

Review



**Cite this article:** Williams H, Lachlan RF. 2021 Evidence for cumulative cultural evolution in bird song. *Phil. Trans. R. Soc. B* **377**: 20200322.  
<https://doi.org/10.1098/rstb.2020.0322>

Received: 13 July 2021  
Accepted: 5 October 2021

One contribution of 17 to a discussion meeting issue ‘The emergence of collective knowledge and cumulative culture in animals, humans and machines’.

**Subject Areas:**  
behaviour, cognition, evolution

**Keywords:**  
bird song, categorization, syntax, learning biases, segmentation, cumulative cultural evolution

**Authors for correspondence:**  
Heather Williams  
e-mail: [hwilliams@williams.edu](mailto:hwilliams@williams.edu)  
Robert F. Lachlan  
e-mail: [robert.lachlan@rhul.ac.uk](mailto:robert.lachlan@rhul.ac.uk)

# Evidence for cumulative cultural evolution in bird song

Heather Williams<sup>1</sup> and Robert F. Lachlan<sup>2</sup>

<sup>1</sup>Biology Department, Williams College, Williamstown, MA 01267, USA  
<sup>2</sup>Department of Psychology, Royal Holloway University of London, London TW20 0EX, UK

HW, 0000-0003-0947-6243

In studies of cumulative cultural evolution in non-human animals, the focus is most often on incremental changes that increase the efficacy of an existing form of socially learned behaviour, such as the refinement of migratory pathways. In this paper, we compare the songs of different species to describe patterns of evolution in the acoustic structure of bird songs, and explore the question of what building blocks might underlie cumulative cultural evolution of bird song using a comparative approach. We suggest that three steps occurred: first, imitation of independent sounds, or notes, via social learning; second, the formation of categories of note types; and third, assembling note types into sequences with defined structures. Simple sequences can then be repeated to form simple songs or concatenated with other sequences to form segmented songs, increasing complexity. Variant forms of both the notes and the sequencing rules may then arise due to copy errors and innovation. Some variants may become established in the population because of learning biases or selection, increasing signal efficiency, or because of cultural drift. Cumulative cultural evolution of bird songs thus arises from cognitive processes such as vocal imitation, categorization during memorization and learning biases applied to basic acoustic building blocks.

This article is part of a discussion meeting issue ‘The emergence of collective knowledge and cumulative culture in animals, humans and machines’.

## 1. Introduction

Bird songs have many of the characteristics of speech and language: they are learned from conspecific models, consist of sounds that can be divided into distinct categories, and are assembled into sequences that may vary according to syntactic rules on the level of syllables, phrases, songs or bouts [1–6]. As do humpback whale songs [7], these socially learned sounds and sequences evolve as a population’s songs change over time [8–13]. This accumulation of changes leads to differentiation of vocal forms across populations [14–16] and to distinct functions for different parts of the repertoire [11,17,18]. As far as we can tell, however, the ‘meanings’ encoded in such songs are relatively restricted: we know that singers impart information about their species, as in the Siberian greenish warblers (the *Phylloscopus trochiloides* complex), whose songs differ in frequency and temporal parameters [19]; about population identity, as is the case for white-crowned sparrows (*Zonotrichia leucophrys*) where geographically separate groups each sing a distinct dialect [20]; and about individual identity, as when red-winged blackbird (*Agelaius phoeniceus*) males respond more aggressively to strangers’ songs from the same population than to their neighbours’ songs [21]. Songs also signal birds’ availability as potential mates and their claim to a territory, as demonstrated by studies that remove males and replace them with speakers broadcasting songs [22]. Information about motivation and aggressive intentions is provided by singing the song at lower amplitude [23], overlapping another’s song [24], or matching another’s song type [25]. The status of the bond between mates is encoded in the

coordination of songs of duetting birds [26]. How proficiently specific sections of a song are performed provides cues to a bird's age, status or quality [27–29]. Variation in the syntax of learned vocalizations can encode information about social context [30–32], the probability of movement [33], the presence of food [34] and the presence and type of predators [35]. Nevertheless, learned bird vocalizations appear to be far less complex and have fewer distinctions of meaning than are present in human language.

As for any socially learned behaviour, imprecise transmission introduces variation into bird songs. Within individuals' songs this variation is observable as innovations and copy errors, either in the basic components or in the assembly of those components into sequences. The variants may have different 'efficiency', either in terms of communication, or in terms of learnability, or in terms of physiological costs [36]. As is the case for genetic evolution, either drift or selection (or both) may operate to increase or decrease the prevalence of particular learned variants within a population [37,38], resulting in cultural evolution. If this process is repeated, the song forms used by a population may continue to change and evolve, but successive changes alone are not cumulative cultural evolution. The criteria for cumulative cultural evolution [39] are satisfied when (a) changes in a population's song accumulate in succession and (b) the new features that are adopted at different time points increase the complexity and efficiency of the song.

Many songbird species learn their songs during a restricted critical period early in life, singing a crystallized song thereafter [40–43]. Because learning is not continuous, critical period learners allow us to distinguish between the social learning processes that underlie cultural evolution and the consequences for the adult of learning a specific form of the behaviour. In this review, we look at how song learning generates variation within acoustic space, how acoustic categories are formed, and how those categories serve as building blocks that are assembled into simple sequences. We suggest that the resulting simple sequences, or syllables, in turn, serve as building blocks that are concatenated into songs with two levels of syntactic structure: syllables and segments. In songs with multiple segments, each segment may then come to play a different signalling function because of different cultural evolution mechanisms.

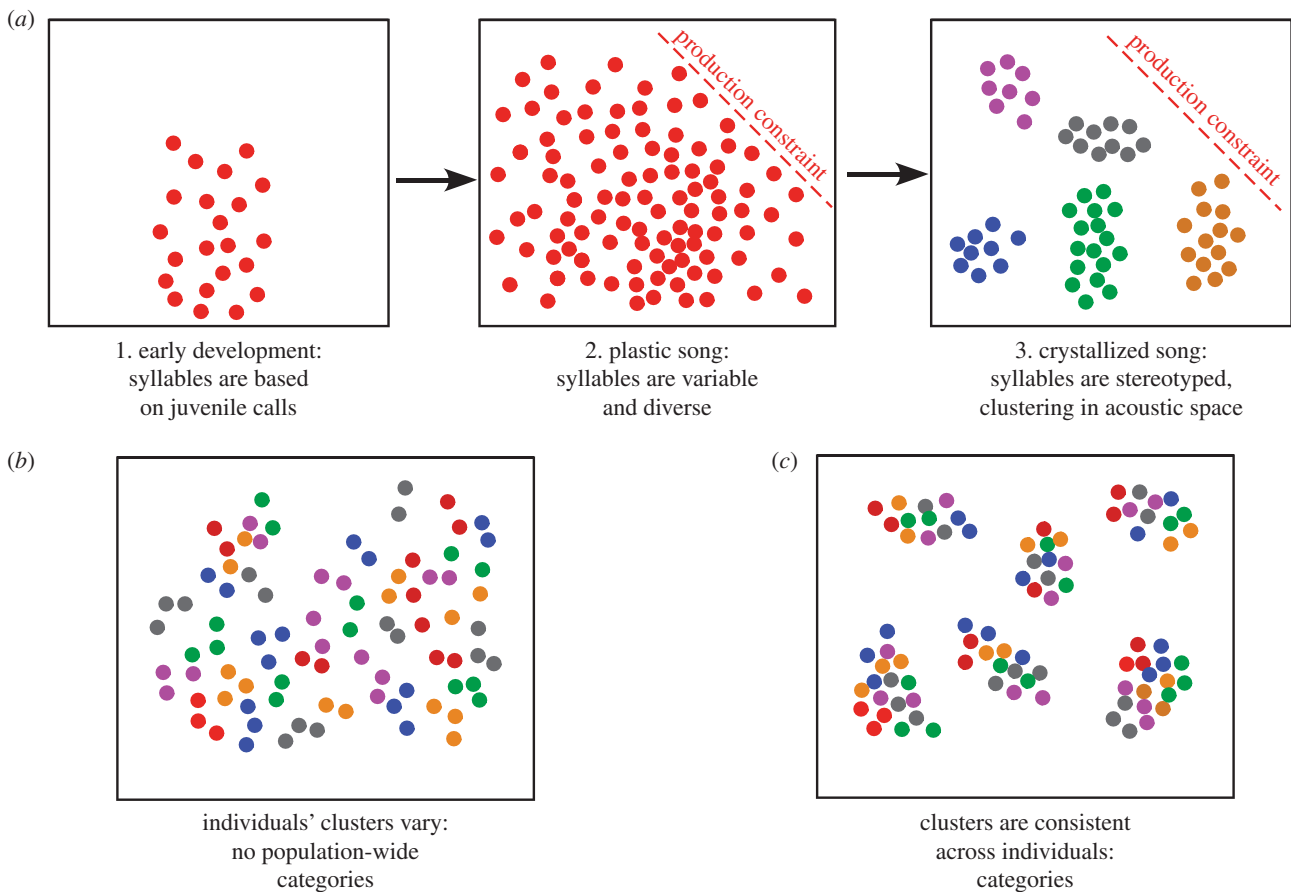
## 2. Song learning, variation and category formation

The developmental stages in the learning of an adult song have been described as (a) memorization, (b) calibrating the vocal organ (subsong) and producing precursor syllables and (c) imitating the memorized sounds (plastic song) [44]. Variation can appear at any stage: as the result of inaccurate storage in memory, because novel sounds arise during the vocal calibration process that then persist through other stages, or as a consequence of inaccurately reproducing the memorized version during plastic song. All of these processes can be considered as introducing 'noise' into the representation of the model for the song, resulting in the use of an expanded acoustic space during the learning process. This variation or noise may or may not be discarded or reconfigured as learning proceeds. Many birds produce more notes, syllables, phrases and songs<sup>1</sup> than they include in

their adult repertoire [45,46], sometimes because they produce more songs than they need during development and sometimes as result of innovations they generate. Which song material is retained by an adult can be influenced by what a bird hears late in the song learning process [42].

