Songbirds can learn flexible contextual control over syllable sequencing

Lena Veit[1,2]*, Lucas Y Tian[1,3], Christian J Monroy Hernandez[1], Michael S Brainard[1]*

[1]Center for Integrative Neuroscience, University of California, San Francisco, USA and
Howard Hughes Medical Institute

[2]Current address: Institute for Neurobiology, University of Tübingen, Germany

[3]Current address: The Rockefeller University, New York, NY, USA

*Correspondence: lena.veit@uni-tuebingen.de, msb@phy.ucsf.edu

Abstract

The flexible control of sequential behavior is a fundamental aspect of speech, enabling endless reordering of a limited set of learned vocal elements (syllables or words). Songbirds are phylogenetically distant from humans but share both the capacity for vocal learning and neural circuitry for vocal control that includes direct pallial-brainstem projections. Based on these similarities, we hypothesized that songbirds might likewise be able to learn flexible, moment-by-moment control over vocalizations. Here, we demonstrate that Bengalese finches (*Lonchura striata domestica)*, which sing variable syllable sequences, can learn to rapidly modify the probability of specific sequences (e.g. ‘ab-c’ versus ‘ab-d’) in response to arbitrary visual cues. Moreover, once learned, this modulation of sequencing occurs immediately following changes in contextual cues and persists without external reinforcement. Our findings reveal a capacity in songbirds for learned contextual control over syllable sequencing that parallels human cognitive control over syllable sequencing in speech.

Introduction

A crucial aspect of the evolution of human speech is the development of flexible control over learned vocalizations (Ackermann, Hage, & Ziegler, 2014; Belyk & Brown, 2017). Humans have unparalleled control over their vocal output, with a capacity to reorder a limited number of learned elements to produce an endless combination of vocal sequences that are appropriate for current contextual demands (Hauser, Chomsky, & Fitch, 2002). This cognitive control over vocal production is thought to rely on the direct innervation of brainstem and midbrain vocal networks by executive control structures in the frontal cortex, which have become more elaborate over the course of primate evolution (Hage & Nieder, 2016; Simonyan & Horwitz, 2011). However, because of the comparatively limited flexibility of vocal production in nonhuman primates (Nieder & Mooney, 2020), the evolutionary and neural circuit mechanisms that have enabled the development of this flexibility remain poorly understood.

Songbirds are phylogenetically distant from humans, but they have proven a powerful model for investigating neural mechanisms underlying learned vocal behavior. Song learning exhibits many parallels to human speech learning (Doupe & Kuhl, 1999); in particular, juveniles need to hear an adult tutor during a sensitive period, followed by a period of highly variable sensory-motor exploration and practice, during which auditory feedback is used to arrive at a precise imitation of the tutor song (Brainard & Doupe, 2002). This capacity for vocal learning is subserved by a well-understood network of telencephalic song control nuclei. Moreover, as in humans, this vocal control network includes strong projections directly from cortical (pallial) to brainstem vocal control centers (Doupe & Kuhl, 1999; Simonyan & Horwitz, 2011). These shared behavioral features and neural specializations raise the question of whether songbirds might also share the capacity to learn flexible control over syllable sequencing.

Contextual variation of song in natural settings, such as territorial counter-singing or female-directed courtship song, indicate that songbirds can rapidly alter aspects of their song, including syllable sequencing and selection of song types (Chen, Matheson, & Sakata, 2016; Heinig et al., 2014; King & McGregor, 2016; Sakata, Hampton, & Brainard, 2008; Searcy & Beecher, 2009; Trillo & Vehrencamp, 2005). However, such modulation of song structure is often described as affectively controlled (Berwick, Okanoya, Beckers, & Bolhuis, 2011; Nieder & Mooney, 2020). For example, the presence of potential mates or rivals elicits a global and unlearned modulation of song intensity (James, Dai, & Sakata, 2018a) related to the singer's level of arousal or aggression (Alcami, Ma, & Gahr, 2021; Heinig et al., 2014; Jaffe & Brainard, 2020). Hence, while prior observations suggest that a variety of ethologically relevant factors can be integrated to influence song production in natural settings, it remains unclear whether song can be modified more flexibly by learned or cognitive factors.

Here, we tested whether Bengalese finches can learn to alter specifically targeted vocal sequences within their songs in response to arbitrarily chosen visual cues, independently of social or other natural contexts. Each Bengalese finch song repertoire includes ~5-12 acoustically distinct elements ('syllables') that are strung together into sequences in variable but non-random order. For a given bird, the relative probabilities of specific transitions between syllables normally remain constant over time (Okanoya, 2004; Warren, Charlesworth, Tumer, & Brainard, 2012), but previous work has shown that birds can gradually adjust the probabilities of alternative sequences in response to training that reinforces the production of some sequences over others. In this case, changes to syllable sequencing develop over a period of hours to days (Warren et al., 2012). In contrast, we investigate here whether birds can learn to change syllable sequencing on a moment-by-moment basis in response to arbitrary visual cues that signal which sequences are adaptive at any given time. Our findings reveal that songbirds can learn to immediately, flexibly, and adaptively adjust the sequencing of selected vocal elements in response to learned contextual cues.

Results

Bengalese finches can learn context-dependent syllable sequencing

For each bird in the study, we first identified variably produced syllable sequences that could be gradually modified using a previously described aversive reinforcement protocol ("single context training"; Tumer & Brainard, 2007; Warren et al., 2012). For example, a bird that normally transitioned from the fixed syllable sequence 'ab' to either 'c' or 'd' (**Figures 1A,B**, sequence probability of ~36% for 'ab-c' and ~64% for 'ab-d'), was exposed to an aversive burst of white noise (WN) feedback immediately after the "target sequence" 'ab-d' was sung. In response, the bird learned over a period of days to gradually decrease the relative probability of that sequence in favor of the alternative sequence 'ab-c' (**Fig. 1C**). This change in sequence probabilities was adaptive in that it enabled the bird to escape from WN feedback. Likewise, when the sequence, 'ab-c' was targeted, the probability of 'ab-d' increased gradually over several days of training (**Fig. 1D**). These examples are consistent with prior work that showed such sequence modifications develop over a period of several days, with the slow time course suggesting a gradual updating of synaptic connections within syllable control networks in response to performance-related feedback (Warren et al., 2012). In contrast, the ability to immediately and flexibly reorder vocal elements in speech must reflect mechanisms that enable contextual factors to exert moment-by-moment control over selection and sequencing of alternative vocal motor programs. Having identified sequences for each bird for which the probability of production could be gradually modified in this manner, we then tested whether birds could be trained to rapidly switch between those same sequences in a context-dependent manner.

To determine whether Bengalese finches can learn to flexibly select syllable sequences on a moment-by-moment basis, we paired WN targeting of specific sequences with distinct contextual cues. In this context-dependent training protocol, WN was targeted to defined sequences in the bird's song as before, but the specific target sequence varied across alternating blocks, signaled by different colored lights in the home cage (see Methods). **Figure 1E** shows an example experiment, with 'ab-d' targeted in yellow light, and 'ab-c' in green light. At baseline, without WN, switches between yellow and green contexts (at random intervals of 0.5-1.5 hours) did not lead to significant changes in the relative proportion of the target sequences, indicating that there was no inherent influence of the light cues on sequence probabilities (**Fig. 1F**, p(ab-d) in yellow vs. green context was 67+/-1.6% vs. 64+/-1.5%, $p = 0.17$, ranksum test, n=53 context blocks from baseline period). Training was then initiated in which WN was alternately targeted to each sequence, over blocks that were signaled by light cues. After 2 weeks of such context-specific training, significant sequencing differences developed between light contexts that were appropriate to reduce aversive feedback in each context (**Fig. 1G,** p(ab-d) in yellow vs. green context shifted to 36.5+/-4.8% vs. 83.1+/3.5%, $p < 0.01$, ranksum test, n=22 context blocks, block duration between 1-2.5 hours). Likewise, for all birds trained on this protocol (n=8), context-dependent sequencing differences developed in the appropriate direction over a period of weeks (27+/-6% difference in probabilities between contexts after a mean of 33 days training, versus 1%+/-2% average difference in probabilities at baseline; $p < 0.01$, n=8, signed rank test, **Fig. 1H**). Thus, Bengalese finches are able to learn context-specific modifications to syllable sequencing.

