Neural Circuit Mechanisms Underlying Skill Learning, Adaptation, and Maintenance

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Otchy, Timothy Matthew. 2016. Neural Circuit Mechanisms Underlying Skill Learning, Adaptation, and Maintenance. Doctoral dissertation, Harvard University, Graduate School of Arts &amp; Sciences.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493332">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33493332</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Neural Circuit Mechanisms Underlying Skill Learning, Adaptation, and Maintenance

A dissertation presented

by

Timothy Matthew Otchy

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Neurobiology

Harvard University

Cambridge, Massachusetts

May 2016
Neural Circuit Mechanisms Underlying Skill Learning, Adaptation, and Maintenance

Abstract

Part I

Mastering a motor skill, such as playing the guitar, requires precisely controlling both spatial and temporal aspects of motor output – that is, what movements to perform when. While it is generally assumed that these aspects are acquired through the same learning processes and in the same circuits, there is also evidence that the brain can control them independently. But if that’s true, how is such modularity in motor control and learning implemented in neural circuitry? To probe this question, we developed a paradigm that ‘trains’ songbirds to change either spatial or temporal aspects of their vocal output and showed that learning in the two domains is implemented in distinct neural circuits. This dissociation extended to premotor nucleus HVC, which we showed encodes changes to temporal but not spectral song structure. Such functional modularity, i.e. different circuits learning and implementing different aspects of motor control, could serve to overcome the limitations of reinforcement learning algorithms in dealing with large task domains.

Having identified key mechanisms by which an acquired motor skill can be modified, we then turned to investigate the mechanisms underlying the formation of circuits during the initial acquisition of a motor skill. The neural circuits controlling learned behaviors develop under genetic constraints and in response to environmental influences. Recent studies have provided an unprecedentedly detailed view of the circuit- and synaptic-level changes that accompany complex motor learning, but have left unexplored how environmental factors influence the formation of the neural circuits underlying motor
skills. To address this, we investigated how the lack of a behavioral model affects normal motor circuit development in songbirds, a question with relevance for developmental disorders associated with deficits in imitation. We found that the primary difference in circuit formation was delayed and decreased pruning at a synapse that is a principal locus of learning. We show that this difference in synapse refinement is consistent with it being the principal mechanism driving reduced temporal precision of song and the underlying motor program. Intriguingly, our finding of impaired synapse formation mirrors what has been suggested in previous studies of autism.

Part II
Assigning function to brain areas is a principal aim of neuroscience that is often pursued by rapidly and reversibly manipulating neural activity in behaving animals. An important assumption underlying this experimental regime is that consequent behavioral changes reflect the function of the targeted circuits. In Part II of this dissertation, we demonstrate that this assumption is problematic in that it fails to account for indirect effects on the independent functions of circuits downstream of the targeted area. Transient inactivation of sensorimotor area Nif in songbirds and motor cortex in rats severely disrupts courtship songs and task-specific movement patterns – learned skills that recover spontaneously after permanent lesions of the same areas. How can a brain area be both essential for behavior execution (as assayed by the now preferred method, transient perturbation) and not (as assayed by the traditional method, lesions)? We resolve this seeming paradox in songbirds, showing that sudden silencing of Nif disrupts song and neural dynamics within HVC, a downstream song control nucleus. In parallel with song recovery, the off-target effects resolved within days of lesion, a recovery consistent with homeostatic regulation of neural activity within HVC. These finding have broad implications for how neural circuit manipulations are interpreted and for understanding the mechanisms supporting functional recovery following brain injury.
Table of Contents

Abstract iii
Table of Contents v
Acknowledgements vi

Part I
Chapter 1: Introduction 2
  Motivation 3
  Background 4
  Research outline 8

Chapter 2: The basal ganglia is necessary for learning spectral not temporal features 10
  Abstract 11
  Introduction 12
  Results 15
  Discussion 28
  Methods 31

Chapter 3: Early sensory deprivation is associated with reduced synaptic pruning 40
  Abstract 41
  Introduction 42
  Results 46
  Discussion 54
  Methods 61

Part II
Chapter 4: Off-target effects of neural circuit manipulations 70
  Abstract 71
  Introduction 72
  Results 74
  Discussion 83
  Methods 87

Appendix A: Supplemental Material for Chapter 2 105
Appendix B: Supplemental Material for Chapter 4 119
Bibliography 126
Acknowledgements

This dissertation is the culmination of many years of curiosity, experimentation, and exploration. Though here I report observations and results made in the last few years, none of it would have been possible without the early preparation and groundwork laid by my parents, Daniel and Kathleen. From encouraging me to breed fruit flies in my bedroom in 5th grade to helping me set up a chemistry lab in the basement during high school, they provided an ideal environment for an inquisitive child and have not ceased in their support since. For every setback or hardship I have met, they have been there to aid and push me forward. In many ways, this dissertation is as much a product of their efforts as it is of mine.

I also must thank my advisor and friend, Bence Ólveczky. Since hiring me as a research assistant eight years ago, he has been an endlessly supportive mentor and cheerleader who very much taught me what it is to be a scientist. I am grateful to have had an advisor that expended significant personal effort and resources on my training; beyond that, though, I am incredibly fortunate to have never – in eight years – felt as though he didn’t care or was otherwise uninterested in what I was doing or thinking. His ability to weave together a variety of observations and results in to a crisp and compelling idea has been an inspiration and is a quality I will strive to replicate. I’ll long cherish his loud chest thumping to signal approval, the incessant half-sarcastic-and-half-hopeful “is it awesome?” inquiries, and the timely reminders that failed experiments can often be reframed as serendipitous “controls”. My experience working with Bence has been extraordinary and provided the most intellectually stimulating and challenging experiences of my life. Thank you.

Thank you also to the members of my Dissertation Advisory Committee – Takao Hensch, Daniel Polley, and Maurice Smith – for the great feedback and sound advice that has helped shape my research. Additional thanks are due to Michale Fee, Venkatesh Murthy, and Nao Uchida for your participation as examiners at my defense.
I am indebted to Florian Engert who was a very early and lasting supporter of both my research and me. I am fortunate to have had the opportunity to work with his lab and will be always appreciative for his long-term commitment to studying fictive navigation in the larva zebra fish.

My time in the Ölveczky Lab has been extraordinary not only for the education and training it afforded me, but also for the people with whom I have been lucky to work. Thank you to my collaborators – Farhan Ali, Yoram Burak, Antoniu Fantana, Jonathan Garst-Orozco, Sharon Gobes, Risa Kawai, Cengiz Pehlevan, Juliana Rhee, and Steffen Wolff – both my work and I have benefited immensely from your efforts and insights. Thank you to other members of the Ölveczky Lab who have helped shape my research and thinking through countless informal discussions: Ashesh Dhawale, Adam Kampff, Raymond Ko, Yohsuke Miyamoto, and Rajesh Poddar. A special thanks to my officemates, Ashesh and Steffen, not only for your friendship but also for adjusting remarkably well to my habit of having audible conversations with myself throughout the day.

I have been lucky to have mentored a few graduate and undergraduate students in our lab: Ester Bosch, Alexandre Kempf, and Pranav Krishnan. I am grateful for having had the chance to introduce you to songbird research and hope you learned as much from me as I from you.

In spite of the many challenges of a life in science, these years have been the most fulfilling and rewarding of my life, and for that I must thank my wife, Nicole. We wed a few weeks before I joined the lab, so she has witnessed firsthand the full range of highs and lows in this endeavor. Her support and understanding have always been there when I most needed them. Thank you, Nicole, for your patience with interrupted vacations, your perseverance through endless projects and my “just one more” thinking, and your love of nerdy science people – especially this one.
Part I
Chapter 1:

Introduction
Motivation

Arguably, the purpose of the brain is to produce movement (Shmuelof and Krakauer, 2011; Wolpert et al., 2001). Movement, at base the patterned contraction of muscle fibers evoked by the nervous system, is the only way we have of interacting with the world. Every form of communication – from written word, to spoken language, to expressive facial and hand gestures – is a product of the motor system. Indeed, it has been suggested that in order to understand whole brain function, we ought to consider all sensory and cognitive processes as inputs that ultimately subserve later motor outputs (Churchland, 1989).

Although all animals are endowed with a range of innate motor behaviors – such as sighing (Li et al., 2016), rudimentary locomotive patterns (Thelen et al., 1981), and facial gestures (Eibl-Eibesfeldt, 1973) – the vast majority of our motor repertoire must be learned. Learning involves changes in behavior that arise from interaction with the environment and is distinct from maturation, which involves changes that occur independent of such interaction (Wolpert et al., 2011). Though some simple species show little to no motor learning, the need for motor learning arises in species in which the organism’s environment, body, or task are subject to rapid or unpredictable changes and so control strategies cannot be pre-specified in the neural substrates (Shadmehr et al., 2010). Thus, learning offers animals enormous flexibility to, for example, explore previously unexploited parts of the environment or use new tools and strategies.

Part I of this dissertation probes changes in brain activity and connectivity that are associated with motor learning and addresses two primary questions. First, can motor skills be dissociated into distinct spatial and temporal domains and if so, does this dissociation extend to its implementation in the underlying neural circuits? Second, how does instructive experience influence the formation of motor learning circuits and can we distinguish this from a background of learning-independent circuit maturation? Though there are several popular model systems in which to study motor learning, the
zebra finch, a songbird, offers many unique advantages, including the experimental tractability of its song and the underlying motor circuits (Brainard and Doupe, 2002; Fee et al., 2004; Ölveczky and Gardner, 2011; Tchernichovski et al., 2001), the sensitivity of the behavior to instructive experience (Eales, 1985; Price, 1979; Tchernichovski et al., 2001), and the similarity of the learning process and associated neural circuits to those of human motor skill acquisition (Brainard and Doupe, 2002; Doupe and Kuhl, 1999; Shmuelof and Krakauer, 2011). In the remainder of this chapter, we introduce the zebra finch and its courtship song, briefly review the physiology of the neural circuits generating it, and outline the main chapters in Part I.

Background

*Zebra finches and their song*

The zebra finch, *Taeniopygia guttata*, is a common estrildine finch native to Central Australia that has been adopted widely as a model system in neuroscience (Zann, 1996). Though many aspects of zebra finch biology and behavior have been investigated, some of the more unique and profound discoveries have focused on the learning and production of vocalizations (Brainard and Doupe, 2002; Doupe and Kuhl, 1999). In zebra finches, singing is a sexually dimorphic behavior that holds special significance for the male, serving as both a means of communicating identity as well as a courtship display (Williams, 2004).

Zebra finch song has a complex, multi-scale structure that is both unique to the individual bird yet highly stereotyped from performance to performance. Each song bout consists of multiple repetitions of a canonical motif that is typically 300-1000 ms in duration (Williams, 2004). The motif is comprised of 2-10 discrete and unique vocal elements, called ‘syllables’, that are themselves tens to hundreds of milliseconds in duration, contain complex spectro-temporal structure, and are separated from each other by brief (10-100 ms) periods of silence (‘gaps’) (Price, 1979; Williams, 2004).
Importantly, each syllable in a motif – both its spectral structure and position in the sequence – is highly stereotyped from rendition to rendition in the adult (Fee et al., 2004; Immelmann, 1969; Ravbar et al., 2012; Tchernichovski et al., 2004; Zann, 1996).

Though the male finch is innately driven to sing, the unique structure of his adult song is not entirely defined by genes and developmental programs (Eales, 1985, 1987; Fehér et al., 2009; Gardner et al., 2005; Ölveczky and Gardner, 2011; Price, 1979). Like many skilled behaviors that are core components of human life, zebra finch courtship song must be learned through imitation of an instructive model provided by a learned conspecific (Eales, 1985; Immelmann, 1969; Tchernichovski et al., 2001). This learning process takes place in two overlapping phases: juvenile male zebra finches listen to and memorize the song of a tutor during the sensory learning phase (~25-60 days post hatch, dph) and then engage in vocal practice to emulate this memorized song model during a prolonged phase of sensorimotor learning (~45-90dph). Over the course of the sensorimotor learning phase, which may involve tens of thousands of song renditions over the course of many weeks, the juvenile uses auditory feedback to match its vocalizations to the memorized song model (Konishi, 1965, 2004; Marler and Sherman, 1983; Tschida and Mooney, 2012b). Sensorimotor learning ends at the onset of sexual maturity (~90dph) with crystallization, wherein the song becomes highly stereotyped and less sensitive to modification (Bottjer and Arnold, 1997; Lombardino and Nottebohm, 2000; Tchernichovski et al., 2001; Woolley and Rubel, 1997; Zann, 1996). Nevertheless even after crystallization, learned song structure slowly deteriorates following deafening (Nordeen and Nordeen, 1992; Woolley and Rubel, 1997) or other disruption of auditory feedback (Konishi, 2004; Leonardo and Konishi, 1999; Zevin et al., 2004), indicating an on-going role for auditory feedback in the maintenance of song as well as a continuing capacity for motor plasticity.

Despite the profound influence the tutor song has on shaping the structure of the tutee’s song, studies of songbirds acoustically isolated from tutors during the sensory learning period (i.e., ‘isolates’)
have revealed the extent to which innate, experience-independent mechanisms can specify a typically learned behavior (Eales, 1985; Fehér et al., 2009; Gardner et al., 2005; Price, 1979; Whaling et al., 1997). Indeed even in the absence of a song model to imitate, adult isolate songbirds can produce stable, species-typically structured song motifs that can be effective in courtship, territory defense, and tutoring offspring (Fehér et al., 2009; Williams et al., 1993). In comparison to those of tutored birds, isolate song syllables are on average of longer duration and lower intra-syllabic frequency modulation, though the magnitude of these differences varies widely between animals (Fehér et al., 2009). Interestingly, the degree to which the songs of isolate and tutored birds diverge is highly species dependent, even among birds with similar song learning behavior (Gardner et al., 2005; Méndez et al., 2010; Ölveczky and Gardner, 2011; Price, 1979; Williams, 2004).

The neural mechanisms of song production and learning

The songbird’s brain contains a set of dedicated, sexually dimorphic and highly specialized nuclei (i.e., the song system) that undergo changes in their size, connectivity and activity throughout the song learning process (Nottebohm et al., 1982; Wild, 1997). At the time of hatching, most song system nuclei have not yet formed and must undergo dramatic neural development concurrent with the process of song learning. The song system is comprised of two major parts: the Motor Pathway (MP), which extends from the telencephalon to the brainstem vocal-respiratory motor neurons, and the Anterior Forebrain Pathway (AFP) that projects from the telencephalon to pallial and thalamic regions (Mooney, 2009a; Nottebohm et al., 1982). These two pathways serve complementary functions: the MP generates the motor commands for adult stereotyped song performance; the AFP underlies vocal exploration and exploitation required for reinforcement-based learning of song (Ali et al., 2013; Aronov et al., 2008; Doya and Sejnowski, 1995; Fee and Goldberg, 2011; Kao et al., 2005; Ölveczky et al., 2005; Warren et al., 2011).
An important part of the MP, and the hypothesized nexus for learning in this system, is the robust nucleus of the arcopallium (RA) (Garst-Orozco et al., 2015; Johnson et al., 1997; Kittelberger and Mooney, 2005; Ölveczky et al., 2011), a nucleus with structural and functional homologies to layer 5 of primary motor cortex (Jarvis et al., 2005) that projects topographically to primary motor neurons in the brainstem (Wild, 1997). During adult singing, RA neurons fire patterned bursts of action potentials which are reproduced precisely each time the bird sings a motif (Leonardo and Fee, 2005; Yu and Margoliash, 1996). This sequence of bursts in turn drives activity in downstream motor neurons and muscles producing the song (Fee et al., 2004; Wild, 1993; Wild et al., 2000). A major input to RA comes from the premotor cortical nucleus HVC (used as a proper name) (Mooney and Rao, 1994; Nottebohm et al., 1982; Vu et al., 1994), which in turn receives input from several smaller telencephalic, pallial, thalamic, and midbrain nuclei (Akutagawa and Konishi, 2010; Hamaguchi and Mooney, 2012; Nottebohm et al., 1982; Scharff and Nottebohm, 1991).

