bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 should be addressed: lena.veit@uni-tuebingen.de, msb@phy.ucsf.edu

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

The flexible control of sequential behavior is a fundamental aspect of speech, enabling endless reordering of a limited set of learned vocal elements (i.e. syllables or words). Songbirds are phylogenetically distant from humans, but share the capacity for vocal learning as well as neural circuitry for vocal control that includes direct cortical-brainstem projections. Based on these similarities, we hypothesized that songbirds might likewise be able to learn flexible, moment-by-moment control over vocal production. Here, we demonstrate that Bengalese finches, 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 in the absence of external reinforcement. Our findings reveal a capacity in songbirds for learned contextual control over syllable sequencing that parallels aspects of human cognitive control over speech.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

A crucial aspect in the evolution of human speech is thought to be 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. This flexibility ingrained in human language stands in striking contrast to the largely innate and stereotypic vocalization patterns of most animal species, including our closest relatives, the non-human primates. One recent hypothesis posits that a marked elaboration of the direct innervation of brainstem and midbrain vocal networks by executive control structures in the frontal cortex underpins the human ability for flexible vocal control (Hage & Nieder, 2016; Simonyan & Horwitz, 2011). However, the evolutionary and neural circuit mechanisms that have enabled the development of this remarkable functionality 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; 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 led us to consider whether birds might also share the human capcity to learn flexible control of syllable sequencing.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

In spite of the similarities between speech and birdsong, song has largely been seen as an affective behavior, elicited instinctually by contact with potential mates, rivals, or performed spontaneously (Doupe & Kuhl, 1999; Nieder & Mooney, 2020; T. N. Suzuki, Wheatcroft, & Griesser, 2020). There are differences between songs produced in distinct social contexts, such as courtship or territorial defense, and these can include alteration to the sequencing of syllables (Chen, Matheson, & Sakata, 2016; Sakata, Hampton, & Brainard, 2008; Searcy & Beecher, 2009). However, these social influences likely reflect a general modulation of song structure related to the animal's affective state (Berwick, Okanoya, Beckers, & Bolhuis, 2011), and do not reveal whether song can be modified more flexibly by different cognitive factors.

Here, we test whether songbirds, like humans, can learn to flexibly control vocal sequences, independently of social or other natural contexts. We specifically asked whether Bengalese finches (*Lonchura striata*) could learn to alter the sequencing of learned song elements in response to arbitrarily chosen visual cues. Each Bengalese finch song consists of ~5-12 acoustically distinct elements ('syllables') that are sung in variable but non-random order (Okanoya, 2004; Warren, Charlesworth, Tumer, & Brainard, 2012). For a given bird, the relative probabilities of specific transitions between syllables normally remain constant over time, but previous work has shown that birds can gradually adjust those probabilities 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 data reveal an unexpected capacity of birds to immediately, flexibly, and adaptively adjust their sequencing of vocal elements in response to learned contextual cues, in a manner that parallels key aspects of human cognitive control over speech.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 (Tumer & Brainard, 2007; Warren et al., 2012). For example, a bird that normally transitioned from the fixed syllable sequence 'ab' to either a 'c' or a 'd' (**Figures 1A,B**, sequence probability of ~36% for 'ab-c' and ~64% for 'ab-d'), was exposed to a aversive white noise (WN) feedback delivered 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 alternative syllable sequence, 'ab-c', was the target sequence, the bird gradually increased the probability of the alternate sequence 'ab-d' 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 also learn to flexibly select specific syllable sequences on a moment-by-moment basis, we paired WN targeting of specific sequences with

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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. **Figure 1E** shows an example, with 'ab-d' targeted in yellow light, 'ab-c' in green light. Light contexts switched after random intervals of 1-2.5 hours. At baseline without WN, such light switches did not lead to significant sequencing changes, indicating that there was no inherent influence of the light context on sequence probabilities (**Fig. 1F**, 67+/-1.6% vs. 64+/-1.5%, $p = 0.17$, ranksum test, n=53 context blocks from baseline period). In contrast, significant sequencing differences that appropriately reduced aversive feedback in each reinforcement context were observed after 2 weeks of training (**Fig. 1G,** 36.5+/-4.8% vs. 83.1+/3.5%, $p < 0.01$, ranksum test, n=22 context blocks). Likewise, for all birds trained on this protocol (n = 8) a significant context-dependent shift in sequence probabilities developed over the period of training (1%+/-2% average difference in probabilities between contexts at baseline increased to 27+/-6% after a mean of 33 days training; $p < 0.01$, n=8, signed rank test, **Fig. 1H**). Thus, Bengalese finches are able to learn context-specific modifications to their syllable sequences.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