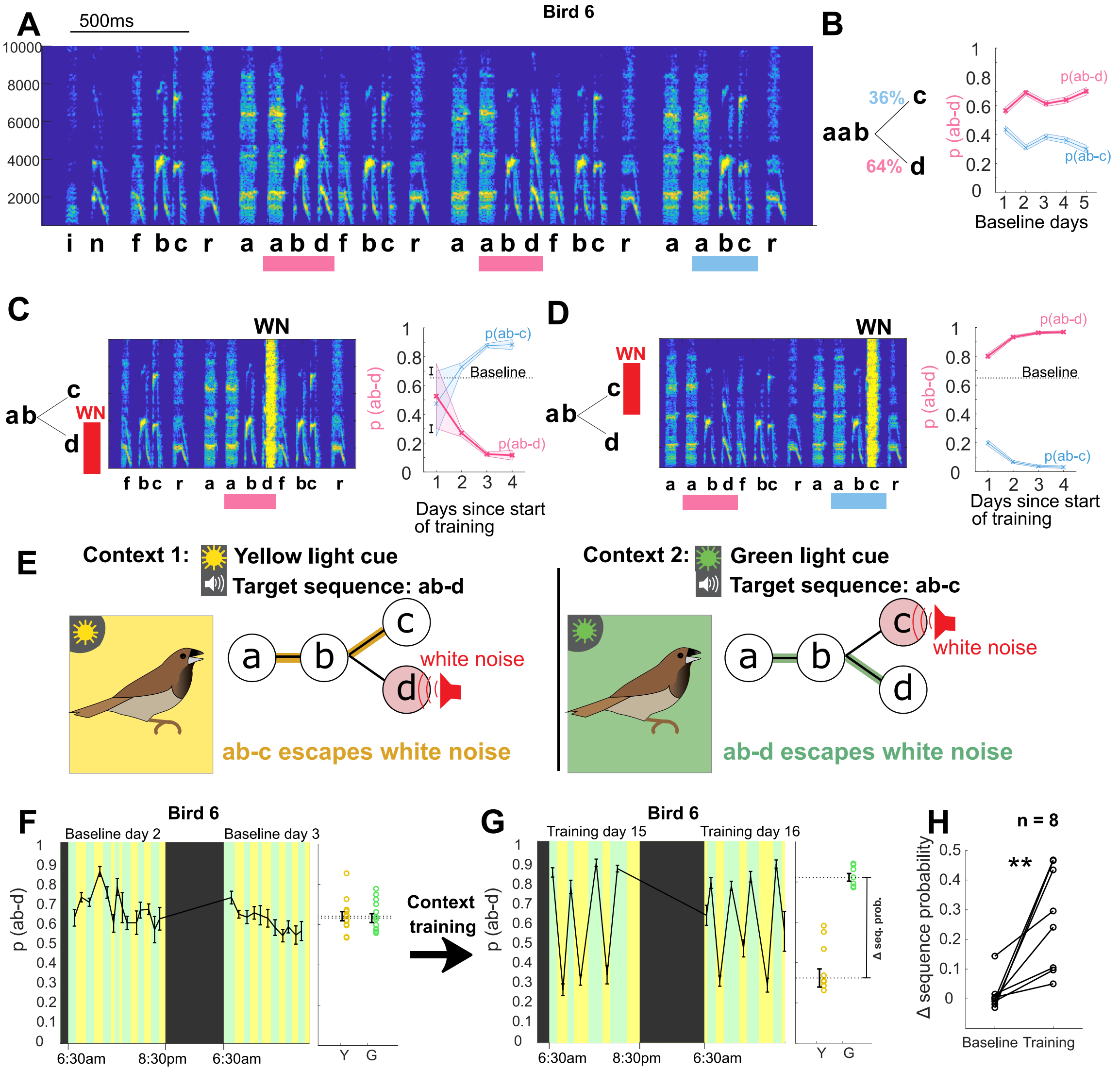


Figure 1. Bengalese finches can learn context-dependent sequencing
A: Example spectrogram highlighting points in the song with variable sequencing. Individual syllables are labeled based on their spectral structure, and target sequences for the different experiments (ab-c and ab-d) are marked with colored bars. Scale for x-axis is 500ms and y-axis shows frequency in Hz. **B**: Transition diagram with probabilities for the sequences ab-c and ab-d. The sequence probability of ab-d (and likewise of the complementary probability ab-c) stayed relatively constant over five days of baseline song recordings. Shaded area shows 95% confidence interval for sequence probability. **C**: Aversive reinforcement training. Schematic showing aversive WN after target sequence ab-d is sung; spectrogram shows actual WN stimulus, covering part of syllable d. WN targeted to the sequence ab-d led to a gradual reduction in the probability of that sequence over several days, and a complementary increase in the probability of ab-c. **D**: In a subsequent experiment, WN targeted to the sequence ab-c led to a gradual increase in the sequence probability of ab-d over several days. **E**: Schematic of the contextual learning protocol, in which the target for WN is signaled by colored lights. **F**: Left: Two example days of baseline without WN but with alternating blocks of green and yellow context. Colors indicate light context (black indicates periods of lights off during the night), error bars indicate SEM across song bouts in each block. Right: Average sequence probability in yellow and green blocks during baseline. Open circles show data from

individual blocks of yellow and green context, error bars indicate SEM across blocks. **G**: Left: Two example days after training (WN on) with alternating blocks of green and yellow context, as in F. Right: Average sequence probability in yellow and green blocks after training, as in F. **H**: Contextual difference in sequence probability for eight trained birds before and after training (** $p < 0.01$ signed rank test).

Syllable sequencing shifts immediately following switches in context

Contextual differences between different blocks could arise through an immediate shift in sequence probabilities upon entry into a new context and/or by rapid learning within each block. We examined whether trained birds exhibited any immediate shifts in their syllable sequencing when entering a new light context by computing the average probability of target sequences across songs aligned with the switch between contexts (**Fig. 2A,B**, example experiment). This "switch-triggered average" revealed that across all birds, switches to the yellow context were accompanied by an immediate decrease in the probability of the yellow target sequence, whereas switches out of the yellow context (and into the green context) led to an immediate increase in the yellow target sequence (**Fig. 2C,D**, $p < 0.05$, signed rank test comparing first and last song, n=8). To quantify the size of these immediate shifts, we calculated the difference in sequence probability from the last five songs in the previous context to the first five songs in the current context; this difference averaged 0.24+/-0.06 for switches to green light and -0.22+/-0.06 for switches to yellow light (**Fig. 2 E,F**). These results indicate that birds could learn to immediately recall an acquired memory of context-appropriate sequencing upon entry into each context, even before having the chance to learn from reinforcing feedback within that context.

We next asked whether training additionally led to an increased rate of learning within each context, which also might contribute to increased contextual differences over time. Indeed, such faster re-learning for consecutive encounters of the same training context, or 'savings', is sometimes observed in contextual motor adaptation experiments (Lee & Schweighofer, 2009). To compare the magnitude of the immediate shift and the magnitude of within-block learning over the course of training, we plotted the switch-aligned sequence probabilities at different points in the training process. **Figure 2G** shows for the example bird that the magnitude of the shift (computed between the first and last five songs across context switches) gradually increased over 11 days of training. **Figure 2H** shows the switch-aligned sequence probability trajectories (as in **Fig. 2A,B**) for this bird early in training (red) and late in training (blue), binned into groups of 7 context switches. Qualitatively, there was both an abrupt change in sequence probability at the onset of each block (immediate shift at time point 0) and a gradual adjustment of sequence probability within each block (within-block learning over the first 80 songs following light switch). Over the course of training, the immediate shift at the onset of each block got larger, while the gradual change within blocks stayed approximately the same (learning trajectories remained parallel over training, **Fig. 2H**). Linear fits to the sequence probabilities for each learning trajectory (i.e. the right side of **Fig. 2H**) reveal that, indeed, the change in sequence probability at the onset of blocks (i.e. intercepts) increased over the training process (**Fig. 2K**), while the rate of change within blocks (i.e. slopes) stayed constant (**Fig. 2I**). To quantify this across birds we measured the change over the course of learning in both the magnitude of immediate shifts (estimated as the intercepts from linear fits) and the rate of within-block learning (estimated as the slopes from linear fits). As for the example bird, we found that the rate of learning within each block stayed constant over time for all five birds (**Fig. 2L**). In contrast, the magnitude of immediate shifts increased over time for all birds (**Fig. 2L**). These analyses indicate that adjustments to sequence probability reflect two dissociable processes, an immediate cue-dependent shift in sequence probability at the beginning of blocks, that increases with contextual training, and a gradual adaptation of sequence probability within blocks, that does not increase with contextual training.

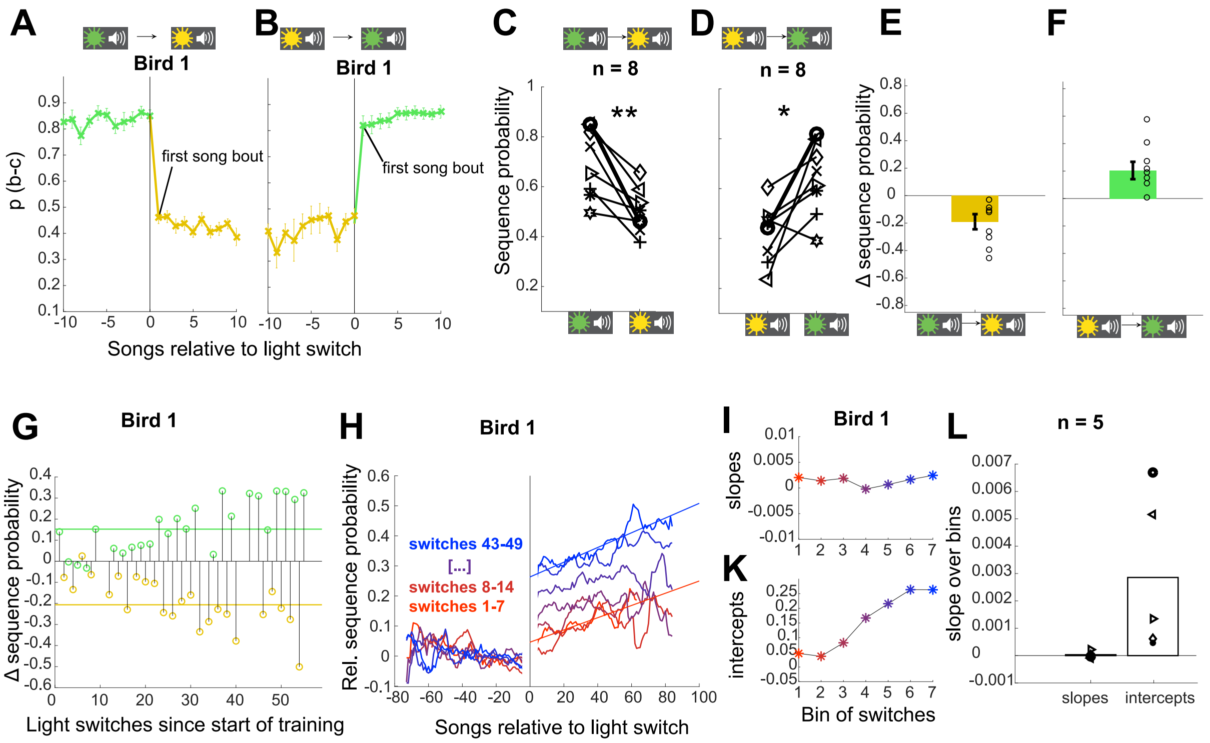


Figure 2. Sequence probabilities shift immediately following a switch in context. **A-B**: Average sequence probability per song for example Bird 1 aligned to switches from green to yellow context (A) and from yellow to green context (B). Error bars indicate SEM across song bouts (n = 35 switches (A), n = 33 switches (B)). **C**: Changes in sequence probability from the last song in green context to the first song in yellow context for all 8 birds. Example bird in A-B highlighted in bold. ** $p < 0.01$ signed-rank test. **D**: Changes in sequence probability from the last song in yellow context to the first song in green context. * $p < 0.05$ signed rank test. **E**: Shift magnitudes for all birds, defined as the changes in sequence probability from the last five songs in the green context to the first five songs in the yellow context. Open circles show individual birds, error bars indicate SEM across birds. **F:** Same as E for switches from yellow to green. **G**: Shift magnitudes over training time for the example bird (11 days and 49 context switches; seven of the original 56 context switches are excluded from calculations of shift magnitudes because at least one of the involved blocks contained only one or two song bouts.). **H**: Trajectory of switch-aligned sequence probabilities for the example bird early in training (red) and late in training (blue). Probabilities are normalized by the sequence probability in preceding block, and plotted so that the adaptive direction is positive for both switch directions (i.e. inverting the probabilities for switches to yellow.) **I**: Slopes of fits to the sequence probability trajectories over song bouts within block. Units in change of relative sequence probability per song bout. **K:** Intercepts of fits to sequence probability trajectories over song bouts within block. Units in relative sequence probability. **L**: Changes in slopes and changes in intercepts for five birds over the training process, determined as the slopes of linear fits to curves as in I,K for each bird.