In addition to this direct projection, HVC also makes an indirect connection to RA via the AFP. A population of efferent HVC neurons, distinct from those projecting to RA, synapse onto the pallidal neurons of Area X (Goldberg and Fee, 2010; Kojima and Doupe, 2008; Person et al., 2008), the primary singing-related region of the songbird basal ganglia. This pathway – which includes stops at the dorsolateral division of the medial thalamus (DLM) (Goldberg and Fee, 2012; Goldberg et al., 2013; Nottebohm et al., 1982; Person et al., 2008), the lateral magnocellular nucleus of the anterior nidopallium (LMAN) (Kao et al., 2005; Ölveczky et al., 2005; Vates and Nottebohm, 1995), and ultimately back to RA (Nottebohm et al., 1982) – bares a strong physiological and functional similarity to the basal ganglia-thalamo-cortical pathways of the mammalian motor system known essential for learning (Jarvis et al., 2005; Person et al., 2008) and suggests common mechanisms of motor learning across species (Brainard and Doupe, 2002; Doupe and Kuhl, 1999). The projection back to the MP at the level of RA is important as it allows the AFP to influence the premotor activity driving song (Charlesworth et al., 2012;
Doya and Sejnowski, 1995; Kao et al., 2005, 2008; Ölveczky et al., 2005), though the character and mechanism of that influence remain to be clearly identified (Charlesworth et al., 2012; Goldberg and Fee, 2010; Kao, 2006; Kojima et al., 2013; Mooney, 2009b).

The architecture of the song system suggests an intuitive model for song production: each HVC neuron that projects to RA generates a single highly precise burst at one specific time in the song motif (Hahnloser et al., 2002; Li and Greenside, 2006; Yu and Margoliash, 1996), and it has been suggested that these drive, at each moment, the population of RA neurons active at that time (Fee et al., 2004; Fiete et al., 2007). In this model, HVC can be thought of as a clock that marches through time in the song sequence (Long et al., 2010), and at each moment a stereotyped pattern of activity in HVC drives a stereotyped pattern of RA neurons, which in turn drive a similarly stereotyped sequence of muscle activity patterns. This suggests that learning the ‘correct’ motor pattern to produce a given song is simply a matter of wiring up HVC to activate the appropriate subset of RA neurons at each moment in time; that is, song learning principally occurs via plasticity in the HVC-RA synapses (Garst-Orozco et al., 2015; Herrmann and Arnold, 1991; Johnson et al., 1997; Kittelberger and Mooney, 2005; Mehaffey and Doupe, 2015).

Research outline

Part I of this dissertation contains two chapters, each describing a separate series of experiments.

Chapter 2 describes experiments aimed at answering a question introduced at the beginning of this chapter, namely: can motor skills be dissociated into distinct spatial and temporal domains and if so, does this dissociation extend to how learning is implemented in the underlying neural circuitry? To probe this question, we developed a conditional auditory feedback paradigm that ‘trains’ songbirds to change either spatial or temporal aspects of their vocal output in a controlled and predictable manner. We found that the pitch and duration of a song ‘syllable’ can be modified independently of each other
and without interference. Importantly, we showed that learning in the two domains is implemented in distinct neural circuits, one basal ganglia dependent, the other not. This dissociation extended to premotor nucleus HVC, which we showed encodes changes to temporal but not spectral song structure.

Chapter 3 takes up the question: how does instructive experience influence the formation of motor learning circuits and can we distinguish this against a background of learning-independent circuit maturation? To address this, we made both extracellular and whole-cell recordings from RA projection neurons in both tutored and isolate zebra finches throughout song development. By comparing the singing-related activity and HVC-input profiles in age-matched birds from each tutoring paradigm, we were able to identify changes in circuit formation – a reduction and delay in synapse pruning – consistent with it being a important mechanism underlying the elevated variability of isolate song. In addition, our results provide key insight into circuit-level effects caused by a failure of synaptic pruning, a pathology associated with a variety of developmental disorders.
Chapter 2:

The basal ganglia is necessary for learning spectral, but not temporal features of birdsong

This chapter was previously published as:


Attribution: This chapter contains collaborative work.

BPÖ, FA and TMO designed the study. FA developed the software for CAF, and performed all behavioral experiments. TMO collected the neural data. CP, YB, FA, TMO, and ALF developed the computational tools for song analysis. FA, CP, TMO, and BPÖ analyzed the data. BPÖ, FA, and TMO wrote the paper with input from the other co-authors.
Abstract

Executing a motor skill requires the brain to control which muscles to activate at what times. How these aspects of control - motor implementation and timing - are acquired, and whether the learning processes underlying them differ, is not well understood. To address this we used a reinforcement learning paradigm to independently manipulate both spectral and temporal features of birdsong, a complex learned motor sequence, while recording and perturbing activity in underlying circuits. Our results uncovered a striking dissociation in how neural circuits underlie learning in the two domains. The basal ganglia was required for modifying spectral, but not temporal structure. This functional dissociation extended to the descending motor pathway, where recordings from a premotor cortex analogue nucleus reflected changes to temporal, but not spectral structure. Our results reveal a strategy in which the nervous system employs different and largely independent circuits to learn distinct aspects of a motor skill.
Introduction

To master a motor skill, both its timing and specific motor implementation must be learned and adaptively refined. Increasing the power of your tennis serve, for example, might mean speeding up certain parts of the service motion (modifying timing), while adding top-spin might require changing the angle of your elbow (modifying motor implementation). Both improvements will require changes to the motor program underlying your serve, but the nature of these changes can be construed as different. Modifying timing equates to changing the temporal progression of the muscle activity patterns to slow down or speed-up certain parts of the action, whereas changing motor implementation means modifying specific muscle commands while maintaining the temporal dynamics of the action (Figures 2.1A-2.1C). Whether this conceptual distinction reflects a dissociation in how the motor system learns and refines motor skills has not been explored.

The zebra finch, a songbird, provides a unique model system for addressing this question. Through a process that resembles human speech learning (Doupe and Kuhl, 1999), juvenile zebra finches gradually improve both temporal (Glaze and Troyer, 2012a; Lipkind and Tchernichovski, 2011) and spectral (Tchernichovski et al., 2001) aspects of their songs (Figures 2.1D-2.1F) until they resemble those of their tutors (Immelmann, 1969). Spectral features of song are largely determined by the activity of vocal muscles (Goller and Suthers, 1996), and thus serve as a proxy for ‘motor implementation’.

The neural circuit architecture underlying song production is well delineated (Figure 2.1G) and suggests a hierarchical organization (Yu and Margoliash, 1996) with a descending motor cortical pathway that encompasses premotor nucleus HVC (proper name) (Vu et al., 1994) and motor cortex analogue robust nucleus of the arcopallium (RA) (Nottebohm et al., 1982). RA projection neurons synapse onto brainstem motor neurons involved in singing (Wild, 1993). HVC and RA are also indirectly connected through the Anterior Forebrain Pathway (AFP), a basal ganglia-thalamo-cortical circuit that is critical for song learning, but not essential for producing learned song (Figure 2.1G) (Bottjer and Arnold,
A separate basal ganglia circuit, medial to the AFP, receives input from and provides output to HVC (Foster et al., 1997; Hara et al., 2007; Williams et al., 2012) (Figure 2.1).
but the role of this circuit in song learning, if any, remains to be elucidated (Foster and Bottjer, 2001).

The analogies and homologies between the AFP and basal ganglia circuits in mammals (Farries and Perkel, 2002; Reiner et al., 2004) have made the songbird a tractable model for exploring how the basal ganglia (used as singular noun, as we refer to it as a functional entity) contributes to motor learning (Doupe et al., 2005; Fee and Goldberg, 2011). Recent models have the AFP implement aspects of a reinforcement learning process that shapes connectivity in motor cortex analogue RA (Doya and Sejnowski, 1995; Fee and Goldberg, 2011; Fiete et al., 2007; Troyer and Doupe, 2000). Besides being the direct target of the AFP, the focus on RA as the nexus for song learning is also motivated by the finding that neurons in premotor nucleus HVC that project to RA encode time in the song (Hahnloser et al., 2002). This ‘clock code’ in HVC has been hypothesized to provide a stable temporal input to RA during learning and production of song (Fee et al., 2004; Fee and Goldberg, 2011).

Given the functional organization of the song circuit (Figure 2.1H), learning can be understood as the process of establishing and refining connections between time-keeper neurons in HVC and muscle-related neurons in RA, and further between RA collaterals (Sizemore and Perkel, 2011), such that the ‘right’ muscles get activated at the appropriate times (Fee and Goldberg, 2011; Fee et al., 2004; Fiete et al., 2007). The AFP is thought to contribute to this process by inducing variability in RA neurons and thus song (Kao et al., 2005; Ölveczky et al., 2005, 2011), and by providing an instructive signal that biases the motor program towards improved performance (Andalman and Fee, 2009; Charlesworth et al., 2012; Fee and Goldberg, 2011; Warren et al., 2011).

While this framework for song learning, i.e. plasticity in RA, can plausibly account for both temporal and spectral changes in song (Figure A.1A), the extent to which other circuits are involved, and whether motor cortical and basal ganglia circuits distinguish learning in the temporal and spectral domains, has not been explored. To address this we developed a reinforcement learning paradigm to
independently modify both temporal and spectral features of zebra finch song. We perturb activity in different parts of the AFP, including its basal ganglia component Area X (Person et al., 2008) and cortical output (LMAN - lateral magnocellular nucleus of the anterior nidopallium), and quantify how these circuit manipulations affect the capacity for learning temporal and spectral aspects of song. To probe whether the descending motor pathway encodes learned changes in the two domains differently, we record from neurons in HVC during modification to both temporal and spectral structure.

**Results**

*Independent modification of temporal and spectral song structure*

Testing whether the song system (Figures 2.1G and 2.1H) differentiates between learning in the temporal and spectral domains requires experimentally modifying both aspects of song. A paradigm in which disruptive auditory feedback is delivered to the bird contingent on the pitch of one of its syllables has been effective in adaptively altering spectral structure of song (pitch-Conditional Auditory Feedback – pCAF) (Tumer and Brainard, 2007). To probe whether temporal structure of adult zebra finch song is similarly plastic, we adapted this method to the temporal domain. This involved delivering loud aversive noise bursts every time the duration of a targeted song segment was below (to lengthen) or above (to shorten) a given threshold value (timing-Conditional Auditory Feedback – tCAF, see Methods and Figure 2.2A). To get precise and reliable on-line estimates of target duration we targeted segments bounded by large and abrupt changes in sound amplitude, which in practice mostly meant intervals between ensuing syllable starts, i.e. ‘syllable + gap’ segments (see Figure 2.2A and Methods).

This paradigm induced rapid and predictable changes in the duration of targeted segments (Figures 2.2B-2.2D), demonstrating a remarkable capacity for changing the temporal structure of zebra finch song even well past song crystallization. Across the population of birds (n=24), the duration of targeted segments changed by, on average, 3.4±1.7 ms/day (mean±SD) across 4-10 days of tCAF (Figure
2.2D; range: 0.9-6.4 ms/day, p=1.8x10^-5). Changes to temporal structure were specific to the targeted segments (Figure 2.2D), with minimal changes to the duration of non-targeted elements (-0.21±0.43 ms/day). When targeting ‘syllable + gap’ segments, both syllables and gaps changed in duration (syllables: 0.7±0.6 ms/day, p=4.6x10^-5; gaps: 2.8±1.6, p=7.7x10^-8, Figures A.2 and A.3C), though gaps changed significantly more than syllables (p=1.3x10^-5). This difference was largely explained by the reinforcer being further removed in time from the syllables (by on average 47.2±13.6 ms). When we experimentally delayed the noise burst by 50 ms relative to the end of the gap, the rate at which gaps changed decreased dramatically (79.7±4.1%, n=3 birds, Figures A.2C and A.2D). The effect was
consistent with the difference in syllable and gap learning rate in our experiments being due to the
differential delay in reinforcement (Figure A.2E), though contribution from other factors cannot be
discounted (Glaze and Troyer, 2012b).

Learning was restricted not only in time, but also to the feature being targeted. Changing the
duration of a syllable did not alter its pitch (Figure 2.2D; pitch change during tCAF=0.2±2.6 Hz/day,
p=0.72). Similarly, modifying the pitch of a syllable using pCAF (Andalman and Fee, 2009; Warren et al.,
2011)(Figure 2.2E; 22.6±16.2 Hz/day; range: 7.3-62.8 Hz/day, n=14 birds, p=1.60x10^{-4}) did not affect its
duration (Figures 2.2C and 2.2E; duration change during pCAF=0.05±0.43 ms/day, p=0.65), suggesting
that the two features, duration and pitch, may be independently learned and controlled (Figure A.3).
Having a method (CAF) for inducing rapid and reproducible changes to both spectral and temporal
aspects of song allowed us to address the neural underpinnings of learning in the two domains and
gauge the extent to which they are indeed distinct.

\textit{Dissecting the role of the AFP, a basal ganglia-thalamo-cortical circuit}

In our paradigm, adaptive changes to both pitch and duration rely on differential reinforcement of
variable actions, and as such are examples of reinforcement learning (Sutton and Barto, 1998). In the
context of motor learning, this process requires two main ingredients: (1) motor variability producing
exploratory actions and (2) a process converting information from this exploration into improved motor
performance. LMAN, the output of the AFP, has been implicated in both aspects. Activity in this nucleus
induces variability in vocal output (Kao et al., 2005; Ölveczky et al., 2005) and, in the spectral domain at
least, drives an error-correcting premotor bias through its action on RA (Andalman and Fee, 2009;
Charlesworth et al., 2012; Warren et al., 2011).

While LMAN has been a convenient proxy for understanding the role of the song-specialized
basal ganglia-thalamo-cortical circuit (AFP), questions of how the basal ganglia itself (Area X) contributes
to song learning (Kojima et al., 2013; Scharff and Nottebohm, 1991), and whether its role - and the role of LMAN - differs for learning in the temporal and spectral domains, have yet to be explored. To address this we lesioned Area X and LMAN in separate experiments and compared variability and learning rates in the spectral and temporal domains before and after lesions.