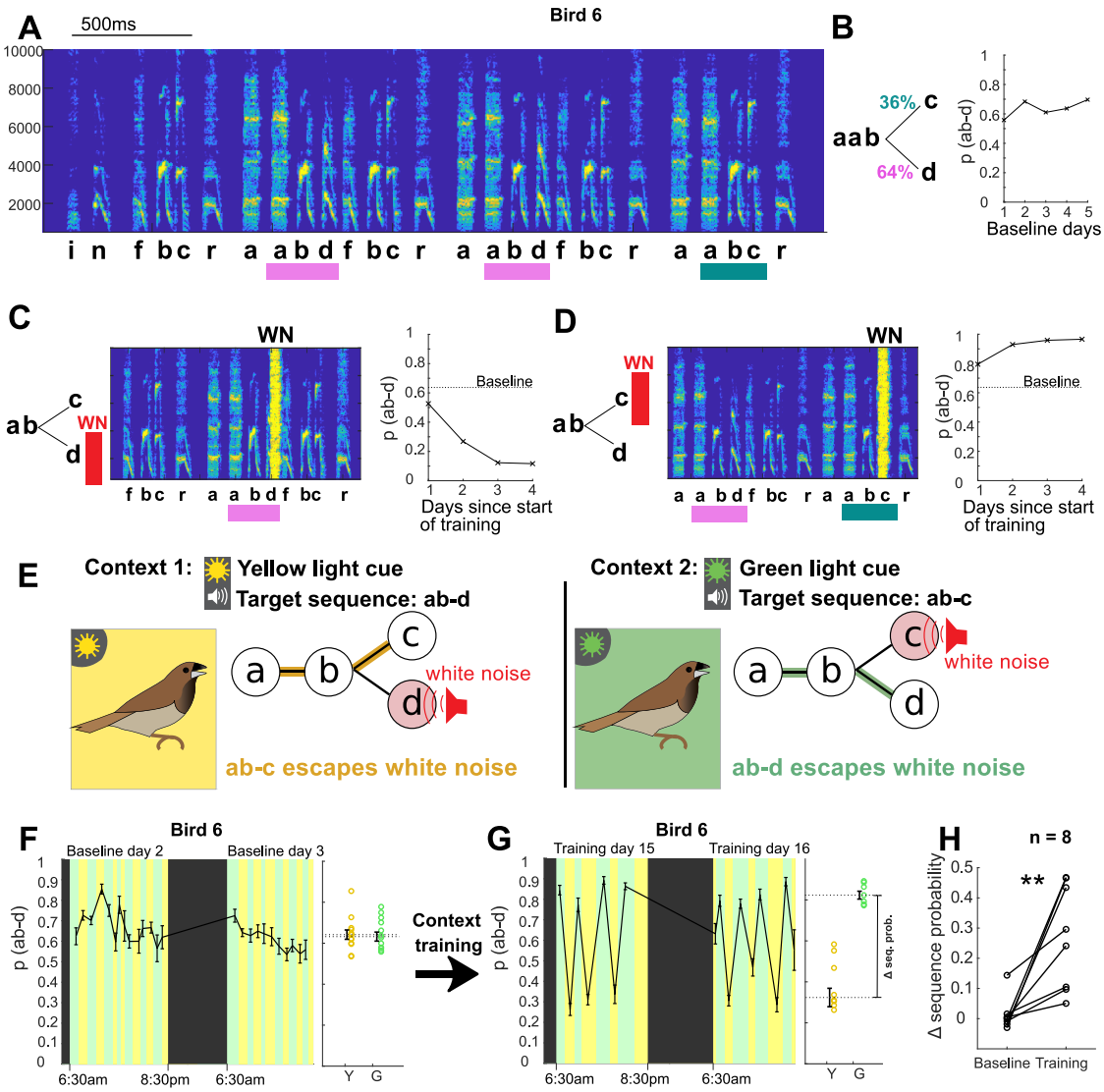

Figure 1. Bengalese finches can learn context-dependent sequencing

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. **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. **C**: Aversive reinforcement training. Schematic showing aversive WN after target syllable 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. **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, where the target for WN depends on colored lights. **F**: Left: Two example days of baseline light switches without WN 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 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