Visual cues in the absence of reinforcement are sufficient to evoke sequencing changes

The ability of Bengalese finches to implement an immediate shift in sequencing on the first rendition in a block – and thus before they have a chance to learn from reinforcing feedback –

argues that they can maintain context-specific motor memories and use contextual visual cues to anticipate correct sequencing in each context. To explicitly test whether birds can flexibly switch between sequencing appropriate for distinct contexts using only visual cues, we included short probe blocks which presented the same light cues without WN stimulation. Probe blocks were interspersed in the sequence of training blocks so that each switch between types of blocks was possible and, on average, every third switch was into a probe block (see Methods, 10 of 34 blocks in sequence). Light switches into probe blocks were associated with similar magnitude shifts in sequence probability as switches into WN blocks of the corresponding color (-0.22 +/- 0.06 to both yellow WN and yellow probe blocks from green WN blocks, $p = 0.94$, signed rank test; 0.24 +/- 0.06 to green WN and 0.23 +/- 0.07 to green probe blocks from yellow WN blocks, $p = 0.64$, signed rank test). As the most direct test of evoked sequencing changes with light cues alone, we compared songs immediately before and after switches between probe blocks without intervening WN training blocks (probe-probe switches). **Figure 3A,B** shows song bouts for one example bird (Bird 2) which were sung consecutively across a switch from yellow probe to green probe blocks. In the first song following the probe-probe switch, the yellow target sequence (‘f-ab’) was more prevalent, and the green target sequence (‘n-ab’) was less prevalent, and such an immediate effect was also apparent in the average sequence probabilities for this bird aligned to probe-probe switches (**Fig. 3C,D**). Similar immediate and appropriately directed shifts in sequencing at switches between probe blocks were observed for all 8 birds (**Fig. 3E,F,** $p < 0.05$ signed rank test, n=8**)**, with average shifts in sequence probabilities of -0.21 +/-0.09 and 0.17 +/-0.08 (**Fig. 3G,H**). The presence of such changes in the first songs sung after probe-probe switches indicates that visual cues alone are sufficient to cause anticipatory, learned shifts between syllable sequences.

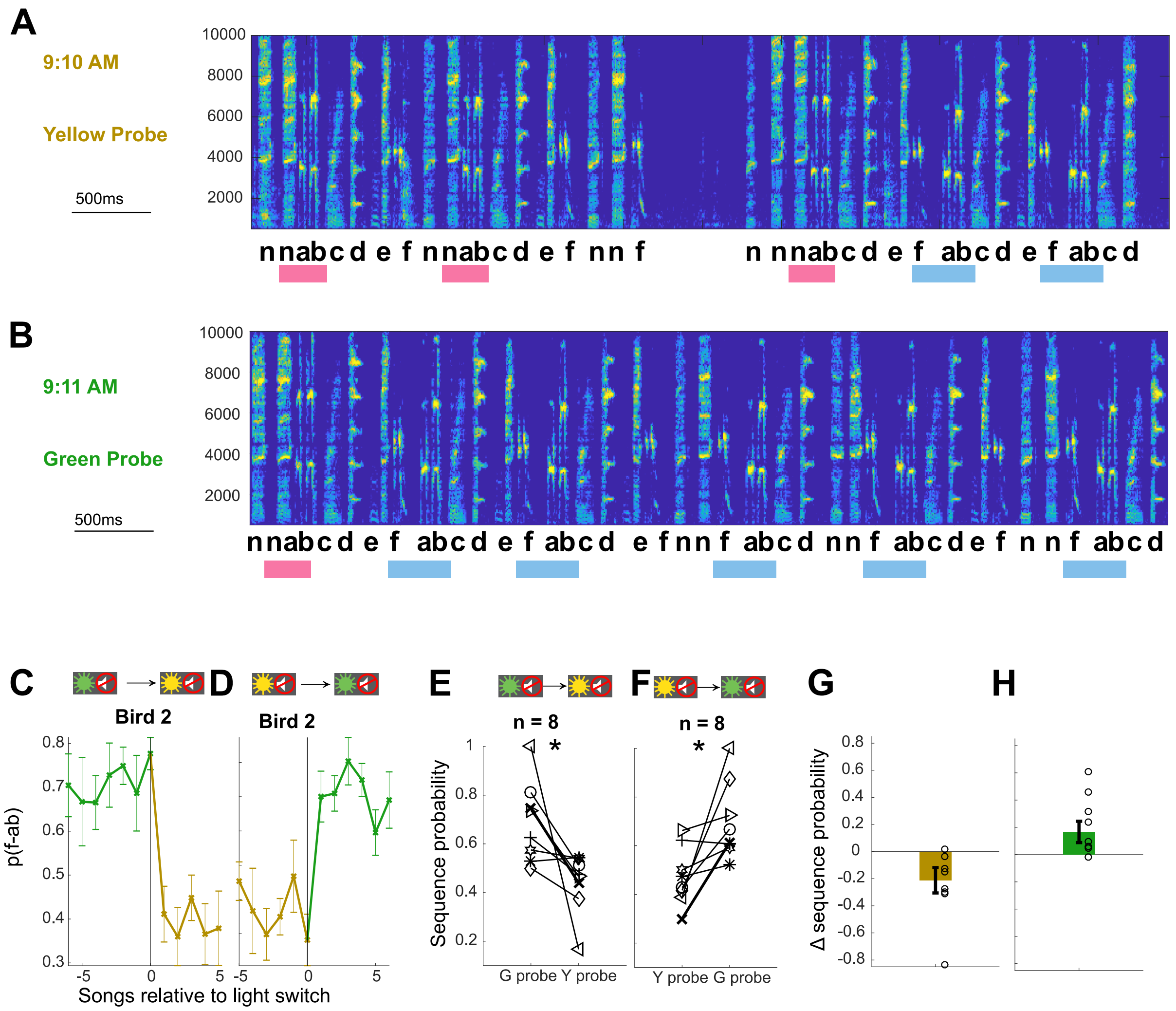
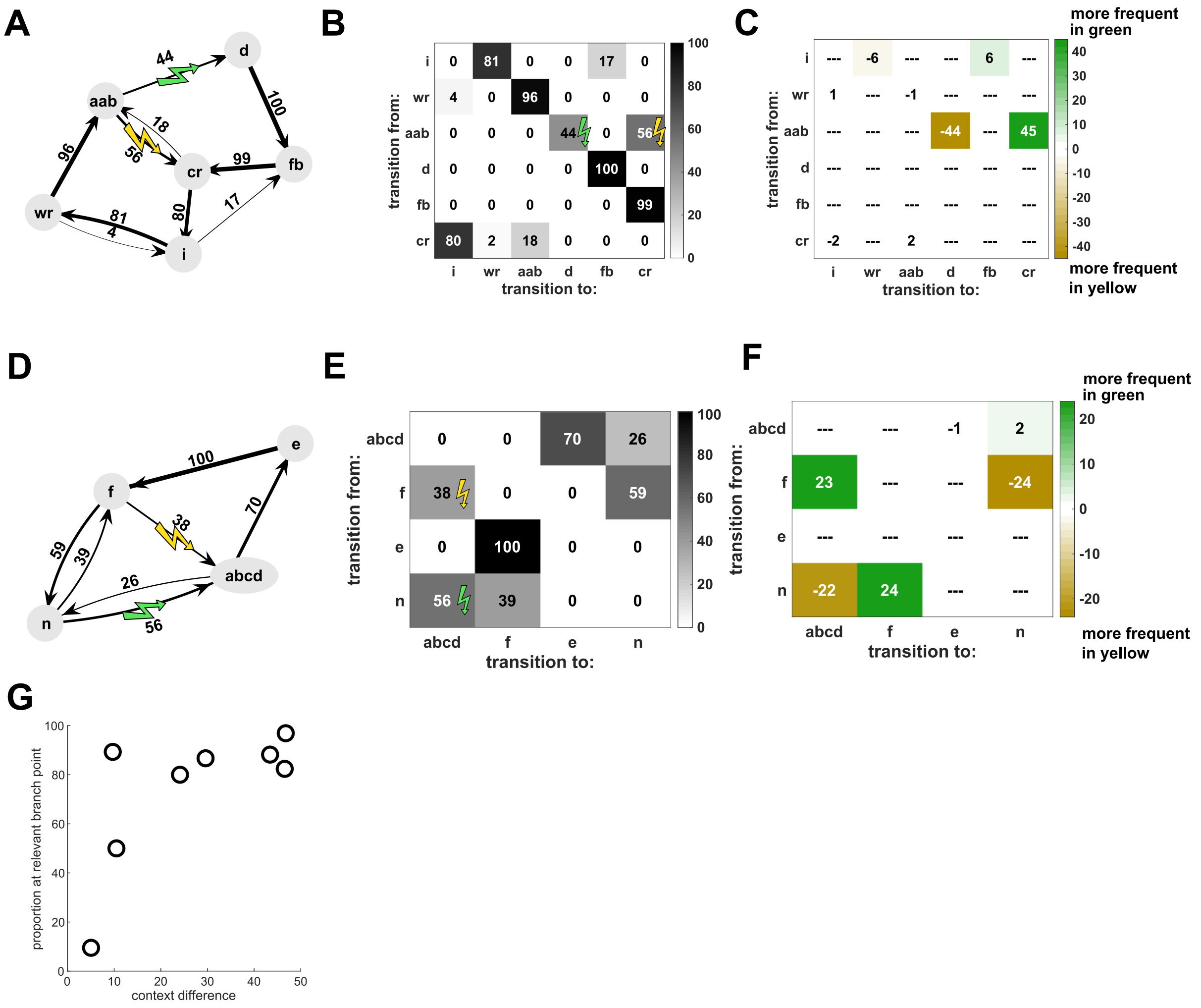


Figure 3. Contextual cues alone are sufficient to enable immediate shifts in syllable sequencing.
A,B: Examples of songs sung by Bird 2 immediately before (A) and after (B) a switch from a yellow probe block to a green probe block (full song bouts in Sup. Fig. 3). Scale for x-axis is 500ms, y-axis shows frequency in Hz. **C,D:** Average sequence probability per song for Bird 2 aligned to switches from green probe to yellow probe blocks (C) and from yellow probe to green probe blocks (D). Error bars indicate SEM across song bouts (n = 14 switches (C), 11 switches (D)).
E,F: Average sequence probabilities for all eight birds at the switch from the last song in green probe context and the first song in yellow probe context, and vice versa. Example Bird 2 is shown in bold. * $p < 0.05$ signed rank test. **G,H**: Shift magnitudes for probe-probe switches for all birds. Open circles show individual birds, error bars indicate SEM across birds.