**Area X is required for learning in the spectral, but not temporal domain**

Bilateral lesions of Area X (Figure 2.3A, Tables A.1 and A.2, and Figure A.5A) revealed a striking dissociation as to its role in learning. In the spectral domain (pCAF), learning was largely abolished following lesions (Figures 2.3B and 2.3E; pitch change $4.52\pm4.05$ Hz/day vs. $32.42 \pm18.97$ Hz/day before lesions, $n=6$ birds; $p=2.03\times10^{-5}$). In fact, pCAF-induced changes to pitch after Area X lesions were not significantly different from normal baseline drift (Figure 2.3E; $p=0.48$). In contrast, the capacity for
modifying temporal structure remained unchanged. Average learning rates in tCAF experiments before and after lesions were indistinguishable, with daily changes to target duration of 3.90±2.03 ms before vs. 3.30±1.72 ms after lesion (Figures 2.3C and 2.3F, p=0.63; n=7 birds, 5 of which were also tested in pCAF).

Though a small (~22%) decrease in the coefficient of variation (CV) of pitch was seen in the first few days after Area X lesions (Figure A.4) (Kojima et al., 2013), the effect was transient. Variability in both temporal and spectral features was unchanged from pre-lesion levels when measured 6±2.5 days post-lesion (range: 3-12 days), consistent with previous studies (Goldberg and Fee, 2011; Scharff and Nottebohm, 1991). The coefficient of variation (CV) in the duration of syllables and inter-syllable gaps (Glaze and Troyer, 2012a) was 2.9±0.9% and 2.8±0.6% before and after lesions respectively (Figure 2.3D; n=9 birds, p=0.89), whereas the CV of pitch was 1.9±1.3% and 1.9±1.5% (Figure 2.3D; n=9 birds, p=0.79). This suggests that Area X is instrumental for learning spectral features not because it produces variability in this domain, but because it is required for generating the instructive signal expressed at the level of LMAN (Fee and Goldberg, 2011).

*Reduction in LMAN activity reveals an error-correcting motor bias in the spectral, but not temporal, domain*

In pCAF experiments, the learning-related instructive signal produced by the AFP manifests as an LMAN-dependent motor bias that shifts the pitch in the direction of learning (Andalman and Fee, 2009; Charlesworth et al., 2012; Warren et al., 2011). This bias can be estimated from the reversion in learned changes upon silencing of LMAN. If, however, learning temporal structure does not require the AFP, as our Area X lesion experiments suggest, then LMAN should also not contribute an error-correcting bias in this domain. To test this, we exposed our experimental subjects to female birds (see Methods), a social manipulation known to dramatically reduce the variability and rate of LMAN firing (Kao et al., 2008) and
thus decrease song variability in a way that mirrors the effect of pharmacological inactivations or lesions of LMAN (Kao et al., 2005; Ölveczky et al., 2005). Suppressing LMAN activity this way after 4-7 hours of pCAF exposure resulted in a 40.1±20.3% mean reversion of that day's learned pitch changes (Figures 2.4A and 2.4B; n=11 birds, 22 experiments, p=6.5x10⁻⁵), an effect very similar to what is seen after LMAN inactivations (Andalman and Fee, 2009; Warren et al., 2011). This reversion was seen both when the pitch was driven away from baseline (reversion towards baseline, 49.1±41.3%) or towards it (reversion away from baseline, 35.2±17.9%). After tCAF however, there was no significant reversion in learned duration changes, consistent with LMAN not contributing an instructive bias in the temporal domain.

Figure 2.4: Distinct roles for LMAN in adaptive modification of temporal and spectral structure.
(A) Female-directed singing, which reduces LMAN activity, caused a reversion in learned changes to pitch, but not interval duration. Example data from the same bird shows the average pitch and duration of the target before and after a day of pCAF (solid blue line) and tCAF (solid red line) respectively; last data point shows the corresponding values after presentation of a female (dashed lines). (B) Average reversion of the day’s learned changes upon presentation of a female bird (i.e. the directed singing-induced change in pitch or duration relative to the day’s total change; n=11 birds for pCAF, n=5 birds for tCAF). Duration values corrected for global tempo changes observed during directed singing (Stepanek and Doupe, 2010) (see Methods). (C) Schematic of the song circuit following LMAN lesions. Grey denotes disrupted pathways/circuits. (D-E) Learning rates in pCAF (n=4 birds) (D) and tCAF (n=8 birds) (E) before and after LMAN lesions. (F) Effects of LMAN lesions on variability in pitch and interval duration.
(Figures 2.4A and 2.4B; n=5 birds, 12 experiments, 10.0±11.2% reversion of the day’s learned duration change, p=0.12; see Methods).

**LMAN lesions affect variability and learning in the temporal domain**

If the AFP is not guiding adaptive changes to temporal structure, we reasoned that the capacity for learning in this domain should be robust to LMAN lesions. To test this, we ablated LMAN bilaterally in a separate group of birds (Figures 2.4C and A.5B, Tables A.1 and A.2). A prior study, using pharmacological inactivation of LMAN in the context of pCAF (Charlesworth et al., 2012), had shown that LMAN is necessary for adaptively modifying pitch in pCAF. We confirmed this result in LMAN lesioned birds, with learning rates in pCAF going from 13.3±5.9 Hz/day before lesions to 0.7±1.1 Hz/day after lesions (p=6.7x10^-4; n=4 birds, 3 of which also tested for tCAF; p=0.11 when comparing LMAN lesioned birds in pCAF to normal drift, Figure 2.4D). In the temporal domain, however, LMAN lesioned birds retained the ability to learn, albeit at a reduced rate compared to pre-lesion (Figure 2.4E, pre-lesion: 2.8±1.6 ms/day, post-lesion: 0.9±0.6 ms/day, p=0.003 when comparing LMAN lesioned birds in tCAF to normal drift). Mean reduction in the learning rate within a bird was 60.7±29.4% (n=8 birds, p=6.3x10^-4).

Since LMAN is known to induce vocal exploration in both the temporal and spectral domains (Thompson et al., 2011), we wondered whether the decreased learning rates in tCAF following lesions could be explained by a reduction in temporal variability. Consistent with this, we found that variability in the duration of song elements (CV of syllable and inter-syllable gaps (Glaze and Troyer, 2012a), see Methods) decreased within a bird by, on average, 38%, from 3.3±1.2 to 2.1±1.1% (Figure 2.4F, p=4.8x10^-4). These results suggest that LMAN contributes to temporal learning by inducing variability in song timing. The process of converting information derived from this variability into improved motor timing, however, is likely implemented outside the AFP, as this process does not require an intact Area X or LMAN.
A basal ganglia-thalamo-cortical circuit parallel to the AFP with projections to HVC is not required for learning temporal structure.

(A) Schematic outline of the basal ganglia loop originating from and projecting back to HVC (brown). mArea X is a basal ganglia region medial to Area X; DMP is the dorsomedial nucleus of the posterior thalamus; MMAN is the medial magnocellular nucleus of the anterior nidopallium. Faint lines denote pathways/circuit disrupted by lesions to MMAN. (B-C) Effect of MMAN lesions on learning rates in tCAF (B) and spectral and temporal variability of song (C) across 3 birds.

A basal ganglia circuit projecting to HVC is not required for learning temporal song structure

Given the architecture of the song circuit, and the assumed role of the basal ganglia in reinforcement learning, a candidate for driving temporal learning is the only other known song-related basal ganglia-thalamo-cortical circuit – a parallel circuit to the AFP that includes a basal ganglia-related structure medial to traditionally defined Area X (mArea X) (Kubikova et al., 2007), (Figure 2.5A). Whereas the AFP projects directly to RA, which encodes spectral features (Sober et al., 2008), MMAN outputs directly to HVC, and could, in analogy to its lateral counterpart (LMAN), provide the instructive signal for altering neural dynamics in HVC and thus temporal structure of song.

To test this, we lesioned MMAN bilaterally (Tables A.1 and A.2, and Figure A.5C), comparing learning rates in our tCAF paradigm before and after lesions. We saw no significant change in the capacity of birds to shift the duration of targeted song segments after MMAN lesions (Figure 2.5B; pre-lesion: 3.4±1.8 ms/day, post-lesion: 3.0±1.3 ms/day, n=3 birds, p=0.34). Neither did MMAN lesions influence variability (CV) in temporal (pre-lesion: 2.6±0.6%, post-lesion: 2.6±0.4%) or spectral (pre-
lesion: 3.0±0.9%, post-lesion: 2.8±0.5%) features of song (Figure 2.5C; p=0.94 and 0.53 respectively), leaving its role in song learning, if any, to be elucidated.

**Song recovery in the temporal domain does not require Area X**

While the CAF paradigm allows us to address how the song system implements reinforcement learning in the spectral and temporal domains, the extent to which the same circuits and neural processes underlie ‘normal’ song learning is unclear. Song learning is thought to be driven by an evaluation of the bird’s vocalizations relative to an auditory template acquired from listening to a tutor early in life (Konishi, 2010). This auditory feedback-dependent learning process maintains stable adult song and restores it after experimental manipulations drive it away from the presumed template (Leonardo and Konishi, 1999; Sober and Brainard, 2009). Such a song recovery process would be expected to interact with CAF-based learning, working against it when the targeted feature (duration or pitch) is driven away from baseline (‘a’ in Figure 2.6A), and in conjunction with it when driven toward it (‘b’ in Figure 2.6A). Consistent with this, learning rates in our CAF experiments were significantly higher when the targeted feature was driven toward baseline than when it was driven away (Figure 2.6B, pCAF: 42.0±25.1 vs. 26.0±15.4 Hz/day, p=0.03; Figure 2.6C, tCAF: 5.4±2.2 vs. 3.4±1.9 ms/day respectively, p=0.04). To compare learning rates in the CAF paradigm to ‘normal’ (i.e. CAF-free) song recovery, we drove the pitch or duration of targeted segments away from baseline by exposing birds to 3-5 days of CAF, and then measured the rate at which the feature returned with and without CAF (Warren et al., 2011). Though both pitch and duration returned towards baseline, the rate of return was much lower without CAF (Figure 2.6B, 12.1±12.4 Hz/day after cessation of pCAF; Figure 2.6C, 0.8±0.3 ms/day after cessation of tCAF).

To test whether the dissociation in basal ganglia function uncovered with the CAF-paradigm (Figure 2.3) extends also to normal song learning, we lesioned Area X in a subset of birds, and compared
spontaneous (i.e. CAF-free) returns toward baseline before and after lesions. Because birds could not predictably alter the spectral structure of their vocal output (pCAF) after Area X lesions (Figures 2.3B and 2.3E), we drove targeted syllables away from their baseline pitch for 4 days (average drive away from baseline: 100.2±76.0 Hz, n=3 birds) before lesioning Area X bilaterally. Consistent with our CAF experiments, we saw no significant return to baseline even after 7 post-lesion days of singing (Figure 2.6D, p=0.17). The spontaneous change in pitch went from 15.3±13.3 Hz/day before lesion to 1.6±1.4 Hz after, suggesting that Area X is required for maintaining the spectral identity of song (Kojima et al., 2013). In the temporal domain, however, lesioning Area X did not affect the spontaneous recovery toward baseline (0.63±0.13 ms before lesion vs. 0.71±0.71 ms after lesion, n=4 birds, p=0.76). Area X lesions also did not affect the difference in learning rates for tCAF drives towards and away from baseline (‘b-a’ in Figure 2.6D, p=0.57), a difference we hypothesize being due, in part at least, to the template-based learning process working in opposite directions in the two cases. These results suggest that the dissociation in how the basal ganglia contributes to learning in the spectral and temporal domains extends to normal CAF-free song learning.
Premotor cortical region HVC encodes changes to temporal, but not spectral, structure

Given the difference in how the AFP contributes to learning in the temporal and spectral domains, we wondered whether learning-related changes in the motor pathway show a similar dissociation. While changes to both temporal and spectral structure can be understood within the existing framework for song learning (i.e. plasticity in RA), significant modifications to the duration of song segments, like those induced by our tCAF paradigm, would require an extensive reorganization of HVC-RA connectivity (Figure A.1A). An alternative, which confers more flexibility on the learning process by capitalizing on the functional organization of the song control circuits (Figure 2.1H), would be for temporal changes to be encoded at the level of HVC (Figure A.1B). Though white-noise feedback does not acutely affect song-related HVC activity (Kozhevnikov and Fee, 2007), we speculated that chronic exposure to the tCAF protocol could alter its dynamics to reflect adaptive changes to temporal structure. This would extend the current framework for song learning (Doya and Sejnowski, 1995; Fiete and Seung, 2006; Fiete et al., 2007; Troyer and Doupe, 2000) to include changes in HVC activity, while also expanding the role of HVC beyond that of a generic ‘clock’ (Fee et al., 2004; Fiete, 2004; Fiete et al., 2007).

Describing the relationship between HVC dynamics and adaptive changes to temporal structure (Figure 2.2C) requires tracking the activity of HVC neurons over the course of learning. Given the difficulty in recording single units in the HVC of freely behaving songbirds for extended periods (i.e. more than a few hours (Kozhevnikov and Fee, 2007; Sakata and Brainard, 2008; Yu and Margoliash, 1996), we recorded multi-unit activity (Crandall et al., 2007; Schmidt, 2003) while exposing birds to the CAF protocols (see Methods). Song-aligned neural signals thus acquired were stable over many days (see Figures 2.7A and 2.7D for examples), allowing us to explore how HVC dynamics changes with significant modifications to the song’s temporal structure.

Relating HVC dynamics to vocal output requires taking into account the temporal lag between premotor activity in HVC and the sound produced. We estimated this lag by cross-correlating the HVC...
signal with sound amplitude and by computing the co-variance in the temporal variability of the two
signals (see Methods). Both analyses showed HVC activity leading sound by, on average, 35 ms (Figure A.6), consistent with the anticipatory premotor nature of HVC reported in previous studies (Fee et al., 2004; Schmidt, 2003; Vu et al., 1994).

In support of HVC encoding temporal changes, modifications to the duration of discrete song segments (mean shift per tCAF drive: 11.3±4.2 ms; 3-5 days per drive; n=13 tCAF drives in 6 birds) were associated with significant and target-specific changes in the underlying HVC signal (Figure 2.7A). Indeed, the correlation between the average song-aligned neural activity pattern before and after tCAF training was 0.50±0.26 and 0.86±0.18 for target and non-target segments respectively (Figure 2.7B, p=0.002; see Methods). Learning-related changes in HVC activity manifested predominantly as a temporal re-scaling of the baseline signal, stretching or shrinking it in segments where the song had experienced lengthening or shortening respectively. Accounting for the temporal changes in song by time-warping the neural traces accordingly yielded a dramatic increase in the correlation between the neural signals before and after tCAF for the targeted segment (0.83±0.09, see Methods), making it not significantly different from the correlation values for time-warped non-targeted segments (0.88±0.07, p=0.24; Figure 2.7B). Time-warping the average neural trace recorded at the end of a tCAF drive to best fit the pre-CAF recordings (see Methods) yielded warping estimates that were very similar to those derived from warping the corresponding average song spectrograms to each other (R = 0.95 for targeted segments, n=23 segments; Figure 2.7C), suggesting a strong mechanistic link between temporal restructuring of behavior and HVC dynamics.