Syllable sequencing shifts immediately following context switches

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 adjustments to their syllable sequencing when entering a new light context by computing the average probability of targeted sequences across songs aligned by the time of transition between contexts (**Fig. 2A,B**, example experiment). This "transition-triggered average" of sequence probabilities revealed that across all birds, transitions to the yellow context were accompanied by an immediate decrease in the probability of the yellow target sequence, whereas transitions 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 immediate shifts, we calculate the difference in sequence probability from the last five songs in the old context to the first five songs in the new context; this difference was on average 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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, it is apparent for each of the context switches that 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 switch-aligned trajectories shifted approximately in parallel, indicating that the immediate shift at the start of the new context got larger, while the gradual change within blocks stayed approximately the same. Linear fits to the sequence probabilities for each trajectory (over the first 80 songs following light switch) 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 change within block (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). Similar to 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 grew over time in all birds (**Fig. 2L)**. This indicates that adjustments to sequence probability are due to two dissociable processes, a gradual adaptation of sequence probability within each block of WN, which is not modified by increased contextual training, and an immediate shift in sequence probability at the beginning of blocks, which is sensitive to contextual cues and increases with contextual training.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

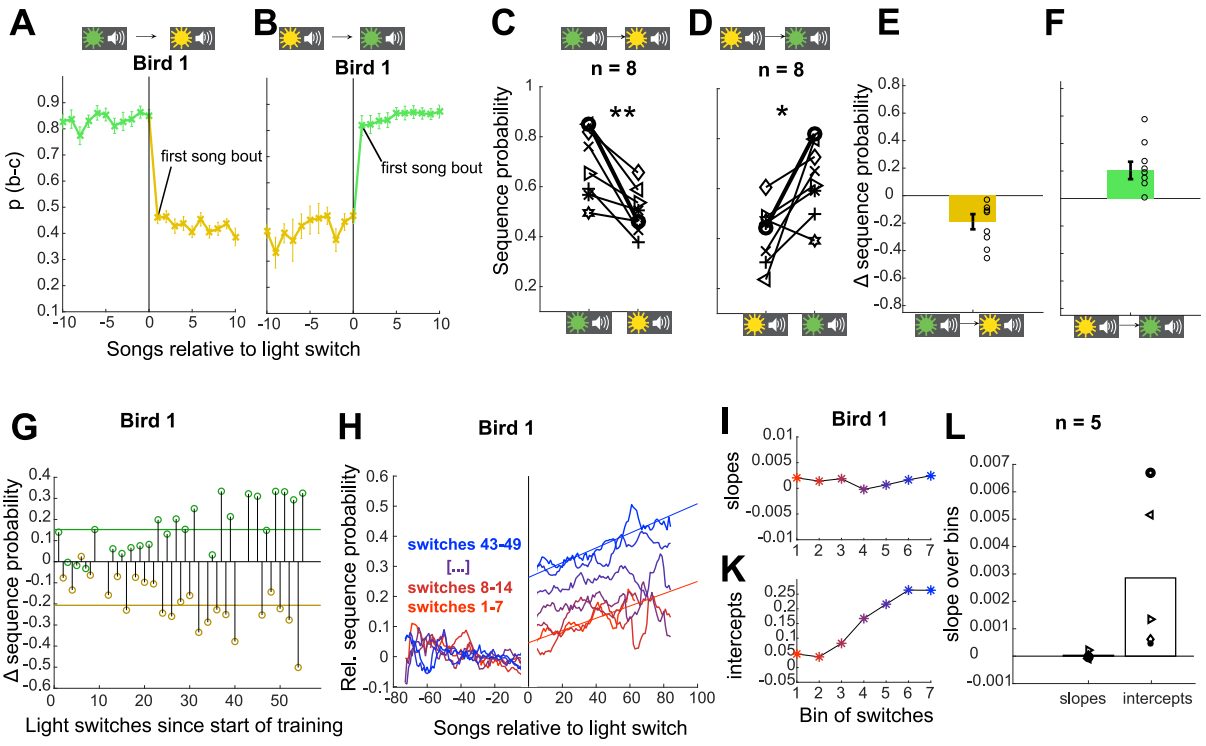

Figure 2. Sequence probabilities can be shifted immediately following a change in context.

