lachlan & Slater, 2003; Slater, 1986; Wheelwright et al., 2008; Youngblood & Lahti, 2022). Such agent-based simulations have been used in conjunction with birdsong data to infer the learning strategies used by swamp sparrows, *Melospiza georgiana*, and house finches, *Haemorhous mexicanus* (Lachlan et al., 2018; Youngblood & Lahti, 2022). With these comparisons of field recordings and results, researchers found evidence for different cultural transmission biases in different species. For example, swamp sparrows showed evidence of conformity bias, a type of frequency bias in which common song variants are disproportionately preferred (Lachlan et al., 2018). However, house finches showed evidence of content bias, in which certain song elements are preferentially learned regardless of their frequency in the population, a form of directional selection on the basis of a feature of the song (Youngblood & Lahti, 2022).

Here, we present an extension to this approach by developing a model of cultural transmission of birdsong and comparing the results of this model to a large-scale song analysis of community science recordings. We suggest that utilizing community science data is a time- and cost-effective supplement to field studies in the study of birdsong evolution. Specifically, community science data, which can cover a large geographical area over many years, can provide a unique insight into patterns of song transmission across large spans of time or space, particularly when these data are considered alongside evolutionary models. Researchers have analysed community science data to examine avian behaviours, whereas other studies have compared spatially explicit models to song recordings from the field to examine evolutionary hypotheses (Bolus, 2014; Dennhardt et al., 2015; Goodenough et al., 2017; Kaluthota et al., 2016; Newson et al., 2016; Robinson et al., 2018; Silvertown et al., 2011). We synthesized these approaches by examining community science data via models of song learning, providing insights into cultural evolutionary patterns.

As a focal species for this study, we chose the chipping sparrow, *Spizella passerina*, which has a simple repertoire of one repeated syllable. As a result, the full vocal repertoire of an adult bird can be captured by a single community science recording. Since cultural transmission includes mechanisms of mutation, selection and drift similar to those found in genetics, we employ techniques from population genetics, in particular, adaptations of site frequency spectra (Bustamante et al., 2001; Nielsen, 2005), to study song evolution. We identify unique syllable types that characterize the songs in a population and use the occurrence and life spans of these syllables to gain a deeper understanding of chipping sparrow learning. Since different learning strategies result in different distributions of syllables, with many replicates of the same syllable persisting over time if birds learn based on a conformity bias and potentially a relatively small number of syllables with desirable characteristics in the case of a directional bias, the frequency at which syllables occur in nature carries information about these biases. We compared the occurrence and longevity of songs to distributions produced by a computational model. This model simulates the transmission of syllables in a population under three types of learning: neutral evolution, conformity bias and directional selection. We demonstrate how analysis of community science data in association with a model can supplement field studies and extend the understanding of birdsong evolution.

METHODS

Categorization of Chipping Sparrow Syllables into Types

In a previous study, we gathered and analysed field site and community science recordings of chipping sparrows across the species' entire breeding range (Appendix, Fig. A1), measuring numerous acoustic features of each song and classifying the syllables into distinct types and categories (Searfoss, Liu et al., 2020; Searfoss, Pino et al., 2020). A number of recorded songs in our previous analysis (Searfoss, Liu et al., 2020) did not have a recording date listed; however, by revisiting the original field recording notes, we were able to find the years for all recordings for our study presented here. We categorized songs as follows: all songs were viewed as spectrograms in Audacity (https://audacityteam.org/) on a fixed frequency and timescale (see Supplementary Table S1). A single syllable was then selected as representative of a song, since chipping sparrow songs are fully characterized by repetition of a single syllable. We manually classified 820 syllables into 112 chipping sparrow syllable types based on the shape of the syllable (Supplementary Table S1; similar to methods of Borror, 1959; Leitner & Catchpole, 2004; Liu, 2001; Vargas-Castro et al., 2012; for examples of spatial syllable distributions, see Supplementary Fig. S14). We further grouped these syllable types into broader categories based on the syllable shape: up–down (up-slur followed by down-slur), down–up (down-slur followed by up-slur), sweep (single up-slur or down-slur), complex (more than two slurs), doubles (a slur with multiple frequencies) and buzz (syllable containing a noisy and/or high-entropy section, generally termed ‘buzzy’). To ensure that we were correctly categorizing the repeated element, particularly in the case of up–down versus down–up syllables, we examined the beginning and end of the song to determine which part of the syllable came first.

In addition, we used the song analysis software Chipper (Searfoss, Pino et al., 2020) to extract eight song features from each recording: mean intersyllable silence duration, mean syllable duration, mean syllable frequency range, mean syllable minimum and maximum frequency, duration of song bout, mean syllable stereotypy and total number of syllables. Chipper allows the user to visualize each song bout and it predicts where syllable boundaries are located using fluctuations in the amplitude of the signal (Searfoss, Pino et al., 2020). The user can change the signal-to-noise threshold, apply low-pass and high-pass filters to exclude high-frequency and low-frequency noise, respectively, and manually correct these syllable boundaries if necessary. Then, Chipper analyses the signal within each identified syllable and outputs a matrix of features for each song (see Searfoss, Pino et al., 2020 for more details on how each song feature is extracted from the spectrogram).

For the catalogue numbers, database, recordist, URL and licence for the 820 song files, see Supplementary Table S1. For the metadata including recording latitudes and longitudes and the 8 song features (all log transformed except mean stereotypy of repeated syllables and the standard deviation of note frequency modulation), see Supplementary Table S2.

Calculating and Analysing the Life Span of Chipping Sparrow Syllable Types

The observed 'life span' of a syllable type was defined as the period between the earliest and latest years in which a syllable type was recorded. To explore the properties of these life spans, we plotted the distribution of syllable life spans and the number of times in which these syllables were identified. We proceeded to compare syllable features between songs that contained short-lived (recorded life span $= 1$ year) versus long-lived (recorded life span ≥ 50 years) syllable types. We performed Wilcoxon rank-sum tests on short- versus long-lived syllables for the eight song features extracted from each recording. For stringency, we conducted a Bonferroni correction for multiple hypothesis testing by dividing the *P* value threshold for significance ($\alpha = 0.05$) by the number of tests. Overall, we performed one test on eight song features, so the threshold for significance was lowered to $\alpha_{\text{adj}} = 6.25 \times 10^{-3}$. In addition to our previously observed geographical patterns in chipping sparrow songs (Searfoss, Liu et al., 2020), we also conducted Wilcoxon rank-sum tests to determine whether short- or long-lived syllables are more frequent on an east–west axis.

Model Design

We developed an agent-based simulation to model song learning in the chipping sparrow population. The entirety of the model was implemented in Python 3.7 and uses the following primary packages: NumPy version 1.16.3, Matplotlib version 3.0.3, Pandas version 0.24.2 and SciPy version 1.2.1 (https://github.com/CreanzaLab/ChippingSparrowCulturalEvolutionModel). With this model, we simulated syllable transmission in a spatially structured population under three learning regimes: neutral transmission, conformity bias and directional selection. Under a neutral model of song learning, a juvenile randomly chooses a tutor's song to imitate; with conformity bias, a juvenile is disproportionately likely to choose a tutor with the most common song; and directional selection operates to favour certain song properties such as rate of syllable production or greater frequency bandwidth (Podos, 1997; Podos & Nowicki, 2004), such that juveniles are more likely to learn songs that exemplify better performance. Here, directional selection is somewhat analogous to the content bias observed in house finches (Youngblood & Lahti, 2022) since juveniles are choosing to learn a syllable based on its properties and not on its frequency of occurrence. Chipping sparrows only learn a single syllable, so in our model, directional selection operates on a continuous feature of a syllable, the rate of syllable production, instead of the selection of certain syllable types to compose a song.

As illustrated in Fig. 1, we initialized each model with a 500 × 500 matrix of syllable types that represented a population of birds (we performed additional analyses with matrix sizes of 400 × 400, 600 × 600 and 700 × 700). Each matrix location represents a single bird that sings a single syllable, a categorical value that was initially assigned randomly from a discrete uniform distribution {1:500}. For the directional selection model, we used an additional matrix of identical size, containing continuous values representing a syllable feature (the rates of syllable production) randomly sampled from a truncated normal distribution confined to the values observed in nature, i.e. a minimum of five syllables/s and a maximum of 40 syllables/s, with mean of 22.5 syllables/s and variance of 25 syllables/s ($\overline{X} \sim N(22.5, 25)$, $5 < \overline{X} < 40$; Searfoss, Liu et al., 2020). In each time step, roughly corresponding to a year, the following steps take place: a fraction of birds die, juvenile birds are tutored and fill the empty territories and a portion of the population disperses (Fig. 1). Note that there is a substantial burn-in period (discussed below; see Sampling the Simulated Bird Population), so initial distributions of syllables and song rates have minimal impact on final sampled values.

More specifically, in each time step, first a fixed percentage of the birds are randomly selected for death. We set this mortality rate to 40% based on similar avian models (Lachlan et al., 2018; Slater, 1986). For every bird that is eliminated, a new juvenile bird replaces it; the new bird's repertoire is either a novel syllable in the case of copy error, or a syllable learned from one of its neighbours (namely, the birds present at the beginning of the time step and adjacent to its hatching location in the two-dimensional matrix, up to eight birds). To maintain a spatial arrangement representative of natural territories, the matrix boundaries do not wrap, so birds at the edges of the matrix have fewer neighbours. Maintaining father–juvenile relationships was not necessary in our model, as oblique transmission of song, rather than vertical song transmission from parent to offspring, appears to be predominant in chipping sparrows. In a well-studied population in the northeastern U.S., juveniles learned a song before their first migration but often changed their song after migration to better match a neighbour, leading to a spatial pattern in which the syllable types of neighbours often differ but occasionally match closely; an individual's song did not change further after the first year (Liu & Kroodsma, 2006; Liu & Nottebohm, 2007). Although the phenomenon of postmigration song modification has not been studied in other chipping sparrow populations, the pattern of occasional neighbour matching was also observed in chipping sparrows in Mexico (Marler & Isaac, 1960). Once all juveniles learn a syllable (see next sections for the three learning strategies), the new syllable types replace the matrix elements of the birds that died, representing juveniles moving into vacant territories. Each new syllable was represented by a new integer, such that all syllable types could be uniquely identified. All territories vacated by birds that died are filled simultaneously, after tutor selection occurred for that time step. Since deaths in nature occur throughout the year and learning takes place over a short time, all birds present at the beginning of the time step can influence the learning of juveniles during that time step.

In addition, birds have an opportunity to disperse: some portion of birds (termed the 'dispersal fraction') is selected to move to a nearby location on the matrix. This promotes the mixing of regionally common syllable types with the larger population. Dispersal may sustain the local syllable diversity seen in chipping sparrow populations (Liu & Kroodsma, 2006). The addition of this dispersal step reflects what has been observed in the field: adults occasionally move to a new location, especially when they share a song with a neighbour (Liu & Kroodsma, 2006). We tested dispersal fractions between 0 and 1 in 0.1 increments, where 0.5 means that half of birds attempt to swap

Figure 1. Model schematic with illustrated representation of learning biases. For neutral learning, the frequency of a syllable type among the juvenile's adjacent neighbours is proportional to the probability that it will be learned by the juvenile (equal weight for all neighbours; identical arrows). Conformity bias modifies this probability by taking probabilities to the power of a conformity factor *c*, such that more common syllables are preferred. In our implementation of directional bias, syllable rate is the song characteristic that is selected for (although this could easily represent any song feature): the neighbour with the highest syllable rate is chosen as the tutor. Both the syllable type (for all learning models) and the syllable rate are learned with some probability of error; if the syllable type is incorrectly copied, a new syllable is invented.

places with a bird that has not yet changed places, chosen randomly from a location within a set radius (up to 11 matrix units). These dispersal fractions and radius values were informed by field studies describing chipping sparrow territory size and dispersal patterns (Liu, 2004; Liu & Kroodsma, 2006; Swanson et al., 2004). For each set of parameters, the simulation is run for 1000 time steps, of which the final 68 are used for sampling (to compare with the 68 years of available community science data) and the first 932 are a long burn-in period prior to sampling.