Inducing shifts in the pitch of targeted syllables (pCAF), on the other hand, yielded no target-specific change in HVC activity (Figure 2.7D; mean total shift per pCAF drive: 52.9±31.3 Hz; 3-5 days per drive; n=8 pCAF drives in 4 birds). Correlations in the neural traces before and after pCAF for target and non-target segments were 0.89±0.13 and 0.87±0.13, respectively (Figure 2.7E; p=0.76). These observations are consistent with the idea that changes to spectral structure are implemented
Discussion

By making reinforcement contingent on variability in either temporal or spectral features of birdsong, we demonstrate the capacity of the nervous system to independently modify timing and motor implementation aspects of a motor skill (Figures 2.1 and 2.2). In dissecting the underlying circuits we discovered a surprising dissociation in how learning is implemented in the two domains, with the basal ganglia essential for modifying spectral, but not temporal features of song (Figure 2.3) and a premotor cortex analogue area (HVC) encoding changes to temporal, but not spectral features (Figure 2.7). The dissociation in how the different aspects of vocal output are learned extended to the normal song maintenance process (Figure 2.6), suggesting that ‘template-based’ song learning (Konishi, 2010) may be an instantiation of reinforcement learning (Doya and Sejnowski, 1995; Fee and Goldberg, 2011). This also further validates the CAF-paradigm (Tumer and Brainard, 2007) as a proxy for normal song learning, though the extent to which the two are similar need to be further explored.

Our results show that reinforcement learning in the spectral and temporal domains is implemented by distinct but partially overlapping circuits. Much of the exploratory variability in both aspects of vocal output is driven by the same thalamo-cortical circuit (DLM-LMAN (Goldberg and Fee, 2011), which outputs directly to RA and indirectly to HVC (Hamaguchi and Mooney, 2012; Schmidt et al., 2004) (Figure 2.4). However, the circuits that convert the information gained from vocal exploration into a learning signal capable of driving changes in motor circuitry differ. For pitch, our results point to Area X as a key locus of reinforcement learning (Fee and Goldberg, 2011; Kojima et al., 2013). This basal ganglia homologue can affect the RA motor program by modulating activity in the downstream thalamo-cortical circuit to produce an error-correcting motor bias at the level of LMAN (Andalman and Fee, 2009; Warren...
et al., 2011; Charlesworth et al., 2012) (Figures 2.4A and 2.4B). For learning in the temporal domain, however, the circuits that translate the consequences of exploration into improved performance do not seem to involve the AFP or, more generally, the song-related basal ganglia circuits (Figures 2.3 and 2.5).

The anatomy of the song circuit together with our results showing learning-related changes in HVC activity, points to this time-keeper circuit as a possible nexus for reinforcement learning of temporal features. This would require variability in motor timing to be expressed within HVC and for a performance-based evaluation signal to reach it – both plausible scenarios: LMAN, which drives much of the temporal variability underlying learning (Figure 2.4F), can influence HVC network dynamics through indirect connections (Schmidt et al., 2004; Roberts et al., 2008; Hamaguchi and Mooney, 2012), while midbrain dopaminergic projection neurons, a common source of reinforcement in vertebrate circuits (Fields et al., 2007), project directly to HVC (Appeltants et al., 2000; Hamaguchi and Mooney, 2012) and, interestingly, also to Area X (Person et al., 2008). Thus the same source of variability (LMAN) and reinforcement (midbrain dopamine neurons) could, in principle, underlie two distinct reinforcement learning processes. While follow-up studies are needed to conclusively establish where and how temporal learning happens within the song system, our result showing basal-ganglia independent changes to HVC activity (Figure 2.7) makes this premotor nucleus a plausible candidate.

The basal ganglia is generally thought to be involved in the acquisition of learned motor behaviors (Doyon et al., 2009; Graybiel, 2005; Turner and Desmurget, 2010), yet the specifics of how it contributes to the learning process remain poorly understood. Our results, showing that the basal ganglia in songbirds is necessary for learning spectral, but not temporal aspects of vocal output, add important nuance to this question. Whether this reflects a general difference in how the basal ganglia contributes to motor skill learning remains to be explored, but our current study strongly suggest that the distinction between timing and motor implementation (Figures 2.1A-2.1C) is a crucial one to make when considering basal ganglia function in the context of motor learning.
Control of motor timing in humans is thought to involve prefrontal regions (Halsband et al., 1993; Harrington and Haaland, 1999), yet little is known about how these circuits represent the temporal structure of motor output, and whether they are involved in learning. HVC, the equivalent structure in songbirds, has been studied in far greater detail. It is thought to control song timing in the form of a synaptically connected chain of neurons, where each node represents a specific time point in the song (Li and Greenside, 2006; Long et al., 2010) (Figure 2.1H). Our HVC recordings during temporal learning, however, show HVC to be more than an immutable time-keeper. We observed activity patterns in this premotor nucleus stretch and shrink with the song (Figure 2.7), suggesting that temporal structure is modified by locally tuning the propagation speed within the network. Thus rather than representing time, our result suggest that neurons in HVC encode specific parts of the song, e.g. the starts and ends of syllabic or sub-syllabic elements, the relative timings of which can be adjusted independently from other features of the song.

Modulating dynamics in HVC by means of temperature has previously been shown to uniformly alter song tempo without interfering with spectral content (Aronov and Fee, 2012; Long and Fee, 2008). Our results show that similar changes to HVC dynamics and song can be induced and consolidated through reinforcement learning. Moreover we show that the temporal changes to song structure can be specific to certain parts of the song. The ability to shape the temporal structure of birdsong in such a specific manner is likely to be ethologically relevant: temporal features, such as syllable duration, distinguish song dialects (Wonke and Wallschläger, 2009) and can be shaped by exposure to different habitats (Kopuchian et al., 2004).

The ability to adaptively modify timing without interfering with other aspects of behavior may be critical to the acquisition and refinement of many motor skills also in humans (Gentner, 1987). Subtle changes to the temporal structure of syllables in human speech, for example, do not unduly change spectral aspects of vocal output (Cai et al., 2011). Furthermore, when a targeted syllable segment is
experimentally lengthened (Cai et al., 2011), subsequent speech patterns are similarly delayed to account for the increase in target duration, i.e. a phenomenology similar to what we see in songbirds (Figures 2.2B and 2.2D). Our results suggest a powerful and potentially very general solution for how this and other processes that alter temporal structure of learned motor output could be instantiated in neural circuitry (Figure A.1B).

Having separate learning processes shape distinct aspects of a motor skill can have several advantages, chief among them the flexibility to modify them independently (Figures 2.1 and 2.2). The success of “slow practice”, a method for training complex motor sequences championed by many music and dance teachers, is one of many examples attesting to this flexibility. Students are first taught proper motor implementation (i.e. which fingers/limbs to move in what sequence and to what extent) before refining the temporal structure of their performance. The underlying premise is that learning in the time domain does not interfere with other learned aspects of motor output. Our results show that this intuition is codified in the organization of the nervous system, which divides up the task of learning precise motor skills into functional modules for timing and motor implementation (Figure 2.1B), each with its distinct circuitry. This modularity may also be necessary to overcome the inherent limitations of reinforcement learning, basic implementations of which do not cope well with large task domains (Botvinick et al., 2009). Indeed, parsing up complex learning tasks into hierarchically connected, but largely independent, modules (Diuk et al., 2013) may have enabled increasingly complex behaviors to evolve by using (and re-using) the same rudimentary learning algorithms.

Methods

Adult male zebra finches (90+ days post hatch, n=40) were obtained from the Harvard breeding facility and housed on a 13:11 hr light/dark cycle in individual sound-attenuating chambers with food and water provided ad libitum. The care and experimental manipulation of the animals were carried out in
accordance with the guidelines of the National Institutes of Health and were reviewed and approved by the Harvard Institutional Animal Care and Use Committee.

*Conditional Auditory Feedback (CAF) protocol*

Custom software (LabVIEW) was used to implement the conditional auditory feedback protocol (CAF) used to manipulate pitch and duration of targeted song segments. The target was detected based on the correlation between the bird's song and a template spectrogram of the preceding 100-500 ms in the bird’s song motif. Average detection rates as quantified by manually examining at least 80 songs both early and late in the CAF drive each were generally high (>80%), and did not differ before or after any of the lesions (98±3% pre-lesion vs. 97±4% post-lesion).

Once a target was detected, its feature (pitch or duration) was computed. If it did not meet the escape threshold, white noise feedback (lasting between 25-100 ms, but constant for a given bird) was played back through a loud speaker with short latency (~1-3 ms). We calibrated the feedback volume to be marginally higher than the bird’s loudest syllable, effectively setting it to ~80-95 dB (A-weighting) 10 cm away from the speaker. The threshold to escape white noise feedback was dynamically updated based on the bird’s performance over the last 200 renditions of the target. If the fraction of escapes exceeded 80%, the threshold was automatically adjusted to the bird's mean in those last 200 renditions, but the adjustment was only made in the direction of learning.

*Target estimates - pCAF.* We chose target syllables with well-defined pitch (i.e. harmonic stacks) that were reliably (>80%) detected. Pitch was computed on a 5 ms sound segment of the target syllable using an algorithm fitting different sets of harmonics (see Supplemental Experimental Procedures). We computed pitch either at the very start of the syllable or 15-50 ms into it (varied between birds, but constant within a bird).

*Target estimates - tCAF.* On-line estimates of targeted segment durations used threshold crossings of the smoothed (5 ms boxcar filter with 1 ms advancement) amplitude envelope. The threshold was set to
~2-10x the background noise levels and kept constant throughout an experiment. Syllable onsets are associated with rapid increases in amplitude, which makes the estimates of their timing more robust to noise. Thus we mostly targeted ‘syllable + gap’ segments and estimated the target duration from the onset of the target syllable to the onset of the following syllable. However, in 1 bird, we made white noise conditional on the duration of a syllable, with the additional contingency that the subsequent gap duration not change significantly. In 4 additional birds, we targeted inter-syllable gaps (offset of last syllable to onset of next syllable). These 5 birds were pooled with the rest because they produced similar effects in response to experimental manipulations (e.g. lesions).

Experimental design. The design for birds that underwent pCAF and tCAF both before and after lesions was as follows: one group did a continuous block of pCAF for at least 6 days, followed by at least a week of no CAF. This was followed by a continuous block of tCAF for at least 6 days. The birds then underwent surgery for lesions and were given at least 1 week to recover before repeating the pCAF and tCAF blocks in the same order. Another group of birds experienced the same protocol but with the order reversed (tCAF followed by pCAF). Because pCAF was impaired after Area X lesions, we wanted to rule out potential short-term effects of lesions on learning. We thus ran pCAF for two birds more than 4 weeks after lesion to confirm abolished learning. We typically exposed birds to CAF for the same number of days before and after lesion and targeted the same song segment. Some birds experienced either tCAF or pCAF only, in which cases we did at least one round of CAF (in both directions). See main text for details of sample sizes for the various experiments. In a subset of birds, we conducted spontaneous return-to-baseline experiments before and after Area X lesions (Figure 2.6). For tCAF experiments we drove the targeted segment duration away from baseline for 3-5 days before removing white-noise feedback. The same protocol was repeated after Area X lesions (as birds can still shift duration). However, since birds cannot shift pitch after Area X lesions (Figure 2.3), for pCAF we drove targeted syllables away from their baseline pitch for 4 days and then turned CAF off, allowing birds to
spontaneously recover to baseline for up to 7 days. The same birds were then driven up again for 4 days before lesioning Area X. The pitch was subsequently monitored for up to 7 days post-lesion to assess any recovery to baseline.

**Lesions**

Birds were anesthetized under 1-3% isoflurane in carbogen and placed in a stereotaxic apparatus. Targeted brain areas were lesioned by injecting 4% (w/v) of N-methyl-DL-aspartic acid (NMA; Sigma, St Louis, MO) at stereotactically defined locations (see Table A.1). Lesions were confirmed histologically using cresyl-violet staining. We identified Area X and LMAN based on regions of stronger staining and/or higher density of cells than surrounding areas and were additionally guided by anatomical landmarks (e.g., lamina pallio-subpallialis and lamina mesopallialis) (Nixdorf-Bergweiler and Bischof, 2007). MMAN was identified based on landmarks and presence of LMAN. Remaining Area X, LMAN, or MMAN volumes were quantified and compared to volumes from adult control birds (n=4) with intact brains. Between 80-100% of LMAN, 72-98% of Area X and 75-100% of MMAN were lesioned (see Figure A.5 and Table A.2).

**Directed singing**

To test LMAN-mediated pre-motor bias, a female bird was presented to the experimental subject after 4-7 hours of CAF. Each female was presented for 2-3 minutes after which it was replaced with a different female. This sequence of single-female presentations continued for 15-30 minutes. All directed songs as well as catch trials just before presentation of females were uncontaminated by white noise (i.e. CAF was turned off).

**Electrophysiology**

**Surgery.** Birds (n=7) were anesthetized with 1-3% isoflurane in carbogen and placed in a stereotaxic apparatus. The location of Area X was estimated (see above) and confirmed by electrophysiological criteria (Kojima and Doupe, 2009). A bipolar electrode was acutely placed in Area X and used to identify
the boundaries of HVC through antidromic stimulation. A custom recording array (4 channels, ~250 um spacing) of 100 kOhm tungsten or platinum electrodes (Microprobes, Inc.) was implanted within the boundaries of HVC and a silver ground reference placed outside of HVC between the dura and the surface of the brain. Implanted components were secured to the skull with dental cement. All birds exhibited normal song output within 3 days of surgery; pre- and post-surgery song spectrograms were similar by visual inspection, suggesting minimal disruption of the targeted tissue. Following completion of the experiment, the animals were sacrificed, their brains harvested, and the placement of recording and stimulating electrodes confirmed by histology.

*Chronic recordings in HVC.* Sound and neural activity were recorded using a custom LabVIEW application. The raw neural signal was amplified (1,000-10,000x) and bandpass filtered (1 Hz - 15 kHz). Multiunit activity was recorded from up to four sites from each bird over four to six weeks. Because multi-day stability of the recordings was crucial for our analysis, all subsequent analysis was done on data collected from the most stable recording site in each bird.

**Data Analysis**

*Song segmentation.* All song and HVC recording analysis was performed off-line using custom-written software (LabVIEW and MATLAB). Songs were sampled at 44.15 kHz and bandpass filtered (0.3-7 kHz). The dominant song motif for each bird was determined by visual inspection. Once a motif was chosen, it was identified in the sound recordings using a semi-automated routine, which included visual inspection of the segmented songs to verify that they indeed matched the chosen motif. These segmented motifs constituted the data for subsequent analysis.

*Catch trials.* Song analysis was done on catch trials, i.e. songs recorded with the CAF protocol turned off, in the early morning (AM session) and evening (PM session). Approximately 100-200 songs per day were analyzed for each bird. Baseline data was analyzed for ~200 songs recorded 1-2 days before the start of CAF at comparable times to the CAF catch trials.
Pitch estimates. Pitch estimates for the catch trials were calculated as described in Supplemental Experimental Procedures. Since pitch can be defined robustly only for harmonic stacks, we computed pitch variability for harmonic stack syllables in birds that had them. If a bird did not have any harmonic stack syllable, we analyzed pitch variability in a sub-syllabic harmonic stack (see the latter half of syllable S4 in Figure 2.1F for an example).