A-B: Sequence probability for example Bird 1 over time, switching from green to yellow context (A) and switching from yellow to green context (B). Error bars indicate SEM across song bouts (n = 35 switches (A), n = 33 switches (B)). **C**: Change 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. **D**: Change in sequence probability from the last song in yellow context to the first song in green context. **E**: Shift magnitude for all birds, defined as the change 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 magnitude 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 in yellow switches.) **I**: Slope of fits to the sequence probability trajectories over song bouts within block. Units in change of relative sequence probability per song bout. **K:** Intercept of fits to sequence probability trajectories over song bouts within block. Units in relative sequence probability. **L**: Change in slope and change in intercept for five birds over the training process, determined as the slope of a linear fit to the curves in I,K.

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 –

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 behaviors appropriate for distinct contexts using solely visual cues, we included short probe blocks which presented the same light colors without WN stimulation. Probe blocks were interspersed in the sequence of training blocks so that each transition between types of blocks was possible and, on average, every third transition 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 transitions directly between probe blocks without intervening WN training blocks (probe-probe transitions). The sequence of context blocks included direct probe-probe transitions approximately every 17 transitions. **Figure 3A,B** shows song bouts for one example bird (Bird 2) which were sung consecutively across a context switch from yellow probe to green probe. In this case, immediately following the switch from the yellow context to the green context, the yellow target sequence ('fab') was more prevalent, and the green target sequence ('nab') was less prevalent. Consistent with this example, this bird exhibited appropriate shifts in syllable sequencing when changing directly between probe blocks in the absence of any WN (**Fig. 3C,D**), demonstrating that the arbitrary visual cues alone had become sufficient to induce an immediate shift in sequence probabilities. Likewise, all eight birds showed appropriate shifts in sequencing for the first song in the new context compared to the old context (**Fig. 3E,F,** $p < 0.05$ signed rank test, n=8**)**, with average shift magnitude -0.21+/-0.09 and 0.17+/-0.08 (**Fig. 3G,H**), confirming that they can indeed maintain context-specific sequencing shifts for consecutive unreinforced switches without immediate experience

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of white noise. Therefore, visual cues alone were sufficient to act as cues for anticipatory shifts between learned context-dependent syllable sequences.