Model Implementation of Neutral Tutor Selection

During each time step, birds that die are replaced by a juvenile bird at each location. For neutral tutor selection, the syllable type learned by this juvenile is chosen at random from its eight immediate neighbours (or fewer, at edges and corners). To account for some likelihood of the new bird producing a novel syllable, we include a probability of error in learning. The probability of learning syllable type *i* in the case of neutral tutor selection can be represented as

$P_i \propto f(i)\,(1-\varepsilon)$,

where $f(i)$ is the frequency of occurrence of a syllable type i among neighbours and ε is the chance the syllable is copied incorrectly. We varied this error rate parameter (ε) to explore the ranges of plausible error rates for each learning model (10^{-6}%, 10^{-5}%, 0.0001%, 0.001%, 0.01%, 0.1% and 1.0%). We also ran several models in which an error meant that, instead of inventing a novel syllable, a bird produced a random syllable from the original set of syllables {1, 500}. For these models of syllable reinvention, called homoplasy, we added a larger error rate of 10%. Since only about 0.5% of observed songs were recorded only once (see Results), we did not test values of new syllable invention higher than 1%.

Model Implementation of Conformity Bias

Under conformity bias, juveniles preferentially learn more frequent songs. Each juvenile surveys the syllable types sung by his neighbours. The probability of learning syllable type i in the case of conformist tutor selection can be represented as

$P_i \propto f(i)^c(1-\varepsilon)$,

where c is the conformity bias factor. In the simulations we report below, a conformity factor $c = 2$ means that each syllable's frequency is squared. These values are then normalized to represent the likelihood of selecting each syllable, such that more common syllable types are learned more often than they appear among neighbours. We tested a series of other conformity factors including less severe conformity biases ($c = \{1.2, 1.4, 1.6, 1.8, 2.0\}$) and a weak novelty bias ($c = 0.8$). The learning error and dispersal were examined identically to those of the neutral tutor selection model.

Model Implementation of Directional Tutor Selection

For directional tutor selection, learning is based not on the frequencies of syllable types, but on a continuous variable representing the rate of syllable production. The probability of learning syllable type i under directional selection is given by

$P_i \propto (1-\varepsilon)$ if $r_i = \max(R)$ and $P_i = 0$ otherwise,

where R is the set of all neighbours' syllable rates r_i. This process mimics a type of directional selection that has been proposed for chipping sparrows, in which the preferred song is most difficult to produce. This model could accommodate directional selection on any continuous song feature; as a case study, we use a simple metric, the rate of syllable production, also called trill rate, as our putative feature that is under selection. This feature has been hypothesized to be relevant in chipping sparrows (Goodwin & Podos, 2014), but we note that our previous analysis showed a relatively wide distribution of trill rates in natural chipping sparrow recordings over time, ~5–40 syllables/s. In the model, the juvenile selects a tutor, the neighbour with the fastest syllable rate, and attempts to copy both the syllable type and the syllable rate of this tutor. The learning error for syllable type operates identically to the neutral and conformity models. With directional selection, there is also a learning error for syllable rate that is weighted such that it is more difficult to replicate or improve upon the tutor's performance than to perform worse than the tutor (as in Henrich, 2004). Thus, juveniles are most likely to sing at a slightly slower rate than the tutor, and the syllable rates are restricted to values observed in our chipping sparrow song database (1–40 syllables/s). Therefore, the learned song is sung at a rate similar to that of the tutor with an error drawn from a uniform distribution between -2 and 0.25 and has either an identical or a novel syllable type. It is important to note that while we describe the song feature under selection as ‘syllable rate’, it can be interpreted that any continuously varying song feature is undergoing directional selection.

Sampling the Simulated Bird Population

The method of data collection from the model population was chosen to replicate the sampling that occurred when songs were recorded by community scientists. For each learning model, the simulated population was sampled such that the number of birds sampled per time step was equal to the number of recordings we had from each year. Our recording data spanned 68 years (1950–2017), with some years having no recordings; thus, some number of birds (possibly 0) were selected at random from the model population for each of the last 68 of 1000 iterations of the model (here, the first 932 time steps served as a burn-in, which minimized the effects of initialization for variables such as the number and distribution of syllables). For each bird sampled from the simulated population, the iteration from which it was collected and the syllable type were documented. This was to ensure the life spans and the counts of syllable types of this sampled model population could be calculated in the same manner as the recording data.

We sampled syllables from our simulations to compare the model to three similarly sized regions of the range chosen for their high sampling density. For regions of approximately 100 000 km^2, we chose three regions with the highest density of song recordings. These regions were defined as the area within a rectangle bounded by latitudes and longitudes; these included the Michigan/Ohio region, U.S.A. (85°–82°W, 39°–43°N, containing 172 song recordings over ~100 000 km^2), the New York region, U.S.A. (77°–73°W, 40°–43°N, containing 88 songs over ~100 000 km^2) and the New England region, U.S.A. (74°–69°W, 41°–45°N, containing 210 songs over ~130 000 km^2). We estimated the chipping sparrow population density in the U.S. to be about 13.3 birds/km^2, calculated as the population size of chipping sparrows (estimated as 100 million chipping sparrows in the U.S.; Will et al., 2020) divided by the area of the continental U.S. (~7.5 million km^2). We assumed that our sampled regions would likely have at least this average density, so we used a range of 13.3–25 birds/km^2 for our calculations. For a breeding region of 100 000 km^2, we thus estimated a population size of about 1.3–2.5 million chipping sparrows. Of these, approximately half are males, and not all males are adults with territories: likely only one in three chipping sparrows is a territorial male. Given their sizes, each region likely has 444 000 to 833 000 territories, compared to the 160 000 territories in the smallest simulations (400 × 400) and 490 000 in the largest (700 × 700). All these regions are concentrated towards the east of the chipping sparrow range, as equivalently sized regions in the west had a much smaller number of recordings (22 in California's Central Valley, and fewer in Oregon, U.S.A.). To estimate the syllable life spans and counts that each model produced for each region, we sampled syllable types from every model 50 times.

Quantitatively Comparing Model Results to Empirical Data

For each simulation, we visualized and compared our results using the site frequency spectrum technique from population genetics (Nielsen, 2005; Pepperell et al., 2013; Zhu & Bustamante, 2005) and created a ‘syllable frequency spectrum’, the frequency of birds that sing various syllable types in the sample, to compare our model results with empirical data. Additionally, we used a similar visualization to compare the frequencies of syllable type lifetimes for each model across a range of learning error rates.

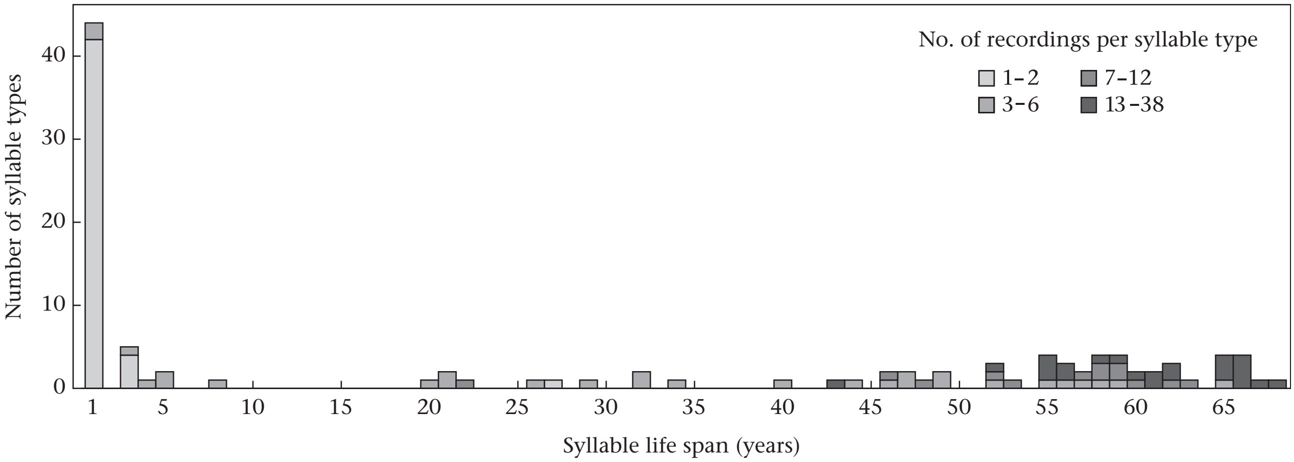


Figure 2. Distribution of chipping sparrow syllable types according to their life spans. For our database of 820 recordings of 112 syllable types, we plot the number of syllable types versus syllable life span across the entire range. Each syllable type is also shaded by the total number of recordings of that syllable type, illustrating that longer-lived syllable types are also more common, although less common long-lived syllables exist.

Specifically, we aimed to identify whether one or more tutor selection models would be able to produce results similar to the empirical data, across both syllable frequency of occurrence and syllable lifetime distributions. These distributions had similarities to the empirical data: there were several frequent long-lasting syllables, many syllables observed very few times and a small number of syllables with intermediate observed life spans (see Results). These intermediate life span syllables were most difficult to replicate among these song-learning models in a way that could be detected by direct comparison of distributions, such as via two-sample Kolmogorov–Smirnov and *k*-sample Anderson–Darling tests; the sparsely populated intermediate values of syllable frequency and life span contained most of the relevant differences between models, whereas most values used in comparisons of distributions were at the extremes (see Results). Instead of manually assigning categories of short longevity, intermediate longevity, etc., we placed the empirical data into bins using an algorithm; it assigned bin edges by minimizing the combined variance of bin count and bin width. First, we placed syllables that were recorded only once (for the syllable frequency spectrum) or only in a single year (for the distribution of syllable life spans) into their own category, and all others were initially placed into six equally spaced bins. Then, the edges of these bins were progressively shifted 10^6 times (the edges were moved, and variances of the resulting bins' sizes and their number of data points were calculated). If the combined variance was lower, this became the new set of bins for the next permutation (see https://github.com/CreanzaLab/ChippingSparrowCulturalEvolutionModel for details).

In this way, we used an algorithm to place the empirical chipping sparrow data into bins by syllable type using the optimally found edges, and the simulated samples were placed into the same bins to compare their distributions with the empirical data. On the binned syllable type and life span spectra, we conducted Fisher's exact tests between the empirical data (null hypothesis) and all simulated data sets to determine which combination of learning strategy, learning error and dispersal fraction resulted in syllable type and life span distributions concordant with patterns found in the chipping sparrow population.

