Temporal variability enhances vocal learning

Logan S. James[1], Olivia Ruge[1], and Jon T. Sakata[1,2,3,*]

1 Department of Biology, McGill University
2 Integrative Program in Neuroscience, McGill University
3 Centre for Research in Brain, Language, and Music, McGill University
* corresponding author: jon.sakata@mcgill.ca

Keywords: zebra finch, spectral, sequencing, timing, complexity

Acknowledgements: We would like to thank F. Bendall, V.Y. Li, and A.S. Wang for their assistance with data collection and processing, and M.S. Brainard and S.C. Woolley for comments on manuscripts and data analysis and interpretation. This work was supported by funding from the National Science and Engineering Research Council (Discovery Grant #05016 to J.T.S.; USRA fellowship to O.R.), the Fonds de Recherche du Quebec – Nature et technologies (PR-299652 to S.C.W.), Smithsonian Institute Postdoctoral Fellowship (L.S.J.), and the Centre for Research on Brain, Language and Music (L.S.J. and J.T.S).

Abstract

Species-typical behaviors are organized into species-typical patterns, and deviations away from these patterns often diminish the strength of behavioral and sensory responses to such stimuli. In songbirds like the zebra finch, species-typical songs consist of acoustic elements (syllables) arranged into stereotyped (i.e., highly predictable) sequences with stereotyped timing. However, the degree to which deviations away from these stereotyped temporal patterns modulate the strength of vocal learning (i.e., the fidelity of vocal imitation) remains unknown. Here we tutored 127 juvenile zebra finches with stimuli that varied in the stereotypy of syllable sequencing and timing. In contrast to the prediction that deviations away from species-typical stereotypy would diminish vocal learning, deviations from sequence or timing stereotypy did not decrease how well juveniles imitated the acoustic structure of syllables. Moreover, presenting syllables in species-atypical sequences (i.e., randomized syllable sequences) enhanced vocal imitation in birds that were tutored later in development. This unexpected enhancement of birdsong learning by sequence variability resembles effects of contextual diversity on speech acquisition and suggests that such variability can benefit learning even for very stereotyped behaviors.

Significance statement

Little is known about the degree to which the temporal patterning of acoustic elements ("syllables") modulates vocal learning and imitation. We tested the contribution of temporal variation in syllable sequencing and timing, hypothesizing that deviations away from species-typical stereotypy (predictability) in syllable sequence and timing patterns would diminish the strength of vocal learning. In contrast to our predictions, deviations from sequence and timing stereotypy did not decrease vocal learning. Moreover, presenting syllables in random sequences enhanced vocal imitation in birds that were tutored later in development. This unexpected finding highlights that sequence complexity can potentiate vocal learning in individuals with more developed vocal control, resembling how complexity and ability interact to modulate various forms of learning in humans.

Introduction

Many important behaviors are organized into species-typical temporal patterns. For example, mice and rats engage in stereotyped sequences of grooming behavior that vary across strains and species (Spruijt et al., 1992; Geuther et al., 2021), and different species of songbirds produce vocal sequences with different statistical (sequence) dependencies (e.g., Jin and Kozhevnikov, 2011; Warren et al., 2012; Lipkind et al., 2013; reviewed in Kershenbaum et al., 2014; Cohen et al., 2020; James et al., 2020; Searcy et al., 2022). In addition to the sequences themselves, the timing of elements within behavioral sequences can demonstrate varying degrees of species specificity or typicality (e.g., Fox et al., 2020; Liberman et al., 1958; Araki et al., 2016; Feng et al., 1990; Gerhardt, 2005; Hedwig, 2006; James et al., 2023).

The sensory systems of organisms are tuned to these temporal patterns, with typical sequences and timing leading to stronger behavioral responses or more robust perception compared to atypical patterns (Gerhardt et al., 2007; Comins and Gentner, 2010; Benichov et al., 2016; Taylor et al., 2017; reviewed in Fishbein et al., 2020; Speck et al., 2020). For instance, California thrashers produce particular sequences of song phrases, and males respond more strongly to stimuli in which the sequence of phrases conforms to typical patterns than to stimuli that violate those patterns (Taylor et al., 2017), and female swamp sparrows prefer acoustic elements ("syllables") in which the order of constituents ("notes") conforms to the pattern typical of their local dialect (Balaban, 1988). Male túngara frogs produce a stereotyped sequence of a "whine" followed by a "chuck" (whine-chucks), and the attractiveness of the call can vary depending on the timing of the whine relative to the chuck (Wilczynski et al., 1999). In addition, increasing or decreasing the silent duration between syllables modulates agonistic responses in indigo buntings (Emlen 1972) and regulates neural responses in various songbird species (Margoliash and Fortune, 1992; Lampen et al., 2014; Araki et al., 2016; Bouchard and Brainard, 2016). While there is substantial evidence that temporal patterning affects the strength of behavioral and neural responses, little is known about how these patterns affect cognitive processes associated with communication.

Here we examined the degree to which species-typical temporal patterns regulate vocal learning in the zebra finch. Songbirds like the zebra finch learn their songs during sensitive periods in development (Doupe and Kuhl, 1999; Brainard and Doupe, 2002; Gobes et al., 2019; Sakata and Woolley, 2020), involving both sensory and sensorimotor learning. Zebra finches produce highly stereotyped songs as adults, with acoustic elements (syllables) arranged into stereotyped sequences (motifs) with stereotyped timing (Sossinka and Böhner, 1980; Glaze and Troyer, 2006; Norton and Scharff, 2016; Murphy et al., 2017; James et al., 2023). It has been hypothesized that such stereotypy could facilitate learning because stereotyped sequences of stimuli are more predictable, and it has been demonstrated that more predictable sequences are learned faster by humans (Henson, 1998; Levy, 2008; Smith and Levy, 2013; Clark, 2013; Pearce, 2018; Varela et al., 2024; Elazar et al., 2022; Siegleman et al., 2018). To test whether stereotypy promotes vocal learning in songbirds, we tutored juvenile zebra finches at various ages in development with acoustic stimuli that varied in their degree of species typicality in syllable sequencing or timing.

Results

We tutored 127 juvenile male zebra finches with stimuli consisting of the same five syllable types using the same controlled, operant tutoring paradigm (see Methods; Figure 1a). This approach provides a powerful opportunity to reveal factors that modulate the strength of vocal learning. We first analyzed how the age and duration of tutoring modulated the strength of vocal learning and then investigate how the temporal variability of the tutor stimulus interacted with the age at tutoring to regulate learning.

Developmental modulation of song learning

Song is learned during a sensitive period in development that gradually closes over time (Eales, 1985; Brainard and Doupe, 2002; Gobes et al., 2019; Sakata and Woolley, 2020). We tutored juvenile male zebra finches with an identical set of species-typical syllables starting from when they were 31-67 days old ("age at tutoring"). The duration of tutoring ranged from 30-82 days and was statistically independent of the age at tutoring. Therefore, we simultaneously analyzed how age at tutoring and tutoring duration explained variation in the strength of song learning. We computed the acoustic similarity of the adult songs of tutored birds to the tutor stimulus using established software and methods [Sound Analysis Pro (SAP); "%similarity" as a proxy for the strength of song learning (see Methods)].

Tutored birds demonstrated a wide range of song learning, producing songs that ranged from 12-97% similarity to the tutor stimulus (Figure 1). Age at tutoring (estimate ± error: -0.786 ± 0.177; t_{117}=4.44, $p<0.0001$) but not tutoring duration (-0.176 ± 0.114; t_{117}=1.55 $p=0.1249$) significantly affected the acoustic similarity to the tutor stimulus (Figure 1b,c). Not surprisingly, the older birds were when tutoring began, the less similar their songs were to the tutor stimulus (i.e., inverse relationship between age at tutoring and song learning).