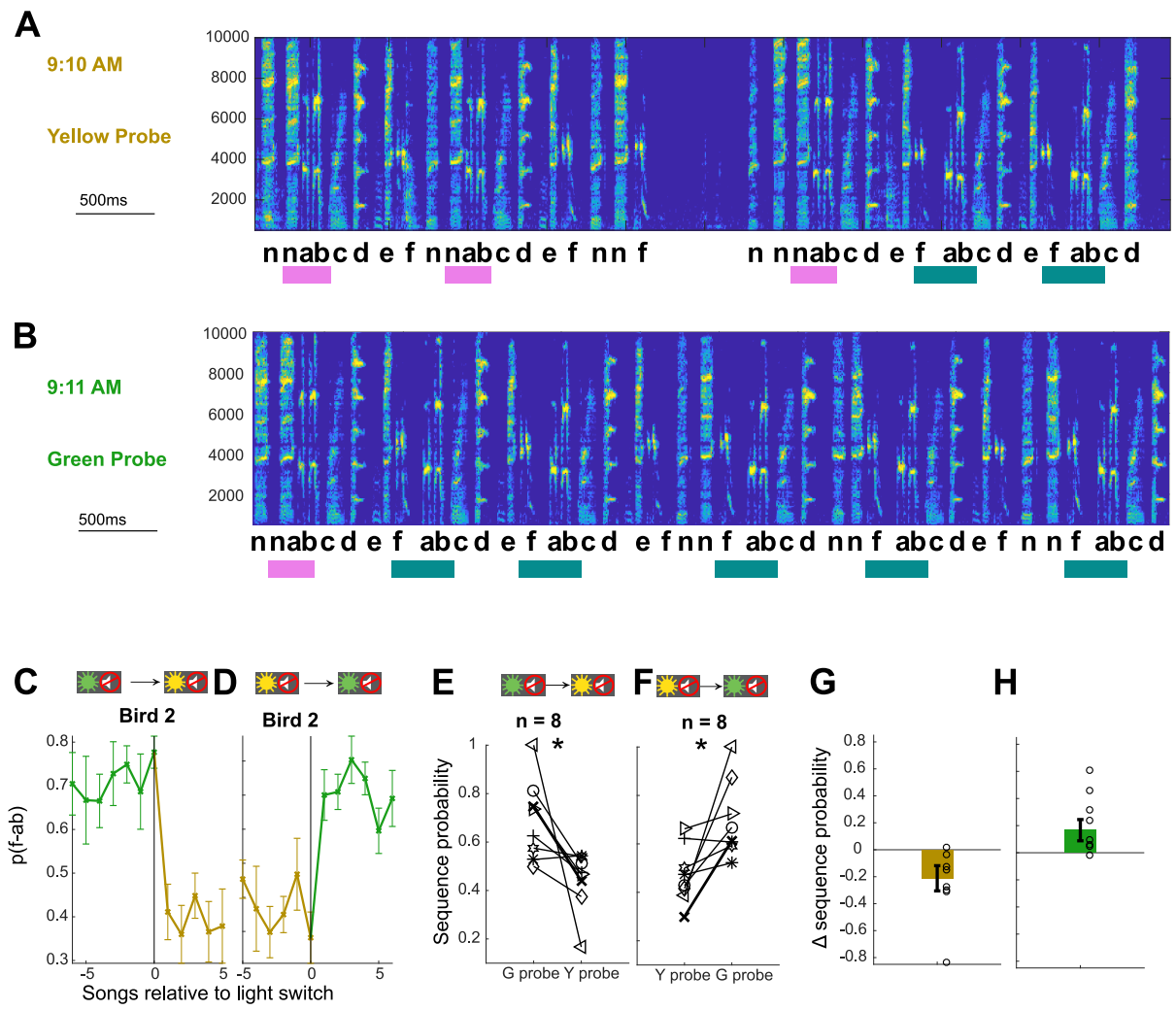

Figure 3. Contextual cues alone are sufficient to enable fast changes to syllable sequencing.
A: Example last song spectrogram in yellow probe context for Bird 2. Scale for x-axis is 500ms, y-axis shows frequency. **B**: Example first song spectrogram in green probe context for Bird 2, sung less than one minute after the song in A. **C,D:** Sequence probability for example Bird 2 over time, switching from green probe to yellow probe context (C) and switching from yellow probe to green probe context (D). Error bars indicate SEM across song bouts (n = 14 switches (C), 11 switches (D)). The display is cut off at the point where less than half of the curves were still contributing data (i.e. in **D**, the median number of song bouts after the switch is 6).
E,F: Sequence probability changes for all eight birds at the transition from the last song in green probe contexts to the first song in yellow probe contexts, and vice versa. Example bird in A-D highlighted in bold. **G,H**: Shift magnitude for direct probe-probe transitions for all birds. Open circles show individual birds, error bars indicate SEM across birds.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Distinct sequence probabilities are specifically associated with different visual cues

The ability to adjust behavior in response to switches between two contexts could potentially depend on a simplified strategy of learning two different motor states and switching between them in response to a detected light switch, independent of the particular colors being switched. In order to test whether birds were using such a strategy or indeed learning specific associations between particular visual cues and particular syllable sequencing states, in a subset of birds we recorded switches between three different color-sequencing pairings. After the conclusion of the other experiments, we introduced blocks of unreinforced context (white light), interleaved with the two previously trained colored contexts. In an example bird (**Fig. 4A),** switching from the unreinforced context elicited changes in opposite directions for the two different light colors. This was true both for switches into WN blocks and into probe blocks (**Fig. 4A**). All birds (n = 3) showed adaptive sequencing changes for the first song bout in probe blocks (**Fig. 4B,C)** as well as fast shifts in the adaptive directions for all color contexts (**Fig. 4D** 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). This indicates that birds can indeed maintain at least two different trained sequencing states separate from a 'neutral' state, and use specific pairings between colored cues and sequencing states to adjust behavior in distinct directions adaptive for each context.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