The mature birdsong of critical period learners is said to be 'crystallized'; the components (notes and syllables) are produced in a stereotyped manner that is consistent across renditions [47]. Within a population, there may typically be hundreds (and frequently thousands) of different song types. However, the notes and syllables that make up a population's songs often fall into clusters: the 'downsweeps' and 'stacks' of zebra finches (*Taenopygia guttata*) [48,49], the distinct note categories of swamp sparrows (*Melospiza georgiana*) [50], the distinctive middle segment notes of Savannah sparrows (*Passerculus sandwichensis*) [11]. These clusters are, to a greater or lesser extent, separated in acoustic space by characteristics such as frequency, bandwidth, modulation and duration (analogous to the acoustic features that define vowels in human speech) [51]. In the case of human speech, the emergence of categories within acoustic space has been modelled using several approaches, with the outcome defined as 'distinguishability' either in perceptual space (how distinct the categories are to the hearer) or in the motor space used to produce speech sounds (how distinct the articulatory gestures are to the producer) [52–54]. When learning and production cycles are iterated, and outputs are retained in the population on the basis of distinguishability, categories of speech sounds such as vowels emerge. However, these approaches to cultural evolution of speech most often either implicitly or explicitly include the assumption that distinct categories map onto separate meanings, and that the functional demands of the signalling system—the need to encode more meanings—drive the addition of more categories to the acoustic space used in the vocalizations. In the case of bird songs, there does not appear to be a strong selective pressure to define additional syllable categories or flexible sequences that are associated with distinct meanings. Nevertheless, bird songs have clearly defined note categories, as well as rules for assembling those categories into higher-order structure. How might such syllable categories and sequences arise in the absence of evolutionary pressure for expanded meaning?

One possible mechanism for the formation of distinct acoustic categories is a combination of noise and learning biases. Noise in the memorization and learning process could first expand the acoustic space a bird uses during vocal learning, and learning biases would then define what is retained and what is discarded as the song matures. For example, in zebra finches the initial units of developing songs are nearly identical and similar to noisy begging calls [55]. These units diverge as song learning progresses, guided by memories of the model's song and becoming less noisy and more acoustically defined in the process [51]. Schematically, the trajectory of song learning thus starts with sounds occupying a broad acoustic space. As learning progresses note types emerge, forming clusters, or categories, within that space (figure 1*a*). In this scenario, each individual's note types might be idiosyncratic. If the note types of several individuals produce overlapping categories, a population's syllables, as a whole, could occupy most of the acoustic space (figure 1*b*), and there would be no distinct acoustic categories within the songs of a population. Alternatively, the notes of all the birds



**Figure 1.** Formation of syllable categories during development. (a) Schematic showing use of acoustic space during song development. Primordial syllables are similar to juvenile calls (often food begging calls). As song learning progresses, the sounds produced diversify and occupy a larger acoustic space, although production constraints may limit which sounds can be sung. As learning proceeds, notes converge and become more stereotyped, forming clusters within the acoustic space. (b) When each individual (denoted by a different colour) sings syllables that fall within narrow regions of acoustic space, and different individuals use different parts of that acoustic space, there are no distinct population-wide categories. (c) Population-wide categories arise when the syllable clusters of different individuals coincide in acoustic space.

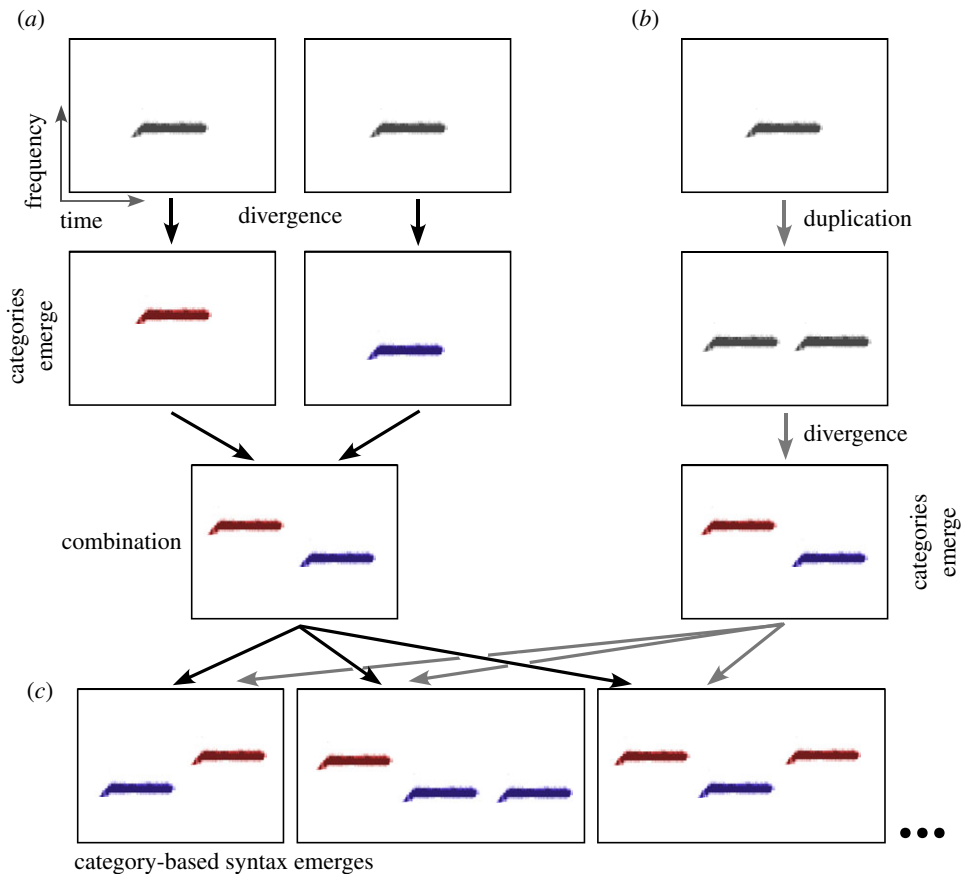
in a population might converge upon the same subsets of acoustic space (figure 1c), forming distinct, population-wide acoustic categories.

In some species, there are innate predispositions for learning sounds that occupy distinct subregions of the acoustic space. Peter Marler's work was the first to investigate such species-specific acoustic categories in detail: swamp sparrows tutored with both swamp and song sparrow (*Melospiza melodia*) syllables include only swamp sparrow syllables in their own songs, regardless of whether those syllables are arranged in species-typical syntax within the tutor songs [56,57]. A bias towards learning the sounds sung by one's own species is also present in other songbirds, such as white-crowned sparrows [58,59]. Birds raised hearing untutored, depauperate songs of their own species [60,61]—or even hearing only recordings of their own developing vocalizations [62]—sing syllables and songs that are closer to the species-typical versions sung by normally reared birds. These results raise the possibility that population-wide categories are the consequence of 'lumpiness' in the acoustic space that a species uses, with birds being genetically predisposed to learn certain subsets or ranges of acoustic features. When categories form they will lie within these species-specific subsets of the possible acoustic space.

Biases in song learning might also be the result of articulatory constraints. Anatomical or neuromuscular constraints on articulatory gestures may delimit the acoustic space a bird uses. For example, birds with larger, more massive beaks tend

to produce smaller frequency ranges within a single sound, as the acoustic filtering provided by changes in beak aperture cannot change as quickly as in birds with smaller beaks [63–65]. Another form of articulatory categorization arises from the fact that songbirds have a two-part syrinx. Each syringeal half is associated with one bronchus, giving the entire syrinx two independent 'voices' [66,67]. The right syringeal half produces high-frequency sounds, while the left side produces low-frequency sounds. This partitioning of frequencies between the two sides of the vocal organ is consistent across a number of species, including canaries (*Serinus canaria*) [68,69], zebra finches [70] and brown-headed cowbirds (*Molothrus ater*) [71]. Specialization for different frequency ranges in the two sides of the syrinx is also present in oilbirds (*Steatornis caripensis*) [72], which are not songbirds and do not learn their vocalizations, suggesting that the distinctive outputs of the two sides of the vocal organ might form a physiological basis for the initial emergence of sound categories.

The examples above describe content-based biases in cultural transmission. Frequency-based learning biases, specifically conformity, guide song development in some songbird species. Lachlan *et al.* [73] compared present-day distributions of syllable types across six swamp sparrow populations with the outcomes of simulations that used different learning biases and parameters. In the absence of a learning bias, neither a low nor a high mutation/innovation rate yielded a good match to field data. A conformist bias



**Figure 2.** Two pathways to category-based syntax. (a) Divergence: notes sung by birds from separate lineages diverge to form different types. When an individual learns and sings both types, an ordered sequence arises. (b) Duplication: a single bird initially sings two duplicate notes, which then diverge from the original form, forming an ordered sequence of two note types. (c) In either case, further sequences arise as inaccurate copying or innovation changes the order or introduces additional copies into the sequence.

resulted in a much better fit to the data than did a demonstrator or a content bias, suggesting that swamp sparrows preferentially learn syllables they have heard sung by more tutors. Such a pattern of transmission tends to remove rare variants from a population and so gives rise to clustered patterns of variation. In such cases, the location of clusters in acoustic space may be arbitrary.