Contextual changes are specific to target sequences

A decrease in the probability of a target sequence in response to contextual cues must reflect changes in the probabilities of transitions leading up to the target sequence. However, such changes could be restricted to the transitions that immediately precede the target sequence, or alternatively could affect other transitions throughout the song. For example, for the experiment illustrated in Figure 1, the prevalence of the target sequence 'ab-d' was appropriately decreased in the yellow context, in which it was targeted. The complete transition diagram and corresponding transition matrix for this bird (**Fig. 4A,B**) reveal that there were four distinct branch points at which syllables were variably sequenced (after 'cr', 'wr', 'i', and 'aab'). Therefore, the decrease in the target sequence 'ab-d' could have resulted exclusively from an increase in the probability of the alternative transition 'ab-c' at the branch point following 'aab'. However, a reduction in the prevalence of the target sequence could also have been achieved by changes in the probability of transitions earlier in song such that the sequence 'aab' was sung less frequently. To investigate how specific contextual changes in probability were to transitions immediately preceding target sequences, we calculated the difference between transition matrices in the yellow and green probe contexts (**Fig. 4C**). This difference matrix indicates that changes to transition probabilities were highly specific to the branch point immediately preceding the target sequences (specificity was defined as the proportion of total changes which occurred at the branch points immediately preceding target sequences; specificity for branch point 'aab' was 83.2%). Such specificity to branch points that immediately precede target sequences was typical across experiments, including cases in which different branch points preceded each target sequence (**Fig. 4D-F**, specificity 96.9%). Across all 8 experiments, the median specificity of changes to the most proximal branch points was 84.95%, and only one bird, which was also the worst learner in the contextual training paradigm, had a specificity of less than 50% (**Fig. 4G**). Hence, contextual changes were specific to target sequences and did not reflect the kind of global sequencing changes that characterize innate social modulation of song structure (Sakata et al., 2008; Sossinka & Böhner, 1980).



B (transition from: rows; transition to: columns)

	i	wr	aab	d	fb	cr
i	0	81	0	0	17	0
wr	4	0	96	0	0	0
aab	0	0	0	44	0	56
d	0	0	0	0	100	0
fb	0	0	0	0	0	99
cr	80	2	18	0	0	0

C (transition from: rows; transition to: columns)

	i	wr	aab	d	fb	cr
i	---	-6	---	---	6	---
wr	1	---	-1	---	---	---
aab	---	---	---	-44	---	45
d	---	---	---	---	---	---
fb	---	---	---	---	---	---
cr	-2	---	2	---	---	---

E (transition from: rows; transition to: columns)

	abcd	f	e	n
abcd	0	0	70	26
f	38	0	0	59
e	0	100	0	0
n	56	39	0	0

F (transition from: rows; transition to: columns)

	abcd	f	e	n
abcd	---	---	-1	2
f	23	---	---	-24
e	---	---	---	---
n	-22	24	---	---

Figure 4. Contextual changes are local to the target sequences

A: Transition diagram for the song of Bird 6 (spectrogram in Fig. 1) in yellow probe context. Sequences of syllables with fixed transition patterns (e.g. 'aab') as well as repeat phrases and introductory notes have been summarized as single states to simplify the diagram. **B**: Transition matrix for the same bird, showing same data as in A. **C:** Differences between the two contexts are illustrated by subtracting the transition matrix in the yellow context from the one in the green context, so that sequence transitions which are more frequent in green context are positive (colored green) and sequence transitions which are more frequent in yellow are negative (colored yellow). For this bird, the majority of contextual differences occurred at the branch point ('aab') which most closely preceded the target sequences ('ab-c' and 'ab-d'), while very little contextual difference occurred at the other three branch points ('i', 'wr', 'cr'). **D-F:** Same for Bird 2 for which two different branch points ('f' and 'n') preceded the target sequences ('f-abcd' and 'n-abcd') (spectrogram in Fig. 3).
G: Proportion of changes at the branch point(s) most closely preceding the target sequences, relative to the total magnitude of context differences for each bird (see Methods). Most birds exhibited high specificity of contextual changes to the relevant branch points.

Distinct sequence probabilities are specifically associated with different visual cues

Our experiments establish that birds can shift between two distinct sequencing states in response to contextual cues. In order to test whether birds were capable of learning to shift to these two states from a third neutral context, we trained a subset of three birds with three different color-cued contexts. For these birds, after completion of training with WN targeted to distinct sequences in yellow and green contexts (as described above), we introduced interleaved blocks cued by white light in which there was no reinforcement. After this additional training, switches from the unreinforced context elicited changes in opposite directions for the green and yellow contexts (example bird **Fig. 5A**). All birds (n = 3) showed adaptive sequencing changes for the first song bout in probe blocks **(Fig. 5B,C)** as well as immediate shifts in the adaptive directions for all color contexts (**Fig. 5D,** 0.11+/-0.04 and 0.19+/-0.05 for switches to green WN and green probe blocks, respectively; -0.15+/-0.06 and -0.09+/-0.02 for switches to yellow WN and yellow probe blocks, respectively). While additional data would be required to establish the number of distinct associations between contexts and sequencing states that can be learned, these findings suggest that birds can maintain at least two distinct sequencing states separate from a 'neutral' state, and use specific associations between cue colors and sequencing states to rapidly shift sequencing in distinct directions for each context.

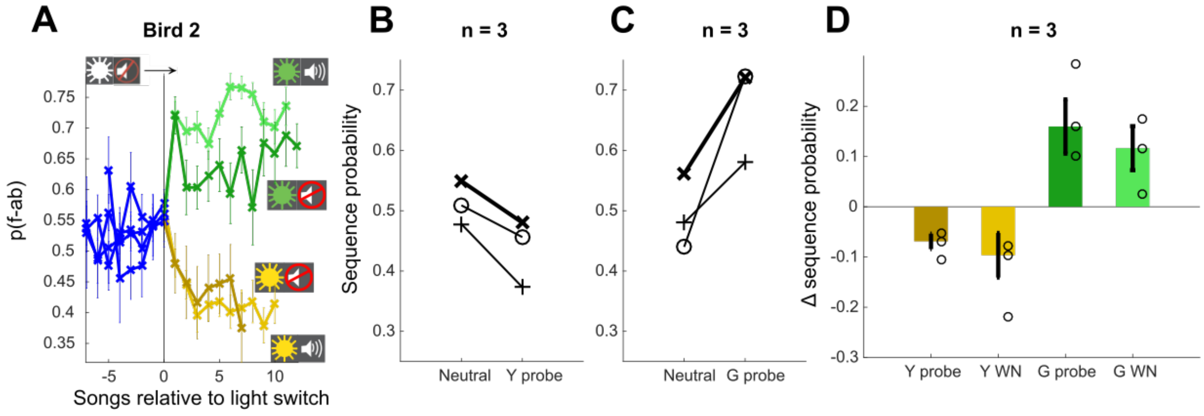


Figure 5. Contextual cues allow shifts in both directions

A: Sequence probability for Bird 2 at the switch from neutral context to yellow and green WN contexts, as well as yellow and green probe contexts (no WN). Error bars indicate SEM across song bouts (n = 68 switches (green WN), 78 switches (yellow WN), 27 switches (green probe), 24 switches (yellow probe)). **B,C**: Sequence probabilities for three birds for the last song in neutral context and the first song in the following probe context. Example bird in A highlighted in bold. **D**: Shift magnitude for three birds at the switch from neutral context to all other contexts. Open circles show individual birds, error bars indicate SEM across birds.

Discussion

Speech, thought, and many other behaviors are composed of ordered sequences of simpler elements. The flexible control of sequencing is thus a fundamental aspect of cognition and motor function (Aldridge & Berridge, 2002; X. Jin & Costa, 2015; Tanji, 2001). While the flexibility of human speech is unrivaled, our contextual training paradigm revealed a simpler, parallel capacity in birds to produce distinct vocal sequences in response to arbitrary contextual cues. The colors of the cues had no prior relevance to the birds, so that their meaning had to be learned as a new association between cues and the specific vocal sequences that were contextually appropriate (i.e. that escaped white noise, given the current cues). Learned modulation of sequencing was immediately expressed in response to changes in cues, persisted following termination of training, and was largely restricted to the targeted sequences, without gross modifications of global song structure. Hence, for song, like speech, the ordering of vocal elements can be rapidly and specifically reconfigured to achieve learned, contextually appropriate goals. This shared capacity for moment-by-moment control of vocal sequencing in humans and songbirds suggests that the avian song system could be an excellent model for investigating how neural circuits enable flexible and adaptive reconfiguration of motor output in response to different cognitive demands.