Interval duration estimates. Off-line duration estimates from the catch trials were obtained by dynamically time-warping (DTW) the songs to an average template (Glaze and Troyer, 2006). We implemented our DTW algorithm on spectrograms, using the $L^2$-norm of the difference in the log-transformed spectrogram at each time point as the local distance metric. Slopes of the warping paths were constrained to be between 0.5 and 2. Template start and end points were not constrained to align to the start and end points in the rendition. For details on how interval durations were estimated using DTW see Supplemental Experimental Procedures.

Temporal variability. Temporal variability in interval (i.e. syllable and gap) durations was estimated as described previously (Glaze and Troyer, 2012a). Briefly, rendition-to-rendition variability of interval durations in the song was parsed into local, global, and jitter components by factor analysis. Local variability refers to independent variations in interval lengths, global variability captures correlated variability across intervals (due to e.g. temperature (Aronov and Fee, 2012; Long and Fee, 2008) or circadian (Glaze and Troyer, 2006) effects), and jitter is the variance in determining an interval’s boundary. Given that we are inducing temporal shifts in the duration of individual segments, we report on the ‘independent’ variability component (Glaze and Troyer, 2012a), but note that the other components showed similar trends after Area X and LMAN lesions. Coefficient of variation was calculated for each interval and averaged over the intervals in the bird’s song.

Pre-and-post lesion variability comparison. We compared both temporal and pitch variability before and after lesion. For pre-lesion, we used songs produced in the mornings up to 2 days preceding the surgery,
grouped into a single catch trial block to increase our sample size. For post-lesion we analyzed morning songs for up to 2 days, at different times after surgery to parse acute (1-3 days post lesion) and persistent (3+ days post lesion) effects (Figures 2.3D and A.4).

**Learning rates.** For tCAF and pCAF respectively, we computed learning rates as the difference in average pitch or duration of the AM catch trials on last day of CAF and that of AM catch trials on the first day of CAF divided by the number of intervening days. We did the same for PM catch trials and the overall learning rate was then averaged across AM and PM catch trials for the whole drive up and down for each bird to obtain a more robust estimate of the learning. For a small number of birds that did not sing during either AM or PM catch trial blocks, we computed learning rate from the remaining block only (e.g., AM only). Comparing the same time periods in the day allowed us to rule out circadian effects.

**Directed singing.** Estimates of pitch and duration were computed as described above. In addition, we corrected duration estimates for global tempo changes during directed singing (Stepanek and Doupe, 2010), estimated as the average change in the duration of non-target intervals during directed songs compared to undirected songs immediately before presentation of females. Reversion was calculated as the difference between the pitch or duration estimate just prior to presentation of the female (undirected PM, see Figure 2.4A) and during directed singing PM, and normalized to the total change in pitch or duration during the 4-7 hours of CAF.

**Song and neural alignment within a block.** Songs during catch trial blocks were segmented and a song template created as described in Supplemental Experimental Procedures. Starts and ends of intervals (syllables and gaps) were extracted for each rendition and linearly warped to the template. The warping path was time-shifted by 35 ms to account for the lag between HVC and sound output (Figure A.6) and then applied to the bandpass filtered HVC voltage trace (0.3-6 kHz, zero-phase, 2-pole Butterworth). The squared voltage was averaged across all renditions in the block and smoothed with a 5 ms boxcar window to generate the mean neural power trace. Spectrograms warped to the common template were
similarly averaged to generate a mean spectrogram for the block. The average warping paths across the renditions was then applied to the mean spectrogram and neural trace to remove any template specific effects.

*Song and neural alignment across blocks.* The mean neural traces and spectrograms were calculated as described above for the start and end of a CAF drive. To account for CAF-induced changes in temporal song structure, the post-CAF spectrogram was warped to the baseline spectrogram, using the same DTW warping routine as described above. Warping estimates for each interval were calculated as the ratio of post-CAF to pre-CAF interval duration. The warping paths thus derived were applied to the average post-CAF neural trace, yielding the green traces in Figure 2.7A. The same DTW routine was also applied to the neural traces to compare the warping in the underlying neural signal to warping in the song (Figure 2.7C). To make the warping estimates for the neural data more reliable we flagged salient points in the neural trace (i.e. well-defined peaks and troughs) and calculated the time-shifts in these points over the course of the CAF drive. Since these points did not always line up with the interval boundaries in the song, we took the weighted average of the time shifts in the points within 10 ms of the interval boundary, each point being weighted inversely to its distance from the boundary. The estimate for the neural warping in a given interval was then derived from the difference in the estimated time shifts corresponding to the start and end points of the interval.

*Correlations in neural power.* To quantify the degree and temporal specificity of the changes in neural power induced by CAF, we calculated running Pearson’s correlations (50 ms boxcar window, 1 ms advance) between the neural power in baseline and post-CAF conditions. For each analyzed CAF drive, we compared the mean correlation of non-targeted song intervals (motif onset to 50-100 ms prior to CAF target) with those in the targeted interval (pCAF) or targeted interval plus 100 ms (tCAF).

*Statistical testing.* All statistics presented in the main text refer to mean±standard deviation (SD), while error bars in the figures all represent standard error of the mean (SEM). All statistical tests assessing
significance across manipulations in the same birds were done using paired samples t-tests or one sample t-tests against mean zero unless otherwise noted.

**Acknowledgements**

We thank Ed Soucy for assistance with the CAF protocol, Stephen Turney, and the Harvard University Neurobiology Department and the Neurobiology Imaging Facility for imaging consultation and equipment use. We acknowledge Jesse Goldberg, Aaron Andalman, Rajesh Poddar, Naoshige Uchida, Markus Meister, Evan Feinberg, Maurice Smith, and Kenneth Blum, for helpful discussions and feedback on the manuscript. This work was supported by a grant from NINDS (R01 NS066408), a McKnight Scholar Award and Klingenstein Fellowship to BPÖ, and a Swartz Foundation post-doctoral fellowship to CP.
Chapter 3:

Early sensory deprivation is

associated with reduced synaptic

pruning in motor circuits

Attribution: This chapter contains collaborative work from Timothy M. Otchy (TMO), Jonathan Garst-Orozco (JGO), and Bence P. Ölveczky (BPÖ).

TMO, JGO and BPÖ designed the study. TMO performed all behavioral experiments, and collected the in vivo neural data. JGO collected all in vitro neural data. TMO and JGO analyzed the data. TMO wrote the chapter with input from BPÖ.
Abstract

Neural circuits implementing learned behaviors develop under genetically specified constraints and in response to environmental influences during critical periods of brain development. However, the mechanisms by which instructive experience shapes the structure and function of learning circuits are poorly understood. Here, we use the songbird to characterize how developmentally prescribed neural circuitry is shaped by environmental input by comparing the vocalizations, premotor activity, and circuit connectivity of tutor-deprived zebra finches (i.e., ‘isolates’) with those of age-matched controls. We found that a principal effect of isolation was a reduction in synaptic pruning in a motor cortex analogue brain region, a plausible mechanism for the aberrant song-related neural activity and vocal output observed in isolates. Our results provide key insight into circuit-level effects caused by a failure of synaptic pruning, a pathology associated with a variety of developmental disorders, including autism, fragile X syndrome, and tuberous sclerosis.
Introduction

Many behaviors are shaped and constrained by the interaction of genes and environmental influences, a process that unfolds in the formation of neural circuits (Hensch, 2005a; Knudsen, 2004; Ölveczky and Gardner, 2011; Turrigiano and Nelson, 2004; Zhang and Poo, 2001). Genetically defined developmental programs drive the formation of neural circuits, which generate intrinsic neural dynamics that influence circuit structure through a recurrent process (Bi and Poo, 2001; Blankenship and Feller, 2010; Feller, 1999). At various stages of development environmental influence, often in the form of sensory experience, intervenes to alter dynamics spurring further refinement of both the underlying circuit and behavior (Berardi et al., 2000; Butz et al., 2009; Hensch, 2005b; Hua and Smith, 2004). A powerful approach for assaying the interactions and effects these factors have on the structure and function of circuits is to experimentally perturb the environment, genes, or targeted circuit mechanisms and identify deviations from the norm (Blankenship and Feller, 2010; Davis and Goodman, 1998; Jessell and Sanes, 2000).

Where this approach has been successful, however, it has been due in no small part to the choice of studying highly specialized and stereotyped circuits close to the periphery where the environmental influences and structural norms are well-defined and experimentally tractable (Kiehn, 2011; Ladle et al., 2007; Tian, 2008). It has been far less successful in probing the mechanisms guiding the formation of circuits underlying learned behaviors, like language (Kuhl and Rivera-Gaxiola, 2008; Schlaggar and McCandliss, 2007) or parental care (Dulac and Kimchi, 2007; Dulac et al., 2014), that are often poorly defined and their development the consequence of intricate and little understood environmental interactions. In particular, it is difficult to identify which of the many parameters changing in a forming learning circuit (e.g., patterns of connectivity, strength of synapses, excitatory-inhibitory balance, etc.) are principally a consequence of specific environmental inputs and which are under tight control of developmental processes, and thus would have proceeded even under deprived...
conditions (Wolpert et al., 2001). That is, how does instructive experience modify the developmentally prescribed architecture and function of learning circuits? Addressing this question requires a model system in which an experimentally tractable learned behavior sensitive to controlled environmental manipulation is realized in discrete and well-characterized neural circuits. The zebra finch, a songbird that learns a complex, species-specific courtship song, is the best-studied – and perhaps only – model system meeting these requirements (Brainard and Doupe, 2002; Mooney, 2009a).

Song learning in zebra finches takes place in two overlapping stages: during an early critical sensory phase (~25-60 days post hatch, dph) pre-vocal juvenile males memorize the song of an adult tutor and over the subsequent sensorimotor phase (~45-90 dph) engage in vocal practice to emulate the memorized model (Figure 3.1A) (Konishi, 2004; Price, 1979; Tchernichovski et al., 2001). Song learning is characterized by the reduction in initially high levels of song variability and slower changes (i.e. plasticity) that render the juvenile’s vocalizations increasingly similar to the memorized tutor song (Figure 3.1B,C) (Konishi, 2004; Ölveczky et al., 2011; Ravbar et al., 2012; Tchernichovski et al., 2004; Tumer and Brainard, 2007). Despite the profound and manifest influence of early sensory experience in defining the structure of adult behavior (Figure 3.1B, left), birds deprived of instructive tutoring experience during the sensory phase of learning (i.e., ‘isolates’) produce stable and species-typical songs (Figure 3.1B, right) albeit with aberrant frequency modulation (Fehér et al., 2009), suggesting that developmental programs can alone go a long way toward defining the structure of adult behavior and the underlying circuits (Eales, 1985; Price, 1979; Williams et al., 1993).

This process plays out in the songbird brain across a chain of highly specialized nuclei (i.e., the song system) that undergo substantial growth and functional reorganization concurrent with learning (Figure 3.1D). The widely hypothesized nexus for learning in this system is the robust nucleus of the arcopallium (RA), a primary motor cortex analogue (Brainard and Doupe, 2002; Doupe and Kuhl, 1999; Fee et al., 2004; Garst-Orozco et al., 2015; Jarvis et al., 2005; Johnson et al., 1997; Leonardo and Fee, 2005;
During adult singing, RA projection neurons fire patterned bursts of action potentials that are reproduced precisely during each motif (Leonardo and Fee, 2005; McCasland, 1987; Yu and Margoliash, 1996), and these in turn drive activity in downstream motor neurons innervating vocal muscles (Fee et al., 2004; Sturdy et al., 2003; Wild, 1993; Wild et al., 2011).
A major input to RA comes from the premotor cortical nucleus HVC (used as a proper name) (Nottebohm et al., 1982; Vu et al., 1994): each neuron projecting to RA generates a brief burst of spikes at one specific time in the song motif (Hahnloser et al., 2002; Kozhevnikov and Fee, 2007; Li and Greenside, 2006; Long et al., 2010; Yu and Margoliash, 1996), and it has been suggested that these burst drive, at each moment, the population of RA neurons active at that time (Fee et al., 2004; Hahnloser et al., 2002; Yu and Margoliash, 1996). Given the functional organization of the song circuit, learning can be understood as the process of establishing and refining the connections between time-keeper neurons in HVC and muscle-related neurons in RA such that the “right” muscles get activated at the appropriate times (Figure 3.1E) (Fee and Goldberg, 2011; Fee et al., 2004; Fiete et al., 2007). And indeed subsequent studies characterizing RA activity and connectivity during learning are consistent with the HVC-RA synapses being a essential locus for this transformation: HVC inputs to RA neurons both strengthen and sparsen during learning (Figure 3.1E) (Garst-Orozco et al., 2015; Herrmann and Arnold, 1991; Johnson et al., 1997; Kittelberger and Mooney, 2005) and this pattern of synapse refinement is consistent with it being a principal mechanism driving the learning-related reduction of motor variability (Fee and Goldberg, 2011; Garst-Orozco et al., 2015; Kao et al., 2005; Ölveczky et al., 2005, 2011).

While this framework for song learning (i.e., plasticity at HVC-RA synapses) can plausibly account for the behavioral changes in song, the extent to which this functional reorganization is sensitive to environmental perturbation, and whether early sensory deprivation generates identifiable, chronic deficits in motor circuit form and function, has been little explored. Here we assess how the environment, in the form of early instructive sensory experience, shapes the developmentally specified structure and function of the neural circuits underlying learned courtship songs by comparing the behavior and underlying neurophysiology of isolates and age-matched tutored controls. To characterize differences in the trajectory of song emergence, we measured the timing and magnitude of the reduction in motor variability indicative of song crystallization. To probe how the song motor program
encoded the observed differences in behavior, we made single unit recordings from RA in singing birds throughout song learning and quantified changes in the statistics of their activity. Then, to identify the synapse-level mechanisms driving differences in motor program structure, we made whole-cell recordings in slice to characterize how input to RA projection neurons is modified over development. We found that a principal effect of isolation was a reduction in motor circuit pruning at the HVC-RA synapses, suggesting that the changes in both singing-related neural activity in RA neurons and vocal output between normally reared and isolate birds could be explained by this deficit in pruning.

**Results**

*Isolate zebra finches show delayed and diminished reduction in motor variability*

Consistent with skill learning in mammals (Churchland et al., 2006; Faisal et al., 2008; Osborne et al., 2005; Wolpert et al., 2011; Wu et al., 2014), studies of song learning have shown that motor variability, measured here as song spectral variability (Methods), is actively regulated by the song system to promote the motor exploration necessary for reinforcement learning (Ali et al., 2013; Andalman and Fee, 2009; Kao, 2006; Kao et al., 2005; Ölveczky et al., 2005; Tumer and Brainard, 2007; Warren et al., 2011). We characterized song variability throughout development and found, consistent with previous reports (Fee et al., 2004; Ölveczky et al., 2011; Ravbar et al., 2012; Sober and Brainard, 2012; Tchernichovski et al., 2001), that the high spectral variability of tutored young juveniles rapidly declines until reaching a plateau just prior to sexual maturity (89 ± 13.9 dph; $\tau = 17 \pm 10.4$ d; $n = 5$ birds; Figure 3.1C). Following song crystallization, average motor variability remained stable for more than 100 days post crystallization. Isolate song development unfolded in stark contrast: though spectral variability of early isolate song ($n = 5$ birds) was not significantly different from that of age-matched tutored controls ($p = 0.64$), we did not find a similar rapid decline leading up to sexual maturity (Figure 3.1C). Rather, isolate song showed a slow, prolonged decline in variability into early adulthood that eventually
plateaued and stabilized at 141 ± 8.2 dph (τ = 77 ± 16 d; Figure 3.1C). Even at this late date, however, isolate song remained significantly more variable than that of tutored controls (p = 0.02).