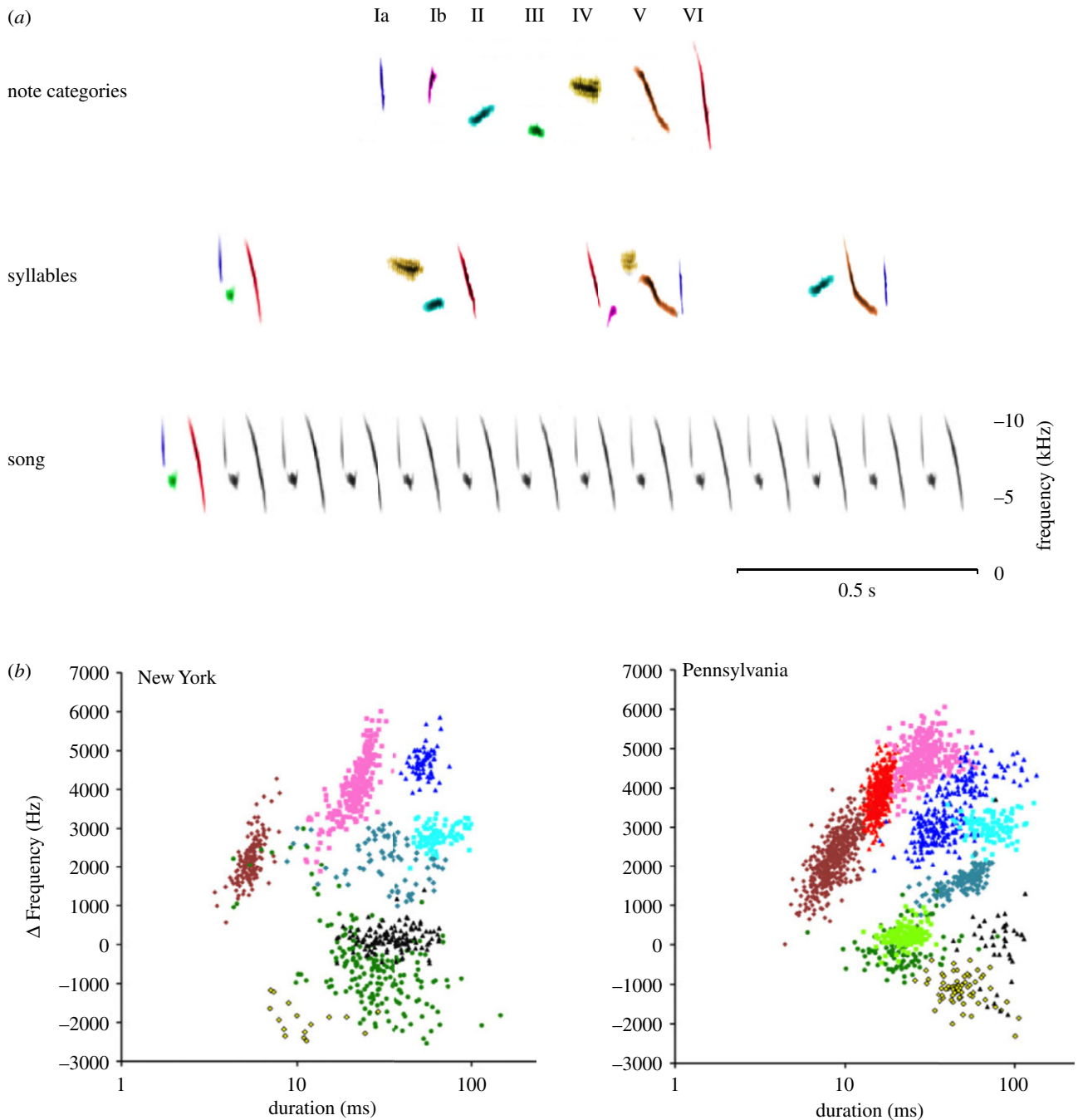
We propose that clusters, or categories, are a building block for the cultural evolution of syntactic structure. Consider a species singing a single learned note type as its vocalization. Two possibilities for the emergence of distinct categories exist: (1) two birds might produce different note types or (2) a single bird could produce more than one note type. Learning, with its attendant variation in the form of noise, copying errors and innovation, makes the first scenario possible: different lineages could produce variations on a single sound (perhaps as simply as one lineage producing a version that emphasizes the output of the left side of the syrinx and another emphasizing a higher frequency produced with the right side). These two variants, diverging over time, could come to distinguish the two lineages (figure 2a). At that point, a bird that by chance learned and sang notes from two different lineages could produce a simple ordered sequence. Alternatively, duplication—repeating a single sound during learning or production—could initially result in a lineage that produces two identical notes (figure 2b). The acoustic characteristics of the two sounds could then diverge within that two-note vocalization. Duplication followed by divergence would result in not only two categories of sounds but also, simultaneously,

in a simple ordered sequence. Once ordered sequences exist, they may be elaborated, either by additional cycles of duplication and divergence, or by learning sequences from more than one tutor and concatenating or mixing them (figure 2c). Either scenario gives rise to the rudiments of syntax.

### 3. Note categories and syllable syntax in swamp sparrow songs

The most complete description of how note categories form and are assembled into higher-order structures within bird songs comes from swamp sparrows. In this species, notes are nearly always simple uninflected frequency sweeps, making it relatively easy to characterize differences between them. Within each population there are a limited number of note clusters: six categories were originally defined by visual scoring of sonograms from a New York population [50], while statistical clustering analyses found evidence for seven or eight categories (figure 3a) [74,75]. While clearly defined within each population, clusters vary between populations in their number and their location in acoustic space [75] (figure 1a). The variation between populations suggests that swamp sparrow note type clusters are not based on underlying innate categories, but emerge from cultural evolutionary processes.

To build syllables from notes, swamp sparrows assemble 2–5 notes into a short sequence (figure 3a) [50]. These syllables are then repeated 10–15 times to form songs. Although



**Figure 3.** Swamp sparrow note categories and syllables. (a) Marler & Pickert [50] originally defined six categories; sub-categories Ia and Ib were split by Lachlan *et al.* [74] based on cluster analyses of acoustic characteristics. Syllables are assembled from 2 to 5 note categories, and songs are formed by repeating a single syllable (individual birds sing 2–5 different songs [74]). (b) Individual notes can be characterized largely by their duration and change in frequency. When notes are clustered using this approach using Gaussian mixture models, populations in New York and Pennsylvania differ in number of clusters and where those clusters are located in acoustic space. After Lachlan *et al.* [73,74].

the algorithm for generating songs from syllables is thus both simple and universal, the rules for assembling notes into syllables can vary across populations. Notes from a given category tend to be used consistently at specific positions within a syllable: for example, in a New York population, short category I notes usually occur only at the beginning of syllables, while longer category VI notes are primarily sung in the final position within syllables [50,74,76]. However, in a Pennsylvania population, there is an additional intermediate-length cluster that can be found either at the start or the end of the syllable [75], while in other populations, the order is reversed [76]. The overarching syntactical rule for generating syllables—concatenate 2–5 notes drawn from different categories—is species-specific,

but the role each constituent note category can play within that structure varies across populations, in a manner reminiscent of human languages.

Swamp sparrow note categories and the assembly of those categories into syllables have both perceptual and motor correlates. In field playback studies, New York birds perceived note types I and VI, which differ in duration, as two distinct categories [77], even though the variation in the stimulus note duration was continuous. When the experiment was repeated in Pennsylvania, the perceptual boundaries varied depending on the position of the note: at the beginning of syllables, short (type I) and intermediate types were perceived as part of the same category, while at the end of the syllable, long

(type VI) and intermediate types were perceived as part of the same category. The perceptual boundaries found in behavioural studies were consistent with those described in parallel electrophysiological studies [78].

Together, these results suggest a hierarchical organization of learning. Social learning of note types first led to the cultural evolution of population-specific clusters. Once these clusters were established, we suggest that birds represented these note clusters as perceptual categories, making sense of regularities in their auditory environment, much as infant humans learn phonemic categories [79]. This categorization of notes then fed back into learning, further separating and defining note type clusters. Then, as some syllable types became more common than others within the population, it provided birds with an additional regularity: the sequence of note categories within the syllable. These ‘phonemic rules’ were then learned as well. Critically, these different aspects of learning appear to build upon each other, sometimes adding complexity to birds’ knowledge of their population’s songs, even when the songs themselves did not become more complex.

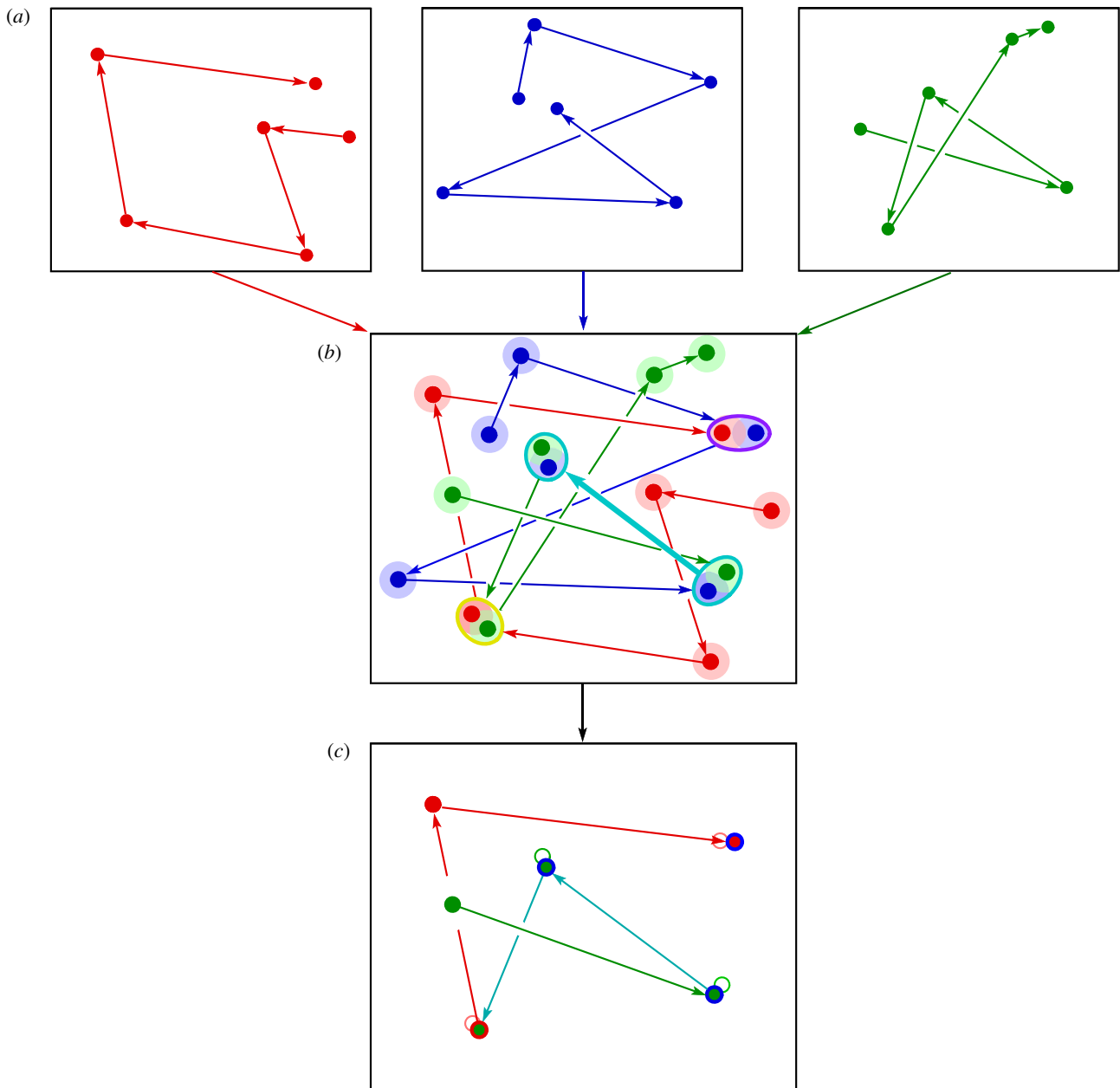
In this scenario, cultural transmission of syllables favours the emergence of note type clusters. Increasing sophistication of swamp sparrow perception, such as conformist biases, would shape the cultural transmission of swamp sparrow song, resulting in a relatively small number of syllable types becoming very common [73], thereby favouring the establishment of note type clusters. This conformist bias may be related to the fact that swamp sparrows recognize and respond most strongly to more ‘typical’ versions of syllables—those that are centrally located in the acoustic space occupied by a syllable type [80]. In field playbacks, males respond more aggressively to typical syllables, and in the laboratory, females primed with oestradiol give more copulation solicitation displays to typical syllables. This finding has been connected to the developmental stress hypothesis, the idea that receivers might assess singers based on how well they were able to learn their songs [81]. Interestingly, a theoretical model found that in order for learning accuracy to serve as a marker of developmental stress, both receivers and signallers need to categorize the songs that they hear early in life [82]. The process of categorization might thus form the basis for the evolution and refinement of units carrying information about fitness within a communication signal. However, as far as we know, neither the different note categories nor the different syllables formed by combinatorial assembly of note categories carry different referential meanings. Note categories and syntactic rules within learned communication systems need not be predicated on the need to generate more signals to encode more meanings—simple learning biases alone may be sufficient to drive the evolution of structure, as is schematically illustrated in figure 4.