Flexible control of vocalizations

Our demonstration of contextual control over the ordering of vocal elements in the songbird builds on previous work showing that a variety of animals can learn to emit or withhold innate vocalizations in response to environmental or experimentally imposed cues. For example, non-human primates and other animals can produce alarm calls that are innate in their acoustic structure, but that are deployed in a contextually appropriate fashion (Nieder & Mooney, 2020; T. N. Suzuki & Zuberbühler, 2019; Wheeler & Fischer, 2012). Similarly, animals, including birds, can be trained to control their vocalizations in an experimental setting, by reinforcing the production of innate vocalizations in response to arbitrary cues to obtain food or water rewards (Brecht, Hage, Gavrilov, & Nieder, 2019; Hage & Nieder, 2013; Nieder & Mooney, 2020; Reichmuth & Casey, 2014). In relation to these prior findings, our results demonstrate a capacity to flexibly reorganize the sequencing of learned vocal elements, rather than select from a fixed set of innate vocalizations, in response to arbitrary cues. This ability to contextually control the ordering, or syntax, of specifically targeted syllable transitions within the overall structure of learned song parallels the human capacity to differentially sequence a fixed set of syllables in speech.

The ability to alter syllable sequencing in a flexible fashion also contrasts with prior studies that have demonstrated modulation of vocalizations in more naturalistic settings. For example, songs produced in the context of courtship and territorial or aggressive encounters (“directed song”) differ in acoustic structure from songs produced in isolation (“undirected song”) (Sakata et al., 2008; Searcy & Beecher, 2009). This modulation of song structure by social context is characterized by global changes to the intensity of song production, with directed songs exhibiting faster tempo, and greater stereotypy of both syllable structure and syllable sequencing, than undirected songs (Sakata et al., 2008; Searcy & Beecher, 2009; Sossinka & Böhner, 1980). This and other ethologically relevant modulation of song intensity may serve to communicate the singer’s affective state, such as level of arousal or aggression (Alcami et al., 2021; Hedley, Denton, & Weiss, 2017; Heinig et al., 2014), and may largely reflect innate mechanisms (James, Dai, & Sakata, 2018b; Kojima & Doupe, 2011) mediated by hypothalamic and neuromodulatory inputs to premotor regions (Berwick et al., 2011; Gadagkar, Puzerey, & Goldberg, 2019; James, Dai, & Sakata, 2018b; Nieder & Mooney, 2020). In contrast, here we show that birds can learn to locally modulate specific features of their songs (i.e. individually targeted syllable transitions) in response to arbitrarily assigned contextual cues that have no prior ethological relevance.

Evolution of control over vocal sequencing

The capacity for moment-by-moment adjustment of vocalizations in response to arbitrary learned cues may depend on similar capacities that evolved to enable appropriate modulation of vocalizations in ethologically relevant natural contexts. For example, some species of songbirds preferentially sing different song types depending on factors such as time of day, location of the singer, or the presence of an audience (Alcami et al., 2021; Hedley et al., 2017; King & McGregor, 2016; Searcy & Beecher, 2009; Trillo & Vehrencamp, 2005). Even birds with only a single song type, such as Bengalese finches, vary parameters of their song depending on social context, including the specific identity of the listener (Chen et al., 2016; Heinig et al., 2014; Sakata et al., 2008). The ability to contextually control vocalizations is also relevant for the customization of vocal signatures for purposes of individual and group recognition (Vignal, Mathevon, & Mottin, 2004), and to avoid overlap and enhance communication during vocal turn-taking and in response to environmental noises (Benichov & Vallentin, 2020; Brumm & Zollinger, 2013). Such capacities for vocal control likely reflect evolutionary advantages of incorporating sensory and contextual information about conspecifics and the environment in generating increasingly sophisticated vocal signaling. Our results indicate a latent capacity to integrate arbitrary sensory signals into the adaptive deployment of vocalizations in songbirds, and suggest that some of the contextual control observed in natural settings may likewise rely on learned associations and other cognitive factors. Perhaps evolutionary pressures to develop nuanced social communication led to the elaboration of cortical (pallial) control over brainstem vocal circuitry (Hage & Nieder, 2016), and thereby established a conduit that facilitated the integration of progressively more abstract cues and internal states in that control.

Neural implementation of context-dependent vocal motor sequencing

The ability of birds to switch between distinct motor programs using visual cues is reminiscent of contextual speech and motor control studies in humans. For example, human subjects in both laboratory studies and natural settings can learn multiple 'states' of vocal motor adaptation, and rapidly switch between them using contextual information (Houde & Jordan, 2002; Keough & Jones, 2011; Rochet-Capellan & Ostry, 2011). Similarly, subjects can learn two separate states of motor adaptation for other motor skills such as reaching, and switch between them using cues or other cognitive strategies (Cunningham & Welch, 1994). Models of such context-dependent motor adaptation frequently assume at least two parallel processes (Abrahamse, Ruitenberg, De Kleine, & Verwey, 2013; Ashe, Lungu, Basford, & Lu, 2006; Green & Abutalebi, 2013; Hikosaka et al., 1999; Lee & Schweighofer, 2009; McDougle, Ivry, & Taylor, 2016; Rochet-Capellan & Ostry, 2011; Wolpert, Diedrichsen, & Flanagan, 2011), one that is more flexible, and sensitive to contextual information (McDougle et al., 2016), and a second that cannot readily be associated with contextual cues and is only gradually updated during motor adaptation (Howard, Wolpert, & Franklin, 2013). Specifically, in support of such a two-process model, Imamizu and colleagues (Imamizu & Kawato, 2009; Imamizu et al., 2007) found that contextual information can drive rapid shifts in adaptation at the beginning of new blocks, without affecting the rate of adaptation within blocks. The similar separation in our study between rapid context-dependent shifts in sequence probability at the onset of blocks, and gradual adaptation within blocks that does not improve with training (**Fig. 2 G-L**), suggests that such contextual sequence learning in the Bengalese finch may also be enabled by two distinct processes.

Humans studies of two-process models suggest that slow adaptation occurs primarily within primary motor structures, while fast context-dependent state switches, including for cued switching between languages in bilinguals, engage more frontal areas involved in executive control (Bialystok, 2017; Blanco-Elorrieta & Pylkkänen, 2016; De Baene, Duyck, Brass, & Carreiras, 2015; Imamizu & Kawato, 2009). In songbirds, the gradual adaptation of sequence probabilities within blocks might likewise be controlled by motor and premotor song control structures, while visual contextual cues could be processed in avian structures analogous to mammalian prefrontal cortex, outside the song system. For example, the association area nidopallium caudolaterale (Güntürkün, 2005), is activated by arbitrary visual cues that encode learned rules (Veit & Nieder, 2013; Veit, Pidpruzhnykova, & Nieder, 2015), and this or other avian association areas (Jarvis et al., 2013) may serve as an intermediate representation of the arbitrary contextual cues that can drive rapid learned shifts in syllable sequencing.

At the level of song motor control, our results indicate a greater capacity for rapid and flexible adjustment of syllable transition probabilities than previously appreciated. Current models of song production include networks of neurons in the vocal premotor nucleus HVC responsible for the temporal control of individual syllables, that are linked together by activity in a recurrent loop through brainstem vocal centers (Andalman, Foerster, & Fee, 2011; Ashmore, Wild, & Schmidt, 2005; Cohen et al., 2020; Hamaguchi, Tanaka, & Mooney, 2016). At branch points in songs with variable syllable sequencing, one influential model posits that which syllable follows a branch point is determined by stochastic processes that depend on the strength of the connections between alternative syllable production networks, and thus dynamics local to HVC (D. Z. Jin, 2009; D. Z. Jin & Kozhevnikov, 2011; Troyer, Brainard, & Bouchard, 2017; Zhang, Wittenbach, Jin, & Kozhevnikov, 2017). Such models could account for a gradual adjustment of sequence probabilities over a period of hours or days (Lipkind et al., 2013; Warren et al., 2012) through plasticity of motor control parameters, such as the strength of synaptic connections within HVC. However, our results demonstrate that there is not a single set of relatively fixed transition probabilities that undergo gradual adjustments, as could be captured in synaptic connectivity of branched syllable control networks. Rather, the song system has the capacity to maintain distinct representations of transition probabilities and can immediately switch between those in response to visual cues. HVC receives a variety of inputs that potentially could convey such visual or cognitive influences on sequencing (Bischof & Engelage, 1985; Cynx, 1990; Seki, Suzuki, Takahasi, & Okanoya, 2008; Ullrich, Norton, & Scharff, 2016; Wild, 1994), and one of these inputs, Nif, has previously been shown to be relevant for sequencing (Hosino & Okanoya, 2000; Vyssotski, Stepien, Keller, & Hahnloser, 2016). It therefore is likely that the control of syllable sequence in Bengalese finches involves a mix of processes local to nuclei of the song motor pathway (Basista et al., 2014; Zhang et al., 2017) as well as inputs that convey a variety of sensory feedback and contextual information. The well-understood circuitry of the avian song system makes this an attractive model to investigate how such top-down pathways orchestrate the kind of contextual control of vocalizations demonstrated in this study, and more broadly to uncover how differing cognitive demands can flexibly and adaptively reconfigure motor output.

Methods

Subjects & Sound recordings

The experiments were carried out on eight adult male Bengalese finches (*Lonchura striata*) obtained from the lab's breeding colony (age range 128-320 days post hatch, median 178 days, at start of experiment). Birds were placed in individual sound-attenuating boxes with continuous monitoring and auditory recording of song. Song was recorded using an omnidirectional microphone above the cage. We used custom software for the online recognition of target syllables and real-time delivery of short 40ms bursts of white noise (WN) depending on the syllable sequence (Tumer & Brainard, 2007; Warren et al., 2012). All procedures were performed in accordance with animal care protocols approved by the University of California, San Francisco Institutional Animal Care and Use Committee (IACUC).

Training procedure and blocks

Bengalese finch song consists of a discrete number of vocal elements, called syllables, that are separated by periods of silence. At the start of each experiment, a template was generated to recognize a specific sequence of syllables (the target sequence) for each bird based on their unique spectral structure. In the context-dependent auditory feedback protocol, the target sequence that received aversive white noise (WN) feedback switched between blocks of different light contexts. Colored LEDs (superbrightleds.com, St. Louis, MO; green 520 nm, amber 600 nm) produced two visually distinct environments (green and yellow) to serve as contextual cues to indicate which sequences would elicit WN and which would 'escape' (i.e. not trigger WN). We wanted to test whether the birds would be able to associate song changes with any arbitrary visual stimulus, therefore there was no reason to choose these specific colors, and the birds' color perception in this range should not matter, as long as they were able to discriminate the colors. The entire day was used for data acquisition by alternating the two possible light contexts. We determined sensitivity and specificity of the template to the target sequence on a randomly selected set of 20 song bouts on which labels and delivery of WN was hand-checked. Template sensitivity was defined as: sensitivity = (number of correct hits)/(total number of target sequences). The average template sensitivity across experiments was 91.3% (range 75.2-100%). Template specificity was defined as: specificity = (number of correct escapes)/(number of correct escapes plus number of false alarms), where correct escapes were defined as the number of target sequences of the currently inactive context that were not hit by WN, and false alarms were defined as any WN that was delivered either on the target sequence of the currently inactive context, or anywhere else in song. The average template specificity was 96.7% (range 90.6-100%).