Ethical Note

All recordings were gathered from open access databases of recordings or published studies from the wild. Playback may have been used to elicit these songs, which can affect bird behaviours in the wild (Harris & Haskell, 2013). The short- and long-term effects of field playbacks are understudied and likely differ by species, but the limited use of playbacks (i.e. when precautions are taken not to unduly disturb birds and relatively few playbacks are presented at a reasonable volume and duration) is thought to be an important part of an ethical birdwatching practice (Sibley, 2011; Watson et al., 2019). Recordists uploading to community science repositories are encouraged to document their use of playbacks, and few of our recordings were noted to be collected in response to playback.

RESULTS

Syllable Types and Cultural Analysis

We categorized syllables from 820 recordings into 112 distinct syllable types (Supplementary Table S1; also see previous analysis Searfoss, Liu et al., 2020). We found that syllable types that continued to exist for much longer than the lifetime of a chipping sparrow (less than 9 years) were also those that were most commonly observed, whereas other syllables were transient and observed rarely (Fig. 2). In other words, syllables that were found in many recordings always existed for a long period. To identify differences between long- and short-lived syllables, we classified syllable types as short-lived (life span $= 1$ year) or long-lived (life span ≥ 50 years) and compared the features of these syllables (Fig. 3). We found that long-lived syllable types were significantly shorter in duration than short-lived syllables ($P < 3.29 \times 10^{-3}$) (Fig. 3a, Table 1), and songs with a long-lived syllable type contained significantly more repetitions of that syllable per bout ($P < 1.33 \times 10^{-3}$) (Fig. 3b). We also found that geographical differences did not explain the trends in longevity, since the longitude of recordings of different syllable types did not significantly differ between the two life span categories ($P = 0.844$). Additionally, buzz syllables tended to be long-lived, whereas double or complex syllables tended to be short-lived, with up–down, down–up and sweep syllables being prominent in both life span groups (Fig. 3c).

Model Results

Here, we ran several models of song transmission in order to test which model produced patterns most similar to empirical data. Since the entire chipping sparrow range contains approximately

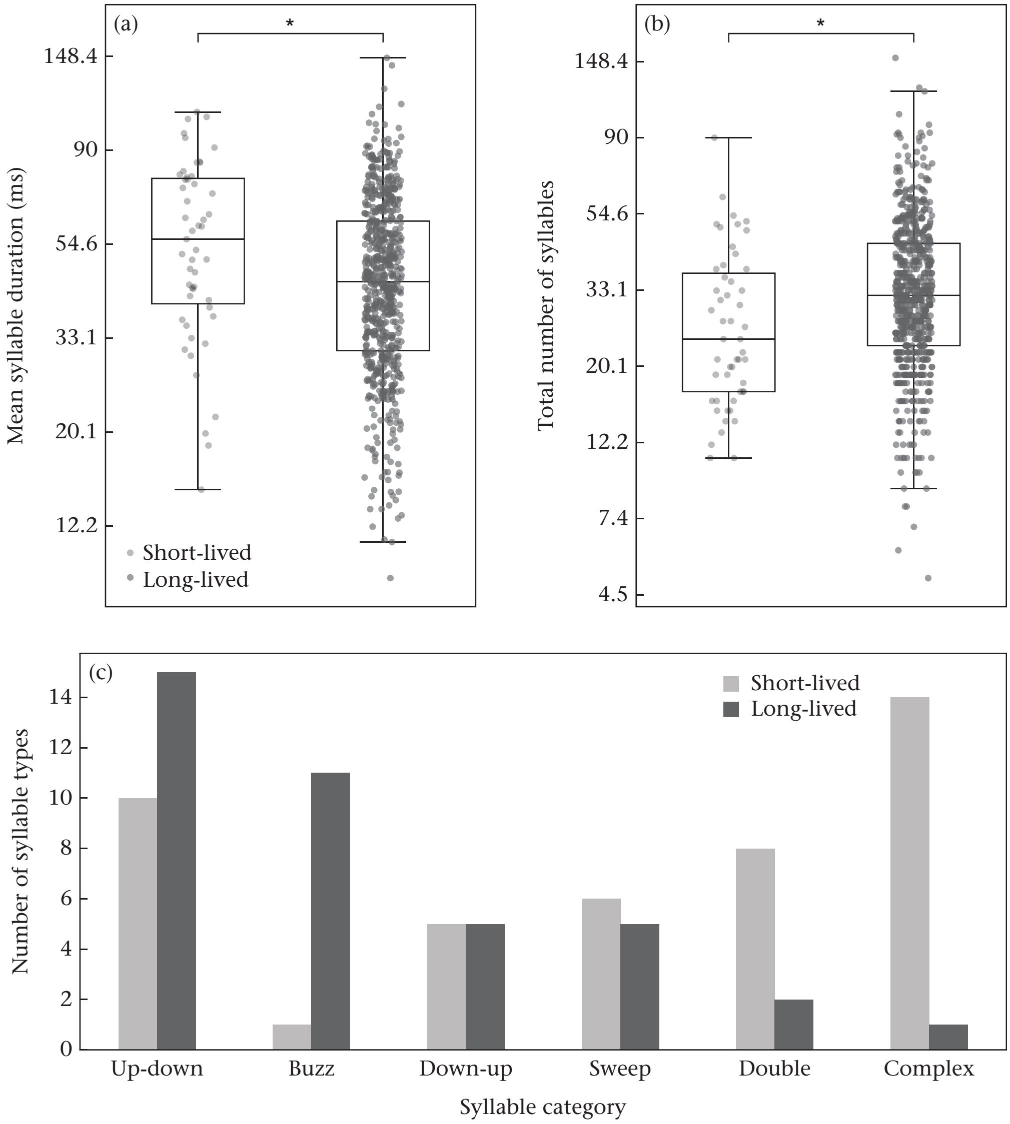


Figure 3. Syllable categories and song properties of short- and long-lived chipping sparrow syllable types. Songs with long-lived syllable types have >50-year life span; songs with short-lived syllable types have 1-year life span; here we show (a) mean syllable duration and (b) total number of syllables for these songs. $*P < 6.25 \times 10^{-3}$ (Wilcoxon rank-sum tests). (c) The number of short- and long-lived syllable types in each syllable category.

240 million individuals (Will et al., 2020), we compared the distributions of syllable occurrences and syllable lifetimes from our model to several focal regions that had high rates of sampling. We conducted a series of parameter sweeps, running the simulation for each of the three learning strategies with multiple error rates (or invention rates) spanning 10^{-6}% to 1.0% and with territory dispersal fractions spanning 0 to 1 (i.e. 0–100% of individuals swap territories every time step with a bird that has not already dispersed). We sampled the simulated syllables at the same frequencies as in the empirical data. We quantitatively determined which combinations of parameters (song-learning strategy, learning error rate, dispersal fraction) produced results that did not deviate from the null hypothesis (empirical data).

Table 1
Results of the Wilcoxon rank-sum tests between short- and long-lived syllable types

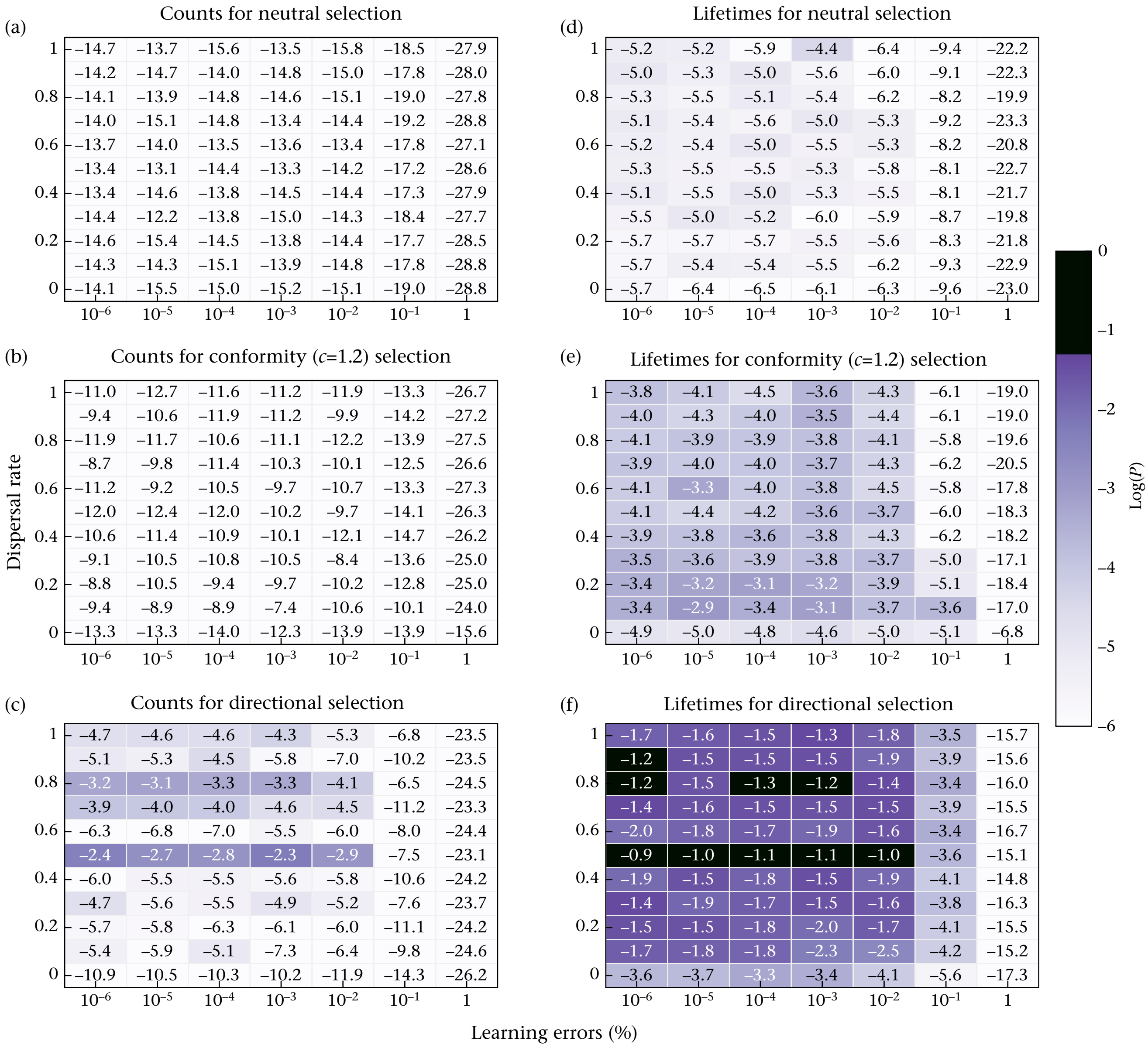
Song features	Short- vs long-lived *P* value
Mean intersyllable silence duration	0.449
Mean syllable duration	**0.003**
Mean syllable frequency range	0.802
Mean syllable minimum frequency	0.476
Mean syllable maximum frequency	0.475
Duration of song bout	0.010
Mean stereotypy of repeated syllables	0.653
Total number of syllables	**0.001**

Bold indicates $P < 6.25 \times 10^{-3}$.