#### 4. Interactions between the formation of note categories and higher-level structure

As we observe them now, 50 Myr after passerine songbirds first emerged, one prominent characteristic of bird songs is their species-characteristic note categories and syntax. For example, chaffinches (*Fringilla coelebs*) [83,84] and white-crowned sparrows [85] sing distinctive whistles, trills and note combinations at the beginnings and ends of their songs,

zebra finches sing longer syllables at the ends of each repeated motif [86], northern mockingbirds (*Mimus polyglottos*) repeat each phrase three times [87], and Savannah sparrows start each song with an accelerating sequence of high-frequency downsweeps [88]. Studies of the development of deaf birds and of birds raised in isolation find that songs with aspects of species-specific sequencing develop in the absence of a model, and so we know that some elements of a species’ normal syntax are not learned [89–92]. These unlearned components of syntax could be based upon perceptual predispositions, motor biases favouring the chaining of particular articulatory gestures, or both. The idea that syllable categories and syntactic rules interact is further supported by work on song development. Young zebra finches use two distinct song learning strategies: (a) simultaneous development of the phonology and sequence of notes or (b) first producing repeated identical syllables, some of which are modified and arranged to form a sequence [93]. Individual birds may use either strategy, or may combine the two strategies. When zebra finches are challenged to learn new material late in song development they first alter existing syllables to match models’ phonology, and then attempt to adjust the order of those syllables [94]. Crystallized canary songs consist of a series of phrases, with each phrase made up of a string of identical syllables [95,96]. When young canaries were tutored with ‘glissando’ songs that consisted of a long series of otherwise identical syllables that varied in frequency, forming a long, smooth transition from high to low, they initially copied these glissandos accurately [97]. However, as song learning progressed, species-specific phrase structure and categories emerged: the glissando string was broken up into phrases, and each phrase became a short repeated string of one syllable drawn from the original continuum of notes. Over the course of song learning, these canaries first accurately copied an atypical song, and then altered it by reducing a continuous series of note types to categories, and repeating each category of note to form species-typical phrases.

As song learning studies illustrate, note and syllable categories do not develop completely independently. A single utterance may be repeated, giving rise to a string of initially identical notes or syllables that differentiate to form a sequence. Alternatively, a continuum of notes may converge to form a single type. Two different sounds can also be integrated during development to form a note: northern cardinals (*Cardinalis cardinalis*) produce long descending or ascending notes that span a broad frequency range and appear to be a single syllable but are, in motor terms, two separate vocalizations, each produced by one half of the syrinx [98]. During development, birds must learn to combine these two separate notes so that they become a single continuous sound. We do not know whether these developmental trajectories parallel the evolution of note categories and the rules used for assembling them. If note types or categories evolved first and sequences emerged later, the seemingly simultaneous emergence of note categories and syntax during development (as seems to be the case in canaries) would be a later adaptation, perhaps to increase the efficiency of learning. Alternatively, song development might recapitulate evolution, with ordered sequences emerging after initially identical repeated notes differentiate into distinct syllables. In either scenario, innate specification of acoustic parameters must combine with social learning to generate the basic building blocks that are



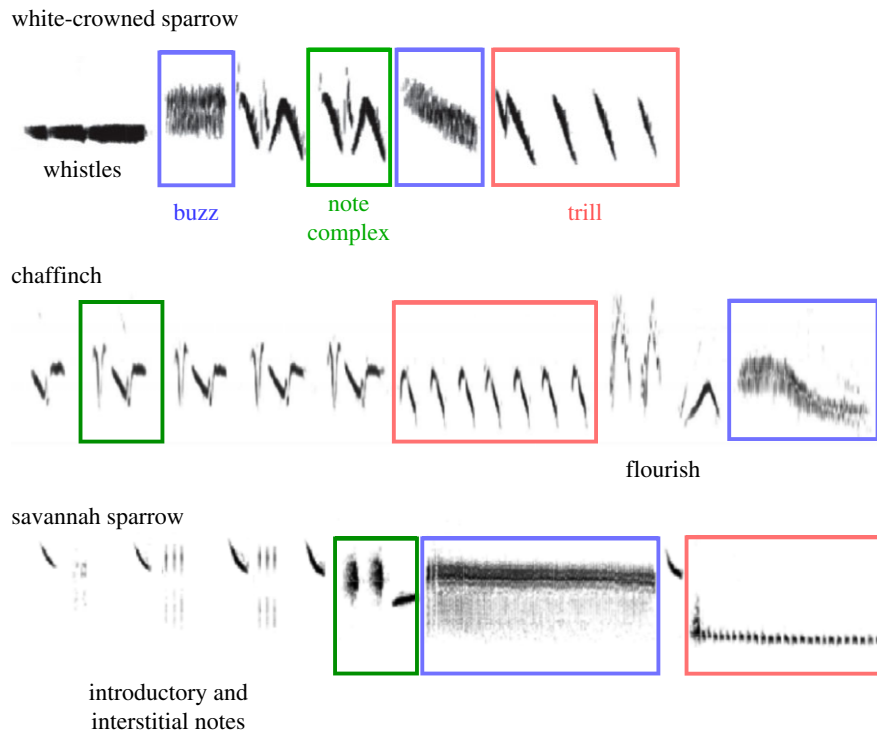
**Figure 4.** Categories and organized syntax emerge from random note sequences. (a) In this schematic, a young bird hears and memorizes three adult songs (red, green and blue), represented in as random strings of six notes. Each note is represented as a point in a two-dimensional acoustic space. (b) Notes with overlapping characteristics combine to form a single representation, or category (circled pairs of notes), as do shared transitions between two note categories (thicker turquoise arrow). If learners use conformist bias, note categories that include more than one memorized example are more likely to persist in the adult song, as are transitions between such categories. (c) In this simple example, the young bird's final song includes four syllables from 'categories' present in more than one model song, connected by transitions between syllables that were present in the model songs. The first four syllables follow the sequence of the original green song (and, in part, the blue song). At the fourth syllable, the order switches to that of the last three syllables of the original red song. In a population of learners that sample many of the same songs and use a simple conformist learning bias, note categories and a systematic syntax will emerge over time.

then assembled, using structural rules that also have innate and learned components, into songs that are species-specific yet vary across populations and individuals.

## 5. Complexity

For the purpose of comparing vocal communication systems, we use the term 'complexity' to describe the number of different sequences that can be generated; complexity increases with the number of sound categories and the number of combinations that syntactic rules can generate from those sounds. Research using iterated learning of artificial languages

suggests that selection for learnability favours systems that are easy to learn but are impoverished in the ability to communicate information [99]. In contrast, selection for the ability to communicate many different meanings favours languages that have distinguishable components that carry different information, but may be hard to learn, while selection for both learnability and communication favours the evolution of languages that are compositional, using combinatorial rules to assemble learnable sound units into different sequences [100]. Like natural human languages, bird songs appear to be under selection for learnability, which may be related to morphological or physiological constraints [101,102] as well as to cognitive mechanisms [73,80]. Although, as we have noted,



**Figure 5.** Segmented song structure, as exemplified by white crowned sparrows, chaffinches and Savannah sparrows. Segments may be defined as: note complexes similar to the syllables of swamp sparrows (green), buzzes consisting of very short broadband notes repeated in rapid succession (blue) and trills made up of a repeated sequence of longer, more tonal notes (red). So that structure can be better compared across the three species' songs, the time ( $x$ -axis) and frequency scales ( $y$ -axis) for the three songs shown here are similar but not identical (white crowned-sparrow song = 2.5 s; chaffinch song = 3 s; Savannah sparrow song = 2.5 s). Spectrograms from Nelson [18], Lachlan *et al.* [74] and Williams *et al.* [112].

the information communicated by bird songs is far less rich than for natural human languages, multiple note categories that are assembled into structured sequences generate enough complexity to signal a variety of meanings.

Songbirds do use simple combinatorial communication systems. Some bird calls (calls are defined as relatively short vocalizations that are not used in courtship or territory defence), such as those of the Paridae [34,103–106] and Australian babblers [107], appear to have a simple combinatorial syntax that signals different meanings in a way that is analogous to human speech. However, the number of different call types and combinations in these systems is relatively small, comparable to the notes and syllables of swamp sparrows. Other species' songs are quite complex: individual birds' repertoires may include dozens of note types or hundreds of different sequences (e.g. grey catbirds, *Dumetella carolinensis* [108]; winter wrens, *Troglodytes troglodytes* [109]; common nightingales, *Luscinia megarhynchos* [110]). Still, other species vary the way songs are sung within a bout, yielding a further layer of sequence complexity [6,111]. In this section, we focus on an intermediate level of song complexity: continuous sequences that last 2–4 s and include at least two segments, each with its own set of notes and structure. Some examples are the songs of Savannah sparrows, chaffinches and white-crowned sparrows (figure 5); they are more amenable to analysis than large repertoires or very long songs. White-crowned sparrow songs begin with one or more whistles, followed by one or more note complexes, and conclude with a trill [85]. Chaffinch songs begin with one or more repeated note complexes or trills and conclude with a flourish that typically includes a terminal buzz [74]. Savannah sparrow songs begin with a series of introductory notes and associated interstitial notes, followed by a note complex, a

buzz and a trill [43]. These songs share the characteristics of being composed of distinct song segments, each consisting of one or more note types organized according to segment-specific rules.

Segmented song structures can be represented as a string of concatenated simpler structures. Each of the segments within these songs is similar in complexity to swamp sparrow songs: one or more notes are assembled into a syllable that may or may not be repeated to form the segment. Two to six segments are then assembled to form the song. As do the note types, segment types fall into categories (buzzes, trills and note complexes), and those categories play consistent roles in the structure of a species' song, in a manner parallel to the role of notes in forming swamp sparrow syllables. A segmented structure can be generated by iterations in which building blocks are assembled into sequences: (1) notes are the building blocks for syllables, (2) syllables can be repeated to form segments and (3) segments are the building blocks for songs. The rules for assembling these structures may have initially arisen so as to allow individuals to learn efficiently (for example, categorization of notes facilitates memorization) and/or to present the components of an individual's vocal repertoire to its hearers efficiently, so that all of an individual's learned notes can be heard within a single vocalization that is organized both to emphasize the characteristics of those notes and to allow for easy comparison of multiple songs. In such songs, notes, syllables, segments and songs may not carry meaning, but instead may form the basis for hearers to assess the accuracy of learning and the size and variety of the repertoire that was learned.