At the start of each experiment, before WN training, songs were recorded during a baseline period in which cage illumination was switched between colors at random intervals. Songs from this baseline period were separately analyzed for each light color to confirm that there was no systematic, unlearned effect of light cues on sequencing before training. During initial training, cage illumination was alternatingly switched between colors at random intervals. Intervals were drawn from uniform distributions which differed between birds (60-150min (4 birds), 10-30min (2 birds), 60-240min (1 bird), 30-150min (1 bird)). Different training schedules were assigned to birds arbitrarily and were not related to a bird's performance. After an extended period of training (average 33 days, range 12-79 days), probe blocks without WN were included, to test whether sequencing changes could be elicited by visual cues alone. During this period, probe blocks were interspersed with WN training blocks. Probe blocks made up approximately one third of total blocks (10 of 34 blocks in the sequence) and 7-35% of total time, depending on the bird. The duration of probe blocks was typically shorter or equal to the duration of WN blocks (10-30min for 6 birds, 30-120min for one bird, 18-46min for one bird). The total duration of the experiment, consisting of baseline, training, and probe periods, was on average 52 days. During this period, birds sang 226 (range 66-356) bouts per day during baseline days and 258 (range 171-368) bouts per day during the period of probe collection at the end of training (14% increase). The average duration of song bouts also changed little, with both the average number of target sequences per bout (8.7 during baseline, 7.7 during probes, 7% decrease) and the average number of syllables per bout (74 during baseline, 71 during probes, 2% decrease) decreasing slightly. In addition to the eight birds that completed this training paradigm, three birds were started on contextual training but never progressed to testing with probe blocks, because they did not exhibit single-context learning (n=1); because of technical issues with consistent targeting at branch points, (n=1); or because they lost sequence variability during initial stages of training (n=1); these birds are excluded from the results. Of the 8 birds that completed training, three birds exhibited relatively small context-dependent changes in sequencing (**Fig. 1H**). We examined several variables to assess whether they could account for differences in the magnitude of learning across birds, including the bird's age, overall transition entropy of the song (Katahira, Suzuki, Kagawa, & Okanoya, 2013), transition entropy at the targeted branch points (Warren et al., 2012), as well as the distance between the WN target and the closest preceding branch point in the sequence. None of these variables were significantly correlated with the degree of contextual learning that birds expressed (**Sup. Fig. 4A-D**) and consequently all birds were treated as a single group in analysis and reporting of results. In a subset of experiments (n=3), after completing measurements with probe blocks, we added a third, neutral context (**Fig. 5**), signaled by white light, in which there was no WN reinforcement.

Syllable sequence annotation

Syllable annotation for data analysis was performed offline. Each continuous period of singing that was separated from others by at least two seconds of silence was treated as an individual 'song', or 'song bout'. Song was bandpass filtered between 500 Hz and 10000 Hz and segmented into syllables and gaps based on amplitude threshold and timing parameters determined manually for each bird. A small sample of songs (approximately 20 song bouts) was then annotated manually based on visual inspection of spectrograms. These data were used to train an offline autolabeler ('hybrid-vocal-classifier', (Nicholson, 2021)) which was then used to label the remaining song bouts. Autolabeled songs were processed further in a semi-automated way depending on each bird's unique song, for example to separate or merge syllables that were not segmented correctly (detected by their duration distributions), to deal with WN covering syllables (detected by its amplitude), and to correct autolabeling errors detected based on the syllable sequence. A subset of songs was inspected manually for each bird to confirm correct labeling.

Sequence probability analyses

Sequence probability was first calculated within each song bout as the frequency of the yellow target sequence relative to the total number of yellow and green target sequences: $p = \frac{n(target\_Y)}{n(target\_Y)+n(target\_G)}$. Note that this differs from transition probabilities at branch points in song in that it ignores possible additional syllable transitions at the branch point, and does not require the targeted sequences to be directly following the same branch point. For example for the experiment in **Figure 3**, the target sequences were 'n-ab' and 'f-ab', so the syllable covered by WN ('b' in both contexts) was 2-3 syllables removed from the respective branch point in the syllable sequence ('n-f' vs. 'n-a', or 'f-n' vs. 'f-a'). Note also that units of sequence probability are in percent, therefore reported changes in percentages (for example Fig. 1H, 2E,F) describe absolute changes in sequence probability, which reflect the proportion of each target sequence, not percent changes. Song bouts that did not contain either of the two target sequences were discarded. In the plots of sequence probability over several days in **Figure 1A-C**, we calculated sequence probability for all bouts on a given day (average n=1854 renditions of both target sequences per day). We estimated 95% confidence intervals by approximation with a normal distribution as $p \pm z * \sqrt{\frac{p*(1-p)}{n}}$ with $n = n(target\_Y) + n(target\_G)$ and z =1.96. Context switches were processed to include only switches between adjacent blocks during the same day, i.e. excluding overnight switches and treating blocks as separate contexts if one day started with the same color that had been the last color on the previous day. If a bird did not produce any song during one block, this block was merged with any neighboring block of the same color (for example, green probe without songs before green WN, where the context switch would not be noticeable for the bird). If the light color switched twice (or more) without any song bouts, those context switches were discarded.

In order to reduce variability associated with changes across individual song bouts, shift magnitude was calculated as the difference between the first five song bouts in the new context and the last five song bouts in the old context. Only context switches with at least 3 song bouts in each adjacent block were included in analyses of shift magnitude. In plots showing songs aligned to context switches, the x-axis is limited to show only points for which at least half of the blocks contributed data (i.e. in **Fig. 2D**, half of the green probe blocks contained at least 6 songs). All statistical tests were performed with MATLAB. We used non-parametric tests to compare changes across birds (Wilcoxon ranksum test for unpaired data, Wilcoxon signed rank test for paired data), because with only 8 birds/data points, it is more conservative to assume that data are not Gaussian distributed.

Analysis of acquisition

In order to investigate how context-dependent performance developed over training (**Fig. 2 G-L**), we quantified changes to sequence probabilities across block switches for 5 birds for which we had a continuous record from the onset of training. Sequence probability curves (e.g. **Fig. 2 H**) for yellow switches were inverted so that both yellow and green switches were plotted in the same direction, aligned by the time of context switches, and were cut off at a time point relative to context switches where fewer than five switches contributed data. We then subtracted the mean pre-switch value from each sequence probability curve. For visual display of the example bird, sequence probability curves were smoothed with a 9 bout boxcar window, and displayed in bins of 7 context switches. To calculate the slope of slopes and slope of intercepts (**Fig. 2 L**), we calculated a linear fit to the post-switch parts of the unsmoothed sequence probability curve for each individual context switch.

Specificity to relevant branch points

To calculate the specificity of the context difference to the targeted branch points in song, we generated transition diagrams for each bird. To simplify the diagrams, introductory notes were summarized into a single introductory state. Introductory notes were defined for each bird as up to three syllables occurring at the start of song bouts before the main motif, which tended to be quieter, more variable, with high probabilities to repeat and to transition to other introductory notes. Repeat phrases were also summarized into a single state. Motifs, or chunks, in the song with fixed order of syllables were identified by the stereotyped transitions and short gap durations between syllables in the motif (Isola, Vochin, & Sakata, 2020; Suge & Okanoya, 2009) and were also summarized as a single state in the diagram. Sometimes, the same syllable can be part of several fixed chunks (Katahira et al., 2013), in which case it may appear several times in the transition diagram. We then calculated the difference between the transition matrices for the two probe contexts at each transition that was a branch point (defined as more than 3% and less than 97% transition probability). These context differences were split into 'targeted branch points' i.e. the branch point or branch points most closely preceding the target sequences in the two contexts, and 'non-targeted branch points' i.e. all other branch points in the song. We calculated the proportion of absolute contextual difference in the transition matrix that fell to the targeted branch points, for example for the matrix in **Figure 4C**

(44+45)/(44+45+6+6+1+1+2+2) = 83.2%. Typically, birds with clear contextual differences at the target sequence also had high specificity of sequence changes to the targeted branch points.

To calculate the transition entropy of baseline song, we again summarized introductory notes into a single introductory state. In addition, the same syllable as part of multiple fixed motifs, or in multiple positions within the same fixed motif, were renamed as different syllables, so as not to count as sequence variability what was really a stereotyped sequence (i.e. a-b 50% and b-c 50% in the fixed sequence 'abbc'). Transition entropy was then calculated as in (Katahira et al., 2013): $H = -\sum_x p(x) \sum_y p(y|x) \log p(y|x)$ with x denoting the preceding syllable and y denoting the current syllable, over all syllables in the song.