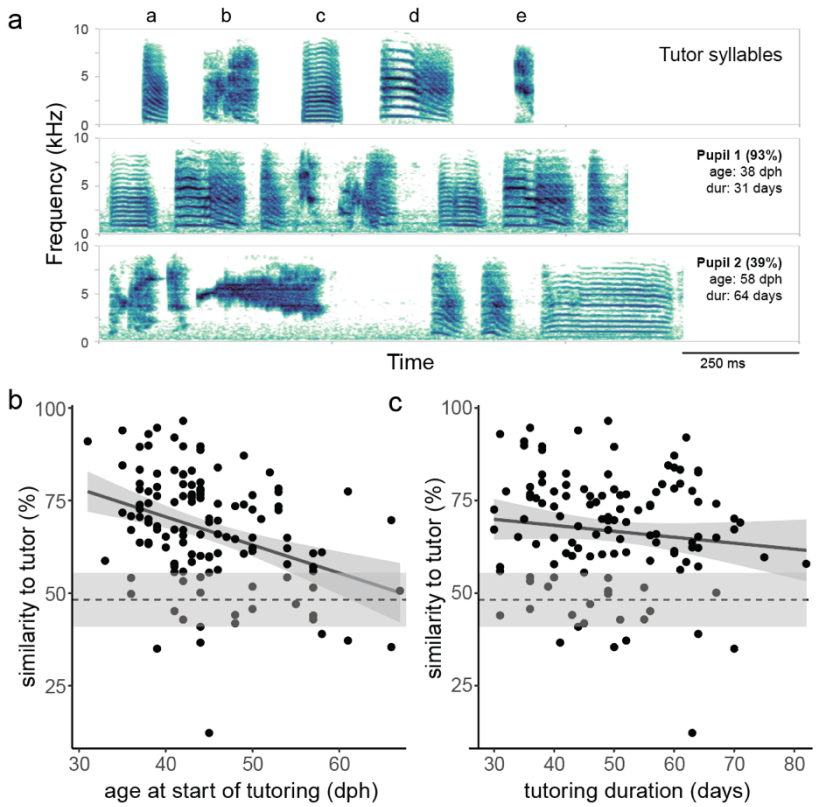

Figure 1: *The strength of song learning is affected by the age at which tutoring started but not by the duration of tutoring. (a) Spectrograms of tutor syllables and the adult songs of pupils. Top: spectrogram of the five tutor syllables; middle and bottom: spectrograms of examples of 1-sec snippets of the adult songs of two pupils. The first pupil was tutored early (<50 dph) and produced a song with high similarity to the tutor (93%), while the second bird was tutored later (>50 dph) and produced a song with lower similarity to the tutor. Both birds were tutored with stereotyped (S) stimuli (see Figure 2). (b,c) The age of pupils at the start of tutoring (b; "age at tutoring") but not the duration of tutoring (c) correlates significantly and negatively with the degree of song learning (i.e., similarity to the tutor song). For both (b) and (c), the dashed horizontal line depicts the mean song similarity [plus 95% confidence interval (CI)] of the songs of untutored birds to the tutor stimulus and represents the chance level of similarity to the tutor song.*

To contextualize the %similarity scores of tutored birds, we computed the acoustic similarity of the adult songs of birds that were not exposed to zebra finch song during the sensitive period for song learning ("untutored birds"; n=11; see Methods) to the tutor stimulus and used these values as a baseline to assess the significance of song learning (i.e., chance level of similarity to the tutor stimulus). When visually comparing the similarity scores of tutored birds to chance [mean±1.96*S.E. (95% confidence interval) of untutored birds], it is evident that the similarity scores of most tutored birds that started their tutoring before ~50 dph were consistently higher than chance but that the similarity scores of many birds that started learning their songs >50 dph fell within the range of chance values (Figure 1b). Because it is difficult to statistically analyze deviation from baseline values in this manner, we categorized birds based on their age at

tutoring, with birds tutored starting from <50 dph classified as "early tutored birds" (n=96) and birds tutored from >50 dph as "late tutored birds" (n=31). Both early and late tutored birds produced songs that were significantly more similar to the tutor stimulus than the songs of untutored birds (Dunnett's test; $p<0.015$ for each), suggesting significant learning at both developmental periods. However, consistent with the previous analysis, early tutored birds produced songs that were significantly more similar to the tutor stimulus than late tutored birds ($t_{125}=3.1$, $p=0.0028$).

General effect of temporal variability of tutor stimuli on song learning

After documenting general learning in response to our stimuli, we next analyzed how variability in syllable timing or sequencing affected song learning. The duration of silent gaps between syllables is highly consistent from rendition-to-rendition for adult zebra finch song (e.g., Glaze and Troyer, 2006; Norton and Scharff, 2016; James et al., 2023) and is encoded by neurons in auditory and sensorimotor circuits of songbirds (Margoliash and Fortune, 1992; Lampen et al., 2014; Araki et al., 2016; Bouchard and Brainard, 2016). Zebra finches also produce stereotyped sequences of syllables, and behavioral responses and neural responses within circuits for song learning are sensitive to the sequential order of syllables (e.g., Cynx, 1993; Lewicki and Konishi, 1995; Lewicki and Arthur, 1996; Braaten et al., 2006; Bouchard and Brainard, 2013, 2016; Chen and ten Cate, 2015; Cazala et al., 2019). Given this species-typical stereotypy of timing and sequencing, one group of birds was tutored with stimuli in which syllables were arranged into stereotyped sequences (i.e., same sequence of syllables across all renditions) and with stereotyped gap durations (i.e., all gap durations within the motif and across renditions set to a particular duration)("stereotyped" or S; n=41). The second group of birds was tutored with a species-atypical stimulus where the gap durations randomly varied from 10-50 ms (in 10-ms increments; with an average of 30-ms gaps)("variable gaps" or Vgap; n=32), but with stereotyped sequences of syllables. This degree of variability in gap durations across renditions is >5 times greater than the variability observed in typical zebra finch song (James et al., 2023). The third and final group of zebra finches was tutored with a species-atypical stimulus that consisted of variable (randomized) sequences of syllables ("variable sequencing" or Vseq) but with stereotyped gap durations (30-ms gaps; n=54).

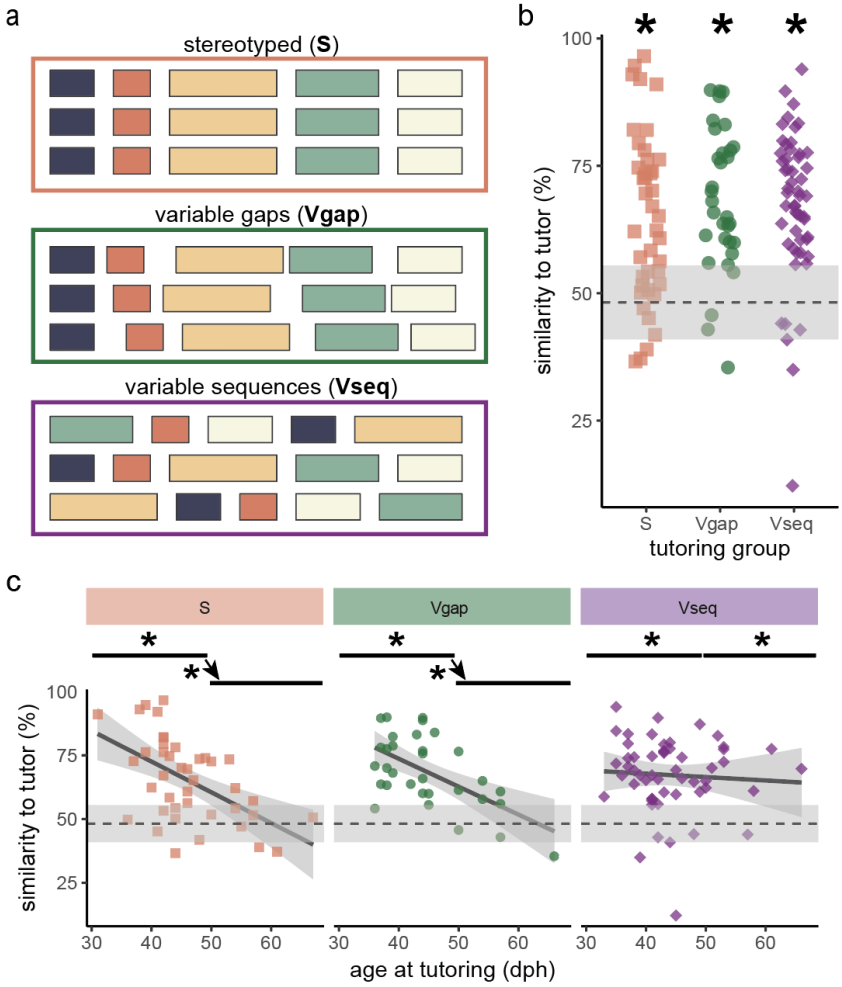

***Figure 2**. Interactive effects of tutor song organization and age at tutoring on the strength of vocal learning (acoustic similarity to the tutor song). (a) Example schematics of the organization of the three tutor stimuli: stereotyped sequences and gap durations (S), stereotyped sequences with variable gap duration (Vgap), and variable sequences with stereotyped gap durations (Vseq). (b) Tutored birds learn significantly better than chance (i.e., vs. untutored birds), regardless of tutor stimulus organization. Each point depicts the average acoustic similarity of a bird's song to the tutor stimulus (see Methods). (c) Tutor stimulus organization interacts significantly with the age at tutoring to affect the strength of song learning, with the slope being shallowest in the Vseq group. The bars above each panel denote the age ranges for early tutored birds and late tutored birds. "*" above the bar indicates when distributions are significantly different than untutored birds, and "*" next to arrows indicates when there is a significant difference in similarity to the tutor between early vs. late tutored birds. For (b) and (c), the dashed grey line depicts the mean (and 95% CI) acoustic similarity of the songs of untutored birds to the tutor stimulus (i.e., chance level of similarity).*

Overall, acoustic similarity scores (%similarity) were not significantly different among S-, Vgap-, and Vseq-tutored birds ($F_{(2,124)}$=0.9, p=0.4163). Compared to untutored birds, S-, Vgap- and Vseq-tutored birds all produced songs that were significantly more similar to the tutor stimulus

($F_{(3,134)}$=5.6, p=0.0012; Dunnett's test: p<0.0030 for each; Figure 2b), indicating significant song learning in response to each type of tutor song organization.

Interactive effects of temporal variability and age at tutoring on song learning

Given the influence of age at tutoring on the strength of vocal learning (acoustic similarity to the tutor stimulus), we next investigated how the relationship between age at tutoring and vocal learning varied across the three types of tutor song organization. Indeed, it is possible that differences in the efficacy of tutor stimuli are more apparent later in development when learning mechanisms are naturally dwindling. Interestingly, there was a significant interaction between tutor stimulus and age at tutoring (ANCOVA: $F_{(2,121)}$=4.2, p=0.0162; Figure 2c). This appeared to be driven by differences in the strength of the negative relationship among S-, Vgap- and Vseq-tutored birds. When the relationship between song learning and age at tutoring was analyzed within each group individually, the relationship was significantly negative for S- and Vgap-tutored birds (p<0.0050 for each) but not for Vseq-tutored birds (p=0.6531). [Note: there was no significant difference in the age at tutoring among S-, Vgap-, and Vseq-tutoring birds.]