As we have seen, learning biases, natural and sexual selection, and random factors such as population size and



drift can result in cultural evolution of a population's song in terms of notes, syllables and structure [113,114]. In some species, separate song types may serve different functions and so evolve in different directions and at different rates [17,115]. When songs are formed of distinct concatenated segments, cultural evolution can act differently on individual segments, and the functions of segments may diverge [116]. For example, male red-winged blackbirds sing a two-segment song; females respond more strongly when the initial whistled segment was present, while the long trill that forms the second part of the song is sufficient to elicit strong male responses [117].

Many songbird species have learned song features that are population markers: segments that are shared by individuals within a population and so define a local dialect. The white-crowned sparrow's terminal trill [18,118] and the Savannah sparrow's buzz [112] are consistent across most or all individuals within a population. Modelling studies suggest several factors and mechanisms that might be responsible for maintenance of dialects [119–127], including drift, philopatry, learning after dispersal, population turnover and conformist learning biases. Playback studies in a number of species [112,128–130] find that males respond more strongly not only to the local population's song but also specifically to song segments that define a dialect or are characteristic of a population.

Other song segments—often note complexes, such as those in the songs of white-crowned and Savannah sparrows—vary substantially between individuals within a population and may serve in part to denote individual or small-group identity [11,18]. Such variability might arise either randomly or because of a relaxation of learning constraints similar to that seen in small or isolated populations [131,132], and the new forms might become established through some combination of (a) cultural drift [37,133] and (b) rare-form bias (anti-conformity) [134].

Because different song segments may serve different signalling functions, they may also be shaped by different learning mechanisms or biases. For example, the Savannah sparrow buzz segment serves as a population marker and is maintained across individuals and time (figure 6, blue) [112], while the middle segment's note complex varies both within a population and over time and appears to denote individual identity or local affiliation (figure 6, green) [11]. The fact that birds apply different learning rules to different song segments adds a level of cognitive complexity to the cultural evolution of bird song.

Another form of signal effectiveness that cultural evolution acts upon is the ability of songs to attract females. An attractive song yields an increase in the singer's reproductive rate, a standard measure of 'fitness' in terms of both natural and sexual selection. If one song segment is attractive to females, singing that segment gives a male clear fitness benefits. Cultural evolution may then act on the sexually selected song segment if males are able to preferentially learn variants they observe to be more effective for attracting females (a form of prestige bias). The interstitial notes in the introductory segment of the songs in one Savannah sparrow population (figure 6, orange) changed from one form, high note clusters, to another, click trains, between 1982 and 2011 [11]. During an intermediate period when both forms were sung within the population, males singing the new click train form fledged more young from their nests, suggesting that the novel form was a more effective signal

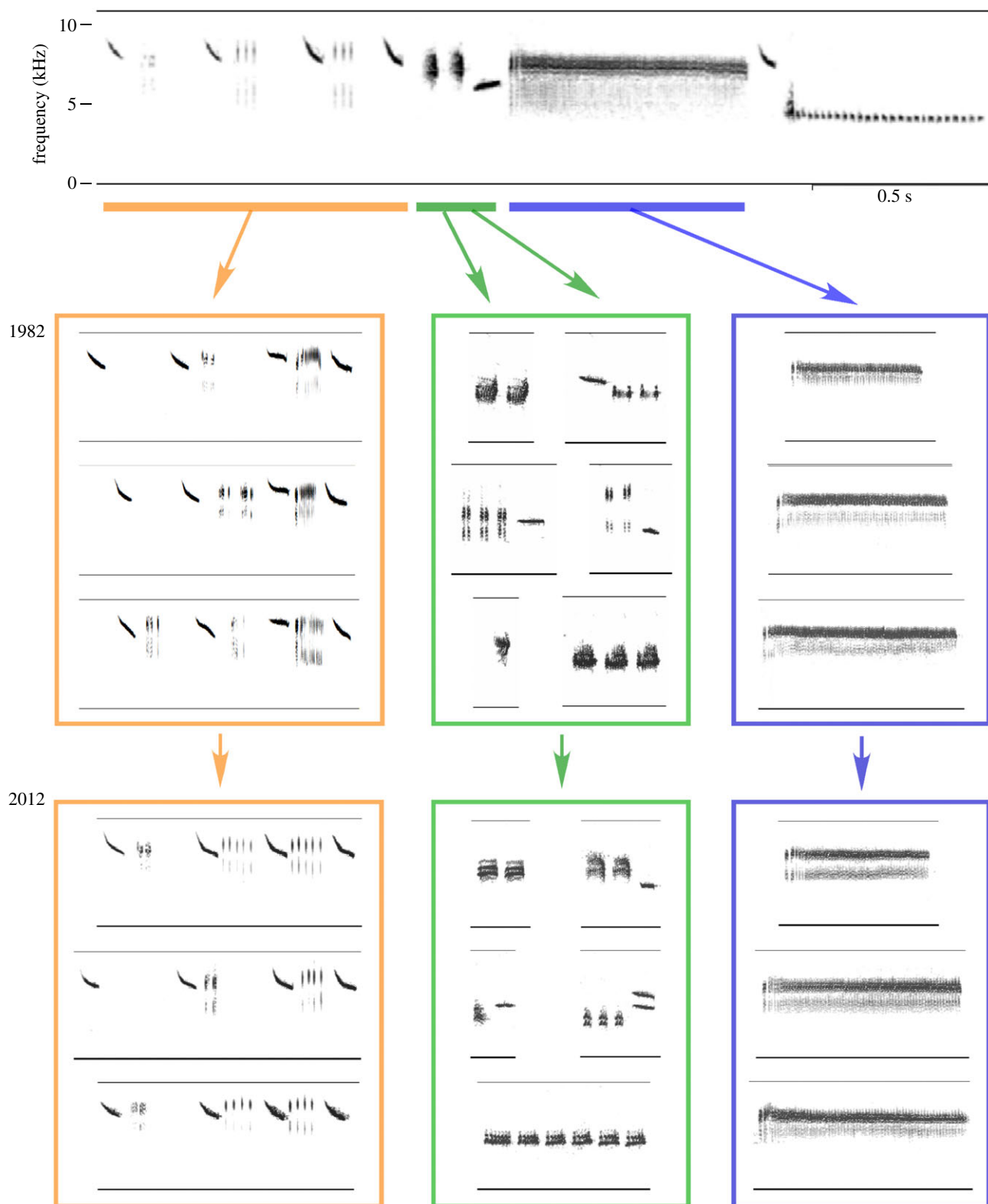
because of a sexual selection advantage. It is not clear why this advantage existed. Perhaps the new version was favoured by a female sensory bias; perhaps females learn their preferences, as in mate choice copying by guppies (*Poecilia reticulata*) [135], and so those preferences are themselves subject to cultural evolution.

These observations highlight several important points about the segments of relatively short learned songs. The segments are (a) made up of different note types organized according to different structural rules, (b) may carry different information, (c) may evolve in different directions and (d) may be shaped by different mechanisms.

## 6. Cumulative cultural evolution: iterated change, effectiveness and complexity

Definitions of cumulative cultural evolution all call for the presence of successive changes in a socially learned behaviour. To meet this criterion, a population's behaviour must have distinctly different forms at distinctly different time periods, perhaps separated by years or generations. At least one additional criterion must also be satisfied, usually defined as either (a) increased behavioural complexity/elaboration [136,137] or (b) increased efficiency in terms of relative costs and benefits [36,39]. Increased complexity (or elaboration) may occur as additional steps in a chain of behaviours, or, in the case of vocalizations, as a more elaborate syntax or an increased number of sound categories. Increased efficiency may take the form of making the behaviour easier to perform (thus reducing physiological costs), easier to learn (reducing developmental costs), or it may make the behaviour itself more efficient: in the case of a communication system, the signals could carry more meanings or provide a more effective signal that better conveys a meaning. We argue that songs based on rules for assembling categories (such as those of swamp sparrows) as well as segmented songs (such as those of Savannah sparrows and white-crowned sparrows) meet these criteria for cumulative cultural evolution. First, iterations of social learning across generations result in systematic changes in a population's songs over time. Second, the successive changes can result in a more effective signal, in terms of reproductive success, as in the case of the interstitial notes of Savannah sparrow song. Third, these songs have evolved at least two layers of organization—categories and syllables for swamp sparrow songs, and syllables and segments for Savannah sparrow song—and so have become more complex and structured over time.

As signalling systems, however, bird songs do not come close to the complexity and combinatoriality of human language, and they also appear to lack referential properties. In many songbird species, learning is completed relatively early in life and new material is not incorporated thereafter, and so the ability of the signal to change during an individual's adult life in order to carry new information is further limited. Nevertheless, variation in the components of the song, the syntax of the song, the performance the song, and the repertoire size carries information about the singer's identity, quality and motivation. Note categories and sequencing rules that are universal within a species' song are likely to have a genetically defined component. Because of these innate predispositions and because individuals within populations that hear each other during development are likely to



**Figure 6.** Cultural evolution differs across Savannah sparrow song segments. A full Kent Island song is shown at the top of the figure; below are examples of song segments recorded from the same population 30 years apart. The buzz (blue) varied in length across individuals but was otherwise similar within the population, and the buzzes recorded at the beginning and end of the 30-year interval were statistically identical. The note complex (green) varied both across individuals and across years, although the structure and many of the notes were consistent over time within the population. The introductory section (orange) consists of several repeated downsweeps that do not vary substantially across populations. The softer interstitial notes that fall between the introductory notes were consistent within the population in both 1982 and 2012 but changed over time. In 1982 that pattern was a three-part cluster consisting of short repeated notes/an unmodulated high note/short high trill. It was replaced over the succeeding decades by a series of clicks, as seen in the 2012 examples. The replacement was gradual, and both of the interstitial note patterns were sung in the 1990s and early 2000s.

learn from each other and to share vocalizations, birds within a population often sing a common dialect. These factors would seem to act in concert to homogenize the learned

songs within a population. However, copying errors and innovation (and perhaps anti-conformity) lead to the production of novel song features, which provide cues to

individual identity and potentially to other attributes associated with that individual. Some of these novel features may prove to be more 'successful' than others: they may be more effective at signalling species, population, or individual identity, more flexible in signalling intention or motivation, or more informative about individual quality. Successful features may then be copied more often and increase within the population because of cultural selection. Random processes, or cultural drift, may also result in spreading novel song forms within a population. In these ways, successive rounds of vocal learning can build upon, change, elaborate, and increase the efficiency of song features, without calling on an evolutionary drive to produce an increased number of distinguishable signals.