We provide the syllable type and syllable lifetime frequency spectra for the best-fit model for each learning type. We determined a ‘best-fit’ model by first ranking the *P* values (from least to greatest) for both the analysis of syllable type frequencies and the analysis of syllable life spans and then combining the two ranks for each parameter set and choosing the one with the highest aggregate ranking, i.e. the largest sum of the two rankings (Appendix, Table A1, Fig. 4). (We note that the results of the Fisher's exact test indicated that the model results were potentially drawn from the same distribution as the real data when the null hypothesis was not rejected; when the *P* value was not significant, the results of the model were statistically indistinguishable from the empirical data). For one of our focal regions, New York, we found that the best-fit model for each learning type was neutral tutor selection with

10^{-5}% learning error and 0.3 dispersal fraction (syllable frequency distribution $P = 5.7 \times 10^{-13}$, life span frequency distribution $P = 9.5 \times 10^{-6}$), conformity bias with 10^{-3}% learning error and 0.1 dispersal fraction ($P = 4.5 \times 10^{-8}$ and $P = 8.2 \times 10^{-4}$, respectively, with conformity factor $c = 1.2$) and directional selection with 10^{-6}% learning error and 0.5 dispersal fraction ($P = 0.0044$ for occurrence spectra and statistically indistinguishable from the empirical lifetime data, $P = 0.123$; Fig. 5, see Supplementary Fig. S1 for corresponding unbinned spectra). For the other two focal regions, the best-fit parameters for each model were similar (Appendix, Table A2, for spectra see Supplementary Fig. S3, for parameter sweeps in other regions see Supplementary Figs S8–S10).

We measured the effect of changing several model parameters, including the strength of conformity selection, matrix size and syllable reinvention. To measure the strength of conformity bias during selection, we varied the conformity factor c. Neutral selection (equivalent to $c = 1$) and strong conformity selection ($c = 2$) performed poorly, whereas an intermediate strength of conformity selection ($c = 1.2$) performed best (Supplementary Figs S6–S7, Appendix, Table A1). However, even the simulation of conformity-biased learning with the best performance ($c = 1.2$, 10^{-3}% learning error and 0.1 dispersal fraction) produced distributions of syllable occurrences and lifetimes dissimilar to those observed in any region. To observe how varying the modelled population would affect the distributions of syllables, we simulated several population matrix sizes. We found that the larger the population matrix, the more difficult it was for our model to approximate the syllable lifetime distributions derived from the community science data



(a) Counts for neutral selection

	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
1	−14.7	−13.7	−15.6	−13.5	−15.8	−18.5	−27.9
	−14.2	−14.7	−14.0	−14.8	−15.0	−17.8	−28.0
0.8	−14.1	−13.9	−14.8	−14.6	−15.1	−19.0	−27.8
	−14.0	−15.1	−14.8	−13.4	−14.4	−19.2	−28.8
0.6	−13.7	−14.0	−13.5	−13.6	−13.4	−17.8	−27.1
	−13.4	−13.1	−14.4	−13.3	−14.2	−17.2	−28.6
0.4	−13.4	−14.6	−13.8	−14.5	−14.4	−17.3	−27.9
	−14.4	−12.2	−13.8	−15.0	−14.3	−18.4	−27.7
0.2	−14.6	−15.4	−14.5	−13.8	−14.4	−17.7	−28.5
	−14.3	−14.3	−15.1	−13.9	−14.8	−17.8	−28.8
0	−14.1	−15.5	−15.0	−15.2	−15.1	−19.0	−28.8

(b) Counts for conformity (c=1.2) selection

	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
1	−11.0	−12.7	−11.6	−11.2	−11.9	−13.3	−26.7
	−9.4	−10.6	−11.9	−11.2	−9.9	−14.2	−27.2
0.8	−11.9	−11.7	−10.6	−11.1	−12.2	−13.9	−27.5
	−8.7	−9.8	−11.4	−10.3	−10.1	−12.5	−26.6
0.6	−11.2	−9.2	−10.5	−9.7	−10.7	−13.3	−27.3
	−12.0	−12.4	−12.0	−10.2	−9.7	−14.1	−26.3
0.4	−10.6	−11.4	−10.9	−10.1	−12.1	−14.7	−26.2
	−9.1	−10.5	−10.8	−10.5	−8.4	−13.6	−25.0
0.2	−8.8	−10.5	−9.4	−9.7	−10.2	−12.8	−25.0
	−9.4	−8.9	−8.9	−7.4	−10.6	−10.1	−24.0
0	−13.3	−13.3	−14.0	−12.3	−13.9	−13.9	−15.6

(c) Counts for directional selection

	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
1	−4.7	−4.6	−4.6	−4.3	−5.3	−6.8	−23.5
	−5.1	−5.3	−4.5	−5.8	−7.0	−10.2	−23.5
0.8	−3.2	−3.1	−3.3	−3.3	−4.1	−6.5	−24.5
	−3.9	−4.0	−4.0	−4.6	−4.5	−11.2	−23.3
0.6	−6.3	−6.8	−7.0	−5.5	−6.0	−8.0	−24.4
	−2.4	−2.7	−2.8	−2.3	−2.9	−7.5	−23.1
0.4	−6.0	−5.5	−5.5	−5.6	−5.8	−10.6	−24.2
	−4.7	−5.6	−5.5	−4.9	−5.2	−7.6	−23.7
0.2	−5.7	−5.8	−6.3	−6.1	−6.0	−11.1	−24.2
	−5.4	−5.9	−5.1	−7.3	−6.4	−9.8	−24.6
0	−10.9	−10.5	−10.3	−10.2	−11.9	−14.3	−26.2

(d) Lifetimes for neutral selection

	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
1	−5.2	−5.2	−5.9	−4.4	−6.4	−9.4	−22.2
	−5.0	−5.3	−5.0	−5.6	−6.0	−9.1	−22.3
0.8	−5.3	−5.5	−5.1	−5.4	−6.2	−8.2	−19.9
	−5.1	−5.4	−5.6	−5.0	−5.3	−9.2	−23.3
0.6	−5.2	−5.4	−5.0	−5.5	−5.3	−8.2	−20.8
	−5.3	−5.5	−5.5	−5.3	−5.8	−8.1	−22.7
0.4	−5.1	−5.5	−5.0	−5.3	−5.5	−8.1	−21.7
	−5.5	−5.0	−5.2	−6.0	−5.9	−8.7	−19.8
0.2	−5.7	−5.7	−5.7	−5.5	−5.6	−8.3	−21.8
	−5.7	−5.4	−5.4	−5.5	−6.2	−9.3	−22.9
0	−5.7	−6.4	−6.5	−6.1	−6.3	−9.6	−23.0

(e) Lifetimes for conformity (c=1.2) selection

	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
1	−3.8	−4.1	−4.5	−3.6	−4.3	−6.1	−19.0
	−4.0	−4.3	−4.0	−3.5	−4.4	−6.1	−19.0
0.8	−4.1	−3.9	−3.9	−3.8	−4.1	−5.8	−19.6
	−3.9	−4.0	−4.0	−3.7	−4.3	−6.2	−20.5
0.6	−4.1	−3.3	−4.0	−3.8	−4.5	−5.8	−17.8
	−4.1	−4.4	−4.2	−3.6	−3.7	−6.0	−18.3
0.4	−3.9	−3.8	−3.6	−3.8	−4.3	−6.2	−18.2
	−3.5	−3.6	−3.9	−3.8	−3.7	−5.0	−17.1
0.2	−3.4	−3.2	−3.1	−3.2	−3.9	−5.1	−18.4
	−3.4	−2.9	−3.4	−3.1	−3.7	−3.6	−17.0
0	−4.9	−5.0	−4.8	−4.6	−5.0	−5.1	−6.8

(f) Lifetimes for directional selection

	10^{-6}	10^{-5}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	1
1	−1.7	−1.6	−1.5	−1.3	−1.8	−3.5	−15.7
	−1.2	−1.5	−1.5	−1.5	−1.9	−3.9	−15.6
0.8	−1.2	−1.5	−1.3	−1.2	−1.4	−3.4	−16.0
	−1.4	−1.6	−1.5	−1.5	−1.5	−3.9	−15.5
0.6	−2.0	−1.8	−1.7	−1.9	−1.6	−3.4	−16.7
	−0.9	−1.0	−1.1	−1.1	−1.0	−3.6	−15.1
0.4	−1.9	−1.5	−1.8	−1.5	−1.9	−4.1	−14.8
	−1.4	−1.9	−1.7	−1.5	−1.6	−3.8	−16.3
0.2	−1.5	−1.5	−1.8	−2.0	−1.7	−4.1	−15.5
	−1.7	−1.8	−1.8	−2.3	−2.5	−4.2	−15.2
0	−3.6	−3.7	−3.3	−3.4	−4.1	−5.6	−17.3

Figure 4. Statistical comparison of (a–c) syllable frequency and (d–f) lifetime spectra between computational models and empirical data for one focal region (New York) with a matrix size of 500 × 500. Sets of parameters (learning error and dispersal rate) for which the models and empirical data produced similar spectra distributions have P values greater than 0.05 (shown in black). All simulations with neutral learning (a, d) and conformity bias (b, e; with conformity factor $c = 1.2$) produced results that were statistically different from the empirical data ($P < 0.05$). Some simulations with directional selection (c, f) produced lifetime spectra statistically indistinguishable from the empirical data (f) ($P > 0.05$), although this was not true for occurrence spectra (c). For other regions, see Supplementary Fig. S5.