To complement the preceding analyses of age at tutoring, we categorically divided tutored birds into early tutored birds (<50 dph) vs. late tutored birds (>50 dph; see above) and examined how song learning varied across tutor song organization and age category (full-factorial model). [It is around 50 dph that one observes the degree of acoustic similarity between the songs of tutored birds and the tutor stimulus approaching chance levels: Figure 1b.] Consistent with the preceding (ANCOVA) analyses, there was a significant interaction between age category (early vs. late) and tutor song organization ($F_{(2,121)}$=4.7, p=0.0107). This was driven by the facts (a) that late tutored birds produced songs that were significantly less similar to the tutor stimulus than early tutored birds for S- and Vgap-tutored birds but not for Vseq-tutored birds and (b) that Vseq-tutored birds produced songs that were significantly more similar to the tutor stimulus than S- or Vgap-tutored birds among late tutored birds (planned contrasts; p<0.05 for each contrast) but not among early tutored birds (Figure 2c).

To investigate which groups of birds demonstrated significant learning, we compared the similarity scores of birds tutored early or later in development to the similarity scores of untutored birds in the model. Early tutored birds demonstrated significant learning, regardless of tutor stimulus organization (ANOVA: $F_{(3,103)}$=7.7, p=0.0001; Dunnett's: p <0.0010 for each; Figure 2c). However, for late tutored birds, only Vseq-tutored birds demonstrated significant learning (ANOVA: $F_{(3,38)}$=6.1, p=0.0017; Dunnett's: p=0.0005; Figure 2c).

It was the case that Vseq-tutored birds were tutored for significantly more days than S- or Vgap-tutored birds ($F_{(2,117)}$=9.4, p=0.0002; Tukey's HSD: p<0.0100 for each). The preceding analysis of all tutored birds suggests an overall non-significant influence of tutoring duration on song learning. We extended this previous analysis and analyzed how tutoring duration and the tutor stimulus organization (S, Vgap, and Vseq) interacted to affect acoustic similarity to the tutor stimulus (strength of vocal learning) and arrived at similar results that tutoring duration does not influence learning. For example, tutoring duration was not significantly related to vocal learning

(-0.185±0.142; $F_{(1,114)}$=1.7, p=0.1963), and there was no significant interaction between tutoring duration and tutor stimulus category on vocal learning ($F_{(2,114)}$=0.1, p=0.9248; i.e., tutoring duration did not affect song learning in S-, Vgap- and Vseq-tutored birds). Other comparisons similarly speak to the lack of effect of tutoring duration on the observed variation in song learning. For example, despite that early-tutored birds produced songs that were more similar to the tutor stimulus than late tutored birds, there were no significant differences in the duration of tutoring between early and late tutored birds, regardless of tutoring method. Relatedly, there was a difference in song learning among S-, Vgap- and Vseq-tutored birds for late tutored birds but for early tutored birds, but differences in tutoring duration among S-, Vgap- and Vseq-tutored birds were observed for both late and early tutored birds. Collectively, these analyses suggest that differences in the duration of tutoring between S-, Vgap-, and Vseq-tutored birds (all were tutored for >30 days) did not contribute to variation in the song learning, a finding supported by previous studies (Eales 1985; Peters et al., 1992; Roper and Zann, 2006; Gobes et al., 2019).

Interactive effects of temporal variability and age at tutoring on syllable learning

Our tutor stimulus consisted of five species-typical syllable "types" that are commonly observed in zebra finch songs (Figure 1a; Zann 1993, 1996): a short, frequency-modulated syllable ("a"), a longer, more complex, and spectrally entropic syllable ("b"; sometimes referred to as a "noise-noise syllable": Zann 1993), a flat, harmonic stack of intermediate duration ("c"), a longer, complex, frequency-modulated syllable ("d", a distance call), and a short, high-frequency syllable ("e"). Despite that these species-typical syllables are commonly observed in zebra finch song, it is not clear whether the acquisition and integration of these syllables are differentially affected by tutor song organization and age at tutoring. Therefore, we analyzed the degree to which birds tutored with S, Vgap, or Vseq stimuli beginning at different ages produced syllables that resembled each of these tutor syllables ("a"-"e").

We computed the average %similarity score across each pupil's song rendition to each individual tutor syllable; this average reflects that the degree to which a syllable similar to the tutor syllable was present in the pupil's song. We first ran a full-factorial model with tutor song organization (S, Vgap, and Vseq), age at tutoring, and syllable ID ("a"-"e") as explanatory variables and found that syllable ID significantly interacted with both tutor song organization and age at tutoring (three-way interaction) as well as just with age at tutoring (p<0.01 for each). Therefore, we next analyzed the interactive effects of tutor song organization and age at tutoring on the %similarity scores for each of the five syllables independently (Figure 3a). For the "a", "b", "d", and "e" syllables, there was a significant effect of age, with %similarity to the syllable decreasing as the age of tutoring increased (i.e., pupils were less likely to produce syllables that matched the tutor syllable the later their tutoring started; p<0.05 for each). Furthermore, there was a significant interaction between tutor song organization and age at tutoring for the "a" ($F_{(2,121)}$=6.4, p=0.0023) and "c" syllables ($F_{(2,121)}$=4.0, p=0.0206), reminiscent of the interaction for song similarity (Figure 2c). For the "a" syllable, this interaction seemed to be driven by a significant negative relationship between %similarity and age at tutoring for S-tutored birds (p=0.0007) but not for Vgap- or Vseq-tutored birds (p>0.15 for each). For the "c" syllable, this interaction seemed to be driven by a negative trend for S-tutored birds (p=0.0683), no significant

relationship for Vgap-tutored birds (p=0.6563), and a significant positive relationship for Vseq-tutored birds (p=0.0461).

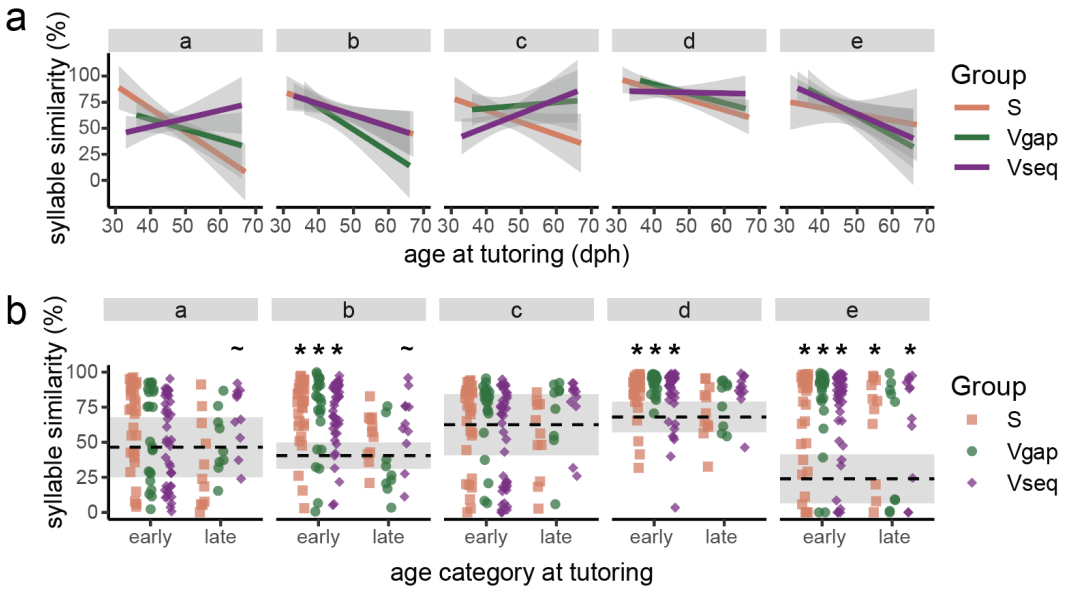

***Figure 3**. Acquisition of tutor syllables. (a) We analyzed the degree to which syllables within the pupil's song matched the tutor syllables ("a"-"e"). Specifically, we computed the average %similarity of each tutor syllable to each pupil's song and correlated this value with the age of tutoring (dph = days post hatch). These relationships were computed for S-, Vgap-, and Vseq-tutored birds independently (see text). Lines plus shaded areas depict the relationship between syllable similarity and age at tutoring. (b). Depiction of raw data for early or late tutored birds according to tutor song organization. The horizontal dashed line represents mean chance level of syllable similarity (i.e., %similarity to each syllable for untutored birds) with the grey area depicting the 95% confidence interval.*

To gain insight into the significance of syllable learning, we computed the %similarity scores of the adult songs of untutored birds to each tutor syllable, and then compared the %similarity scores for each individual syllable in S-, Vgap-, or Vseq-tutored birds to these baseline values (see above for similar analysis of song). Among early tutored birds (<50 dph; Figure 3b), there were significant differences among the groups for the "b", "d" and "e" syllables (ANOVA: p<0.05 for each), with S-, Vgap- and Vseq-tutored birds demonstrating significantly higher similarity scores for these syllables than untutored birds (Dunnett's test; p<0.05 for each). Among late tutored birds (Figure 3b), there was evidence of significant learning of the "b" and "e" syllables (ANOVA: p<0.05 for each): for the "b" syllable, Vseq-tutored birds tended to demonstrate higher similarity scores than untutored birds (Dunnett's test; p=0.0550), and for the "e" syllable, S- and Vseq-tutored birds demonstrated significantly higher similarity scores than untutored birds (Dunnett's test; p<0.05 for each). There was also a trend for group differences in the similarity to the "a" syllable for older birds (ANOVA: p=0.0576), with Vseq-tutored birds tending to demonstrate higher similarity scores than untutored birds.