## 7. Conclusion

Reduced to their most basic building blocks, bird songs consist of a set of notes. General cognitive functions such as species-specific sensory predispositions, memorization and simple learning biases also serve as building blocks, working together to group the notes into categories during social learning. A third type of building block, sequencing, forms simple strings, or syllables, that include notes from two or more categories. In the case of swamp sparrows, a bird may produce several syllable types; each song consists of a single syllable type repeated several times. In other species, different syllable types are concatenated, forming a learned sequence with two levels of structure—an elaboration that increases the complexity of the song. Different song segments may then diverge in function over time, with different pressures defining how each segment evolves, increasing the efficacy of each segment in terms of the communication role it plays. Although bird songs do not, as far as we know, have combinatorial structure that can be used to generate referential meaning, they do meet the criteria for cumulative cultural evolution: the songs are socially learned, a population's songs change systematically over time, and a

comparative approach reveals that structures—the note categories and rules for assembling notes into syllables and syllables into songs—are elaborated in ways that increase complexity. In songs such as those of the Savannah sparrow, different segments carry different information, and cultural evolution of those song segments increases their efficiency. Although bird songs lack the evolutionary pressure for combinatorial structure that referential meaning engenders, simple building blocks (acoustically variable notes and general cognitive mechanisms) are sufficient to give rise to vocal signals with enough attributes for social learning and evolutionary mechanisms to incrementally produce increased complexity and efficacy—the hallmarks of cumulative cultural evolution.

**Ethics.** This article reports on previously published animal research; there are no new data derived from animals.

**Data accessibility.** This article has no additional data.

**Authors' contribution.** The two authors conferred and planned the content and organization of the paper. H.W. wrote the paper, R.F.L. rewrote §3, and both authors edited the paper.

**Competing interests.** We declare we have no competing interests.

**Funding.** H.W. was funded by a grant from the Groff Foundation, and R.F.L. was funded by the BBSRC (grant no. BB/R008736/2).

## Endnote

<sup>1</sup>Although terminology differs by author and species, in general 'notes' are continuous sound elements, 'syllables' are clusters of notes separated by short (less than 10 ms) silent intervals delivered in a stereotyped order and demarcated by being repeated or having longer (but consistent) silent intervals at the beginning and end of the syllable than within the syllable, 'phrases' are groups of syllables that are sung in a consistent (sometimes repeated) order within a song and may be demarcated by slightly longer (but again consistent) silent intervals at the boundaries of the phrase, and 'songs' are groups of syllables or phrases separated by longer (greater than 0.5 s) and more variable silent intervals (see examples in figures 2 and 5). 'Bouts' are groups of songs that are separated by still longer and more variable silent intervals.

## References

- Marler P. 1970 Birdsong and speech development: could there be parallels? *Am. Sci.* **58**, 669–673.
- Doupe AJ, Kuhl PK. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
- Wilbrecht L, Nottebohm F. 2003 Vocal learning in birds and humans. *Ment. Retard. Dev. Disabil. Res. Rev.* **9**, 135–148. (doi:10.1002/mrdd.10073)
- Searcy WA, Nowicki S. 2019 Birdsong learning, avian cognition and the evolution of language. *Anim. Behav.* **151**, 217–227. (doi:10.1016/j.anbehav.2019.01.015)
- Hyland Bruno J, Jarvis ED, Liberman M, Tchernichovski O. 2021 Birdsong learning and culture: analogies with human spoken language. *Annu. Rev. Linguist.* **7**, 449–472. (doi:10.1146/annurev-linguistics-090420-121034)
- Hyland Bruno J, Tchernichovski O. 2019 Regularities in zebra finch song beyond the repeated motif. *Behav. Processes* **163**, 53–59. (doi:10.1016/j.beproc.2017.11.001)
- Garland EC, Garrigue C, Noad MJ. 2021 When does cultural evolution become cumulative culture? A case study of humpback whale song. *Phil. Trans. R. Soc. B* **377**, 20200313. (doi:10.1098/rstb.2020.0313)
- Mundinger PC. 1980 Animal cultures and a general theory of cultural evolution. *Ethol. Sociobiol.* **1**, 183–223. (doi:10.1016/0162-3095(80)90008-4)
- Trainer JM. 1989 Cultural evolution in song dialects of yellow-rumped caciques in Panama. *Ethology* **80**, 190–204. (doi:10.1111/j.1439-0310.1989.tb00739.x)
- Slater PJB, Ince SA. 1979 Cultural evolution in chaffinch song. *Behaviour* **71**, 146–166. (doi:10.1163/156853979X00142)
- Williams H, Levin II, Norris DR, Newman AEM, Wheelwright NT. 2013 Three decades of cultural evolution in Savannah sparrow songs. *Anim. Behav.* **85**, 213–223. (doi:10.1016/j.anbehav.2012.10.028)
- Payne RB, Payne LL, Doehlert SM. 1988 Biological and cultural success of song memes in indigo buntings. *Ecology* **69**, 104–117. (doi:10.2307/1943165)
- Ju C, Geller FC, Mundinger PC, Lahti DC. 2019 Four decades of cultural evolution in House Finch songs. *Auk* **136**, 1–18. (doi:10.1093/auk/uky012)
- Nelson DA, Hallberg KL, Soha JA. 2004 Cultural evolution of Puget Sound white-crowned sparrow song dialects. *Ethology* **110**, 879–908. (doi:10.1111/j.1439-0310.2004.01025.x)
- Wright TF, Dahlin CR. 2017 Vocal dialects in parrots: patterns and processes of cultural evolution. *Emu* **118**, 50–66. (doi:10.1080/01584197.2017.1379356)
- Riebel K, Lachlan RF, Slater PJB. 2015 Learning and cultural transmission in chaffinch song. *Adv. Study Behav.* **47**, 181–227. (doi:10.1016/bs.asb.2015.01.001)