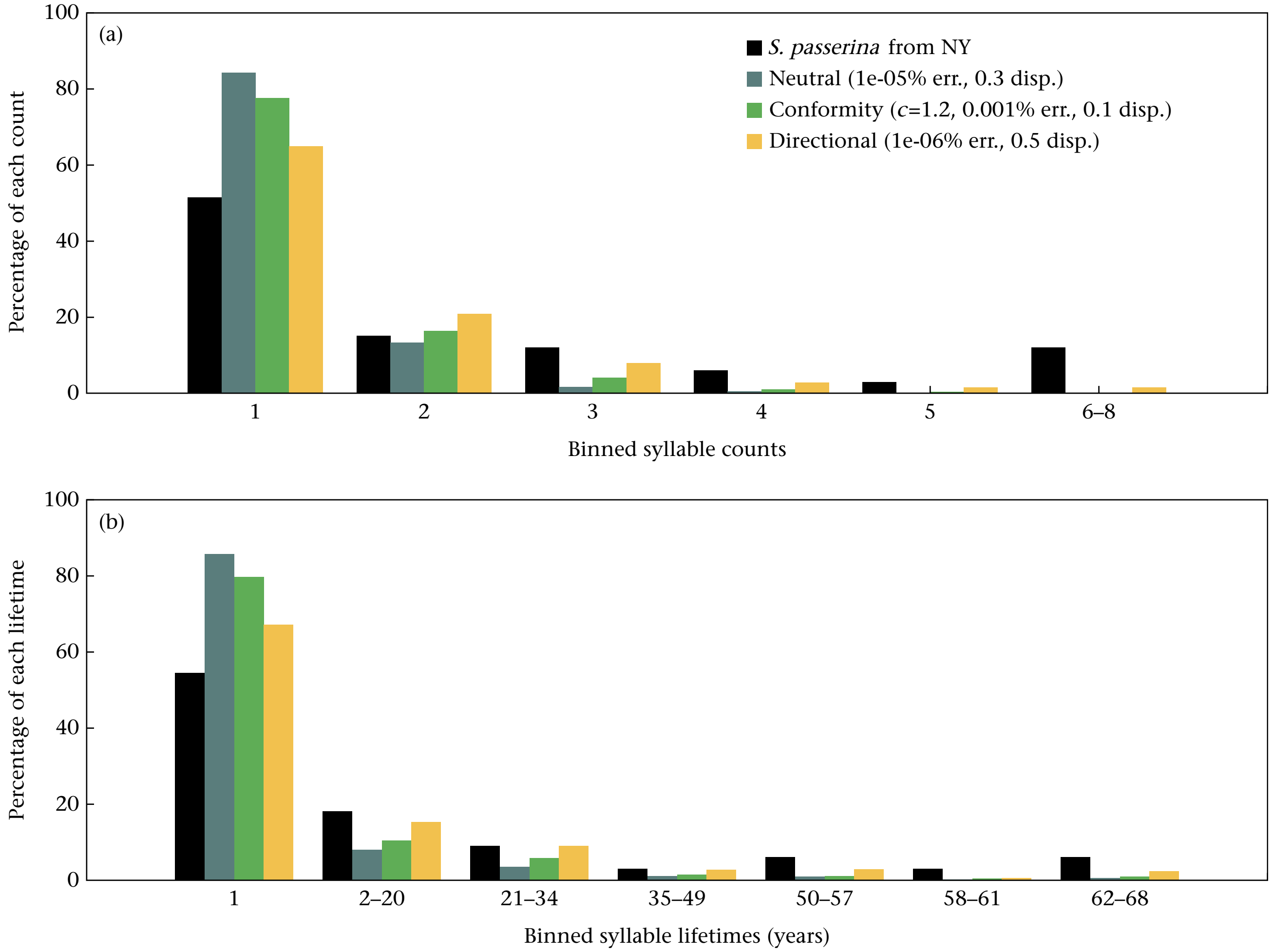


Figure 5. Comparison of binned syllable frequency and syllable life span spectra between empirical data and samples from best-fit models from one focal region (NY: New York): (a) The number of times a syllable was sampled and (b) syllable lifetimes (i.e. 1 + (the last year) − (the first year in which the syllable type was sampled)). Each panel includes data from the best-fit models of each of the three song-learning strategies: neutral tutor selection, conformity bias and directional selection. Data from community science recordings are also provided. (See Methods for how bins were calculated.) For unbinned data, see Supplementary Figs S1–S2; for frequency and life span spectra from other regions, see Supplementary Figs S3–S4.

(Appendix, Table A3, Supplementary Figs S12g, h, S13g, h, S14g, h) and that smaller matrices reproduced the empirical syllable lifetime distribution at a wide array of values (Supplementary Figs S12e–S14e). Finally, we modelled whether homoplasy, the reinvention of syllable types, affects these distributions. Since homoplasy could only occur as the result of an error in song learning, the differences between models with and without syllable reinvention were greatest when error rates were high (Supplementary Figs S8–S11). The best-fit models with homoplasy had higher dispersal rates (Appendix, Table A4, Supplementary Figs S8–S11), but models with homoplasy did not describe the observed parameters better than models without homoplasy: models with high learning error, which homoplasy affected the most, did not fit the empirical data for any combination of parameters.

Our regional models could not reproduce both the syllable counts and lifetimes found in the empirical data (Fig. 4a–c, Appendix, Tables A1–A2). The best-fit model (of those with a matrix size of 500 × 500 territories), which relied on directional selection, low learning error (10^{-6}% in New York and Ohio/Michigan and 10^{-4}% in New England) and intermediate dispersal fraction (0.5 in all regions), only reproduced the life spans found from the community science sampling (New York: $P = 0.123$; Ohio/Michigan: $P = 0.076$; New England: $P = 0.055$; Appendix, Table A2, Supplementary Fig. S3). These models did not reproduce the empirical occurrence spectrum in any region (New York: $P = 0.0044$; Ohio/Michigan: $P = 0.0037$; New England: $P = 1.4 \times 10^{-4}$; Fig. 4, Appendix, Table A2). The simulation with directional selection of tutors appeared to most closely match the frequency of syllable types found in our empirical data, similar to the distribution of singletons and a long, flat tail (Supplementary Fig. S2e–h). While long-lived syllable types arose in all three song-learning models, directional selection had an enrichment for these syllable types compared to the neutral model, whereas conformity bias had an abundance of long-lived syllables and few syllables with intermediate life spans; thus, directional selection's frequency spectrum of syllable type life spans best reflected what we observed in our community science data sample (Fig. 5b).

DISCUSSION

Here, we performed an analysis of chipping sparrow song recordings across nearly seven decades to identify long-term patterns. We extended the use of computational approaches to cultural evolution (Youngblood, 2019) with techniques from population genetics and rapidly growing community science data (Searfoss, Liu et al., 2020) to assess cultural change and stability in birdsong.

Community science recordings provide broad spatiotemporal coverage of a species' range, resulting in a data set in which temporal changes could be identified across the entire population. By comparing these rich data sets with the predictions of cultural evolutionary models, we were able to evaluate the possible strategies underlying the social transmission of song. Specifically, we constructed a spatially explicit model of cultural transmission of chipping sparrow songs with different types of learning bias: neutral evolution (unbiased transmission), directional selection (favouring a certain characteristic of song) and conformity bias (favouring locally common songs). By comparing our empirical and simulated results, we found that a directional model most closely replicated the patterns of both syllable longevity and counts produced by chipping sparrow song learning. In addition, our computational analyses agreed with evidence from field research in finding that chipping sparrows had high-fidelity song learning (predicted new syllable invention rate of less than 0.1% in focal region analyses) and likely dispersed to new territories (a dispersal rate of 0.1 or greater, most likely near 0.5) after initial learning (Fig. 4f, Appendix, Table A2, Supplementary Fig. S5).

Past studies have examined the diversity in syllables within the chipping sparrow population. For example, in the 1950s, Borror classified chipping sparrow syllables from 58 recordings into categories and further subdivided the 58 recordings into 28 syllable types, demonstrating great song diversity and few observations of each syllable type (Borror, 1959). In a later analysis, the syllables of 157 chipping sparrows from the eastern U.S. were analysed and placed into around 30 distinct syllable types by eye (Liu, 2001). With 820 songs, we identified 112 syllable types (Supplementary Table S1). While our larger community scientist-informed sampling was far smaller than the current chipping sparrow population, which is on the scale of 240 million, our analysis contained syllables that existed for decades (in the entire range, see Fig. 2, and in focal regions, see Fig. 5b, Supplementary Fig. S3e–f) and captured variation in birdsong that could not be identified via field studies of a species with such a large range (Supplementary Fig. S15). It is possible that the sampled chipping sparrow syllables appeared identical by chance rather than by common descent as a result of syllable reinvention, a sort of cultural homoplasy. However, when we modelled syllable reinvention, models with homoplasy did not fit the empirical data better than models without it for any form of selection (Appendix, Table A4, Supplementary Figs S8–S11, (a)–(f) versus (g)–(l)). Given the high fidelity of pupil learning (Liu & Kroodsma, 2006), the presence of geographically clustered syllables (Supplementary Fig. S15) (Searfoss, Liu et al., 2020) and the results of our models, we posit that learning errors rarely result in birds reproducing an existing syllable elsewhere in the chipping sparrow range. These data have allowed us to explore trends in chipping sparrow song over time that will inform future studies of their song and cultural evolution.

With our analysis of chipping sparrow syllables sampled from their entire range, we found that many syllable types appeared to be rare and short-lived, whereas others were quite common and persisted for decades (Fig. 2). This was true both for the entire region and when dividing the entire range into focal regions (Fig. 5b, Supplementary Fig. S3e–f). Furthermore, we found evidence that some broad characteristics were associated with longer syllable life spans. Buzz syllables tended to be long-lived whereas complex syllables tended to be short-lived, and songs with long-lived syllable types had more repetitions of shorter syllables, which would be consistent with predictions that songs with faster syllable repetitions might be favoured in birds (Byers et al., 2010). Notably, this pattern of shorter and faster syllables being long-lived was geographically distributed: the distribution of short- versus long-lived syllables was independent of longitude (Wilcoxon rank-sum test: $P = 0.992$) despite songs having more shorter syllables on average than songs in the western U.S./Canada than in the eastern U.S./Canada (Searfoss, Liu et al., 2020). Our results demonstrate that the diversity of chipping sparrow syllable types was not fully sampled in previous studies, and it is likely that other syllable types will be discovered as contributions of song recordings to community science databases become more widespread. These results raise an important question: are syllables common and long-lived because of neutral transmission (similar to genetic drift), culturally favourable properties (i.e. certain syllables are inherently salient or associated with successful birds) or conformity bias (i.e. common syllables are preferred when learning song)?

Selectively neutral processes of song learning, such as unbiased learning of a song with a relatively high rate of error, are predicted to result in a simple pattern of syllable prevalence: most sampled syllable types would be sung by only one bird, fewer syllables would be sung by two birds, even fewer by three birds, and so on, until only a small handful of syllables might be sung by many birds (Slater, 1986). Slater observed this distribution of syllables in chaffinches, *Fringilla coelebs*: in a population of 36 chaffinches, most songs were sung by only one bird, but one song was sung by 22 birds. Furthermore, he modelled the song-learning process with a simulation in which newly settled birds learned a random nearby song with some error; this simulation demonstrated that a neutral learning process with a predictable rate of copy error was sufficient to replicate the observed distribution of chaffinch syllables. A similar pattern is regularly observed in genetic data in a stable population in the absence of selection pressures: most genotypes are rare, and few genotypes predominate (Nielsen, 2005). Thus, for both genotypes and song types, one does not need to invoke selection pressures to explain a pattern in which one or very few types are widespread but most are observed only once.

The question of whether directional selection plays a strong role in chipping sparrow song evolution has been a topic of debate in the literature (Akçay & Beecher, 2015; Goodwin & Podos, 2014, 2015; Kroodsma, 2017). In chipping sparrows, syllable rate in particular has been shown to be correlated with displays of territory defence: 'birds responded more vigorously when simulated intruders sang the more difficult to produce, faster songs, and also when there was a stronger disparity between intruder trill rates and their own' (Goodwin & Podos, 2014, p. 4). Some evidence suggests that chipping sparrows are subject to a performance constraint, specifically one in which there is a trade-off between large sweeps in frequency (Hz) and a high rate of syllable delivery (Goodwin & Podos, 2014; Podos, 1997). It has been proposed that physiological constraints contribute to this balance in song performance (Podos, 1996, 1997). Other studies have suggested that a performance trade-off between frequency bandwidth and syllable rate could be meaningful: under the stress of competing with the noise of an urban environment, chipping sparrows underperformed, singing 'twice as far below the trade-off frontier' than those in less noisy environments (Davidson et al., 2017). Kroodsma presented a contrary view based on results from field studies demonstrating that juvenile chipping sparrows imitate their neighbours with great success, and he suggested that physiological constraints do not inhibit juveniles from performing fast songs (Kroodsma, 2017). Instead, he suggested that their performance is determined by that of their neighbour.

Our analysis is a step towards resolving the debate between performance-driven and neighbour-dependent hypotheses. These results suggest that chipping sparrows select which of their neighbours will be a tutor based on some aspect of their song performance: certain tutors may be preferred for reasons other than how frequently their song is heard locally. Our analysis of recorded songs and song-learning models points to directional selection as the best explanation for chipping sparrow song diversity. In nature, juvenile chipping sparrows sing several neighbours' songs before selecting a final song, which suggests that a selective process is taking place (Liu & Nottebohm, 2007). This selective process, along with juveniles' modification of their song during the plastic phase of song learning, have been proposed to play a part in determining their final song (Podos, 2017). The extent to which these potential selective processes affect song learning is controversial, suggesting that the combination of song data with learning simulations could shed light on the evolutionary dynamics of vocal learning.