Discussion

The sounds that animals use for communication are organized into species-typical patterns. A number of species learn the spectral features of acoustic units for communication ("vocal production learning"), and we tested the hypothesis that deviations away from species-typical temporal patterns would attenuate spectral learning in zebra finches. In addition to previous experiments in songbirds demonstrating such an effect (e.g., Marler and Peters, 1988), we also hypothesized this because the stereotyped nature of syllable sequencing and timing for zebra finch song creates predictability that enhances various forms of learning in humans, including music and speech acquisition (Garrido et al., 2018; Henson, 1998; Levy, 2008; Smith and Levy, 2013; Clark, 2013; Hansen and Pearce, 2014; Bianco et al., 2020; Otsuka and Saiki, 2016; Benitez and Saffran, 2018; Gold et al., 2019; Bianco et al., 2019). Contrary to our expectations, deviations away from species-typical stereotypy in syllable timing or sequencing did not attenuate vocal learning. Moreover, increasing the variability of syllable sequencing (i.e., decreasing stereotypy) enhanced vocal learning. For example, among birds that started tutoring ≥50 dph (late tutored birds), only birds tutored with randomized sequences (and not birds tutored with stereotyped sequences or random gap durations) demonstrated significant song learning. These novel findings not only link the temporal patterning of elements to the fidelity with which individuals learn or imitate those elements (e.g., Marler and Peters, 1988; Plamondon et al., 2008; Mol et al., 2021) but also highlight the role of temporal variability or complexity in engaging mechanisms underlying vocal learning.

Given that sequence variability enhanced the ability of juvenile zebra finches to learn the acoustic structure of syllables, we propose that sequence variability within a tutor stimulus could enhance a juvenile's attention to the tutor stimulus and, consequently, enhance the sensory encoding of song. Attention has been postulated to play a major role in vocal learning in songbirds as well as humans (ten Cate, 1986; Meltzoff et al., 2009; Chen et al., 2016; Carouso-Peck et al., 2020), and the random sequencing of syllables could prevent habituation to the tutor stimulus, maintain the salience of syllables, and enhance the consolidation of auditory stimuli. Enhancing attention could be particularly important for birds that start to learn songs later in development and could resemble the enhancing effects of social interactions on song learning (Ljubičic et al., 2016; Sakata and Yazaki-Sugiyama, 2020). From a mechanistic perspective, sequence variability can modulate sensory encoding or processing in the sensorimotor nucleus HVC (Bouchard and Brainard, 2016), an area important for the sensory learning of song (Tanaka et al., 2018; reviewed in Ikeda et al., 2020), and it is possible that stimuli consisting of variable sequences lead to patterns of activity or gene expression that promote synaptic plasticity in brain areas underlying song learning.

Despite that many studies highlight the importance of predictability to learning, studies also underscore the importance of contextual diversity on speech acquisition and other forms of learning (reviewed in Raviv et al., 2022). The impact of many types of variability (e.g., diversity of training exemplars, of locations and timing of instruction, of contexts, etc.) on perceptual and cognitive processes have been investigated, but one form of variability pertinent to our study is contextual variability. In the domain of word learning, for example, contextual variability can

represent the diversity of other words a focal word is produced with. Observations of natural word learning reveal that words emerging first are those that are produced in a wide range of word contexts (Hills et al., 2010; Jones et al., 2012). Further, in experiments of novel word learning, adults are significantly better at recalling and recognizing new words when they appear in a greater diversity of contexts (e.g., in paragraphs on different topics; Johns et al., 2016; Frances et al., 2020). Collectively, these data highlight that contextual variability could not only benefit the acquisition of behaviors as sequentially complex as language but also impact the learning of sequentially simple (stereotyped) behaviors like zebra finch song.

What is particularly interesting about previous findings (and relevant to our findings) is that contextual variability differentially affects learning depending on the age or proficiency of the learner. Low levels of variability (e.g., lower contextual diversity) tend to be more useful in the early stages of learning whereas higher levels of variability can be useful for individuals who have already acquired some proficiency (reviewed in Raviv et al., 2022). For example, in contrast to students with relatively little math training, more experienced math students benefit from exposure to arithmetic problems with high rather than low variability (Likourezos et al., 2019). While it is not the case that variability in sequencing or timing hindered song learning in early tutored juveniles, it is possible that late tutored juveniles benefited from sequence variability in the tutor stimulus because they have already acquired some mastery of vocal control. Juvenile songbirds will practice singing even if they are not provided with the opportunity to hear and memorize songs to imitate, and this practice has been hypothesized to allow juveniles to create an internal model for sound production (Brainard and Doupe, 2013; Mooney, 2022). In this respect, it is possible that older juveniles (>50 dph) are more adept at vocal control and could be in a state to take advantage of the sequence variability in the tutor stimulus.

Our experiments suggest that learning the acoustic properties of syllables might not be significantly affected by variability in gap timing. Variability in gap durations for Vgap-tutored birds was ~5X as high as that observed in normal zebra finch song, and this degree of variability has been found to affect auditory responses in the zebra finch brain (Lampen et al., 2014, 2019). Despite this variability, Vgap- and S-tutored birds learned the tutor song to a similar degree and displayed similar relationships between song learning and age at tutoring. This lack of effect of gap variability on learning resembles findings from some experiments reporting that auditory perception in zebra finches is less dependent on gap durations than on syllable structure (e.g., Nagel et al., 2010; Dooling and Prior 2017).

While we hypothesize that variability in syllable sequencing underlies the enhancement of song learning, our results could also highlight the importance of timing variability. Gap durations were fixed in the Vseq stimulus, and tutor syllables differed in duration; consequently, sequence variability leads to variability in the intervals between the onsets of consecutive syllables (inter-syllable interval, or ISI). Syllable durations ranged from 45-150 ms, causing ISIs to vary to this extent, and this variability in ISIs was larger in the Vseq stimulus than in the Vgap stimulus. As such, it is possible that this larger degree of ISI variability in the Vseq stimulus could have contributed to the enhancement of learning.

In addition to the interactive effects of the temporal variability of tutor song and the age at tutoring on song learning, we also noted variation in the relationship between the age at tutoring and syllable learning across different tutor syllables ("a"-"e"). Although similarity to each tutor syllable generally decreased as the age of tutoring increased for most syllables, the slope was steepest for the "b" and "e" syllables. Zann (1993) classified 2711 elements (from 402 zebra finch males) into 14 different syllable categories or types, and the five syllables used in the tutor stimulus represent the top five categories. These syllable types were found in the songs of 40-90% of the males, with the noise-noise ("b") and high pitch ("e") syllable types being the least prevalent of the five. Therefore, the tutor syllable types that were least prevalent in wild birds were the syllable types that showed the steepest diminution in learning with age. This finding is consistent with the experimental finding that zebra finches are more likely to learn common syllable types over uncommon types (ter Haar et al., 2014).

In summary, our findings reveal that the temporal organization of acoustic elements can affect vocal learning in novel ways. In contrast to previous studies describing that syllable timing and sequencing have negligible effects on auditory perception (Lawson et al., 2018; Kriengwatana et al. 2016; Fishbein et al., 2019; Geberzahn and Deregnaucourt, 2020), our expansive study emphasizes that temporal organization can modulate the strength of vocal learning and is consistent with studies highlighting a significant contribution of syllable timing and sequencing to sensory and sensorimotor processes across vertebrate taxa (Bee and Klump, 2005; Verzijden et al., 2007; van Heijningen et al. 2013; Chen and ten Cate 2015, 2017; Benichov et al., 2016; Chen et al. 2016; Spierings and ten Cate 2016; Knowles et al. 2018; Santolin and Saffran, 2018; Rouse et al., 2021; Mol et al., 2021; Ning et al,. 2023). Because randomized sequences of tutor syllables seemed to potentiate learning, we propose that sequence variability could prevent neural adaptation, allow for greater attention to acoustic features of syllables, and act upon the same cognitive processes as contextual diversity in humans.

Methods

Animals

Song learning in male zebra finches was analyzed. Birds were provided food and water *ad libitum* and housed on a 14 L:10 D light cycle. All procedures were approved by the McGill University Animal Care and Use Committee in accordance with the guidelines of the Canadian Council on Animal Care.