17. Byers BE, Belinsky KL, Bentley RA. 2010 Independent cultural evolution of two song traditions in the chestnut-sided warbler. *Am. Nat.* **176**, 476–489. (doi:10.1086/656268)
18. Nelson DA. 2017 Geographical variation in song phrases differs with their function in white-crowned sparrow song. *Anim. Behav.* **124**, 263–271. (doi:10.1016/j.anbehav.2016.05.016)
19. Irwin DE, Bensch S, Price TD. 2001 Speciation in a ring. *Nature* **409**, 333–337. (doi:10.1038/35053059)
20. Marler P, Tamura M. 1962 Song ‘dialects’ in three populations of white-crowned sparrows. *Condor* **64**, 368–377. (doi:10.2307/1365545)
21. Yasukawa K, Bick EI, Wagman DW, Marler P. 1982 Playback and speaker-replacement experiments on song-based neighbor, stranger, and self discrimination in male red-winged blackbirds. *Behav. Ecol. Sociobiol.* **10**, 211–215. (doi:10.1007/BF00299687)
22. Collins S. 2004 Vocal fighting and flirting: the functions of birdsong. In *Nature’s music: The science of birdsong* (eds P Marler, H Slabbekoorn), pp. 39–79. San Diego, CA and London, UK: Elsevier Academic Press.
23. Akçay Ç, Anderson RC, Nowicki S, Beecher MD, Searcy WA. 2015 Quiet threats: soft song as an aggressive signal in birds. *Anim. Behav.* **105**, 267–274. (doi:10.1016/j.anbehav.2015.03.009)
24. Naguib M, Mennill DJ. 2010 The signal value of birdsong: empirical evidence suggests song overlapping is a signal. *Anim. Behav.* **80**, e11–e15. (doi:10.1016/j.anbehav.2010.06.001)
25. Burt JM, Campbell SE, Beecher MD. 2001 Song type matching as threat: a test using interactive playback. *Anim. Behav.* **62**, 1163–1170. (doi:10.1006/anbe.2001.1847)
26. Logue DM, Krupp DB. 2016 Duetting as a collective behavior. *Front. Ecol. Evol.* **4**, 7. (doi:10.3389/fevo.2016.00007)
27. Ballentine B. 2009 The ability to perform physically challenging songs predicts age and size in male swamp sparrows, *Melospiza georgiana*. *Anim. Behav.* **77**, 973–978. (doi:10.1016/j.anbehav.2008.12.027)
28. Woodgate JL, Mariette MM, Bennett ATD, Griffith SC, Buchanan KL. 2012 Male song structure predicts reproductive success in a wild zebra finch population. *Anim. Behav.* **83**, 773–781. (doi:10.1016/j.anbehav.2011.12.027)
29. Sung H-C, Handford P. 2020 Song characters as reliable indicators of male reproductive quality in the Savannah sparrow (*Passerculus sandwichensis*). *Can. J. Zool.* **98**, 32–38.
30. Clucas BA, Freeberg TM, Lucas JR. 2004 Chick-a-dee call syntax, social context, and season affect vocal responses of Carolina chickadees (*Poecile carolinensis*). *Behav. Ecol. Sociobiol.* **57**, 187–196. (doi:10.1007/s00265-004-0847-9)
31. Ciaburri I, Williams H. 2019 Context-dependent variation of house finch song syntax. *Anim. Behav.* **147**, 33–42. (doi:10.1016/j.anbehav.2018.11.001)
32. Sakata JT, Hampton CM, Brainard MS. 2008 Social modulation of sequence and syllable variability in adult birdsong. *J. Neurophysiol.* **99**, 1700–1711. (doi:10.1152/jn.01296.2007)
33. Marler P. 2006 Bird calls: their potential for behavioral neurobiology. *Ann. NY Acad. Sci.* **1016**, 31–44. (doi:10.1196/annals.1298.034)
34. Freeberg TM, Lucas JR. 2002 Receivers respond differently to chick-a-dee calls varying in note composition in Carolina chickadees, *Poecile carolinensis*. *Anim. Behav.* **63**, 837–845. (doi:10.1006/anbe.2001.1981)
35. Templeton CN, Greene E, Davis K. 2005 Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science* **308**, 1934–1937. (doi:10.1126/science.1108841)
36. Gruber T, Chimento M, Aplin LM, Biro D. 2021 Efficiency fosters cumulative culture across species. *Phil. Trans. R. Soc. B* **377**, 20200308. (doi:10.1098/rstb.2020.0308)
37. Potvin DA, Clegg SM. 2015 The relative roles of cultural drift and acoustic adaptation in shaping syllable repertoires of island bird populations change with time since colonization. *Evolution* **69**, 368–380. (doi:10.1111/evo.12573)
38. Lynch A. 1996 The population memetics of bird song. In *Ecology and evolution of acoustic communication in birds* (eds DE Kroodsma, EH Miller), pp. 181–197. Ithaca, NY and London, UK: Cornell University Press.
39. Mesoudi A, Thornton A. 2018 What is cumulative cultural evolution? *Proc. R. Soc. B* **285**, 20180712. (doi:10.1098/rspb.2018.0712)
40. Marler P. 1970 A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol.* **71**, 1–25. (doi:10.1037/h0029144)
41. Immelmann K. 1969 Song development in the zebra finch and other estrildid finches. In *Bird vocalizations* (ed. RA Hinde), pp. 61–74. Cambridge, UK: Cambridge University Press.
42. Thorpe WH. 1954 The process of song-learning in the chaffinch as studied by means of the sound spectrograph. *Nature* **173**, 465–469. (doi:10.1038/173465a0)
43. Wheelwright NT, Swett MB, Levin II, Kroodsma DE, Freeman-Gallant CR, Williams H. 2008 The influence of different tutor types on song learning in a natural bird population. *Anim. Behav.* **75**, 1479–1493. (doi:10.1016/j.anbehav.2007.08.030)
44. Nottebohm F. 1970 Ontogeny of bird song. *Science* **167**, 950–956. (doi:10.1126/science.167.3920.950)
45. Nelson DA, Marler P. 1994 Selection-based learning in bird song development. *Proc. Natl Acad. Sci. USA* **91**, 10 498–10 501. (doi:10.1073/pnas.91.22.10498)
46. Nelson DA. 1992 Song overproduction and selective attrition lead to song sharing in the field sparrow (*Spizella pusilla*). *Behav. Ecol. Sociobiol.* **30**, 415–424. (doi:10.1007/BF00176177)
47. Marler P, Peters S. 1982 Structural changes in song ontogeny in the swamp sparrow *Melospiza georgiana*. *Auk* **99**, 446–458. (doi:10.1093/auk/99.3.446)
48. Williams H, Staples K. 1992 Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J. Comp. Psychol.* **106**, 278–286. (doi:10.1037/0735-7036.106.3.278)
49. Lachlan RF, van Heijningen CAA, ter Haar SM, ten Cate C. 2016 Zebra finch song phonology and syntactical structure across populations and continents—a computational comparison. *Front. Psychol.* **7**, 980. (doi:10.3389/fpsyg.2016.00980)
50. Marler P, Pickert R. 1984 Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Anim. Behav.* **32**, 673–689. (doi:10.1016/S0003-3472(84)80143-8)
51. Tchernichovski O, Lints TJ, Derégnaucourt S, Cimenser A, Mitra PP. 2004 Studying the song development process: rationale and methods. *Ann. N. Y. Acad. Sci.* **1016**, 348–363. (doi:10.1196/annals.1298.031)
52. Zuidema W, de Boer B. 2009 The evolution of combinatorial phonology. *J. Phon.* **37**, 125–144. (doi:10.1016/j.wocn.2008.10.003)
53. Little H, Eryilmaz K, de Boer B. 2017 Signal dimensionality and the emergence of combinatorial structure. *Cognition* **168**, 1–15. (doi:10.1016/j.cognition.2017.06.011)
54. Wedel A, Fatkullin I. 2017 Category competition as a driver of category contrast. *J. Lang. Evol.* **2**, 77–93. (doi:10.1093/jole/lzx009)
55. Tchernichovski O, Mitra PP, Lints T, Nottebohm F. 2001 Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* **291**, 2564–2569. (doi:10.1126/science.1058522)
56. Marler P, Peters S. 1977 Selective vocal learning in a sparrow. *Science* **198**, 519–521. (doi:10.1126/science.198.4316.519)
57. Marler P, Peters S. 1988 The role of phonology and syntax in vocal learning preferences in the song sparrow, *Melospiza melodia*. *Ethology* **77**, 125–149. (doi:10.1111/j.1439-0310.1988.tb00198.x)
58. Nelson DA. 2000 A preference for own-subspecies’ song guides vocal learning in a song bird. *Proc. Natl Acad. Sci. USA* **97**, 13 348–13 353. (doi:10.1073/pnas.240457797)
59. Nelson DA, Marler P. 1993 Innate recognition of song in white-crowned sparrows. *Anim. Behav.* **46**, 806–808. (doi:10.1006/anbe.1993.1258)
60. Fehér O, Wang H, Saar S, Mitra PP, Tchernichovski O. 2009 *De novo* establishment of wild-type song culture in the zebra finch. *Nature* **459**, 564–568. (doi:10.1038/nature07994)
61. Belzner S, Voigt C, Catchpole CK, Leitner S. 2009 Song learning in domesticated canaries in a restricted acoustic environment. *Proc. R. Soc. B* **276**, 2881–2886. (doi:10.1098/rspb.2009.0669)
62. Fehér O. 2017 Atypical birdsong and artificial languages provide insights into how communication systems are shaped by learning, use, and transmission. *Psychon. Bull. Rev.* **24**, 97–105. (doi:10.3758/s13423-016-1107-5)
63. Podos J. 1996 Motor constraints on vocal development in a songbird. *Anim. Behav.* **51**, 1061–1070. (doi:10.1006/anbe.1996.0107)

64. Podos J. 2001 Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* **409**, 185–188. (doi:10.1038/35051570)
65. Hoese WJ, Podos J, Boetticher NC, Nowicki S. 2000 Vocal tract function in birdsong production: experimental manipulation of beak movements. *J. Exp. Biol.* **203**, 1845–1855. (doi:10.1242/jeb.203.12.1845)
66. Greenewalt CH. 1968 *Bird song: acoustics and physiology*. Washington, DC: Smithsonian Institution Press.
67. Suthers RA. 1990 Contributions to birdsong from the left and right sides of the syrinx. *Nature* **347**, 473–477. (doi:10.1038/246170a0)
68. Suthers RA, Vallet E, Tanvez A, Kreuzer M. 2004 Bilateral song production in domestic canaries. *J. Neurobiol.* **60**, 381–393. (doi:10.1002/neu.20040)
69. Nottebohm F, Nottebohm ME. 1976 Left hypoglossal dominance in the control of canary and white-crowned sparrow song. *J. Comp. Physiol. A* **108**, 171–192. (doi:10.1007/BF02169047)
70. Williams H, Crane LA, Hale TK, Esposito MA, Nottebohm F. 1992 Right-side dominance for song control in the zebra finch. *J. Neurobiol.* **23**, 1006–1020. (doi:10.1002/neu.480230807)
71. Allan SE, Suthers RA. 1994 Lateralization and motor stereotype of song production in the brown-headed cowbird. *J. Neurobiol.* **25**, 1154–1166. (doi:10.1002/neu.480250910)
72. Suthers RA. 1994 Variable asymmetry and resonance in the avian vocal tract: a structural basis for individually distinct vocalizations. *J. Comp. Physiol. A* **175**, 457–466. (doi:10.1007/BF00199253)
73. Lachlan RF, Ratmann O, Nowicki S. 2018 Cultural conformity generates extremely stable traditions in bird song. *Nat. Commun.* **9**, 2417. (doi:10.1038/s41467-018-04728-1)
74. Lachlan RF, Verhagen L, Peters S, ten Cate C. 2010 Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). *J. Comp. Psychol.* **124**, 92–108. (doi:10.1037/a0016996)
75. Lachlan RF, Nowicki S. 2015 Context-dependent categorical perception in a songbird. *Proc. Natl Acad. Sci. USA* **112**, 1892–1897. (doi:10.1073/pnas.1410844112)
76. Balaban E. 1988 Cultural and genetic variation in swamp sparrows (*Melospiza georgiana*): I Song variation, genetic variation, and their relationship. *Behaviour* **105**, 250–291. (doi:10.1163/156853988X00043)
77. Nelson DA, Marler P. 1989 Categorical perception of a natural stimulus continuum: Birdsong. *Science* **244**, 976–978. (doi:10.1126/science.2727689)
78. Prather JF, Nowicki S, Anderson RC, Peters S, Mooney R. 2009 Neural correlates of categorical perception in learned vocal communication. *Nat. Neurosci.* **12**, 221–228. (doi:10.1038/nn.2246)
79. Maye J, Werker JF, Gerken LA. 2002 Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* **82**, 101–111. (doi:10.1016/S0010-0277(01)00157-3)
80. Lachlan RF, Anderson RC, Peters S, Searcy WA, Nowicki S. 2014 Typical versions of learned swamp sparrow song types are more effective signals than are less typical versions. *Proc. R. Soc. B* **281**, 1–7. (doi:10.1098/rspb.2014.0252)
81. Nowicki S, Searcy WA, Peters S. 2002 Brain development, song learning and mate choice in birds: a review and experimental test of the 'nutritional stress hypothesis'. *J. Comp. Physiol. A. Neuroethol. Sens. Neural. Behav. Physiol.* **188**, 1003–1014. (doi:10.1007/s00359-002-0361-3)
82. Lachlan RF, Nowicki S. 2012 How reliable is song learning accuracy as a signal of male early condition? *Am. Nat.* **180**, 751–761. (doi:10.1086/668010)
83. Thorpe WH. 1958 The learning of song patterns by birds, with especial reference to the song of the chaffinch *Fringilla coelebs*. *Ibis* **100**, 535–570. (doi:10.1111/j.1474-919X.1958.tb07960.x)
84. Marler P. 1952 Variation in the song of the chaffinch *Fringilla coelebs*. *Ibis* **94**, 458–472. (doi:10.1111/j.1474-919X.1952.tb01845.x)
85. Baptista LF. 1977 Geographic variation in song and dialects of the Puget Sound white-crowned sparrow. *Condor* **79**, 356–370. (doi:10.2307/1368014)
86. Zann R. 1993 Structure, sequence and evolution of song elements in wild Australian zebra finches. *Auk* **110**, 702–715. (doi:10.2307/4088626)
87. Wildenthal JL. 1965 Structure in the primary song of the mockingbird (*Mimus polyglottos*). *Auk* **82**, 161–189. (doi:10.2307/4082931)
88. Chew L. 1981 Geographic and individual variation in the morphology and sequential organization of the song of the Savannah sparrow (*Passerculus sandwichensis*). *Can. J. Zool.* **59**, 702–713. (doi:10.1139/z81-099)
89. Marler P, Sherman V. 1985 Innate differences in singing behaviour of sparrows reared in isolation from adult conspecific song. *Anim. Behav.* **33**, 57–71. (doi:10.1016/S0003-3472(85)80120-2)
90. Marler P, Sherman V. 1983 Song structure without auditory feedback: emendations of the auditory template hypothesis. *J. Neurosci.* **1**, 517–531. (doi:10.1523/JNEUROSCI.03-03-00517.1983)
91. Volman SF, Khanna H. 1995 Convergence of untutored song in group-reared zebra Finches (*Taenopygia guttata*). *J. Comp. Psychol.* **109**, 211–221. (doi:10.1037/0735-7036.109.3.211)
92. Güttinger H-R. 1981 Self-differentiation of song organization rules by deaf canaries. *Z. Tierpsychol.* **56**, 323–340. (doi:10.1111/j.1439-0310.1981.tb01305.x)
93. Liu W, Gardner TJ, Nottebohm F. 2004 Juvenile zebra finches can use multiple strategies to learn the same song. *Proc. Natl Acad. Sci. USA* **101**, 18 177–18 182. (doi:10.1073/pnas.0408065101)
94. Lipkind D, Zai AT, Hanuschkin A, Marcus GF, Tchernichovski O, Hahnloser RHR. 2017 Songbirds work around computational complexity by learning song vocabulary independently of sequence. *Nat. Commun.* **8**, 1247. (doi:10.1038/s41467-017-01436-0)
95. Waser MS, Marler P. 1977 Song learning in canaries. *J. Comp. Physiol. Psychol.* **91**, 1–7. (doi:10.1037/h0077299)
96. Markowitz JE, Ivie E, Kligler L, Gardner TJ. 2013 Long-range order in canary song. *PLoS Comput. Biol.* **9**, e1003052. (doi:10.1371/journal.pcbi.1003052)
97. Gardner TJ, Naef F, Nottebohm F. 2005 Freedom and rules: the acquisition and reprogramming of a bird's learned song. *Science* **308**, 1046–1049. (doi:10.1126/science.1108214)
98. Zollinger SA, Suthers RA. 2004 Motor mechanisms of a vocal mimic: implications for birdsong production. *Proc. R. Soc. B* **271**, 483–491. (doi:10.1098/rspb.2003.2598)
99. Kirby S, Cornish H, Smith K. 2008 Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl Acad. Sci. USA* **105**, 10 681–10 686. (doi:10.1073/pnas.0707835105)
100. Kirby S, Tamariz M, Cornish H, Smith K. 2015 Compression and communication in the cultural evolution of linguistic structure. *Cognition* **141**, 87–102. (doi:10.1016/j.cognition.2015.03.016)
101. Podos J, Lahti DC, Moseley DL. 2009 Vocal performance and sensorimotor learning in songbirds. *Adv. Study Behav.* **40**, 159–195. (doi:10.1016/S0065-3454(09)40005-6)
102. Prather JF, Peters S, Mooney R, Nowicki S. 2012 Sensory constraints on birdsong syntax: neural responses to swamp sparrow songs with accelerated trill rates. *Anim. Behav.* **83**, 1411–1420. (doi:10.1016/j.anbehav.2012.03.012)
103. Hailman JP, Ficken MS, Ficken RW. 1985 The 'chick-a-dee' calls of *Parus atricapillus*: a recombinant system of animal communication compared with written English. *Semiotica* **56**, 191–224. (doi:10.1515/semi.1985.56.3-4.191)
104. Charrier I, Bloomfield LL, Sturdy CB. 2004 Note types and coding in parid vocalizations. I: the chick-a-dee call of the black-capped chickadee (*Poecile atricapillus*). *Can. J. Zool.* **82**, 769–779. (doi:10.1139/Z04-045)
105. Suzuki TN, Griesser M, Wheatcroft D. 2019 Syntactic rules in avian vocal sequences as a window into the evolution of compositionality. *Anim. Behav.* **151**, 267–274. (doi:10.1016/j.anbehav.2019.01.009)
106. Salis A, Léna JP, Lengagne T. 2021 Great tits (*Parus major*) adequately respond to both allopatric combinatorial mobbing calls and their isolated parts. *Ethology* **127**, 213–222. (doi:10.1111/eth.13111)
107. Engesser S, Holub JL, O'Neill LG, Russell AF, Townsend SW. 2019 Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proc. Natl Acad. Sci. USA* **116**, 19 579–19 584. (doi:10.1073/pnas.1819513116)
108. Kroodsma DE, Houlihan PW, Fallon PA, Wells JA. 1997 Song development by grey catbirds. *Anim. Behav.* **54**, 457–464. (doi:10.1006/anbe.1996.0387)