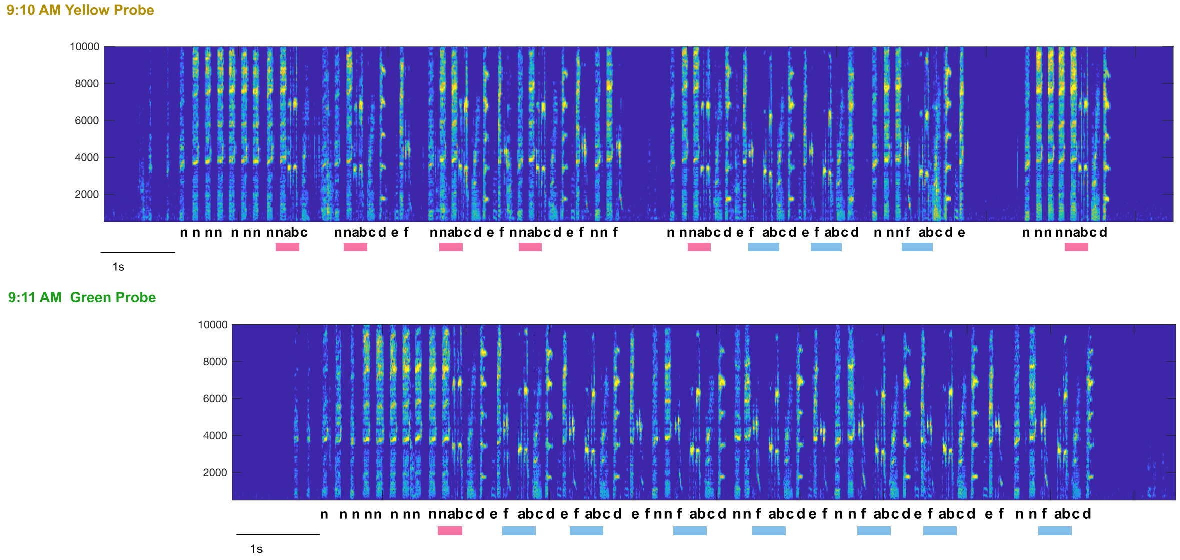
Acknowledgements

We thank Alla Karpova, Jon Sakata, Dave Mets, William Mehaffey, Assaf Breska, Guy Avraham for helpful discussions and comments on earlier versions of this manuscript. This work was supported by the Howard Hughes Medical Institute. Lena Veit was supported as a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation and by a postdoctoral fellowship from Leopoldina German National Academy of Sciences. Christian J Monroy Hernandez was supported by an HHMI EXROP summer fellowship.

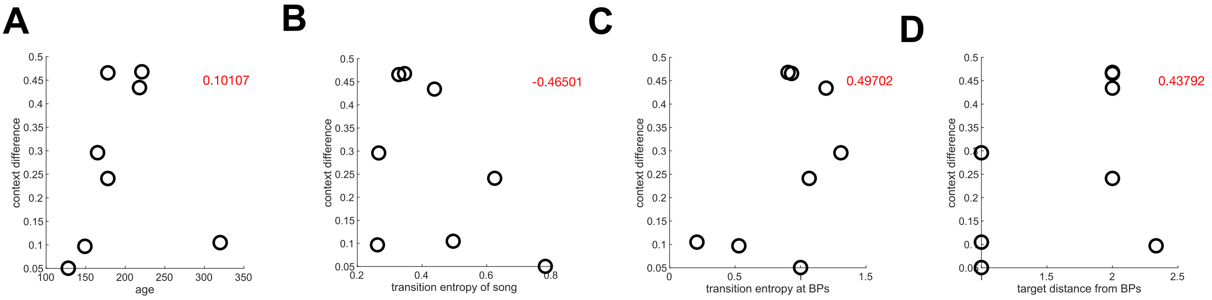
Competing interests:

The authors declare no competing interests.

Supplementary Figures



Supplement to Figure 3: Example song bouts surrounding a probe-probe context switch Full song bouts (same as in Fig.3 A,B) sung by Bird 2 immediately before and after a switch from a yellow probe block to a green probe block. Scale for x-axis is 1s, y-axis shows frequency in Hz. The recording program was set to never switch lights during an ongoing song recording, so the time of light switch in between these two recordings cannot be shown.



Supplement to Figure 4: Possible explanations for differences in contextual learning

Correlations of the magnitude of contextual differences with the birds' age (A), transition entropy of the entire song (B), transition entropy at the branch points preceding the target sequences (C) and distance of the target sequence from the immediately preceding branch points (D). None of these variables were significantly correlated with the degree of contextual learning that birds expressed, p=0.81 (A), p=0.24 (B), p=0.21 (C), p=0.28 (D). Red numbers indicate correlation coefficients.

References

Abrahamse, E., Ruitenberg, M., De Kleine, E., & Verwey, W. B. (2013). Control of automated behavior: insights from the discrete sequence production task. *Frontiers in Human Neuroscience*, *7*. http://doi.org/10.3389/fnhum.2013.00082

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: an evolutionary perspective. *The Behavioral and Brain Sciences*, *37*(6), 568–9– discussion 577–604. http://doi.org/10.1017/S0140525X13003099

Alcami, P., Ma, S., & Gahr, M. (2021). Telemetry reveals rapid duel-driven song plasticity in a naturalistic social environment. *bioRxiv*, 803411.

Aldridge, J. W., & Berridge, K. C. (2002). Coding of Behavioral Sequences in the Basal Ganglia. In *The Basal Ganglia VII* (Vol. 52, pp. 53–66). Boston, MA: Springer, Boston, MA. http://doi.org/10.1007/978-1-4615-0715-4_7

Andalman, A. S., Foerster, J. N., & Fee, M. S. (2011). Control of vocal and respiratory patterns in birdsong: dissection of forebrain and brainstem mechanisms using temperature. *PloS One*, *6*(9), e25461. http://doi.org/10.1371/journal.pone.0025461

Ashe, J., Lungu, O. V., Basford, A. T., & Lu, X. (2006). Cortical control of motor sequences. *Current Opinion in Neurobiology*, *16*(2), 213–221. http://doi.org/10.1016/j.conb.2006.03.008

Ashmore, R. C., Wild, J. M., & Schmidt, M. F. (2005). Brainstem and forebrain contributions to the generation of learned motor behaviors for song. *The Journal of Neuroscience*, *25*(37), 8543–8554. http://doi.org/10.1523/JNEUROSCI.1668-05.2005

Basista, M. J., Elliott, K. C., Wu, W., Hyson, R. L., Bertram, R., & Johnson, F. (2014). Independent Premotor Encoding of the Sequence and Structure of Birdsong in Avian Cortex. *The Journal of Neuroscience*, *34*(50), 16821–16834. http://doi.org/10.1523/JNEUROSCI.1940-14.2014

Belyk, M., & Brown, S. (2017). The origins of the vocal brain in humans. *Neuroscience and Biobehavioral Reviews*, *77*, 177–193. http://doi.org/10.1016/j.neubiorev.2017.03.014

Benichov, J. I., & Vallentin, D. (2020). Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches. *Nature Communications*, *11*(1), 1–10. http://doi.org/10.1038/s41467-019-13938-0

Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences*, *15*(3), 113–121. http://doi.org/10.1016/j.tics.2011.01.002

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, *143*(3), 233–262. http://doi.org/10.1037/bul0000099

Bischof, H. J., & Engelage, J. (1985). Flash evoked responses in a song control nucleus of the zebra finch (Taeniopygia guttata castanotis). *Brain Research*, *326*(2), 370–374.

Blanco-Elorrieta, E., & Pylkkänen, L. (2016). Bilingual Language Control in Perception versus Action: MEG Reveals Comprehension Control Mechanisms in Anterior Cingulate Cortex and Domain-General Control of Production in Dorsolateral Prefrontal Cortex. *The Journal of Neuroscience*, *36*(2), 290–301. http://doi.org/10.1523/JNEUROSCI.2597-15.2016

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, *417*(6886), 351–358. http://doi.org/10.1038/417351a

Brecht, K. F., Hage, S. R., Gavrilov, N., & Nieder, A. (2019). Volitional control of vocalizations in corvid songbirds. *PLoS Biology*, *17*(8), e3000375. http://doi.org/10.1371/journal.pbio.3000375

Brumm, H., & Zollinger, S. A. (2013). Avian Vocal Production in Noise. In *Animal Communication and Noise* (2nd ed., Vol. 2, pp. 187–227). Berlin, Heidelberg: Springer, Berlin, Heidelberg. http://doi.org/10.1007/978-3-642-41494-7_7

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), 6641–6646. http://doi.org/10.1073/pnas.1522306113

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

Cunningham, H. A., & Welch, R. B. (1994). Multiple concurrent visual-motor mappings: implications for models of adaptation. *Journal of Experimental Psychology. Human Perception and Performance*, *20*(5), 987–999. http://doi.org/10.1037//0096-1523.20.5.987

Cynx, J. (1990). Experimental determination of a unit of song production in the zebra finch (Taeniopygia guttata). *Journal of Comparative Psychology*, *104*(1), 3–10.

De Baene, W., Duyck, W., Brass, M., & Carreiras, M. (2015). Brain Circuit for Cognitive Control Is Shared by Task and Language Switching. *Journal of Cognitive Neuroscience*, *27*(9), 1752–1765.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: common themes and mechanisms. *Annual Review of Neuroscience*, *22*(1), 567–631. http://doi.org/10.1146/annurev.neuro.22.1.567

Gadagkar, V., Puzerey, P. A., & Goldberg, J. H. (2019). Dopamine neurons change their tuning according to courtship context in singing birds. *bioRxiv*, 822817. http://doi.org/10.1101/822817

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology (Hove, England)*, *25*(5), 515–530. http://doi.org/10.1080/20445911.2013.796377

Güntürkün, O. (2005). The avian “prefrontal cortex”and cognition. *Current Opinion in Neurobiology*, *15*(6), 686–693.

Hage, S. R., & Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nature Communications*, *4*(1), 2409. http://doi.org/10.1038/ncomms3409

Hage, S. R., & Nieder, A. (2016). Dual Neural Network Model for the Evolution of Speech and Language. *Trends in Neurosciences*, *39*(12), 813–829. http://doi.org/10.1016/j.tins.2016.10.006

Hamaguchi, K., Tanaka, M., & Mooney, R. (2016). A Distributed Recurrent Network Contributes to Temporally Precise Vocalizations. *Neuron*, *91*(3), 680–693. http://doi.org/10.1016/j.neuron.2016.06.019

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579. http://doi.org/10.1126/science.298.5598.1569

Hedley, R. W., Denton, K. K., & Weiss, R. E. (2017). Accounting for syntax in analyses of countersinging reveals hidden vocal dynamics in a songbird with a large repertoire. *Animal Behaviour*, *131*, 23–32. http://doi.org/10.1016/j.anbehav.2017.06.021

Heinig, A., Pant, S., Dunning, J., Bass, A., Coburn, Z., & Prather, J. F. (2014). Male mate preferences in mutual mate choice: finches modulate their songs across and within male-female interactions. *Animal Behaviour*, *97*, 1–12. http://doi.org/10.1016/j.anbehav.2014.08.016

Hikosaka, O., Nakahara, H., Rand, M. K., Sakai, K., Lu, X., Nakamura, K., et al. (1999). Parallel neural networks for learning sequential procedures. *Trends in Neurosciences*, *22*(10), 464–471. http://doi.org/10.1016/S0166-2236(99)01439-3