Birds were raised by both parents in a sound-attenuating chamber ("soundbox"; TRA Acoustics, Ontario, Canada) up to 5-7 days of age, at which time their father was removed. Only male zebra finches learn and produce courtship song, and the sensitive period for song learning starts at approximately 20 days post-hatching (Brainard and Doupe, 2013; Roper and Zann, 2006). Thus, juveniles were not exposed to song from a live adult during the sensitive period for song learning and were naïve to song before experimental tutoring. When these juveniles were nutritionally independent (i.e., could feed themselves; ~30–40 days old), they were housed individually in a soundbox for a period of song tutoring (see below) and thereafter until they were mature (~4

months of age; Derégnaucourt et al., 2005; James and Sakata, 2017; Lipkind et al., 2013; Lipkind et al., 2017; Mets and Brainard, 2018; Mets and Brainard, 2019; Tchernichovski et al., 2001; James et al., 2023).

Song tutoring and stimuli

Birds were operantly tutored for >1 month with song playback operantly triggered by perch hops (n = 125) or string pulls (n = 2) using custom-built perches or strings connected to a National Instruments PCI-6503 I/O card (National Instruments, TX). Each hop or string pull triggered the playback of one song (i.e., four concatenated motifs of the tutor stimulus; see below). Song playbacks were spaced in time such that juveniles could hear a maximum of 10 operantly triggered playbacks within each of three time periods per day (morning, noon, and afternoon); this schedule of song exposure has been found to enhance song learning (Tchernichovski et al. 1999; Derégnaucourt et al. 2005). We used Sound Analysis Pro 2011 (SAP; http://soundanalysispro.com) for song tutoring, and stimuli were played out of an Avantone Pro Mixcube speaker (Avantone, NY) connected to a Crown XLS 1000 amplifier (Crown Audio, IN).

Birds were between 31-67 dph when tutoring started ("age at tutoring"). Most of the birds were tutored between 30-50 dph because birds learn best during this period and because many of these birds were tutored for experiments that investigated common song patterns in learned song (i.e., we aimed to maximize learning: James and Sakata, 2017; James et al., 2020, 2023). Importantly, the age at tutoring was not significantly different among S-, Vgap-, and Vseq-tutored birds ($F_{(2,124)}$=0.9, p=0.4163).

Birds were tutored, on average, for 49.6±11.4 days (mean±S.D.; range: 30-82). The duration of tutoring was unavailable for 7 birds (i.e., the start of tutoring was noted but not the end of tutoring). All of these birds were Vseq-tutored birds that started their tutoring between 35-42 dph (i.e., early tutored birds), and their similarity scores were well within the range of Vseq-tutored birds with information on tutoring duration (range: 55-90%). Given the sample size of the current analysis correlating song learning with tutoring duration (n=120 birds), the absence of relationship between the duration of tutoring and similarity scores in the existing dataset, and previous studies documenting a lack of a relationship between tutoring duration on learning for birds tutored for at least a week (Eales, 1985; Roper and Zann, 2006; Gobes et al., 2019), we do not anticipate that this lack of information for 7 birds had substantial effects on the reported results regarding the effect of tutoring duration on song learning.

All song stimuli consisted of five canonical zebra finch song elements ("syllables"; hereafter, "a", "b", "c", "d", and "e"; Figure 1). Zebra finch song is typically organized into bouts that contain multiple renditions of a motif (stereotyped sequence of syllables); song stimuli therefore consisted of four renditions of a motif to emulate species-typical song bouts. In addition, because gaps are longer between motifs than within motifs (Glaze and Troyer 2006; Hyland Bruno and Tchernichovski 2019), between-motif gap durations were fixed at 100 ms for all stimuli. The variability of syllable sequencing and timing across motifs distinguished the experimental groups.

Birds that were exposed to stereotyped (S) stimuli were tutored with stimuli in which the sequence of syllables and duration of silent gaps between syllables within the motif were fixed across all renditions of song (n=41). For instance, a bird might be tutored solely with the sequence "becad" with 30-ms gaps between each syllable within the motif (see Supplementary Table 1). Ten different motif sequences were used to tutor birds in the S group, and inter-syllable gaps were fixed at 10, 20, 30, 40, or 50 ms (with each bird being tutored with only one of these sequences and generally a single inter-syllable interval: Supplementary Table 1). These durations were chosen to cover the species-typical range of gaps observed in zebra finches (Price 1979; Norton and Scharff 2016; Cooper and Goller 2006; Norton and Scharff 2016; Lachlan et al. 2016; Araki et al. 2016; James and Sakata 2017). Three S-tutored birds were tutored with gap durations that did not vary across renditions of the motif but varied between different syllables in the motif; specifically, two birds were tutored with the sequence "becad" and one bird with the sequence "dacbe" in which the four gaps between syllables within the motif were fixed, respectively, at 20, 50, 10, and 40 (Supplementary Table 1).

Another group of birds was tutored with stereotyped sequences of syllables in which the duration of the gaps varied randomly within and between motifs ("variable gap" or Vgap stimulus; n=32). We manipulated gap durations because neurons in auditory and sensorimotor circuits of songbirds are sensitive to the duration of silent gaps between syllables (Margoliash and Fortune, 1992; Lampen et al., 2014; Araki et al., 2016; Bouchard and Brainard, 2016) and manipulated the variability of gap durations because gap durations are typically highly consistent from rendition-to-rendition for adult zebra finch song (Glaze and Troyer, 2006; Norton and Scharff, 2016; James et al., 2023); for example, across 43 adult zebra finches, median gap durations were ~40 ms and the median IQR for these gaps was 3-4 ms: L.S.J. & J.T.S, unpublished data). Each bird was presented with a single sequence of syllables (e.g., "becad"), but the gap durations between syllables within the motif were randomly selected from 5 possible gap durations: 10, 20, 30, 40, or 50 ms. Zebra finches attend to, discriminate among, and imitate gap durations in this range of gap durations (e.g., James et al., 2023). Gap durations were pseudo-randomly chosen such that each gap duration was presented the same number of times; as such, the average gap duration was 30 ms for these stimuli. Twelve different motif sequences were used to tutor birds in the Vgap group (Supplementary Table 1), only two of which differed from those used as stimuli for S-tutored birds.

The last group of birds were tutored with stimuli in which the sequence of the syllables within the motif varied from rendition to rendition (including within the same song bout; "variable sequencing" or Vseq stimulus; n=54). Specifically, we first synthesized 120 different 5-syllable sequence variants in which each syllable is produced only once. Sequences were created in this manner because zebra finches do not typically repeat syllables within the motif; therefore, sequences in which each syllable is produced only once is typical of this species. We then pseudo-randomly selected and concatenated four different sequence variants to construct a song bout, ensuring that each of the 120 sequence variants was presented equally. Gap durations within the motif were fixed at 30 ms for all birds in the Vseq group; this was the average gap

duration within the stimuli of Vgap-tutored birds and the most common gap duration for the stimuli of S-tutored birds.

In addition to tutored birds, some birds were raised without exposure to conspecific song ("untutored birds"; n=11) during the sensitive period for song learning; they were raised with their mother and then housed individually until their songs were recorded in adulthood (see below). These birds served to compute chance levels of similarity to the tutor stimuli (i.e., to assess the significance of song learning).

Aspects of the sequencing or timing of the songs of many of these birds have been published previously (James and Sakata, 2017; James et al., 2020, 2023). For this manuscript, data from 8, 4, and 5 additional birds were added to the S-, Vgap-, and Vseq-tutored groups, respectively. Importantly, we analyzed the spectral similarity of the song snippets each of these tutored birds to the tutor stimulus, which has not been previously measured or reported.

Song recording and analysis

When tutored birds were sexually mature (mean±S.D.: 121.4±8.6 dph; range: 87-139), they were individually housed in a soundbox for song recording. Ages of recording were not significantly different among the three groups of tutored birds ($F_{(2,124)}$=0.2, p=0.8348). Untutored birds were slightly older at the time of song recording (mean±S.D.: 132.0±20.2 dph; range: 92-164).

For all song recordings, birds were housed individually in a soundbox, and song was recorded using an omnidirectional microphone (Countryman Associates, Inc, Menlo Park, CA) positioned above the bird's cage. A computerized, song-activated recording system was used to detect and digitize song (SAP; digitized at 44.1 kHz). Recorded songs were digitally filtered (0.3–10 kHz) for off-line analysis using software custom-written in the Matlab programming language (MathWorks, Natick, MA). All songs recorded and analyzed were spontaneous songs produced in isolation ("undirected song").

After identifying recordings with songs, we annotated introductory notes and song syllables within song bouts. Zebra finches repeat a series of short, acoustically simple syllables called introductory notes before singing their song. Even untutored birds and birds tutored without introductory notes in the stimulus produce readily identifiable introductory notes in their songs (e.g., Price, 1979; Kalra et al., 2021). Many studies quantify song learning by first identifying a stereotyped sequence of syllables (motifs) in pupils. However, a number of pupils in this study produced adult songs with variable acoustic and temporal structure (i.e., some birds did not produce a stereotyped motif at the time of their adult song recording); this made motif identification difficult and subjective. Consequently, we extracted the first 1-sec of song (excluding the repetition of introductory notes) ensuring that we retained the entire syllable that first crossed the one second mark (i.e., we did not clip any syllables). We compared the acoustic similarity of syllables in this snippet to the tutor syllables, providing a relatively unbiased approach to examine song learning across birds, regardless of the stereotypy of their adult song.

The focus on song structure following introductory notes is useful because tutor stimuli were not preceded by the repetition of introductory notes.