Previous studies have suggested that early life stress can have a detrimental effect on a variety of behaviors (Bale et al., 2010; Park et al., 2001; Teicher et al., 2003), including song learning (Nowicki et al., 2002; Spencer et al., 2003). To rule out the possibility that the deficits in song development were a general effect of removing the father (e.g., thereby increasing early life stress) rather than due to the specific lack of early instructive song experience, we quantified the trajectory of spectral variability in finches raised in social isolation from a male tutor during the sensory phase but were artificially tutored via speaker playback of pre-recorded adult song (Tchernichovski et al., 2001) (Figure 3.1A, n = 3 birds). We found no significant difference in spectral variability between artificially tutored and age-matched “live” tutored control birds either early or late in song learning (Figure 3.1C; p = 0.90 and 0.63, respectively; Methods), suggesting that the song deficits observed in isolates were principally a consequence of their lack of early instructive song experience.

*Early sensory deprivation is associated with reduced RA firing rates and temporal precision in vivo*

Song production involves the coordinated activity of the syrinx and respiratory muscles (Goller and Suthers, 1996; Suthers and Margoliash, 2002) which are innervated by motor neurons in the brainstem that are themselves directly controlled by RA projection neurons (Ashmore et al., 2005; Wild, 1997). Given that song spectral structure has been tightly linked to syringeal and respiratory motor gestures (Gardner et al., 2001; Laje et al., 2002; Méndez et al., 2010; Sober et al., 2008; Suthers and Margoliash, 2002), we reasoned that the prolonged elevated variability observed in isolates (Figure 3.1C) may be associated with decreased precision of song motor gestures, and thus of RA projection neuron activity during singing. To test these predictions, we made single-unit extracellular recordings of RA neurons during singing in both tutored (223 cells in 14 birds) and isolate (92 cells in 8 birds) zebra finches.
Figure 3.2: Song development in tutored and isolate birds is characterized by changes in the statistics of RA projection neuron activity.

(A) Schematic of the experiment: making single unit recordings from RA projection neurons in singing zebra finches. (B-E) Changes in spike train statistics for RA projection neurons as a function of age. Blue: tutored birds (n = 216 neurons in 16 birds). Red: isolate birds (n = 90 neurons in 8 birds). Points represent individual cells. (B) Average firing rate during singing. (C) The fraction of spikes in a burst (i.e., > 150 Hz instantaneous firing rate). (D) Sparseness index. (E) Average pairwise cross correlation of song-aligned spike trains. (F-I) Comparison of the spike train statistics across tutoring paradigms (blue: tutored; red: isolate) for discrete age groups: 50-55 dph, 60-65 dph, 90-130 dph and +200 dph. Error bars represent s.e.m. Starring convention: * p < 0.05, ** p < 0.01, and *** p < 0.001.

throughout song learning and into late adulthood (43 - 218 dph; Figure 3.2A). The recorded cells fell into two functional classes based on their spontaneous firing patterns and spike shapes, consistent with previous classification as putative projection neurons and interneurons (Leonardo and Fee, 2005; Ölveczky et al., 2011; Spiro et al., 1999). Here, we report only on the putative projection neurons recorded in tutored and isolate birds (216 and 90 cells, respectively).

Consistent with previous reports in tutored birds (Chi and Margoliash, 2001; Leonardo and Fee, 2005; Ölveczky et al., 2011), at all stages of vocal development RA projection neurons in tutored and
isolate birds were tonically active while birds were silent and transitioned to patterned, higher frequency spiking during singing. In juvenile birds, these neurons discharged sporadically during singing and only occasionally generated spike bursts (i.e., >150 Hz instantaneous firing rates; Figure 3.2B,C). By early adulthood, however, a larger fraction of RA spikes were discharged as part of bursts that achieved considerably higher peak firing rates – a trend that continued into late adulthood. Though we found significant correlation between age and increases in firing rate and bursting in both tutored (R = 0.55, p = 10^{-18}; R = 0.68, p = 10^{-30}, respectively) and isolate birds (R = 0.47, p = 10^{-6}; R = 0.67, p = 10^{-13}, respectively), projection neurons recorded from isolates had on average lower firing rates than those from age-matched tutored controls (Figure 3.2B,C; Methods).

We quantified the temporal precision of RA activity patterns from the time the birds sang the first recognizable motifs (~50 dph) until late adulthood. Motif aligned spike rasters were created to visualize how song-related patterns change over development, and the correlation of neural data and song was quantified in two ways. To quantify how a neuron’s spiking was distributed in time, we computed a sparseness index (SI) in which a value of 1 indicates a neuron whose activity was entirely restricted to a single part of the motif and 0 indicates an even distribution across the motif (Methods). We found that the SI increased throughout development in both tutored and isolate birds (Figure 3.2D), showing significant correlation with age (R = 0.42, p = 10^{-9}; R = 0.34, p = 10^{-3}, respectively). In addition, to quantify trial-to-trial variability in spiking across multiple motif renditions, we computed pairwise cross correlations (PCC) between RA activity patterns during different motif renditions (Methods). We found that PCC increased over development in both tutored and isolate birds (Figure 3.2D,E), and the correlation between age and song-aligned spike train variability was again significant (R = 0.47, p = 10^{-10}; R = 0.76, p = 10^{-18}, respectively). Similar to the firing rate, however, the average temporal precision of projection neurons in isolates remained at lower levels compared to age-matched tutored controls.
The divergent trends observed in the evolution of RA activity reveal differences in the functional reorganization of the premotor circuit that accompanies song development in each tutoring paradigm. To identify when the most significant changes occur, we compared within each tutoring paradigm the RA spike train statistics at four song developmental stages: group I, 50-55 dph (n = 15, n = 7 neurons; tutored and isolate birds, respectively); group II, 60-65 dph (n = 18, n = 11 neurons); group III, 90-130 dph (n = 43, n = 26 neurons); and group IV, +200 dph (n = 28, n = 16 neurons). In both tutored and isolate birds, we found significant increases in average firing rates between age groups II and III (p < 10^{-3}), but not between any other adjacent groups (Figure 3.2F). There was a statistically significant increase in burstiness from groups I to II and II to III (p < 10^{-6}) in tutored birds, but in isolates significant changes were only found from groups II to III and III to IV (p < 0.002) (Figure 3.2G). Consistent with the large reduction of song variability observed in tutored birds, we found that SI and PCC increased significantly for every adjacent group (p < 0.03 and p < 0.01, respectively) (Figure 3.2H,I). In contrast, PCC in isolates only showed significant increases between groups II to III and III to IV (p < 0.003). Though SI in isolates increased by more than 36% between groups I and IV, the incremental differences between adjacent age groups were not significant.

To understand how development of the RA motor programs differ in isolates and tutored birds, we compared RA spike train statistics between tutoring paradigms at each of the four age groups. We found no significant difference in any of the four metrics for the earliest stage. That there is little divergence in group I suggests that deprivation from early tutor experience has minimal effect on motor circuit function at the earliest stages of development. In contrast, burstiness, SI, and PCC showed maintained divergence between isolate and tutored controls for groups II through IV (p < 10^{-4}, Figure 3.2G-I), and mean firing rates differed for groups III and IV (p < 10^{-3}, Figure 3.2F). Together, our results show that in both isolate and tutored birds the neural correlate of song development at the level of RA neurons can be described as a gradual change toward higher firing rates in more precisely time-locked
bursts, though in isolates these changes tend to occur later and are of smaller magnitude even late in adulthood. This demonstrates that depriving birds of early instructive experience produces lasting effects on the development of a precise song motor program. Furthermore, these results are consistent with increased song variability being a result of both a less temporally precise RA motor program as well as reduced total excitatory drive to brainstem motorneurons (Stein, 1965).

Figure 3.3: The change in HVC input to RA projection neurons over song development is sensitive to early instructive experience. (A) Bright-field image of the experimental setup used for probing HVC input. In an acute parasaggital brain slice containing both RA and the input tract, HVC fibers are stimulated and the evoked currents in RA recorded in voltage clamp. (B) Increasing the stimulation intensity activates increasingly more input fibers, allowing us to measure both SF and MAX currents. Shown is an example of HVC-evoked currents at increasing stimulation intensities with the SF and MAX currents for this cell labeled. (C) Distributions of SF currents in tutored birds for the four age groups tested. The blue line shows the lognormal fit to the data (Methods). (D) Same as in C, but for isolate birds. Red line shows lognormal fit. For comparison, inset plots show overlaid lognormal fits for isolates and age-matched tutored controls. (E-G) Comparison of the change in HVC input to RA projection neurons in tutored (blue) and isolate (red) birds in each age group. (E) Box and whisker plots showing the median IQR (box), +/- 1.5 standard deviations (whiskers), and outliers (points) of the SF current distributions. (F) Average MAX currents. (G) Average number of HVC inputs to RA projection neurons. Error bars in F and G denote standard error. Starring convention: * p < 0.05, ** p < 0.01, and *** p < 0.001.
Early isolation from a tutor is associated with reduced pruning of HVC-RA synapses

We sought to identify key differences in HVC-RA synapse reorganization that underlie the observed divergence in RA activity patterns in tutored and isolate finches (Figure 3.2) by making whole-cell recordings from RA projection neurons in acute brain slice. Because the HVC projections form a dense tract as they enter RA (Kittelberger and Mooney, 1999; Mooney and Rao, 1994), it is possible to section the brain to preserve both the input tract and a large fraction of RA (Figure 3.3A). This preparation allows for probing the strength of HVC input by stimulating the fiber tract at different intensities and recording evoked currents in RA projection neurons (Figure 3.3A,B, Methods) (Garst-Orozco et al., 2015).

Inspired by efforts to characterize experience-dependent maturation of the retinogeniculate synapse (Hong and Chen, 2011; Hooks and Chen, 2008), we quantified Single Fiber (SF) and Maximal (MAX) currents to estimate the strength of HVC inputs to RA projection neurons (Methods). SF currents from individual fibers were evoked by electrically stimulating the fiber tract at intensities that produced a mix of stable EPSCs and failures (Methods). We used the peak of the single fiber evoked EPSC (SF current; Figure 3.3B) to estimate the summed strength of inputs arising from a single HVC axon. MAX currents, measured as the maximum EPSC at saturating stimulus amplitudes, were used to estimate the total drive from all HVC axons to single RA projection neurons (Figure 3.3B, Methods). A prior study (Garst-Orozco et al., 2015) quantifying HVC input strength during song learning in tutored finches found significant age-related increases in SF currents paired with initial increases in MAX currents, followed by later reversal in adulthood. To assay whether early instructive experience affects this pattern of HVC input refinement, we extended this earlier study and compared SF and MAX currents in tutored and isolate birds with ages corresponding to groups I through IV.

We first probed the strength of individual HVC inputs to RA projection neurons by recording HVC fiber evoked SF currents in 134 RA projection neurons in tutored birds (n = 124 birds) and 82 neurons in
isolates (n = 30 birds). By stimulating the HVC fiber tract using a multi-channel stimulation electrode (Methods), we were often able to selectively and stably activate more than one SF input per cell (mean SFs per cell: 2.2 ±/ 1.0). Consistent with synaptic weight distributions reported in other central nervous systems (Barbour et al., 2007; Song et al., 2005), we found that the distribution of SF currents within each age group were well approximated by a lognormal fit (Figure 3.3C,D). In both tutored (Figure 3.3C) and isolate (Figure 3.3D) birds, mean SF currents became progressively larger with age until sexual maturity (group III, 90-130 dph), at which point we found no further significant changes (Figure 3.3E, top). We found that the distribution of SF currents did not differ significantly between tutoring paradigms at any age (Figure 3.3D, inset).

In addition to increases in SF input, our experiments revealed significant but non-monotonic changes in total HVC drive to RA projection neurons. In both tutored and isolate birds, MAX currents more than doubled between age groups I and II, and we found no significant differences between tutoring paradigms at either age (Figure 3.3F). Recordings from older animals, however, revealed a striking divergence: whereas tutored birds showed a 38% reduction in MAX currents between age groups II and III that then remained stable into late adulthood, those of isolates continued to increase, becoming more than three times MAX currents in controls of similar age. Though isolate MAX currents subsequently decreased 43% by late adulthood (group IV), they remained significantly larger than those of age-matched tutored controls (Figure 3.3F).

Given unchanged SF currents, the divergence in MAX currents in adulthood suggests differences in the timing and extent of pruning of HVC inputs to RA. To characterize the change in the number of functional HVC inputs to projection neurons, we estimated the relative number of inputs by dividing the MAX current for each cell by the age-matched mean SF current (Methods). Both tutored and isolate birds showed significant increases in the average number of HVC inputs from group I to II (41% and 70%, respectively), though this did not differ between tutoring paradigms within an age group (Figure 3.3G).
As we observed with the MAX currents, the mean number of inputs in tutored birds decreased then leveled off, while those in isolate birds continued to increase in early adulthood before showing a decrease in late adulthood. The differences in average input number between tutoring paradigms were highly significant for each of the adult age categories, consistent with a prior EM study showing elevated synapse density in the RA of adult isolate Bengalese finches (Peng et al., 2012a).

Within the context of prominent models of learning within the songbird motor pathway (Fee and Goldberg, 2011; Fiete, 2004; Fiete et al., 2007; Garst-Orozco et al., 2015; Ölveczky et al., 2005), the increased number of HVC inputs to RA neurons likely reduces temporal precision of RA firing patterns in two ways. First, the increased number of inputs during a motif is capable of reducing the sparseness of singing-related activity by directly triggering more RA spiking events during a motif. Second, the increased number of depolarizing events during a motif leaves RA neurons more susceptible to NMDA-receptor mediated variable input from LMAN by more frequently removing the Mg-block during the motif (Mooney and Rao, 1994; Ölveczky et al., 2005; Stark and Perkel, 1999). Together, our results suggest that one mechanism by which early instructive experience shapes developmentally specified motor circuit architecture is through either promoting or permitting increased synapse elimination in the later stages of song learning, which ultimately reduces drive to RA during singing.

**Discussion**

We set out to identify the mechanisms by which early instructive experience shapes the formation and function of neural circuits underlying complex learned behaviors. We used zebra finches because courtship song is an experimentally tractable learned behavior implemented in well-characterized neural circuits. We found that the aberrant vocalizations characteristic of isolation-reared birds (Figure 3.1) were generated by RA projection neurons spiking patterns that were of lower average firing rate and less temporally precise than those in age-matched tutored controls (Figure 3.2). Subsequent
characterization of the number and strength of HVC input to RA neurons revealed that isolate motor circuits showed reduced incidence of synapse elimination in adulthood (Figure 3.3), which is a plausible mechanism for the decreased precision and firing rates of singing-related RA activity. Together, these findings begin to delineate a categorization of the changes in motor circuit formation over song development: those driven by genetically defined developmental programs (i.e., synapse strengthening) and those that are consequence of early instructive sensory experience (i.e., synapse elimination).