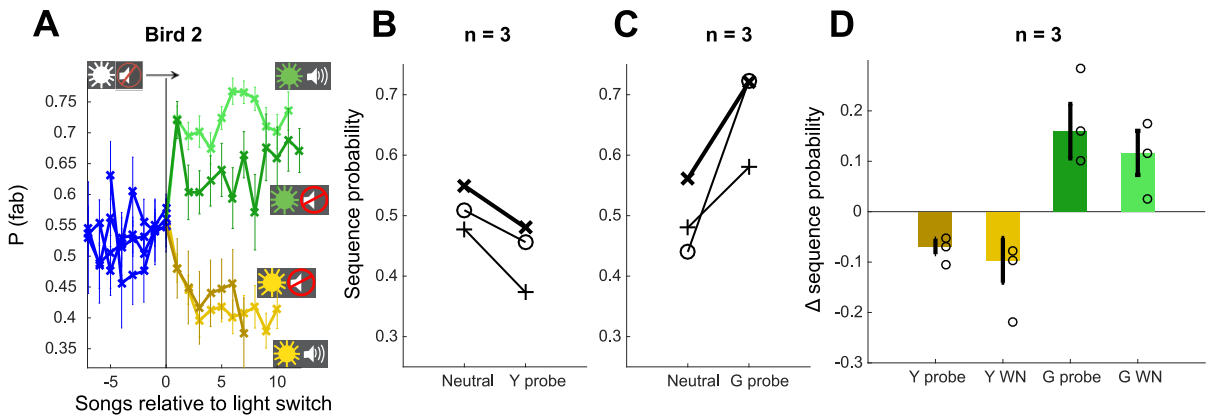

Figure 4. Contextual cues allow shifts in both directions
A: Sequence probability for Bird 2 at the transition from neutral context to yellow and green WN contexts, as well as yellow and green probe contexts. Error bars indicate SEM across song bouts (n = 68 switches (green WN), 78 switches (yellow WN), 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 I highlighted in bold. **D**: Shift magnitude for three birds at the transition 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 in health and disease (Aldridge & Berridge, 2002; X. Jin & Costa, 2015; Tanji, 2001). While the flexibility of communication in human speech is unparalleled, our contextual training paradigm revealed a striking capacity in birds to produce distinct vocal sequences in response to arbitrary cues. Hence, for song, like speech, the ordering of vocal elements can be rapidly reconfigured to achieve 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Flexible control of vocalizations

Our finding that Bengalese finches can learn to exert contextual control over the ordering of vocal elements 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 exert contextual control over 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 particularly sophisticated 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 is striking in its parallel to the human capacity to differentially sequence a fixed set of syllables to create distinct meanings.

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"). This modulation of song structure by social context is characterized by global changes to the intensity of song production, with directed songs exhibiting greater tempo and greater stereotypy of both syllable structure and syllable sequencing than undirected songs (Hedley,

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Denton, & Weiss, 2017; Sakata et al., 2008; Searcy & Beecher, 2009). Moreover, the modulation of song and other simpler vocalizations in natural settings may be largely innate and affective in nature, with influences of conspecific signaling potentially hardwired via hypothalamic and neuromodulatory inputs to premotor regions (Berwick et al., 2011; Gadagkar, Puzerey, & Goldberg, 2019; James, Dai, & Sakata, 2018; Nieder & Mooney, 2020; Wheeler & Fischer, 2012). 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 stimuli that have no prior ethological relevance.

Evolution of control over vocal sequencing

While birds are phylogenetically distant from humans, our results raise the question of whether similar evolutionary forces may have led to a shared capacity for flexible control of vocal sequencing, and whether such forces might have contributed to the eventual development of human speech and language. The capacity for moment-by-moment adjustment of vocalizations may have developed in part to enable the customization of vocal signatures for purposes of individual and group recognition (Vignal, Mathevon, & Mottin, 2004), to enhance communication, or to avoid overlap or 'jamming' during vocal turn-taking and in response to environmental noises (Benichov & Vallentin, 2020; Brumm & Zollinger, 2013). Such ethologically relevant 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. Perhaps evolutionary pressures to develop more nuanced social communication led to an elaboration of cortical control over vocalizations in both songbirds and humans, establishing a conduit that facilitated the integration of progressively more abstract cues and internal states in

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that control. The avian song system provides an ideal opportunity to investigate mechanistically how such top-down pathways orchestrate the flexible control of vocalizations.