We evaluated the acoustic similarity between song snippets and a motif of the tutor stimulus using SAP. Specifically, we computed the %similarity score (using the "asymmetric" and "time course" settings) for each song snippet of a pupil, with a tutor motif as a reference. This measurement uses acoustic features (pitch, frequency modulation, amplitude modulation, goodness of pitch, and Wiener entropy) to compute how much of the tutor stimulus is present in the pupil's song. Importantly, this similarity metric is useful to quantify song learning here because it is not affected by the sequence of syllables in the tutor stimulus. We used the default settings on SAP that are optimized for zebra finch song (p=0.05, interval=70 ms; minimum duration=15 ms). We measured the SAP similarity for, on average, 33.1±10.6 song snippets (mean±S.D.; range: 8 - 77), and computed the mean %similarity score across all song snippets for each pupil for analyses. The number of snippets was not significantly related to the mean %similarity score for tutored birds ($F_{(1,125)}$=1.0, p=0.3295). To identify variation in the learning of each of the five individual tutor syllables, we ran the same similarity measurements across all song snippets except with the tutor stimulus limited to a single syllable (e.g., "a", "b", "c", etc).

Statistical analyses

In general, we used ANOVAs to analyze the effects of tutor stimulus (S, Vgap, and Vseq) on song learning. We also used regressions, multiple regressions, and ANCOVAs to analyze how age at tutoring and duration of tutoring independently affected song learning or interacted with tutor song organization to affect song learning. All analyses were run in JMP v.17 for the Mac, with α=0.05 for all analyses. Figures were generated in R (v.4.2.1) using RStudio v.2023.12.1.

References

Araki, M., Bandi, M. M., & Yazaki-Sugiyama, Y. (2016). Mind the gap: Neural coding of species identity in birdsong prosody. *Science*, *354*(6317), 1282–1287. https://doi.org/10.1126/science.aah6799

Balaban, E. (1988). Bird song syntax: Learned intraspecific variation is meaningful. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(10), Article 10. https://doi.org/10.1073/pnas.85.10.3657

Beckers, G. J. L., Goossens, B. M. A., & Ten Cate, C. (2003). Perceptual salience of acoustic differences between conspecific and allospecific vocalizations in African collared-doves. *Animal Behaviour*, *65*(3), 605–614. https://doi.org/10.1006/anbe.2003.2080

Beckers, G. J. L., & Ten Cate, C. (2001). Perceptual relevance of species-specific differences in acoustic signal structure in Streptopelia doves. *Animal Behaviour*, *62*(3), 511–518. https://doi.org/10.1006/anbe.2001.1768

Bee, M. A., & Klump, G. M. (2005). Auditory stream segregation in the songbird forebrain: Effects of time intervals on responses to interleaved tone sequences. *Brain, Behavior and Evolution*, *66*(3), 197–214. https://doi.org/10.1159/000087854

Benichov, J. I., Globerson, E., & Tchernichovski, O. (2016). Finding the beat: From socially coordinated vocalizations in songbirds to rhythmic entrainment in humans. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00255

Benitez, V. L., & Saffran, J. R. (2018). Predictable events enhance word learning in toddlers. *Current Biology*, *28*(17), Article 17. https://doi.org/10.1016/j.cub.2018.06.017

Bianco, R., Gold, B. P., Johnson, A. P., & Penhune, V. B. (2019). Music predictability and liking enhance pupil dilation and promote motor learning in non-musicians. *Scientific Reports*, *9*(1), 17060. https://doi.org/10.1038/s41598-019-53510-w

Bianco, R., Ptasczynski, L. E., & Omigie, D. (2020). Pupil responses to pitch deviants reflect predictability of melodic sequences. *Brain and Cognition*, *138*, 103621. https://doi.org/10.1016/j.bandc.2019.103621

Bouchard, K. E., & Brainard, M. S. (2013). Neural encoding and integration of learned probabilistic sequences in avian sensory-motor circuitry. *J Neurosci*, *33*(45), Article 45. https://doi.org/10.1523/JNEUROSCI.2181-13.2013

Bouchard, K. E., & Brainard, M. S. (2016). Auditory-induced neural dynamics in sensory-motor circuitry predict learned temporal and sequential statistics of birdsong. *Proceedings of the National Academy of Sciences*, *113*(34), 9641–9646. https://doi.org/10.1073/pnas.1606725113

Braaten, R. F., Petzoldt, M., & Colbath, A. (2006). Song perception during the sensitive period of song learning in zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, *120*(2), 79–88. https://doi.org/10.1037/0735-7036.120.2.79

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, *417*(May), Article May. https://doi.org/10.1038/417351a

Brainard, M. S., & Doupe, A. J. (2013). Translating birdsong: Songbirds as a model for basic and applied medical research. *Annual Review of Neuroscience*, *36*(1), Article 1. https://doi.org/10.1146/annurev-neuro-060909-152826

Burgering, M. A., Ten Cate, C., & Vroomen, J. (2018). Mechanisms underlying speech sound discrimination and categorization in humans and zebra finches. *Animal Cognition*, *21*(2), 285–299. https://doi.org/10.1007/s10071-018-1165-3

Burgering, M. A., Vroomen, J., & Ten Cate, C. (2019). Zebra finches (*Taeniopygia guttata*) can categorize vowel-like sounds on both the fundamental frequency ("pitch") and spectral envelope. *Journal of Comparative Psychology*, *133*(1), 106–117. https://doi.org/10.1037/com0000143

Carouso-Peck, S., Menyhart, O., DeVoogd, T. J., & Goldstein, M. H. (2020). Contingent parental responses are naturally associated with zebra finch song learning. *Animal Behaviour*, *165*, 123–132. https://doi.org/10.1016/j.anbehav.2020.04.019

Cate, C. T. (1986). Listening behaviour and song learning in zebra finches. *Animal Behaviour*, *34*(4), 1267–1268. https://doi.org/10.1016/S0003-3472(86)80189-0

Cazala, A., Giret, N., Edeline, J.-M., & Del Negro, C. (2019). Neuronal encoding in a high-level auditory area: From sequential order of elements to grammatical structure. *The Journal of Neuroscience*, *39*(31), 6150–6161. https://doi.org/10.1523/JNEUROSCI.2767-18.2019

Chen, J., Jansen, N., & Ten Cate, C. (2016). Zebra finches are able to learn affixation-like patterns. *Animal Cognition*, *19*(1), 65–73. https://doi.org/10.1007/s10071-015-0913-x

Chen, J., & Ten Cate, C. (2015). Zebra finches can use positional and transitional cues to distinguish vocal element strings. *Behavioural Processes*, *117*, 29–34. https://doi.org/10.1016/j.beproc.2014.09.004

Chen, J., & Ten Cate, C. (2017). Bridging the gap: Learning of acoustic nonadjacent dependencies by a songbird. *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*(3), 295–302. https://doi.org/10.1037/xan0000145

Chen, Y., Matheson, L. E., & Sakata, J. T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences*, *113*(24), Article 24. https://doi.org/10.1073/pnas.1522306113

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cohen, Y., Shen, J., Semu, D., Leman, D. P., Liberti, W. A., Perkins, L. N., Liberti, D. C., Kotton, D. N., & Gardner, T. J. (2020). Hidden neural states underlie canary song syntax. *Nature*, *582*(7813), 539–544. https://doi.org/10.1038/s41586-020-2397-3

Comins, J. A., & Gentner, T. Q. (2010). Working memory for patterned sequences of auditory objects in a songbird. *Cognition*, *117*(1), 38–53. https://doi.org/10.1016/j.cognition.2010.06.009

Cooper, B. G., & Goller, F. (2006). Physiological insights into the social-context-dependent changes in the rhythm of the song motor program. *Journal of Neurophysiology*, *95*(6), 3798–3809. https://doi.org/10.1152/jn.01123.2005

Cynx, J. (1993). Conspecific song perception in zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, *107*(4), 395–402. https://doi.org/10.1037/0735-7036.107.4.395

Derégnaucourt, S., Mitra, P. P., Fehér, O., Pytte, C., & Tchernichovski, O. (2005). How sleep affects the developmental learning of bird song. *Nature*, *433*(7027), 710–716. https://doi.org/10.1038/nature03275

Dooling, R. J., & Prior, N. H. (2017). Do we hear what birds hear in birdsong? *Animal Behaviour*, *124*, 283–289. https://doi.org/10.1016/j.anbehav.2016.10.012

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*(1), Article 1. https://doi.org/10.1146/annurev.neuro.22.1.567

Eales, L. A. (1985). Song learning in zebra finches: Some effects of song model availability on what is learnt and when. *Animal Behaviour*, *33*(4), Article 4. https://doi.org/10.1016/S0003-3472(85)80189-5

Eens, M. (1997). Understanding the complex song of the European starling: An integrated ethological approach. In *Advances in the Study of Behavior* (Vol. 26, pp. 355–434). Elsevier. https://doi.org/10.1016/S0065-3454(08)60384-8

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the "tabula" is anything but "rasa:" What determines performance in the auditory statistical learning task? *Cognitive Science*, *46*(2), e13102. https://doi.org/10.1111/cogs.13102

Emlen, S. T. (1972). An experimental analysis of the parameters of bird song eliciting species recognition. *Behaviour*, *41*(1–2), 130–171. https://doi.org/10.1163/156853972X00248

Feng, A., Hall, J., & Gooler, D. (1990). Neural basis of sound pattern recognition in anurans. *Progress in Neurobiology*, *34*(4), 313–329. https://doi.org/10.1016/0301-0082(90)90008-5