Our 70-year sampling time span gave us the opportunity to analyse the observed longevity of chipping sparrow syllable types and their frequency of occurrence. We found that it was difficult to reproduce the distribution of syllable occurrences in our regional analyses (Figs 4–5) and the distribution of syllable life spans was only reproduced by models of directional selection. This divergence seems to be driven by the models predicting a large number of uncommon, short-lived syllables. Overall, these spectra of syllable properties favour the directional model of song transmission in chipping sparrows, such that some quality of the song, rather than the song's frequency of occurrence in the local population, is under selection. Lachlan et al. (2018) demonstrated a model of conformity bias in swamp sparrows leading to a qualitatively similar life span distribution as ours, in which certain syllables tended to be longer-lived, even predicting that these syllables were maintained for upwards of 500 years. In contrast to their model of swamp sparrows, our chipping sparrow model supports directional selection as the more likely source of the observed patterns of syllable life span (New York: Figs 4–5; all other regions: Supplementary Figs S3–S4). A major difference between our model and that of Lachlan et al. is that we explicitly modelled the spatial structure of songbird populations, such that conformity bias only acted on the syllables found among neighbours. As a result, the conformity factor that we found to be most appropriate ($c = 1.2$) cannot be directly compared to the parameter α used by Lachlan et al., which they found fitted their swamp sparrow data best at $\alpha = 1.316$.

We compared the results of our model to empirical data from three focal regions, each having a high density of song recording coverage, and we found that the same patterns applied to all of these regions. Directional selection produced the best result in all three regions, consistently favouring low learning error rates (<0.1% error) and some amount of dispersal (dispersal rate $\geq$0.1) (for a heatmap of the New York region, see Fig. 4; for Ohio/Michigan and New England, see Supplementary Fig. S5g–r; for best-fit results for all regions, see Appendix, Table A2). The comparison of the model to the entire range produced different results: in this case, conformity-biased learning can also reproduce the empirical distribution of syllable lifetimes (Supplementary Fig. S5e). The directional model of selection consistently produced the syllable lifetimes found in all regions, including a number of long-lived syllables. However, the directional learning strategy never produced a good fit for the empirical frequencies of syllable occurrence for a matrix size of 500 × 500. Even the best-fit models tended to underestimate the number of very common syllable types (Fig. 5, Supplementary Fig. S3). Stronger selective pressures may cause syllables to be more common in these models, leading to better estimates of syllable occurrences and life spans.

Several reasons for the differences between the model results and empirical data are suggested by the patterns in our results. We found that smaller models of directional selection (with 160 000 territories) effectively described the empirical distributions of song occurrences and lifetimes for a wide range of parameters (Supplementary Fig. S14), whereas models with population sizes closer to our estimates (up to 490 000 territories, compared to 250 000 in most of our models, and from 444 000 to 833 000 in these regions) were less effective (Appendix, Table A3, Supplementary Figs S12–S14). A major factor that can explain this discrepancy is the difference in sampling: the community science samples were not randomly distributed, whereas those of our model were. Song recordings were most common at the intersection of high human and high chipping sparrow population densities. This sampling discrepancy could mean that the empirical samples capture a much smaller effective population of chipping sparrows than exists in the entire region. In addition, the range of song rates (~36.5–40 syllables/s) observed in the entire simulated population (before sampling) was much higher and narrower than that observed in the chipping sparrow population (~5–38 syllables/s) (Searfoss, Liu et al., 2020). This supports our intuition that syllable rate is not under directional selection on its own, since we previously observed a wide range of syllable rates in chipping sparrow songs that persisted over many years in nature (Searfoss, Liu et al., 2020).

This model did not reproduce the entire song-learning process and, since there has been a single detailed study on the chipping sparrow song-learning process (Liu & Kroodsma, 2006), we do not know whether chipping sparrows learn identically across their range. However, our results suggest that chipping sparrows learn songs with a preference for one or several song features in at least part of their range. The presence of significant local diversity (Liu & Kroodsma, 2006; Marler & Isaac, 1960) and the distribution of multiple syllables across the country and overlapping in the same region (Searfoss, Liu et al., 2020) also suggests that chipping sparrows do not have a strong conformist bias in their learning. Our results can be compared to those in house finches, which demonstrate content bias, certain syllables are more likely to be learned because of their acoustic features, not because of their frequency of occurrence in the population (Youngblood & Lahti, 2022). In chipping sparrows, since only one syllable is learned per bird, we tracked potential selection on the acoustic features themselves to test whether directional selection favouring the learning of faster songs could explain the observed distribution of syllables.

Sparrow species such as the white-throated sparrows, *Zonotrichia albicollis*, sing in their wintering grounds, allowing for rapid transmission of birdsong after these birds migrate north (Otter et al., 2020). We did not include this effect, since all of our songs from breeding ranges were recorded outside the winter months (Searfoss, Liu et al., 2020), and chipping sparrows are not known to sing regularly during winter months (Liu & Kroodsma, 1999). Song learning on the wintering grounds may explain some of the observed song variation, including songs that were widely dispersed (Supplementary Fig. S15), as birds may have more tutors to learn from. This additional learning step may homogenize the population or increase the strength of a conformity bias (causing common songs to become more common among birds sharing wintering grounds).

The divergences between our model and the community science data suggests that more complex evolutionary pressures or cultural transmission biases might be at play, such as performance trade-offs or differing selection pressures for tutor selection compared to mate choice, which could be integrated into the model for future research. One such explanation is a hypothesized performance trade-off in chipping sparrow song between frequency bandwidth and the rate of syllable delivery (Podos, 1996, 1997); due to physiological constraints, a high-performance song might have a large frequency bandwidth but a slower syllable rate, or a faster syllable rate but a reduced frequency bandwidth. In this case, directional selection likely occurs on multiple axes and operates on both traits at once. If there is a trade-off between two song parameters under selection, we would not expect to see tight distributions of a single syllable parameter (syllable rate or frequency modulation) as there will be a boundary along which the properties are balanced. Furthermore, although long-lived syllables had significantly shorter durations overall, we found a wide distribution of mean syllable durations, implying that both long and short syllables can persist over time.

It is difficult to determine whether a certain song feature is being favoured by directional selection without corresponding field experiments. We framed directional tutor selection in our model such that a parameter was the determining factor for the learned syllable type. As a result, our simulations only suggest that a song characteristic could be under directional selection, not that syllable rate in particular is under selection. To test which song features might be favoured in learning and tutor selection, we propose playback experiments to determine whether there is a difference in juveniles' responses to recordings of different song rates, frequency bandwidths and syllable complexities as well as to historical versus current song recordings (as in Derryberry, 2007). These results could then be compared to females' responses to determine whether tutor selection and mate choice favour similar song characteristics. Ideally, this would be carried out at multiple locations across the chipping sparrow's range, given the geographical patterns observed earlier. We aim to extend this model to incorporate content bias more broadly, allowing selection on the syllables themselves rather than on aspects of syllable production such as syllable rate, as this may better align with empirical data in which buzz syllable types are long-lived. To execute this extension to the model, it would be necessary to create a measure of syllable quality to drive tutor choice.

We demonstrate that coupling an agent-based model with analyses of community science data is a tool to better understand the evolution of behaviour in a songbird. By developing a model of cultural transmission of song and comparing patterns produced by three learning strategies to those found in our empirical data, we demonstrate that the observed distributions of chipping sparrow syllable types show evidence of transmission bias. In particular, our results are indicative of a song-learning strategy in which tutor selection is under directional selection pressure in the chipping sparrow population, with juveniles preferentially selecting tutors with certain song features, and in which copy errors or invention rates are quite low (<0.1%). While our simulation does not specify the specific features of syllables or songs that are under selection, we found that neutral song-learning processes and conformity-biased learning, both of which have been observed in other species (Lachlan et al., 2018; Slater, 1986), cannot explain the distribution of songs observed in chipping sparrows. Despite their deceptively simple song, our computational analyses suggest that chipping sparrows appear to be exhibiting learning biases and complex cultural transmission patterns, warranting further investigation in the field.

Author Contributions

N.C., Y.P. and A.M.S. conceived and designed the project. Y.P. and A.M.S. coded the model and conducted all analyses with assistance from N.C. Y.P., A.M.S. and N.C. wrote the manuscript.

Declaration of Interest

We declare no competing interests.

Data Availability

All code is available at https://github.com/CreanzaLab/ChippingSparrowCulturalEvolutionModel. For the catalogue numbers, database, recordist, URL and licence for the 820 song files, see Supplementary Table S1. For the metadata including recording latitudes and longitudes and the eight song features (all log-transformed except mean stereotypy of repeated syllables and the standard deviation of note frequency modulation), see Supplementary Table S2.

Acknowledgments

We are indebted to the contributors and maintainers of the following databases of birdsong: Xeno-canto, Macaulay Library at the Cornell lab of Ornithology and Borror Laboratory of Bioacoustics. We are grateful to Wan-chun Liu for his previous studies on chipping sparrows and the contribution of his field recordings of chipping sparrow songs. We thank the members of the Creanza Laboratory for their feedback. Y.P and N.C. were supported by the John Templeton Foundation (62187). N.C. was supported by the U.S. National Science Foundation (IOS-2327982). A.M.S was supported by the U.S. National Science Foundation Graduate Research Fellowship Program under Grant DGE-1445197, and all authors received support from Vanderbilt University.

Supplementary Material

Supplementary material associated with this article is available, in the online version, at https://doi.org/10.1016/j.anbehav.2024.01.009.

References

Akçay, Ç., & Beecher, M. D. (2015). Team of rivals in chipping sparrows? A comment on Goodwin & Podos. *Biology Letters, 11*(7), Article 20141043. https://doi.org/10.1098/rsbl.2014.1043

Akashi, H., & Schaeffer, S. W. (1997). Natural selection and the frequency distributions of ‘silent’ DNA polymorphism in *Drosophila. Genetics, 146*(1), 295–307.

Aplin, L. (2016). Understanding the multiple factors governing social learning and the diffusion of innovations. *Current Opinion in Behavioral Sciences, 12*, 59–65. https://doi.org/10.1016/j.cobeha.2016.09.003

Bamshad, M., & Wooding, S. P. (2003). Signatures of natural selection in the human genome. *Nature Reviews Genetics, 4*(2), 99–110. https://doi.org/10.1038/nrg999

Bolus, R. T. (2014). Geographic variation in songs of the common yellowthroat. *Auk, 131*(2), 175–185. https://doi.org/10.1642/auk-12-187.1

Borror, D. J. (1959). Songs of the chipping sparrow. *Ohio Journal of Science, 59*(6), 347–356. http://hdl.handle.net/1811/4655.

Bustamante, C. D., Wakeley, J., Sawyer, S., & Hartl, D. L. (2001). Directional selection and the site-frequency spectrum. *Genetics, 159*(4), 1779–1788.