While our previous study (Garst-Orozco et al., 2015) showed how HVC inputs to RA are modified over the course of song learning, that study was not able to assess the degree to which the identified changes were a result of experience-dependent learning versus developmentally prescribed maturation (Bottjer and Arnold, 1997; Knudsen, 2004; Ölveczky and Gardner, 2011). Our results here add new insight to the mechanism by which the baseline neural architecture of the song system is sculpted by instructive sensory experience. In addition, our work establishes a baseline against which future studies may evaluate the effects of targeted manipulations to further delineate the mechanisms guiding the formation of the song circuits.

Alternative mechanisms of instructive-experience dependent differences in RA activity

We argue that the reduction in HVC-RA synapse pruning in isolate birds is consistent with it being the primary mechanism underlying lower average firing rates and temporal precision of singing-related RA projection neuron activity. Given prior studies identifying HVC-RA synapses as a key locus of plasticity for shaping the song motor program (Garst-Orozco et al., 2015; Herrmann and Arnold, 1991; Johnson et al., 1997; Kittelberger and Mooney, 2005; Ölveczky et al., 2011) and a report of isolation-related changes in HVC-RA synapse density (Peng et al., 2012a), we reasoned that this was a likely location to observe instructive experience-dependent changes as well. Nevertheless, our findings do not preclude other
experience-dependent changes in motor circuit function or structure that could contribute to the observed differences in the development of RA projection neuron activity patterns.

Though a recent study characterized the development of sparse and precise HVC activity in tutored birds (Okubo et al., 2015), how the emergence of such activity may differ in tutor-deprived birds is unknown. Differences in singing-related HVC activity in the two tutoring paradigms (i.e., changes in variability, sparseness, spikes-per-burst, etc.) could plausibly contribute to the observed reduction of RA firing rates and temporal precision in at least two important ways. First, the singing-related activity of RA neurons is principally driven by strong AMPA-mediated input from HVC (Aronov et al., 2008; Kittelberger and Mooney, 2005; Ölveczky et al., 2011; Stark and Perkel, 1999; Yu and Margoliash, 1996), and thus it is highly likely that changes in this input would be directly reflected in the pattern of RA activity (Garst-Orozco et al., 2015; Spiro et al., 1999; Stark and Perkel, 1999). Second, aberrant patterns of HVC activity in isolate birds could themselves disrupt mechanisms driving HVC-RA synapse refinement that are thought sensitive to the relative timing of inputs to RA (Doya and Sejnowski, 1995; Fiete, 2004; Fiete et al., 2007; Mehaffey and Doupe, 2015; Sizemore and Perkel, 2011).

In juvenile isolates, hearing a tutor song for the first time led to rapid dendritic spine stabilization and enlargement in HVC (Roberts et al., 2010), demonstrating a role for instructive experience in influencing the development of HVC; however, the same study found that spine dynamics within the HVC of young adult isolates were indistinguishable from those of tutored adults, suggesting that experience-independent mechanisms (i.e., developmental regulation of hormones or gene expression) may be capable of compensating for the effects of deprivation. Furthermore, given the stable temporal structure of isolate song (Figure 3.1B) (Eales, 1985; Fehér et al., 2009; Immelmann, 1969) and HVC’s central role in establishing that temporal structure (Ali et al., 2013; Hahnloser et al., 2002; Jin et al., 2007; Li and Greenside, 2006; Long and Fee, 2008; Long et al., 2010; McCasland, 1987; Yu and Margoliash, 1996), we think it likely that the pattern of HVC activity in isolates is broadly similar
that that of tutored controls. Nevertheless, additional recordings in isolates would be required to exclude aberrant HVC activity as contributing to the deficits in motor circuit function and structure we have identified.

In addition, the observed differences singing-related RA projection neuron activity may also be shaped by tutoring-paradigm dependent changes in local inhibition. Previous work has shown that RA inhibitory interneurons are driven by HVC and LMAN input (Mooney and Rao, 1994; Spiro et al., 1999; Stark and Perkel, 1999) and make feedforward GABAergic circuits within RA (Mooney, 1992; Sakaguchi, 1996). Feed-forward inhibition of this type has been hypothesized to reduce the temporal jitter of projection neurons by narrowing the time window during which correlated afferent activity drives RA neurons to fire (Pouille and Scanziani, 2001; Spiro et al., 1999), making it a potentially important mechanism for regulating the temporal precision of the RA activity patterns. Furthermore, both electrophysiological (Ölveczky et al., 2011) and immunohistological (Sakaguchi, 1996) evidence suggests that inhibition within RA is regulated coincident with song learning and sexual maturation, but whether this regulation is sensitive to instructive experience has not been directly shown. Nevertheless, in other systems exhibiting critical periods of neurodevelopment, maturation of local inhibitory circuits and a recalibration of the excitatory-inhibitory balance is a recurring theme (Berardi et al., 2000; Gogolla et al., 2009; Hensch, 2004; Hensch and Stryker, 2004; Huang et al., 1999; Knudsen, 2004) and disruption of these processes is a consequence typical of critical experience deprivation (Hensch, 2005a). Given this context, deprivation-induced changes in the development of local inhibitory circuitry are indeed possible, though our experiments here are not well positioned to adequately characterize these differences. Additional in vivo and in vitro recordings targeting RA interneurons in both tutored and isolate birds would be particularly helpful in addressing the functional significance of altered inhibitory circuits.
The determinants of song structure in isolation-reared birds

Our results identifying differences in the refinement of HVC-RA synapses speak to long-debated questions about the determinants of song structure in isolation-reared songbirds (Eales, 1985; Konishi, 1965; Price, 1979). Given the profound influence of the tutor in determining the final form of the tutee’s song a looming question has been how isolates, having never heard conspecific vocalizations, are able to sculpt initially variable subsong into adult songs with strikingly species-typical structure (Figure 3.1B). A common explanation hypothesizes that in the absence of an ‘acquired’ (i.e., tutor-provided) song model, learning is guided by an ‘innate’ (i.e., genetically determined) song template (Konishi, 2004) that either directly specifies isolate song structure or establishes a species-specific song grammar, depending on the specific version of the theory (Konishi, 1965; Marler and Sherman, 1983; Troyer and Bottjer, 2001). The principal assumption is that song development in isolates and tutored birds are not fundamentally different processes; rather, access to an innate template allows isolate song and the underlying neural substrates to develop under the same reinforcement learning mechanisms shaping tutored song (Fee and Goldberg, 2011; Fiete et al., 2007; Marler, 1997). Thus, the innate template hypothesis predicts that synapse- and circuit-level differences between isolate and tutored birds ought to be no more different than those between tutored birds.

That the motor circuits of tutored and isolate birds develop along different trajectories with distinct structural and functional end states is inconsistent with these assumptions. Our findings support the idea that isolate song structure is not the result of “learning” from a poorly specified song model, but is instead the product of developmental processes that result in hyperinnervation of RA and altered patterns of song-related activity. Hence, the primary determinants of isolate song structure are more likely to be related to biophysical constraints on neural dynamics within a hyperinnervated RA and the biomechanical limits of the songbird vocal apparatus.
Unique specificity of tutor song deprivation

As noted above, our finding of aberrant synapse refinement at the HVC-RA synapses is reminiscent of deficits observed in animals deprived of sensory experience during sensitive periods of brain development (Hensch, 2005a). In visual cortex, altered spine and synapse number has been observed following prolonged manipulation of input, such as dark-rearing or continuous high-intensity illumination from birth (Cragg, 1975; Mataga et al., 2004; Valverde, 1971, 1967; Wallace, 2004). Similar changes in synapse number have been reported in barrel cortex of whisker-trimmed mice (Zuo et al., 2005) and thalamus-recipient auditory cortex of noise-reared mice (Barkat et al., 2011). Thus, it is plausible that the basic circuit and cellular mechanisms underlying decreased motor circuit pruning will parallel those found in these deprived sensory circuits.

There are, however, important differences between these previous studies and the experiments presented here – chief among them the type of disruption to the ‘natural’ sensory environment. Though the deprivation protocols vary among studies, all specify continuous total blockade of the sensory modality for a prolonged period for malformation of the circuit (Berardi et al., 2000). Indeed, even brief restoration of naturalistic sensory input is sufficient to redirect development toward a typical (i.e., unmanipulated) outcome (Hooks and Chen, 2006, 2008; Trachtenberg et al., 2002). In contrast, we show similar disruption of circuit form and function as a consequence of removing from the sensory environment one particular type of experience – tutor song – while leaving the birds’ auditory experience otherwise intact and open to ambient noise from the environment, non-song vocalizations from mother and siblings, and even birds’ own vocalizations. Furthermore, the results we describe are markedly different from the behavior (Funabiki and Konishi, 2003; Lombardino and Nottebohm, 2000; Nordeen and Nordeen, 1992) and motor synapse-level (Peng et al., 2012a, 2012b; Tschida and Mooney, 2012a) consequences of deafening. This suggests remarkable specificity of the sensory filters mediating acquisition of a song model for sounds matching the spectro-temporal features zebra finch tutor song.
Relevance to neurodevelopmental disorders

Heightened synaptogenesis at very early ages and a subsequent period of elimination prior to adulthood is a common motif in developing neural circuits (Belmonte et al., 2004; Bourgeois and Rakic, 1993; Huttenlocher and Dabholkar, 1997; Purves and Lichtman, 1980; Zecevic et al., 1989). Given the centrality of these processes to the emergence of normal brain function, it is perhaps unsurprising that differences in synapse shape, size, or number accompany a large number of disorders, suggesting that synapses may serve as a common substrate for many neuropsychiatric disorders (Penzes et al., 2011).

Interestingly, reduced synapse pruning and an abundance of excitatory synapses, comparable to what we report here in isolation-reared zebra finches, has been associated with neurodevelopmental disorders that involve deficits in imitation including autism (Tang et al., 2014), fragile X syndrome (Comery et al., 1997), and tuberous sclerosis (Bateup et al., 2013). Though much has been uncovered about the genetic and molecular mechanisms underlying these disorders (Belmonte and Bourgeron, 2006; Bourgeron, 2015), our understanding of how neural dynamics and computations are altered by these pathological connectivity patterns remains limited. Nevertheless, the breadth of the anatomical and functional abnormalities associated with these disorders suggests that the core dysfunction may involve some pervasive disruption of neural processing (Doll and Broadie, 2014), perhaps due to abnormally low signal-to-noise in developing circuits as a result of heightened connectivity (Belmonte et al., 2004). Our finding that an increase in the number of inputs reduces the fidelity of an RA projection neuron’s response to precise HVC input is consistent with the hypothesis of pervasive disruption and highlights the potential for the songbird as a tractable model for future investigations of circuit-level consequences of these synapse-pruning disorders.
Methods

Animal care

The care and experimental manipulation of all animals were reviewed and approved by the Harvard Institutional Animal Care and Use Committee. Adult male zebra finches (Taeniopygia guttata) between 40-243 dph (n = 186 birds) were obtained from the Harvard University zebra finch breeding facility and housed on a 13:11 hr light/dark cycle in sound-attenuating chambers with food and water provided ad libitum. Because the effect sizes of our results could not be pre-specified before the experiments, we chose sample sizes that would allow for identification of outliers and for validation of experimental reproducibility. No animals were excluded from experiments post-hoc.

Tutoring manipulations

Irrespective of tutoring treatment, all birds were hatched in the aviary and reared by both parents in dedicated breeding cages.

Tutored birds. Young hatchlings remained housed with both parents until 35dph, affording hatchlings unrestricted access to the song produced by the father. Thereafter, males were removed and housed singly in sound isolation boxes for the remainder of the experiment.

Isolation rearing. For birds targeted for either complete tutor isolation (n = 43 birds) or delayed artificial tutoring (n = 3), the father was removed from the breeding cage at day 10, and the remaining mother and her clutch were housed in sound attenuating chambers in a separate nursery room. Previous studies have shown that zebra finches do not imitate songs heard before 20 dph (Eales, 1985; Roper and Zann, 2006), therefore possible exposure to adult song prior to 10 dph is unlikely to have any effect. The mother and clutch were housed together until 30 dph, at which point males were removed and housed singly for the remainder of the experiment. Prior to ~35 dph, zebra finches do not produce vocalizations.
other than begging calls (Price, 1979), so by separating males by day 30 it is unlikely that vocalizations of other siblings could have an effect on vocal development.

Artificial tutoring. Recordings from artificially tutored birds generously provided by Ofer Tchernichovski (Hunter College) and trained as described in ref (Tchernichovski et al., 2001). Briefly, birds (n = 3) completed the isolation rearing protocol described above. From 43 until 90 dph, a pre-recorded adult song was played through a speaker inside the sound isolation box in response to key pecks by the bird. The birds were restricted to hearing 10 repetitions of a two-motif song bout twice per day (once in the morning and once at night), which totaled ~28 s of song exposure per day. All birds successfully copied the song provided.

Song recordings

Sound was recorded via an in cage microphone using a custom LabVIEW application (National Instruments) as previously described (Ali et al., 2013).

In vivo electrophysiology

Implantation of recording device. Zebra finches (n = 22) were anesthetized with 1-3% isoflurane and placed in a stereotaxic apparatus. The location of RA was identified by stereotactically and further confirmed by electrophysiological criteria (Ölveczky et al., 2011; Spiro et al., 1999). A custom microdrive recording device (Otchy and Ölveczky, 2012) with 4 platinum-iridium electrodes (10 MΩ) was implanted targeting RA in the right hemisphere. The device was permanently secured to the skull with anchor pins and dental cement. All birds exhibited normal song output within 7 days of surgery.

Chronic recordings in RA. Recordings were made using a motorized recording device as described previously (Otchy and Ölveczky, 2012). Cells were isolated by searching for spontaneous activity, and putative projection neurons and interneurons identified by their distinct electrophysiological profiles
(Leonardo and Fee, 2005; Ölveczky et al., 2011; Spiro et al., 1999). Neural and acoustic data were recorded using a custom LabVIEW application (National Instruments), as previously described (Ali et al., 2013). Single units were recorded from each bird over two to seven weeks. At the end of the experiment, the electrodes were positioned in RA, and electrolytic microlesions (30 µA @ 20 s) were made to mark the recording position.

**Histological verification.** At the end of the experiments, birds were anesthetized with natriumpentobarbital (Nembutal, IM) and subsequently transcardially perfused with PBS, followed by fixation with 4% paraformaldehyde (PFA) in PBS. Brains were dissected out and post-fixed in 4% PFA overnight. Parasagittal sections (100 µm) were cut on a Vibratome (Leica), mounted, and stained with cresyl violet to reconstruct the location of implanted electrodes. Histology was done blind to the identity of the animal.

**In vitro electrophysiology**

**Slice Preparation.** Slices were prepared as in ref (Garst-Orozco et al., 2015). Briefly, birds (n = 154) were anesthetized with isoflourane and subsequently decapitated. Brains were harvested and placed in 4°C oxygenated (bubbled with 0.4 liters per minute 95% O₂/5% CO₂) artificial cerebrospinal fluid (ACSF). Osmolarity of all ACSF solutions was elevated to 350 mosM with sucrose (Bottjer, 2005). Experiments were performed in 300µm thick acute brain slices (Mooney and Konishi, 1991), cut using a vibrating microtome (Leica VT1000 S).