Fishbein, A. R., Idsardi, W. J., Ball, G. F., & Dooling, R. J. (2020). Sound sequences in birdsong: How much do birds really care? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1789), 20190044. https://doi.org/10.1098/rstb.2019.0044

Fishbein, A. R., Prior, N. H., Brown, J. A., Ball, G. F., & Dooling, R. J. (2021). Discrimination of natural acoustic variation in vocal signals. *Scientific Reports*, *11*(1), 916. https://doi.org/10.1038/s41598-020-79641-z

Fox, N. P., Leonard, M., Sjerps, M. J., & Chang, E. F. (2020). Transformation of a temporal speech cue to a spatial neural code in human auditory cortex. *eLife*, *9*, e53051. https://doi.org/10.7554/eLife.53051

Frances, C., Martin, C. D., & Duñabeitia, J. A. (2020). The effects of contextual diversity on incidental vocabulary learning in the native and a foreign language. *Scientific Reports*, *10*(1), 13967. https://doi.org/10.1038/s41598-020-70922-1

Garrido, M. I., Rowe, E. G., Halász, V., & Mattingley, J. B. (2018). Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimuli. *Cerebral Cortex*, *28*(5), 1771–1782. https://doi.org/10.1093/cercor/bhx087

Geberzahn, N., & Derégnaucourt, S. (2020). Individual vocal recognition in zebra finches relies on the song syllable structure rather than on song syllable order. *Journal of Experimental Biology*, jeb.220087. https://doi.org/10.1242/jeb.220087

Gerhardt, H. C. (2005). Advertisement-call preferences in diploid-tetraploid treefrogs (*Hyla chrysoscelis* and *Hyla versicolor*): Implications for mate choice and the evolution of communication systems. *Evolution*, *59*(2), 395–408. https://doi.org/10.1111/j.0014-3820.2005.tb00998.x

Gerhardt, H. C., Humfeld, S. C., & Marshall, V. T. (2007). Temporal order and the evolution of complex acoustic signals. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1619), 1789–1794. https://doi.org/10.1098/rspb.2007.0451

Geuther, B. Q., Peer, A., He, H., Sabnis, G., Philip, V. M., & Kumar, V. (2021). Action detection using a neural network elucidates the genetics of mouse grooming behavior. *eLife*, *10*, e63207. https://doi.org/10.7554/eLife.63207

Glaze, C. M., & Troyer, T. W. (2006). Temporal structure in zebra finch song: Implications for motor coding. *Journal of Neuroscience*, *26*(3), Article 3. https://doi.org/10.1523/JNEUROSCI.3387-05.2006

Gobes, S. M. H., Jennings, R. B., & Maeda, R. K. (2019). The sensitive period for auditory-vocal learning in the zebra finch: Consequences of limited-model availability and multiple-tutor paradigms on song imitation. *Behavioural Processes*, *163*, 5–12. https://doi.org/10.1016/j.beproc.2017.07.007

Gold, B. P., Mas-Herrero, E., Zeighami, Y., Benovoy, M., Dagher, A., & Zatorre, R. J. (2019). Musical reward prediction errors engage the nucleus accumbens and motivate learning. *Proceedings of the National Academy of Sciences*, *116*(8), 3310–3315. https://doi.org/10.1073/pnas.1809855116

Gómez Varela, I., Orpella, J., Poeppel, D., Ripolles, P., & Assaneo, M. F. (2024). Syllabic rhythm and prior linguistic knowledge interact with individual differences to modulate phonological statistical learning. *Cognition*, *245*, 105737. https://doi.org/10.1016/j.cognition.2024.105737

Hansen, N. Chr., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01052

Hedwig, B. (2006). Pulses, patterns and paths: Neurobiology of acoustic behaviour in crickets. *Journal of Comparative Physiology A*, *192*(7), 677–689. https://doi.org/10.1007/s00359-006-0115-8

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*(2), 73–137. https://doi.org/10.1006/cogp.1998.0685

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hyland Bruno, J., & Tchernichovski, O. (2019). Regularities in zebra finch song beyond the repeated motif. *Behavioural Processes*, *163*(October), Article October. https://doi.org/10.1016/j.beproc.2017.11.001

Ikeda, M. Z., Trusel, M., & Roberts, T. F. (2020). Memory circuits for vocal imitation. *Current Opinion in Neurobiology*, *60*, 37–46. https://doi.org/10.1016/j.conb.2019.11.002

James, L. S., & Sakata, J. T. (2017). Learning biases underlie "universals" in avian vocal sequencing. *Current Biology*, *27*(23), Article 23. https://doi.org/10.1016/j.cub.2017.10.019

James, L. S., Sun, H., Wada, K., & Sakata, J. T. (2020). Statistical learning for vocal sequence acquisition in a songbird. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-58983-8

James, L. S., Wang, A. S., Bertolo, M., & Sakata, J. T. (2023). Learning to pause: Fidelity of and biases in the developmental acquisition of gaps in the communicative signals of a songbird. *Developmental Science*, *26*(5), e13382. https://doi.org/10.1111/desc.13382

Jin, D. Z., & Kozhevnikov, A. A. (2011). A compact statistical model of the song syntax in Bengalese finch. *PLoS Computational Biology*, *7*(3), Article 3. https://doi.org/10.1371/journal.pcbi.1001108

Johns, B. T., Dye, M., & Jones, M. N. (2016). The influence of contextual diversity on word learning. *Psychonomic Bulletin & Review*, *23*(4), 1214–1220. https://doi.org/10.3758/s13423-015-0980-7

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *66*(2), 115–124. https://doi.org/10.1037/a0026727

Kalra, S., Yawatkar, V., James, L. S., Sakata, J. T., & Rajan, R. (2021). Introductory gestures before songbird vocal displays are shaped by learning and biological predispositions. *Proceedings. Biological Sciences*, *288*(1943), Article 1943. https://doi.org/10.1098/rspb.2020.2796

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: Not the Markov chains we thought they were. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1792), Article 1792. https://doi.org/10.1098/rspb.2014.1370

Knowles, J. M., Doupe, A. J., & Brainard, M. S. (2018). Zebra finches are sensitive to combinations of temporally distributed features in a model of word recognition. *The Journal of the Acoustical Society of America*, *144*(2), 872–884. https://doi.org/10.1121/1.5050910

Kriengwatana, B., Spierings, M. J., & Ten Cate, C. (2016). Auditory discrimination learning in zebra finches: Effects of sex, early life conditions and stimulus characteristics. *Animal Behaviour*, *116*, 99–112. https://doi.org/10.1016/j.anbehav.2016.03.028

Lachlan, R. F., van Heijningen, C. A. A., ter Haar, S. M., & ten Cate, C. (2016). Zebra finch song phonology and syntactical structure across populations and continents—A computational comparison. *Frontiers in Psychology*, *7*, 980. https://doi.org/10.3389/fpsyg.2016.00980

Lampen, J., Jones, K., McAuley, J. D., Chang, S.-E., & Wade, J. (2014). Arrhythmic song exposure increases ZENK expression in auditory cortical areas and nucleus taeniae of the adult zebra finch. *PLoS ONE*, *9*(9), e108841. https://doi.org/10.1371/journal.pone.0108841

Lampen, J., McAuley, J. D., Chang, S.-E., & Wade, J. (2019). Neural activity associated with rhythmicity of song in juvenile male and female zebra finches. *Behavioural Processes*, *163*, 45–52. https://doi.org/10.1016/j.beproc.2017.12.003

Lawson, S. L., Fishbein, A. R., Prior, N. H., Ball, G. F., & Dooling, R. J. (2018). Relative salience of syllable structure and syllable order in zebra finch song. *Animal Cognition*, *21*(4), 467–480. https://doi.org/10.1007/s10071-018-1182-2

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Lewicki, M. S., & Arthur, B. J. (1996). Hierarchical organization of auditory temporal context sensitivity. *The Journal of Neuroscience*, *16*(21), 6987–6998. https://doi.org/10.1523/JNEUROSCI.16-21-06987.1996

Lewicki, M. S., & Konishi, M. (1995). Mechanisms underlying the sensitivity of songbird forebrain neurons to temporal order. *Proceedings of the National Academy of Sciences*, *92*(12), 5582–5586. https://doi.org/10.1073/pnas.92.12.5582

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, *1*(3), 153–167. https://doi.org/10.1177/002383095800100301

Likourezos, V., Kalyuga, S., & Sweller, J. (2019). The Variability Effect: When instructional variability is advantageous. *Educational Psychology Review*, *31*(2), 479–497. https://doi.org/10.1007/s10648-019-09462-8

Lipkind, D., Marcus, G. F., Bemis, D. K., Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., Okanoya, K., & Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, *498*(7452), Article 7452. https://doi.org/10.1038/nature12173

Lipkind, D., Zai, A. T., Hanuschkin, A., Marcus, G. F., Tchernichovski, O., & Hahnloser, R. H. R. (2017). Songbirds work around computational complexity by learning song vocabulary independently of sequence. *Nature Communications*, *8*(1), Article 1. https://doi.org/10.1038/s41467-017-01436-0

Ljubičić, I., Hyland Bruno, J., & Tchernichovski, O. (2016). Social influences on song learning. *Current Opinion in Behavioral Sciences*, *7*, 101–107. https://doi.org/10.1016/j.cobeha.2015.12.006

Margoliash, D., & Fortune, E. S. (1992). Temporal and harmonic combination-sensitive neurons in the zebra finch's HVc. *The Journal of Neuroscience*, *12*(11), Article 11.