Byers, J., Hebets, E., & Podos, J. (2010). Female mate choice based upon male motor performance. *Animal Behaviour, 79*(4), 771–778. https://doi.org/10.1016/j.anbehav.2010.01.009

Cardoso, G. C., & Atwell, J. W. (2016). Shared songs are of lower performance in the dark-eyed junco. *Royal Society Open Science, 3*(7), Article 160341.

Catchpole, C. K., & Slater, P. J. B. (2003). *Bird song: Biological themes and variations*. Cambridge University Press.

Crozier, G. K. D. (2010). A formal investigation of cultural selection theory: Acoustic adaptation in bird song. *Biology and Philosophy, 25*(5), 781–801. https://doi.org/10.1007/s10539-010-9194-6

Davidson, B. M., Antonova, G., Dlott, H., Barber, J. R., & Francis, C. D. (2017). Natural and anthropogenic sounds reduce song performance: Insights from two emberizid species. *Behavioral Ecology, 28*(4), 974–982. https://doi.org/10.1093/beheco/arx036

Dennhardt, A. J., Duerr, A. E., Brandes, D., & Katzner, T. E. (2015). Integrating citizen-science data with movement models to estimate the size of a migratory golden eagle population. *Biological Conservation, 184*, 68–78. https://doi.org/10.1016/j.biocon.2015.01.003

Derryberry, E. P. (2007). Evolution of bird song affects signal efficacy: An experimental test using historical and current signals. *Evolution, 61*(8), 1938–1945.

Derryberry, E. P. (2011). Male response to historical and geographical variation in bird song. *Biology Letters, 7*(1), 57–59.

Ellers, J., & Slabbekoorn, H. (2003). Song divergence and male dispersal among bird populations: A spatially explicit model testing the role of vocal learning. *Animal Behaviour, 65*(4), 671–681.

Ford, M. J. (2002). Applications of selective neutrality tests to molecular ecology. *Molecular Ecology, 11*(8), 1245–1262.

Goodenough, A. E., Little, N., Carpenter, W. S., & Hart, A. G. (2017). Birds of a feather flock together: Insights into starling murmuration behaviour revealed using citizen science. *PLoS One, 12*(6), Article e0179277.

Goodwin, S. E., & Podos, J. (2014). Team of rivals: Alliance formation in territorial songbirds is predicted by vocal signal structure. *Biology Letters, 10*(2), Article 20131083.

Goodwin, S. E., & Podos, J. (2015). Reply to Akçay & Beecher: Yes, team of rivals in chipping sparrows. *Biology Letters, 11*(7), Article 20150319. https://doi.org/10.1098/rsbl.2015.0319

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics, 5*(10), Article e1000695.

Harris, J. B. C., & Haskell, D. G. (2013). Simulated birdwatchers' playback affects the behavior of two tropical birds. *PLoS One, 8*(10), Article e77902.

Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes can produce maladaptive losses – The Tasmanian case. *American Antiquity, 69*(2), 197–214. https://doi.org/10.2307/4128416

Hoppitt, W., & Laland, K. N. (2013). *Social learning: An introduction to mechanisms, methods, and models*. Princeton University Press.

Ju, C., Geller, F. C., Mundinger, P. C., & Lahti, D. C. (2019). Four decades of cultural evolution in house finch songs. *Auk, 136*(1), Article uky012. https://doi.org/10.1093/auk/uky012

Kaluthota, C., Brinkman, B. E., dos Santos, E. B., & Rendall, D. (2016). Transcontinental latitudinal variation in song performance and complexity in house wrens (*Troglodytes aedon*). *Proceedings of the Royal Society B: Biological Sciences, 283*(1824), Article 20152765. https://doi.org/10.1098/rspb.2015.2765

Kandler, A., & Laland, K. N. (2009). An investigation of the relationship between innovation and cultural diversity. *Theoretical Population Biology, 76*(1), 59–67.

Kroodsma, D. (2017). Birdsong performance studies: A contrary view. *Animal Behaviour, 125*, e1–e16. https://doi.org/10.1016/j.anbehav.2016.06.015

Kroodsma, D., & Miller, E. H. (Eds.). (1996). *Ecology and evolution of acoustic communication in birds*. Cornell University Press.

Kroodsma, D., Miller, E. H., & Ouellet, H. (1982). *Acoustic communication in birds: Song learning & its consequences*. Academic Press.

Kryazhimskiy, S., & Plotkin, J. B. (2008). The population genetics of dN/dS. *PLoS Genetics, 4*(12), Article e1000304.

Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications, 9*(1), Article 2417.

Lachlan, R. F., & Slater, P. J. B. (2003). Song learning by chaffinches: How accurate, and from where? *Animal Behaviour, 65*(5), 957–969.

Leitner, S., & Catchpole, C. K. (2004). Syllable repertoire and the size of the song control system in captive canaries (*Serinus canaria*). *Journal of Neurobiology, 60*(1), 21–27.

Liu, W.-C. (2001). *Development, variation, and use of songs by chipping sparrows* (Ph.D. thesis). University of Massachusetts Amherst. Doctoral Dissertations Available from: Proquest. AAI3012159 https://scholarworks.umass.edu/dissertations/AAI3012159.

Liu, W.-C. (2004). The effect of neighbours and females on dawn and daytime singing behaviours by male chipping sparrows. *Animal Behaviour, 68*(1), 39–44. https://doi.org/10.1016/j.anbehav.2003.06.022

Liu, W.-C., & Kroodsma, D. E. (1999). Song development by chipping sparrows and field sparrows. *Animal Behaviour, 57*(6), 1275–1286.

Liu, W.-C., & Kroodsma, D. E. (2006). Song learning by chipping sparrows: When, where, and from whom. *Condor, 108*(3), 509–517.

Liu, W.-C., & Nottebohm, F. (2007). A learning program that ensures prompt and versatile vocal imitation. *Proceedings of the National Academy of Sciences of the United States of America, 104*(51), 20398–20403.

Lynch, A., & Baker, A. J. (1993). A population memetics approach to cultural evolution in chaffinch song: Meme diversity within populations. *American Naturalist, 141*(4), 597–620. https://doi.org/10.1086/285493

Lynch, A., & Baker, A. J. (1994). A population memetics approach to cultural evolution in chaffinch song: Differentiation among populations. *Evolution, 48*(2), 351–359. https://doi.org/10.2307/2410097

Lynch, A., Plunkett, G. M., Baker, A. J., & Jenkins, P. F. (1989). A model of cultural evolution of chaffinch song derived with the meme concept. *American Naturalist, 133*(5), 634–653.

Marler, P., & Isaac, D. (1960). Physical analysis of a simple bird song as exemplified by the chipping sparrow. *Condor, 62*(2), 124–135.

Marler, P., & Peters, S. (1982). Developmental overproduction and selective attrition: New processes in the epigenesis of birdsong. *Developmental Psychobiology, 15*(4), 369–378.

Marler, P., & Tamura, M. (1962). Song 'dialects' in three populations of white-crowned sparrows. *Condor, 64*(5), 368–377. https://doi.org/10.2307/1365545

Mason, N. A., Burns, K. J., Tobias, J. A., Claramunt, S., Seddon, N., & Derryberry, E. P. (2017). Song evolution, speciation, and vocal learning in passerine birds. *Evolution, 71*(3), 786–796. https://doi.org/10.1111/evo.13159

Mcgregor, P. K., & Krebs, J. R. (1982). Song types in a population of great tits (*Parus major*): Their distribution, abundance and acquisition by individuals. *Behaviour, 79*(2–4), 126–152.

Newson, S. E., Moran, N. J., Musgrove, A. J., Pearce-Higgins, J. W., Gillings, S., Atkinson, P. W., Miller, R., Grantham, M. J., & Baillie, S. R. (2016). Long-term changes in the migration phenology of UK breeding birds detected by large-scale citizen science recording schemes. *Ibis, 158*(3), 481–495. https://doi.org/10.1111/ibi.12367

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics, 39*(1), 197–218.

Otter, K. A., Mckenna, A., LaZerte, S. E., & Ramsay, S. M. (2020). Continent-wide shifts in song dialects of white-throated sparrows. *Current Biology, 30*(16), 3231–3235.

Parker, K. A., Anderson, M. J., Jenkins, P. F., & Brunton, D. H. (2012). The effects of translocation-induced isolation and fragmentation on the cultural evolution of bird song. *Ecology Letters, 15*(8), 778–785.

Pepperell, C. S., Casto, A. M., Kitchen, A., Granka, J. M., Cornejo, O. E., Holmes, E. C., Birren, B., Galagan, J., & Feldman, M. W. (2013). The role of selection in shaping diversity of natural *M. tuberculosis* populations. *PLoS Pathogens, 9*(8), Article e1003543. https://doi.org/10.1371/journal.ppat.1003543

Podos, J. (1996). Motor constraints on vocal development in a songbird. *Animal Behaviour, 51*(5), 1061–1070. https://doi.org/10.1006/anbe.1996.0107

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution, 51*(2), 537–551. https://doi.org/10.1111/j.1558-5646.1997.tb02441.x

Podos, J. (2017). Birdsong performance studies: Reports of their death have been greatly exaggerated. *Animal Behaviour, 125*, e17–e24. https://doi.org/10.1016/j.anbehav.2016.12.010

Podos, J., & Nowicki, S. (2004). Performance limits on birdsong. In P. Marler, & H. Slabbekoorn (Eds.), *Nature's music: The science of birdsong* (pp. 318–342). https://doi.org/10.1016/b978-012473070-0/50014-1

Robinson, O. J., Ruiz-Gutierrez, V., Fink, D., Meese, R. J., Holyoak, M., & Cooch, E. G. (2018). Using citizen science data in integrated population models to inform conservation. *Biological Conservation, 227*, 361–368. https://doi.org/10.1016/j.biocon.2018.10.002

Searcy, W. A., & Andersson, M. (1986). Sexual selection and the evolution of song. *Annual Review of Ecology and Systematics, 17*(1), 507–533. https://doi.org/10.1146/annurev.es.17.110186.002451

Searfoss, A. M., Liu, W.-C., & Creanza, N. (2020). Geographically well-distributed citizen science data reveals range-wide variation in the chipping sparrow's simple song. *Animal Behaviour, 161*, 63–76.

Searfoss, A. M., Pino, J. C., & Creanza, N. (2020). Chipper: Open-source software for semi-automated segmentation and analysis of birdsong and other natural sounds. *Methods in Ecology and Evolution, 11*(4), 524–531. https://doi.org/10.1111/2041-210x.13368

Sibley, D. (2011). *The proper use of playback in birding*. https://www.sibleyguides.com/2011/04/the-proper-use-of-playback-in-birding/. (Accessed 6 December 2019).

Silvertown, J., Cook, L., Cameron, R., Dodd, M., McConway, K., Worthington, J., Skelton, P., Anton, C., Bossdorf, O., Baur, B., Schilthuizen, M., Fontaine, B., Sattmann, H., Bertorelle, G., Correia, M., Oliveira, C., Pokryszko, B., Ożgo, M., Stalažs, A., … Juan, X. (2011). Citizen science reveals unexpected continental-scale evolutionary change in a model organism. *PLoS One, 6*(4), Article e18927. https://doi.org/10.1371/journal.pone.0018927

Slater, P. J. B. (1986). The cultural transmission of bird song. *Trends in Ecology & Evolution, 1*(4), 94–97.