109. Van Horne B 1995 Assessing vocal variety in the winter wren, a bird with a complex repertoire. *The Condor* **97**, 39–49. (doi:10.2307/1368981)
110. Hultsch H, Todt D. 1981 Repertoire sharing and song-post distance in nightingales (*Luscinia megarhynchos* B.). *Behav. Ecol. Sociobiol.* **8**, 183–188. (doi:10.1007/BF00299828)
111. Weiss M, Hultsch H, Adam I, Scharff C, Kipper S. 2014 The use of network analysis to study complex animal communication systems: a study on nightingale song. *Proc. R. Soc. B* **281**, 20140460. (doi:10.1098/rspb.2014.0460)
112. Williams H, Robins CW, Norris DR, Newman AEM, Freeman-Gallant CR, Wheelwright NT, Mennill DJ. 2019 The buzz segment of Savannah sparrow song is a population marker. *J. Ornithol.* **160**, 217–227. (doi:10.1007/s10336-018-1611-7)
113. Whiten A. 2019 Cultural evolution in animals. *Annu. Rev. Ecol. Syst.* **50**, 27–48. (doi:10.1146/annurev-ecolsys-110218-025040)
114. Whiten A. 2021 The burgeoning reach of animal culture. **372**, eabe6514. (doi:10.1126/science.abe6514)
115. Kroodsmas DE, Bereson RE, Byers BE, Minear E. 1989 Use of song types by the chestnut-sided warbler: evidence for both intra- and inter-sexual interactions. *Can. J. Zool.* **67**, 447–456. (doi:10.1139/z89-065)
116. Williams H. 2021 Mechanisms of cultural evolution in the songs of wild bird populations. *Front. Psychol.* **12**, 1–18. (doi:10.3389/fpsyg.2021.643343)
117. Searcy WA, Brenowitz EA. 1988 Sexual differences in species recognition of avian song. *Nature* **332**, 152–154. (doi:10.1038/332152a0)
118. Zimmerman HD, Ramsay SM, Mesias V, Mora M, Murray BW, Otter KA. 2016 Cultural evolution in white-throated sparrow song: regional variation through shift in terminal strophe type and length. *Behaviour* **153**, 1839–1861. (doi:10.1163/1568539X-00003394)
119. Planqué R, Britton NF, Slabbekoorn H. 2014 On the maintenance of bird song dialects. *J. Math. Biol.* **68**, 505–531. (doi:10.1007/s00285-012-0632-8)
120. Lachlan RF, Janik VM, Slater PJB. 2004 The evolution of conformity-enforcing behaviour in cultural communication systems. *Anim. Behav.* **68**, 561–570. (doi:10.1016/j.anbehav.2003.11.015)
121. Strigul N. 2009 Can imitation explain dialect origins? *Ecol. Model.* **220**, 2624–2639. (doi:10.1016/j.ecolmodel.2009.07.005)
122. Burridge J, Kenney S. 2016 Birdsong dialect patterns explained using magnetic domains. *Phys. Rev. E* **93**, 1–10. (doi:10.1103/PhysRevE.93.062402)
123. Eilers J, Slabbekoorn H. 2003 Song divergence and male dispersal among bird populations: a spatially explicit model testing the role of vocal learning. *Anim. Behav.* **65**, 671–681. (doi:10.1006/anbe.2003.2081)
124. Rowell JT, Servedio MR. 2012 Vocal communications and the maintenance of population specific songs in a contact zone. *PLoS ONE* **7**, e35257. (doi:10.1371/journal.pone.0035257)
125. Lachlan RF, Slater PJB. 1999 The maintenance of vocal learning by gene–culture interaction: the cultural trap hypothesis. *Proc. R. Soc. Lond. B* **266**, 701–706. (doi:10.1098/rspb.1999.0692)
126. Lynch A, Plunkett GM, Baker AJ, Jenkins PF. 1989 A model of cultural evolution of chaffinch song derived with the meme concept. *Am. Nat.* **133**, 634–653. (doi:10.1086/284942)
127. Bentley RA, Hahn MW, Shennan SJ. 2004 Random drift and culture change. *Proc. R. Soc. B* **271**, 1443–1450. (doi:10.1098/rspb.2004.2746)
128. Nelson DA. 1998 Geographic variation in song of Gambel's white-crowned sparrow. *Behaviour* **135**, 321–342. (doi:10.1163/156853998793066258)
129. Nelson DA, Soha JA. 2004 Perception of geographical variation in song by male Puget Sound white-crowned sparrows, *Zonotrichia leucophrys pugetensis*. *Anim. Behav.* **68**, 395–405. (doi:10.1016/j.anbehav.2003.08.027)
130. Podos J. 2007 Discrimination of geographical song variants by Darwin's finches. *Anim. Behav.* **73**, 833–844. (doi:10.1016/j.anbehav.2006.11.001)
131. Lachlan RF, Verzijden MN, Bernard CS, Jonker PP, Koese B, Jaarsma S, Spoor W, Slater PJB, Ten Cate C. 2013 The progressive loss of syntactical structure in bird song along an island colonization chain. *Curr. Biol.* **23**, 1896–1901. (doi:10.1016/j.cub.2013.07.057)
132. Gammon DE, Baker MC, Tipton JR. 2005 Cultural divergence within novel song in the black-capped chickadee (*Poecilia atricapillus*). *Auk* **122**, 853. (doi:10.1642/0004-8038(2005)122[0853:CDWNSI]2.0.CO;2)
133. Bentley RA, Lipo CP, Herzog HA, Hahn MW. 2007 Regular rates of popular culture change reflect random copying. *Evol. Hum. Behav.* **28**, 151–158. (doi:10.1016/j.evolhumbehav.2006.10.002)
134. Sprau P, Mundry R. 2010 Song type sharing in common nightingales, *Luscinia megarhynchos*, and its implications for cultural evolution. *Anim. Behav.* **80**, 427–434. (doi:10.1016/j.anbehav.2010.05.028)
135. Dugatkin LA. 1996 Interface between culturally based preferences and genetic preferences: female mate choice in *Poecilia reticulata*. *Evolution* **93**, 2770–2773. (doi:10.1073/pnas.93.7.2770)
136. Claidière N, Messer EJE, Hoppitt W, Whiten A. 2013 Diffusion dynamics of socially learned foraging techniques in squirrel monkeys. *Curr. Biol.* **23**, 1251–1255. (doi:10.1016/j.cub.2013.05.036)
137. Boyd R, Richerson PJ, Henrich J. 2013 The cultural evolution of technology: facts and theories. In *Cultural evolution: society, technology, language, and religion* (eds PJ Richerson, MH Christiansen), pp. 119–142. Cambridge, MA, and London, UK: MIT Press.