Neural implementation of context-dependent motor sequencing

The ability of birds to switch between distinct motor programs using visual cues is reminiscent of contextual motor control in humans, where subjects can learn two separate states of motor adaptation, and switch between them using contextual 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; Hikosaka et al., 1999; Lee & Schweighofer, 2009; McDougle, Ivry, & Taylor, 2016; Wolpert, Diedrichsen, & Flanagan, 2011). One is more flexible, and sensitive to contextual information (McDougle et al., 2016), while a slower process, related to basic motor adaptation, might not be easily associated with arbitrary cues (Howard, Wolpert, & Franklin, 2013). Specifically, in a test of one two-process model, Imamizu and colleagues (Imamizu & Kawato, 2009; Imamizu et al., 2007) found that contextual information can be used for rapid adaptation at the beginning of new blocks, without affecting the rate of adaptation within blocks. In striking parallel to these results in a human motor adaptation study, our analysis of the progression of learning (**Fig. 2 G-L)** reveals two separate processes: fast context-dependent shifts in sequence probability, which increased over the course of training, and gradual adaptation within blocks, which did not change over training. For humans, the slow processes are thought to rely on primary motor structures, while fast context-dependent adaptation has been associated with activation of prefrontal and parietal association areas (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 might be processed in avian structures analogous to mammalian prefrontal cortex outside the song system. For example, the association area

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 drive rapid contextual shifts.

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 for song production include networks of neurons in the vocal premotor nucleus HVC responsible for the temporal control of individual syllables, which 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 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 (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 could potentially convey such visual or cognitive influences on sequencing (Bischof & Engelage, 1985; Cynx, 1990; Seki, Suzuki, Takahasi, & Okanoya, 2008; Ullrich, Norton, & Scharff, 2016; Wild, 1994). One of these inputs, Nif, has previously been shown to be relevant for sequencing (Hosino &

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 HVC (Basista et al., 2014; Zhang et al., 2017) as well as inputs that convey a variety of sensory feedback and contextual information. Our behavioral studies pave a way for investigating how neural circuits enable flexible and adaptive reconfiguration of motor output in response to different cognitive demands. Together with the quantifiable nature of syllable sequencing and well-understood neural circuitry of the bird song system, the ability to elicit moment-by-moment contextual adjustment should provide insight into contextual learning and execution of skilled motor behavior and the neural underpinnings of cognitive flexibility.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

Subjects & Sound recordings

The experiments were carried out on eight adult 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

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 target syllables for each bird based on their unique spectral structure. In the context-dependent auditory feedback protocol, the specific sequence of syllables that received aversive white noise (WN) feedback (the target sequence) 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).

At the start of each experiment, each color of cage illumination was alternatingly turned on for a random duration within a range of several minutes to hours (example bird 6: 30-90 min)

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

without WN playback to gather baseline data on sequence probability in the two contexts, in order to confirm that sequencing changes we observed after training were not just unlearned responses to the light cues. After gathering baseline data, we started WN training. During initial training, each color of cage illumination was alternatingly active and then automatically switched to the other color at a random time within an interval of one to several hours. After several days of training (average 33), we started including probe blocks in the sequence of block transitions, to confirm that sequencing changes could be elicited by visual cues alone. Probe blocks presented the individual light cues without WN. Presentation sequence was set so that approximately one third of blocks were probe blocks (10/34 blocks) and probes made up on average between 7-35% of total time. In additional to the eight birds that completed this training paradigm, three birds were started on training but never progressed to the full probe sequence either because they did not exhibit single-context learning or because of technical issues with consistent targeting at branch points; these birds are excluded from the results. In a subset of experiments, after completing measurements with probe transitions, we added a third, neutral context (**Fig. 4**), 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 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, 2016)), 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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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'). Song bouts that did not contain either of the two target sequences were discarded. Context switches were then 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 the previous day ended. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 transitions 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. 3 L**), we calculated a linear fit to the post-switch parts of the unsmoothed sequence probability curve for each individual context switch.

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.

Lena Veit has been a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation and was supported by a postdoc fellowship from Leopoldina German National Academy of Sciences. Christian J Monroy Hernandez was supported by a HHMI EXROP summer fellowship.

Competing interests:

The authors declare no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

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.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

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.org*. http://doi.org/https://doi.org/10.1101/822817

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

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

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.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

James, L. S., Dai, J. B., & Sakata, J. T. (2018). 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

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

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, A. A. (2016). Comparison of machine learning methods applied to birdsong element classification. *Proceedings of the 15th Python in Science Conference*.

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

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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

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

Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2020). The syntax–semantics interface in animal vocal communication. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20180405. http://doi.org/10.1098/rstb.2018.0405

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.08.05.238717; this version posted August 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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