Swanson, H. M., Kinney, B., & Cruz, A. (2004). Breeding biology of the chipping sparrow in ponderosa pine forests of the Colorado Front Range. *Wilson Bulletin, 116*(3), 246–251. https://doi.org/10.1676/03-105

Thorpe, W. H. (1958). The learning of song patterns by birds, with especial reference to the song of the chaffinch *Fringilla coelebs*. *Ibis, 100*(4), 535–570. https://doi.org/10.1111/j.1474-919x.1958.tb07960.x

Tomasello, M., Kruger, A. C., & Ratner, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences, 16*(3), 495–511.

Vargas-Castro, L. E., Sánchez, N. V., & Barrantes, G. (2012). Repertoire size and syllable sharing in the song of the clay-colored thrush (*Turdus grayi*). *Wilson Journal of Ornithology, 124*(3), 446–453. https://doi.org/10.1676/11-044.1

Watson, D. M., Znidersic, E., & Craig, M. D. (2019). Ethical birding call playback and conservation. *Conservation Biology, 33*(2), 469–471.

Wheelwright, N. T., Swett, M. B., Levin, I. I., Kroodsma, D. E., Freeman-Gallant, C. R., & Williams, H. (2008). The influence of different tutor types on song learning in a natural bird population. *Animal Behaviour, 75*(4), 1479–1493.

Whiten, A. (2017). A second inheritance system: The extension of biology through culture. *Interface Focus, 7*(5), Article 20160142.

Will, T., Stanton, J. C., Rosenberg, K. V., Panjabi, A., Camfield, A., Shaw, A., Thogmartin, W. E., & Blancher, P. J. (2020). *Handbook to the partners in flight population estimates database* (Vol. 7.1). Partners in Flight Science Committee Version 3.1. .

Williams, H., Levin, I. I., Ryan Norris, D., Newman, A. E. M., & Wheelwright, N. T. (2013). Three decades of cultural evolution in Savannah sparrow songs. *Animal Behaviour, 85*(1), 213–223. https://doi.org/10.1016/j.anbehav.2012.10.028

Williamson, S. H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R., & Bustamante, C. D. (2005). Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proceedings of the National Academy of Sciences of the United States of America, 102*(22), 7882–7887.

Youngblood, M. (2019). Conformity bias in the cultural transmission of music sampling traditions. *Royal Society Open Science, 6*(9), Article 191149.

Youngblood, M., & Lahti, D. C. (2022). Content bias in the cultural evolution of house finch song. *Animal Behaviour, 185*, 37–48.

Zhu, L., & Bustamante, C. D. (2005). A composite-likelihood approach for detecting directional selection from DNA sequence data. *Genetics, 170*(3), 1411–1421.

Appendix

Table A1
Best parameters for different models in the New York region

Model type	Best parameters		Counts *P* value	Lifetimes *P* value
	Learning error rate	Dispersal fraction		
Neutral	10^{-5}	0.3	5.63×10^{-13}	9.46×10^{-6}
Conformity (c=1.2)	0.001	0.1	4.56×10^{-8}	8.20×10^{-4}
Conformity (c=1.4)	0.01	0.6	3.22×10^{-11}	1.48×10^{-4}
Conformity (c=1.6)	0.1	0.9	6.53×10^{-13}	6.02×10^{-5}
Conformity (c=1.8)	10^{-5}	0.3	3.44×10^{-13}	3.15×10^{-5}
Conformity (c=2)	0.01	0.4	3.44×10^{-13}	3.15×10^{-5}
Novelty (c=0.8)	10^{-5}	0.1	1.1×10^{-15}	1.09×10^{-6}
Directional	10^{-6}	0.5	0.00441	**0.0123**

P values representing simulated data statistically indistinguishable from the empirical data in bold. For best parameters of other regions, see Table A2.

Table A2
Best parameters for different models and regions

Model type	Region	Best parameters		Counts *P* value	Lifetimes *P* value
		Learning error rate	Dispersal fraction		
Neutral	Entire range	0.001	1.0	1.76×10^{-42}	9.25×10^{-4}
	NY	10^{-5}	0.3	5.63×10^{-13}	9.46×10^{-6}
	OH/MI	0.0001	0.1	6.34×10^{-22}	3.16×10^{-8}
	New England	0.0001	0.6	4.71×10^{-28}	3.63×10^{-9}
Conformity (c=1.2)	Entire range	0.001	0.1	6.91×10^{-17}	**0.109**
	NY	0.01	0.4	3.44×10^{-13}	3.15×10^{-5}
	OH/MI	0.001	0.1	1.07×10^{-10}	1.35×10^{-4}
	New England	0.001	0.1	5.82×10^{-13}	3.63×10^{-5}
Directional	Entire range	0.01	0.5	3.08×10^{-4}	**0.136**
	NY	10^{-6}	0.5	0.00441	**0.123**
	OH/MI	10^{-6}	0.5	0.00367	**0.0762**
	New England	0.0001	0.5	1.41×10^{-4}	**0.0545**

OH/MI: Ohio/Michigan. All models are for a matrix size of 500 × 500. *P* values representing simulated data statistically indistinguishable from the empirical data in bold.

Table A3
Best parameters for different models and different matrix sizes for the New York region

Model type	Matrix dimension	Best parameters		Counts *P* value	Lifetimes *P* value
		Learning error rate	Dispersal fraction		
Neutral	400	0.0001	1.0	1.62×10^{-12}	1.73×10^{-4}
	500	10^{-5}	0.3	5.63×10^{-13}	9.46×10^{-6}
	600	0.001	0.6	1.84×10^{-14}	2.50×10^{-6}
	700	10^{-5}	0.8	2.15×10^{-15}	8.32×10^{-7}
Conformity (c=2)	400	0.001	0.7	1.67×10^{-10}	4.36×10^{4}
	500	0.01	0.4	3.44×10^{-13}	3.15×10^{-5}
	600	10^{-5}	0.2	3.37×10^{-14}	5.22×10^{-6}
	700	0.001	0.6	1.82×10^{-15}	6.15×10^{-7}
Directional	400	10^{-4}	0.4	**0.189**	**0.345**
	500	10^{-6}	0.5	0.00441	**0.123**
	600	10^{-5}	0.5	1.73×10^{-5}	0.00871
	700	0.001	0.5	7.90×10^{-9}	0.00222

P values representing simulated data statistically indistinguishable from the empirical data in bold.

Table A4
Best parameters for different models of syllable learning error

Model type	Homoplasy	Best parameters		Counts *P* value	Lifetimes *P* value
		Learning error rate	Dispersal fraction		
Neutral	No	10^{-5}	0.3	5.63×10^{-13}	9.46×10^{-6}
	Yes	0.0001	0.6	8.58×10^{-14}	1.134×10^{-5}
Conformity (c=2)	No	0.01	0.4	3.44×10^{-13}	3.15×10^{-5}
	Yes	10^{-5}	0.7	1.20×10^{-14}	2.31×10^{-5}
Directional	No	10^{-6}	0.5	0.00441	**0.123**
	Yes	0.01	1.0	0.00270^{3}	**0.0809**

In the standard model, errors result in novel syllable invention; in the model with homoplasy, syllables are chosen from a fixed set of 500 syllables, identical with the syllables from the model initialization. All values represent those for the New York region with a matrix size of 500 × 500. *P* values representing simulated data statistically indistinguishable from the empirical data in bold.

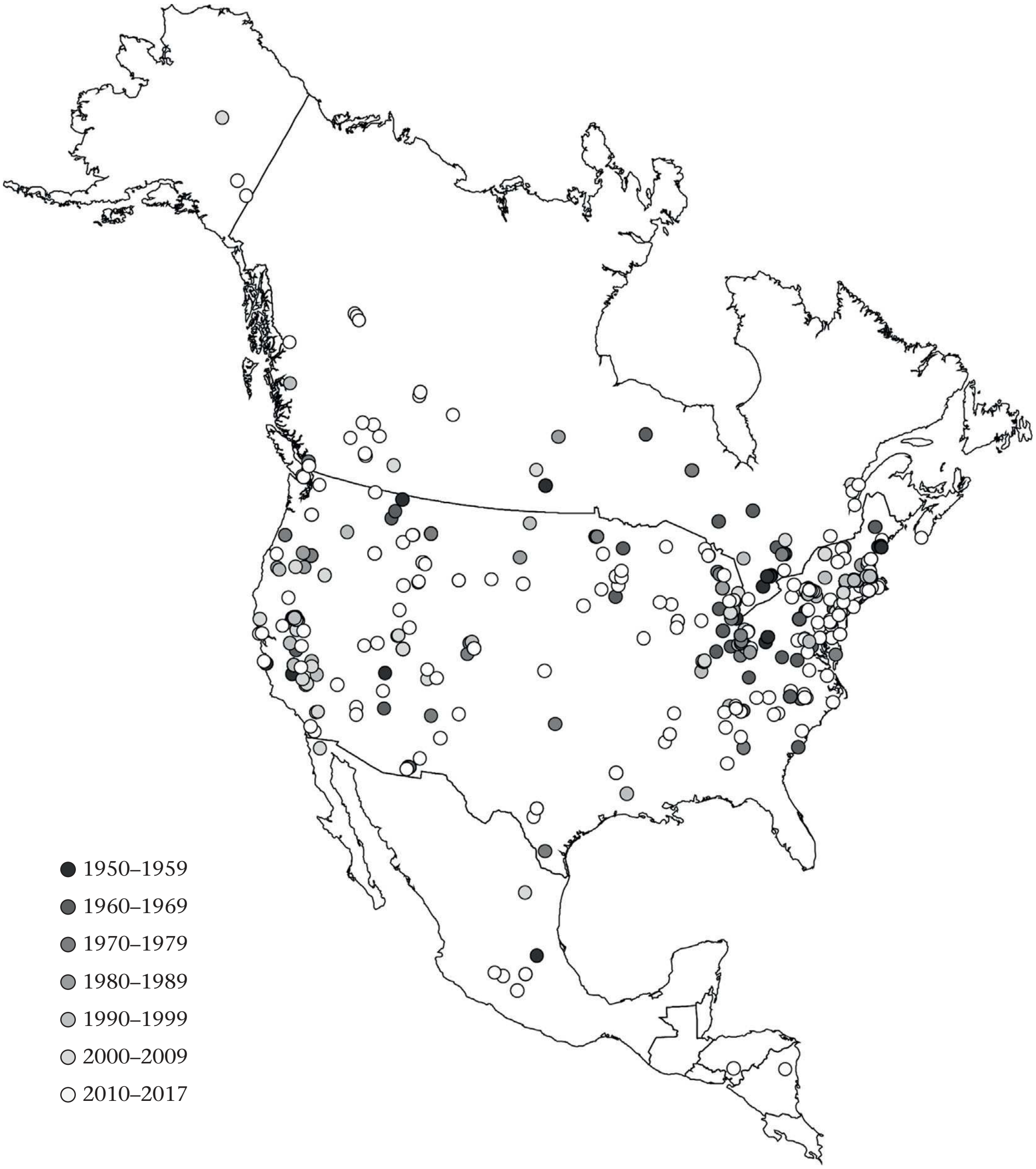


Figure A1. Locations and decades of collection for the 820 chipping sparrow recordings. The geographical map was made using ArcMap v.10.7 and country outlines are from ESRI (Environmental Systems Research Institute, Redlands, CA, U.S.A., map projection, North_America_Lambert_Conformal_Conic, WKID: 102009).