Hosino, T., & Okanoya, K. (2000). Lesion of a higher-order song nucleus disrupts phrase level complexity in Bengalese finches. *NeuroReport*, *11*(10), 2091–2095.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor Adaptation of Speech I. http://doi.org/10.1044/1092-4388(2002/023)

Howard, I. S., Wolpert, D. M., & Franklin, D. W. (2013). The effect of contextual cues on the encoding of motor memories. *Journal of Neurophysiology*, *109*(10), 2632–2644. http://doi.org/10.1152/jn.00773.2012

Imamizu, H., & Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions. *Psychological Research*, *73*(4), 527–544. http://doi.org/10.1007/s00426-009-0235-1

Imamizu, H., Sugimoto, N., Osu, R., Tsutsui, K., Sugiyama, K., Wada, Y., & Kawato, M. (2007). Explicit contextual information selectively contributes to predictive switching of internal models. *Experimental Brain Research*, *181*(3), 395–408. http://doi.org/10.1007/s00221-007-0940-1

Isola, G. R., Vochin, A., & Sakata, J. T. (2020). Manipulations of inhibition in cortical circuitry differentially affect spectral and temporal features of Bengalese finch song. *Journal of Neurophysiology*. http://doi.org/10.1152/jn.00142.2019

Jaffe, P. I., & Brainard, M. S. (2020). Acetylcholine acts on songbird premotor circuitry to invigorate vocal output. http://doi.org/10.7554/eLife.53288

James, L. S., Dai, J. B., & Sakata, J. T. (2018a). Ability to modulate birdsong across social contexts develops without imitative social learning. *Biology Letters*, *14*(3), 20170777. http://doi.org/10.1098/rsbl.2017.0777

James, L. S., Dai, J. B., & Sakata, J. T. (2018b). Ability to modulate birdsong across social contexts develops without imitative social learning. *Biology Letters*, *14*(3), 20170777. http://doi.org/10.1098/rsbl.2017.0777

Jarvis, E. D., Yu, J., Rivas, M. V., Horita, H., Feenders, G., Whitney, O., et al. (2013). Global view of the functional molecular organization of the avian cerebrum: Mirror images and functional columns. *The Journal of Comparative Neurology*, *521*(16), 3614–3665. http://doi.org/10.1002/cne.23404

Jin, D. Z. (2009). Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *80*(5 Pt 1), 051902. http://doi.org/10.1103/PhysRevE.80.051902

Jin, D. Z., & Kozhevnikov, A. A. (2011). A compact statistical model of the song syntax in Bengalese finch. *PLoS Computational Biology*, *7*(3), e1001108–e1001108. http://doi.org/10.1371/journal.pcbi.1001108

Jin, X., & Costa, R. M. (2015). Shaping action sequences in basal ganglia circuits. *Current Opinion in Neurobiology*, *33*, 188–196. http://doi.org/10.1016/j.conb.2015.06.011

Katahira, K., Suzuki, K., Kagawa, H., & Okanoya, K. (2013). A simple explanation for the evolution of complex song syntax in Bengalese finches. *Biology Letters*, *9*(6), 20130842–20130842. http://doi.org/10.1098/rsbl.2013.0842

Keough, D., & Jones, J. A. (2011). Contextual cuing contributes to the independent modification of multiple internal models for vocal control. *Journal of Neurophysiology*. http://doi.org/10.1152/jn.00291.2010

King, S. L., & McGregor, P. K. (2016). Vocal matching: the what, the why and the how. *Biology Letters*, *12*(10). http://doi.org/10.1098/rsbl.2016.0666

Kojima, S., & Doupe, A. J. (2011). Social performance reveals unexpected vocal competency in young songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1687–1692. http://doi.org/10.1073/pnas.1010502108

Lee, J.-Y., & Schweighofer, N. (2009). Dual Adaptation Supports a Parallel Architecture of Motor Memory. *The Journal of Neuroscience*, *29*(33), 10396–10404. http://doi.org/10.1523/JNEUROSCI.1294-09.2009

Lipkind, D., Marcus, G. F., Bemis, D. K., Sasahara, K., Jacoby, N., Takahasi, M., et al. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature*, *498*(7452), 104–108. http://doi.org/10.1038/nature12173

McDougle, S. D., Ivry, R. B., & Taylor, J. A. (2016). Taking Aim at the Cognitive Side of Learning in Sensorimotor Adaptation Tasks. *Trends in Cognitive Sciences*, *20*(7), 535–544. http://doi.org/10.1016/j.tics.2016.05.002

Nicholson, D. (2021). NickleDave/hybrid-vocal-classifier:. http://doi.org/10.5281/zenodo.4678768

Nieder, A., & Mooney, R. (2020). The neurobiology of innate, volitional and learned vocalizations in mammals and birds. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20190054. http://doi.org/10.1098/rstb.2019.0054

Okanoya, K. (2004). The Bengalese finch: a window on the behavioral neurobiology of birdsong syntax. *Annals of the New York Academy of Sciences*, *1016*, 724–735. http://doi.org/10.1196/annals.1298.026

Reichmuth, C., & Casey, C. (2014). Vocal learning in seals, sea lions, and walruses. *Current Opinion in Neurobiology*, *28*, 66–71. http://doi.org/10.1016/j.conb.2014.06.011

Rochet-Capellan, A., & Ostry, D. J. (2011). Simultaneous acquisition of multiple auditory-motor transformations in speech. *The Journal of Neuroscience*, *31*(7), 2657–2662. http://doi.org/10.1523/JNEUROSCI.6020-10.2011

Sakata, J. T., Hampton, C. M., & Brainard, M. S. (2008). Social Modulation of Sequence and Syllable Variability in Adult Birdsong. *Journal of Neurophysiology*, *99*(4), 1700–1711. http://doi.org/10.1152/jn.01296.2007

Searcy, W. A., & Beecher, M. D. (2009). Song as an aggressive signal in songbirds. *Animal Behaviour*, *78*(6), 1281–1292. http://doi.org/10.1016/j.anbehav.2009.08.011

Seki, Y., Suzuki, K., Takahasi, M., & Okanoya, K. (2008). Song Motor control organizes acoustic patterns on two levels in Bengalese finches (Lonchura striata var. domestica). *J Comp Physiol a*, *194*(6), 533–543. http://doi.org/10.1007/s00359-008-0328-0

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *The Neuroscientist : a Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *17*(2), 197–208. http://doi.org/10.1177/1073858410386727

Sossinka, R., & Böhner, J. (1980). Song Types in the Zebra Finch Poephila guttata castanotis1. *Zeitschrift Für Tierpsychologie*, *53*(2), 123–132. http://doi.org/10.1111/j.1439-0310.1980.tb01044.x

Suge, R., & Okanoya, K. (2009). Perceptual chunking in the self-produced songs of Bengalese finches (Lonchura striata var. domestica). *Animal Cognition*, *13*(3), 515–523. http://doi.org/10.1007/s10071-009-0302-4

Suzuki, T. N., & Zuberbühler, K. (2019). Animal syntax. *Current Biology*, *29*(14), R669–R671. http://doi.org/10.1016/j.cub.2019.05.045

Tanji, J. (2001). Sequential organization of multiple movements: involvement of cortical motor areas. *Annual Review of Neuroscience*, *24*(1), 631–651. http://doi.org/10.1146/annurev.neuro.24.1.631

Trillo, P. A., & Vehrencamp, S. L. (2005). Song types and their structural features are associated with specific contexts in the banded wren. *Animal Behaviour*, *70*(4), 921–935. http://doi.org/10.1016/j.anbehav.2005.02.004

Troyer, T. W., Brainard, M. S., & Bouchard, K. E. (2017). Timing during transitions in Bengalese finch song: implications for motor sequencing. *Journal of Neurophysiology*, *118*(3), 1556–1566. http://doi.org/10.1152/jn.00296.2017

Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of “crystallized” adult birdsong. *Nature*, *450*(7173), 1240–1244. http://doi.org/10.1038/nature06390

Ullrich, R., Norton, P., & Scharff, C. (2016). Waltzing Taeniopygia: integration of courtship song and dance in the domesticated Australian zebra finch. *Animal Behaviour*, *112*, 285–300. http://doi.org/10.1016/j.anbehav.2015.11.012

Veit, L., & Nieder, A. (2013). Abstract rule neurons in the endbrain support intelligent behaviour in corvid songbirds. *Nature Communications*, *4:2878*. http://doi.org/10.1038/ncomms3878

Veit, L., Pidpruzhnykova, G., & Nieder, A. (2015). Associative learning rapidly establishes neuronal representations of upcoming behavioral choices in crows. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(49), 15208–15213. http://doi.org/10.1073/pnas.1509760112

Vignal, C., Mathevon, N., & Mottin, S. (2004). Audience drives male songbird response to partner's voice. *Nature*, *430*(6998), 448–451. http://doi.org/10.1038/nature02645

Vyssotski, A. L., Stepien, A. E., Keller, G. B., & Hahnloser, R. H. R. (2016). A Neural Code That Is Isometric to Vocal Output and Correlates with Its Sensory Consequences. *PLoS Biology*, *14*(10), e2000317. http://doi.org/10.1371/journal.pbio.2000317

Warren, T. L., Charlesworth, J. D., Tumer, E. C., & Brainard, M. S. (2012). Variable sequencing is actively maintained in a well learned motor skill. *The Journal of Neuroscience*, *32*(44), 15414–15425. http://doi.org/10.1523/JNEUROSCI.1254-12.2012

Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: a promising paradigm whose time has passed. *Evolutionary Anthropology*, *21*(5), 195–205. http://doi.org/10.1002/evan.21319

Wild, J. M. (1994). Visual and somatosensory inputs to the avian song system via nucleus uvaeformis (Uva) and a comparison with the projections of a similar thalamic nucleus in a nonsongbird,columbia livia. *The Journal of Comparative Neurology*, *349*(4), 512–535. http://doi.org/10.1002/cne.903490403

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), 739–751. http://doi.org/10.1038/nrn3112

Zhang, Y. S., Wittenbach, J. D., Jin, D. Z., & Kozhevnikov, A. A. (2017). Temperature Manipulation in Songbird Brain Implicates the Premotor Nucleus HVC in Birdsong Syntax. *The Journal of Neuroscience*, *37*(10), 2600–2611. http://doi.org/10.1523/JNEUROSCI.1827-16.2017