Marler, P., & Peters, S. (1988). The role of song phonology and syntax in vocal learning preferences in the song sparrow, *Melospiza melodia*. *Ethology*, *77*(2), Article 2. https://doi.org/10.1111/j.1439-0310.1988.tb00198.x

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science*, *325*(5938), 284–288. https://doi.org/10.1126/science.1175626

Mets, D. G., & Brainard, M. S. (2018). Genetic variation interacts with experience to determine interindividual differences in learned song. *Proceedings of the National Academy of Sciences*, *115*(2), Article 2. https://doi.org/10.1073/pnas.1713031115

Mets, D. G., & Brainard, M. S. (2019). Learning is enhanced by tailoring instruction to individual genetic differences. *eLife*, *8*, 1–15. https://doi.org/10.7554/eLife.47216

Mol, C., Bolhuis, J. J., & Moorman, S. (2021). Vocal learning in songbirds: The role of syllable order in song recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200248. https://doi.org/10.1098/rstb.2020.0248

Mooney, R. (2022). Birdsong. *Current Biology*, *32*(20), R1090–R1094. https://doi.org/10.1016/j.cub.2022.07.006

Murphy, K., James, L. S., Sakata, J. T., & Prather, J. F. (2017). Advantages of comparative studies in songbirds to understand the neural basis of sensorimotor integration. *Journal of Neurophysiology*, *118*(2), 800–816. https://doi.org/10.1152/jn.00623.2016

Nagel, K. I., McLendon, H. M., & Doupe, A. J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *Journal of Neurophysiology*, *104*(3), 1426–1437. https://doi.org/10.1152/jn.00028.2010

Ning, Z.-Y., Honing, H., & Ten Cate, C. (2023). Zebra finches (*Taeniopygia guttata*) demonstrate cognitive flexibility in using phonology and sequence of syllables in auditory discrimination. *Animal Cognition*, *26*(4), 1161–1175. https://doi.org/10.1007/s10071-023-01763-4

Norton, P., & Scharff, C. (2016). "Bird song metronomics": Isochronous organization of zebra finch song rhythm. *Frontiers in Neuroscience*, *10*(July), Article July. https://doi.org/10.3389/fnins.2016.00309

Otsuka, S., & Saiki, J. (2016). Gift from statistical learning: Visual statistical learning enhances memory for sequence elements and impairs memory for items that disrupt regularities. *Cognition*, *147*, 113–126. https://doi.org/10.1016/j.cognition.2015.11.004

Pearce, M. T. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences*, *1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

Peters, S., Marler, P., & Nowicki, S. (1992). Song sparrows learn from limited exposure to song models. *The Condor*, *94*(4), 1016–1019. https://doi.org/10.2307/1369302

Plamondon, S. L., Goller, F., & Rose, G. J. (2008). Tutor model syntax influences the syntactical and phonological structure of crystallized songs of white-crowned sparrows. *Animal Behaviour*, *76*(6), Article 6. https://doi.org/10.1016/j.anbehav.2008.07.029

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *Journal of Comparative and Physiological Psychology*, *93*(2), Article 2. https://doi.org/10.1037/h0077553

Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences*, *26*(6), 462–483. https://doi.org/10.1016/j.tics.2022.03.007

Roper, A., & Zann, R. (2006). The onset of song learning and song tutor selection in fledgling zebra finches. *Ethology*, *112*(5), Article 5. https://doi.org/10.1111/j.1439-0310.2005.01169.x

Rouse, A. A., Patel, A. D., & Kao, M. H. (2021). Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proceedings of the National Academy of Sciences*, *118*(29), e2026130118. https://doi.org/10.1073/pnas.2026130118

Sakata, J. T., & Woolley, S. C. (2020). Scaling the levels of birdsong analysis. In J. T. Sakata, S. C. Woolley, R. R. Fay, & A. N. Popper (Eds.), *The Neuroethology of Birdsong* (pp. 1–27). Springer.

Sakata, J. T., & Yazaki-Sugiyama, Y. (2020). Neural circuits underlying vocal learning in songbirds. In J. T. Sakata, S. C. Woolley, R. R. Fay, & A. N. Popper (Eds.), *The Neuroethology of Birdsong* (pp. 29–63). Springer.

Santolin, C. and Saffran, J.R. (2018). Constraints on statistical learning across species. *Trends in Cognitive Sciences*, 22(1), 52-63. https://doi.org/10.1016/j.tics.2017.10.003

Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2022). Long-distance dependencies in birdsong syntax. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1967), 20212473. https://doi.org/10.1098/rspb.2021.2473

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, *128*(3), 302–319. https://doi.org/10.1016/j.cognition.2013.02.013

Sossinka, R., & Böhner, J. (1980). Song types in the zebra finch. *Z. Tierpsychol.*, *53*, 123–132. https://doi.org/10.1111/j.1439-0310.1980.tb01044.x

Speck, B., Seidita, S., Belo, S., Johnson, S., Conley, C., Desjonquères, C., & Rodríguez, R. L. (2020). Combinatorial signal processing in an insect. *The American Naturalist*, *196*(4), 406–413. https://doi.org/10.1086/710527

Spierings, M. J., & Ten Cate, C. (2016). Budgerigars and zebra finches differ in how they generalize in an artificial grammar learning experiment. *Proceedings of the National Academy of Sciences*, *113*(27). https://doi.org/10.1073/pnas.1600483113

Spruijt, B. M., Van Hooff, J. A., & Gispen, W. H. (1992). Ethology and neurobiology of grooming behavior. *Physiological Reviews*, *72*(3), 825–852. https://doi.org/10.1152/physrev.1992.72.3.825

Tanaka, M., Sun, F., Li, Y., & Mooney, R. (2018). A mesocortical dopamine circuit enables the cultural transmission of vocal behaviour. *Nature*, *563*(7729), 117–120. https://doi.org/10.1038/s41586-018-0636-7

Taylor, C. E., Brumley, J. T., Hedley, R. W., & Cody, M. L. (2017). Sensitivity of California thrashers (*Toxostoma redivivum*) to song syntax. *Bioacoustics*, *26*(3), 259–270. https://doi.org/10.1080/09524622.2016.1274917

Tchernichovski, O., Lints, T., Mitra, P. P., & Nottebohm, F. (1999). Vocal imitation in zebra finches is inversely related to model abundance. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(22), Article 22. https://doi.org/10.1073/pnas.96.22.12901

Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science*, *291*(5513), Article 5513. https://doi.org/10.1126/science.1058522

ter Haar, S. M., Kaemper, W., Stam, K., Levelt, C. C., & ten Cate, C. (2014). The interplay of within-species perceptual predispositions and experience during song ontogeny in zebra finches (*Taeniopygia guttata*). *Proceedings of the Royal Society B: Biological Sciences*, *281*(1796), Article 1796. https://doi.org/10.1098/rspb.2014.1860

Van Heijningen, C. A. A., Chen, J., Van Laatum, I., Van Der Hulst, B., & Ten Cate, C. (2013). Rule learning by zebra finches in an artificial grammar learning task: Which rule? *Animal Cognition*, *16*(2), 165–175. https://doi.org/10.1007/s10071-012-0559-x

Verzijden, M. N., Etman, E., Van Heijningen, C., Van Der Linden, M., & Ten Cate, C. (2007). Song discrimination learning in zebra finches induces highly divergent responses to novel songs. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1607), 295–301. https://doi.org/10.1098/rspb.2006.3728

Warren, T. L., Charlesworth, J. D., Tumer, E. C., & Brainard, M. S. (2012). Variable sequencing is actively maintained in a well learned motor skill. *J Neurosci*, *32*(44), Article 44. https://doi.org/10.1523/JNEUROSCI.1254-12.2012

Wilczynski, W., Rand, S. A., & Ryan, M. J. (1999). Female preferences for temporal order of call components in the túngara frog: A Bayesian analysis. *Animal Behaviour*, *58*(4), 841–851. https://doi.org/10.1006/anbe.1999.1208

Zann, R. (1993). Structure, sequence and evolution of song elements in wild Australian zebra finches. *The Auk*, *110*(4), 702–715. https://doi.org/10.2307/4088626

Zann, R. A. (1996). *The zebra finch: A synthesis of field and laboratory studies*. Oxford University Press.

Supplementary Table 1

TABLE 1 Number of birds tutored with specific syllable sequences (motifs) and gap durations between each syllable in the motif.

Sequence	10	20	30	40	50	Different gaps between syllables	Variable gaps across renditions (Vgap)
Variable (Vseq)			n = 54				
abced							n = 1
abdce							n = 1
abdec			n = 1				n = 3
aecbd			n = 2				n = 1
bacde		n = 1	n = 2		n = 1		n = 3
beadc	n = 1		n = 4		n = 1		n = 4
becad		n = 1	n = 1		n = 1	n = 2	n = 4
cdaeb	n = 1		n = 2	n = 1	n = 2		n = 4
cedba							n = 1
dacbe	n = 2	n = 1	n = 1	n = 1	n = 1	n = 1	n = 5
decab			n = 2				
eabcd			n = 1				n = 2
edbca			n = 1		n = 2		n = 3

Note: The gap duration of 30 ms is used in the Vseq group because this was the average gap duration between syllables for the Vgap-tutored, and the most common gap duration for the stimuli of S-tutored birds.