**Whole-cell recordings.** Whole-cell electrophysiological recordings in RA were made as in ref (Garst-Orozco et al., 2015). Voltage-clamp recordings were carried out at room temperature (20-23°C) using electrodes of 2.5-3.5 MΩ resistance, filled with an internal solution containing (in mM): 35 CsF, 100 CsCl, 10 EGTA, and 10 HEPES, with pH adjusted to 7.32 with CsOH. 50µM picrotoxin was added to the external solution to block fast feed-forward GABAergic inhibition (Kittelberger and Mooney, 1999; Stark and
HVC-evoked EPSCs were recorded at -70mV in order to minimize contamination from predominantly NMDAR-mediated LMAN inputs (Spiro et al., 1999; Stark and Perkel, 1999) by means of hyperpolarization induced Mg$^{2+}$ block. RA projection neurons were identified based on their spontaneous tonic activity and characteristic spike waveform recorded in cell-attached configuration prior to breaking into the cell (Spiro et al., 1999; Stark and Perkel, 1999). The data presented are from identified projection neurons. Electrophysiological data were recorded with a MultiClamp 700B (Axon Instruments) using mafPC software (courtesy of M.A. Xu-Friedman) and custom macros in Igor Pro. Signals were digitized at 100kHz and Bessel filtered at 10kHz.

**Fiber-tract Stimulation.** Inputs were stimulated as in ref (Garst-Orozco et al., 2015): either with a pair of saline-filled glass electrodes or a pair of tungsten electrodes (MicroProbes) selected from an array of four evenly spaced electrodes (Figure 3.3A). Using the array allowed us to switch the electrodes across which stimulation was delivered and thus activate different parts of the fiber tract independently without moving the electrodes and risk losing the cell. Stimulation currents were delivered using an ISO-Flex Stimulator (A.M.P.I.). The current pulse was 0.2ms in width and varied in amplitude. The stimulating electrodes were placed on parasagittal slices ~1mm dorsal to RA on the fiber tract connecting HVC to RA (tractus archistriatalis)(Figure 3.3A). The stimulus intensity was gradually increased from 10μA to 1mA. At the lowest intensities no current was typically evoked, but as stimulation intensity gradually increased, the evoked postsynaptic currents (EPSCs) also increased in amplitude, until reaching a maximum.

**Data Analysis**

**Song analysis**

All song recording, filtering, and segmentation was performed as previously described (Otchy et al., 2015) using custom written software in MATLAB. Briefly, songs were segmented and vocalization types
(i.e. syllables and calls) annotated using a semi-automated routine. The dominant song motif for each bird was determined by visual inspection of song spectrograms.

Song variability. Quantification of acoustic variability across renditions was done as in as in (Ölveczky et al., 2005, 2011), using a custom MATLAB implementation of a previously described method for calculating acoustic similarity (Tchernichovski et al., 2000). Briefly, pairwise comparisons of the similarity of acoustic features of identified syllables were made. This score ($S$, ranging from 0-100) was converted, through a linear remapping, to a variability score ($V$) by the following formula:

$$V = \frac{S_{\text{max}} - \langle S \rangle}{S_{\text{max}} - \langle S_{\text{min}} \rangle}$$

$\langle S_{\text{min}} \rangle$ is the average similarity score of randomly chosen pairs of syllables from unrelated birds. The similarity of identical syllables, $S_{\text{max}}$, is 100 by definition of the similarity measure. Thus, a variability score of 1 means that syllables are as different as two unrelated syllables, while variability score of 0 means that the syllables are identical. Error bars for in Figure 3.1 denote standard error of the mean.

Extracellular data analysis

Spike sorting and alignment. Spikes were sorted off-line using custom Matlab software. All units included in the analysis had signal-to-noise ratios greater than 5:1. Single-unit signals were verified by a spike refractory period in the interspike interval histogram. Spike were aligned to the dominant song motif as described in (Otchy et al., 2015). Briefly, dynamic time warping (DTW) algorithm was used to align individual song motifs to a common template as previously described (Ali et al., 2013). The warping path derived from this alignment was then applied to the corresponding RA recording spike times with a premotor lead of 15 ms (Leonardo and Fee, 2005; Ölveczky et al., 2011).
Spike train correlation. Precision of the song-aligned spike trains was measured using average pair-wise correlation across all pairs of spike train for a given condition. Spike trains were converted into instantaneous firing rates $R(t)$ as follows:

$$R(t) = \frac{1}{t_{i+1} - t_i};$$

for $t_i < t \leq t_{i+1}$,

where $t_i$ is the $i^{th}$ spike. These instantaneous firing rates were then convolved with a 8 ms Gaussian function (Leonardo and Fee, 2005), yielding a smoothed firing rate function $r(t)$. The correlation coefficient (CC) was then calculated between these firing rate functions for all pairs of spike trains as follows:

$$CC = \frac{1}{N_{pairs}} \sum_{i=1}^{N} \sum_{j \neq i}^{N} CC_{ij},$$

$$CC_{ij} = \frac{\langle \hat{r}_i(t) \cdot \hat{r}_j(t) \rangle_t}{\sqrt{\langle \hat{r}_i^2(t) \rangle_t \cdot \langle \hat{r}_j^2(t) \rangle_t}},$$

where $\hat{r}(t)$ is the mean-subtracted smoothed firing rate function.

Burstiness. To quantify burstiness, we looked at fraction of spikes for which the instantaneous firing rate was above 150 Hz.

Sparseness. To quantify sparseness of a firing pattern we used the entropy method (Lehky et al., 2005; Tolhurst et al., 2009), which gives a measure of how selective neural activity is for specific times in the song. For each neuron we calculated song-aligned rate histogram (3 ms bins). Histograms were normalized to generate a time-varying probability spike density function, $p_i$, where the $i^{th}$ value indicates the normalized firing probability for that time bin, such that $\sum_{i=1}^{N} p_i = 1$. We the computed the sparseness index (SI) as follows (Lehky et al., 2005):
This index is 1 (maximal sparseness) when the activity is restricted to a single time bin, and 0 if the spikes are evenly distributed across the time bins.

**Intracellular data analysis**

All analysis was performed as in (Garst-Orozco et al., 2015).

**Single fiber currents.** The recorded current traces were smoothed using a 1ms sliding window. The threshold for detecting a unitary EPSC single fiber (SF) input was two times the RMS of the signal acquired without any stimulation. The RMS (or ‘noise’) of our recordings was 2-5pA, meaning that SF current < 4pA were not registered. SF currents were measured at the stimulus intensity that produced 25-75% failures (a pre-set criteria) relative to EPSCs of consistent amplitude. The SF was defined as the average peak EPSC (N>3) evoked at that minimal stimulus intensity (Figure 3.3B).

**Maximal currents.** MAX currents were characterized at the stimulus intensity (MAX-stim) where EPSC peak amplitude no longer increased despite a greater than 3-fold increase in stimulus intensity. MAX current was the average peak response evoked at MAX-stim (N>3; CV≤0.2). MAX currents were always evoked by stimulating the fiber tract across the outermost stimulating electrodes.

**Input number.** The relative number of inputs to a cell (Figures 3.3G) was estimated from the ratio of the MAX current of a cell to the mean SF current for the given age category. The numbers from this analysis represent a lower bound on the average number of inputs to cells at a given age. Thus rather than represent the number of inputs, the analysis was used to compare relative changes in connectivity across different ages.

**Statistical analysis**
All statistics on data pooled across animals is reported in the main text as mean ± SD and depicted in figure error bars as mean ± SEM. To determine significance of the trends we observed, we calculated the correlation coefficient between the given measure (e.g. sparseness, firing rate, etc.) and age of the birds. The p-value of the correlation is the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero. All other tests of significance were student’s t-test unless otherwise noted. Figure starring schema: *p<0.05, **p<0.01, ***p<0.001.
Part II
Chapter 4:

Off-target effects of neural circuit manipulations

This chapter was previously published as:


Attribution: This chapter contains collaborative work.

B.P.Ö. and T.M.O. designed the study with input from all authors. T.M.O. collected and analyzed the data from songbirds with help from A.K. S.M.H.G. did initial pilot experiments in songbirds that inspired the study. C.P. implemented the HVC network model. S.B.E.W. performed the optogenetics experiments in rats and analysed the data. J.Y.R. and R.K. performed the pharmacological inactivation experiments in rats and analysed the data. B.P.Ö. supervised and coordinated the project. B.P.Ö., T.M.O., and S.B.E.W. wrote the paper with input from the other authors.
Abstract

Rapid and reversible manipulations of neural activity in behaving animals are transforming our understanding of brain function. An important assumption underlying much of this work is that evoked behavioral changes reflect the function of the manipulated circuits. We show that this assumption is problematic because it disregards indirect effects on the independent functions of downstream circuits. Transient inactivations of motor cortex in rats and nucleus interface (Nif) in songbirds severely degraded task-specific movement patterns and courtship songs respectively, learned skills that recover spontaneously after permanent lesions of the same areas. We resolve this discrepancy in songbirds, showing that Nif silencing acutely affects the function of HVC, a downstream song control nucleus. Paralleling song recovery, the off-target effects resolved within days of Nif lesions, a recovery consistent with homeostatic regulation of neural activity in HVC. These results have implications for interpreting transient circuit manipulations and for understanding recovery after brain lesions.
Introduction

Understanding how the brain generates behavior is a daunting task often simplified by studying anatomically distinct brain regions in isolation. The underlying assumption is that different parts of the brain are specialized for different functions that can be understood by monitoring and altering activity in local circuits. An increasingly powerful and widely used approach is to transiently silence or otherwise perturb - by optogenetic, pharmacological, or other means - neural activity in specific circuits and observe the consequences on behavior (Lomber, 1999; Zhang et al., 2007). If there is an effect, the conclusion is that the circuit under investigation is causally 'involved' in the behavior. But what does such a causal link actually tell us?

In a densely interconnected dynamical system like the brain, sudden perturbations to one node (e.g. brain area) could send ripples through the system, compromising the capacity of downstream circuits to perform computations on other inputs or generate patterned activity from internal dynamics (Carrera and Tononi, 2014; Honey and Sporns, 2008). Given the reliance on transient circuit manipulations for localizing computations and memory functions (Lomber, 1999; Zhang et al., 2007), the caveats and limitations of these methods should be scrutinized.

If inactivating a brain area interferes with the independent functions of downstream circuits, an important next question is whether those functions remain compromised after the silencing is made permanent (e.g. through lesions). For example, deficits caused by changes in the excitability of downstream neurons could plausibly resolve through homeostatic regulation of neural activity (Golowasch et al., 1999; Keck et al., 2013; Marder and Goaillard, 2006; Thoby-Brisson and Simmers, 2002; Turrigiano, 1999). Spontaneous recalibration of neural dynamics, more generally, could help explain why chronic effects of permanent lesions are often far less severe than those induced by transient inactivations (Bender and Baizer, 1990; Talwar et al., 2001; Wilke et al., 2010), and why
patients with strokes and other brain injuries can overcome some of their initial deficits without rehabilitation (Peppen et al., 2004).

Functional recovery after brain lesions, however, is thought to be driven predominantly by the adoption of new behavioral strategies and the adaptive repurposing of non-lesioned circuits (Newsome and Pare, 1988; Talwar et al., 2001), processes contingent on renewed experience with the affected tasks (Maldonado et al., 2008). Therefore, demonstrating acute off-target effects of inactivations and spontaneous recovery after permanent lesions requires showing that task-specific behaviors sensitive to transient inactivations can recover after lesions without additional task experience. Since experience-dependent recovery is difficult to rule out for basic sensory or motor functions that are central to many behaviors and hence naturally ‘practiced’ after lesions (Bender and Baizer, 1990; Newsome and Pare, 1988; Wilke et al., 2010), our study utilized behaviors for which such incidental practice can be withheld. We chose learned movement sequences of rats (Kawai et al., 2015) and courtship songs of zebra finches (Immelmann, 1969) because they are task-specific skills associated with complex, stereotyped, and idiosyncratic motor patterns that can be precisely quantified and compared across various manipulations.

To probe the effects of transient manipulations on the independent functions of downstream circuits, we targeted brain areas – motor cortex in rats and sensorimotor area Nif in songbirds – that are known, based on lesion studies, to be dispensable for storing and executing the skills we study (Cardin, 2004; Kawai et al., 2015). Despite this, transient manipulations severely degraded the learned behaviors in both systems. These discrepancies were consistent with acute disruptions of downstream circuit function. Though we saw similar behavioral effects immediately after permanent lesions, they resolved spontaneously, leading to full recovery of the initially affected behaviors.
Results

Motor cortex inactivation disrupts skill execution

In our motor learning task, rats are rewarded for pressing a lever in a precise temporal sequence (two presses 700 ms apart, Figure 4.1a). Animals solve this task by acquiring spatiotemporally precise movement patterns that produce the prescribed lever-press sequence (Kawai et al., 2015). Though the learned skills are robust to motor cortical lesions (Kawai et al., 2015), motor cortex projects to subcortical motor structures whose independent functions could be sensitive to sudden changes in motor cortical input (Figure 4.1b). To probe this, we inactivated primary forelimb motor cortex of rats that had learned the task (n=5 rats) by injecting 100 nL of the GABA_\_agonist muscimol (1-25 mM) into the hemisphere contralateral to the dominant paw (Huber et al., 2012; Peters et al., 2014) (Methods). Based on previous studies (Martin, 1991) and injections of 100 nL fluorescein into motor cortex (Figure 4.1c), we estimate that the direct effects of our injections were confined to a volume far smaller than our previous lesions (Kawai et al., 2015) (Figure 4.1c; Methods).

In stark contrast to animals tested for the first time 5-10 days after lesions (Kawai et al., 2015), muscimol-injected rats had severe deficits in skill execution with dramatic drops in performance and disrupted paw kinematics (Figures 4.1d,e). These effects were evident even in the rat receiving the lowest concentration of muscimol (1 mM) (rat in Figure 4.1d). We later lesioned motor cortex in this rat, and, as previously reported (rat ‘Kansas’ in (Kawai et al., 2015)), saw no effect on skill execution when the rat was tested again 10 days post-lesion (Figure 4.1d).

To explore dose-dependence, we injected larger volumes of muscimol in two of the rats (200 nL and 400 nL respectively). We found task performance to be even more affected with lever-interactions restricted to a few single presses with no rewarded trials.
Optogenetic stimulation of motor cortex

Transient stimulation of neural activity is an alternative method for disrupting ongoing circuit dynamics, especially well-suited for interrogating processes associated with precise and reproducible neural

Figure 4.1: Motor skills that survive motor cortical lesions are acutely affected by transient activity manipulations.
(a) Our task trains rats to press a lever twice with a specified inter-press interval (IPI), typically 700 ms. (b) Schematic of the mammalian motor system. Motor cortex (MC-black) provides input to subcortical circuits (red shaded regions; BG-basal ganglia, CB-cerebellum, BS-brainstem, Th-Thalamus, SC-spinal cord). (c) Coronal sections comparing the spread of a fluorescent tracer matched in concentration and volume to our muscimol injections (left) with motor cortex lesions that leave the learned skills intact (right, Methods). (d) (Left) Forepaw trajectories for five consecutive trials in PBS and muscimol sessions for the rat that received the lowest dose of muscimol (100 nl, 1 mM) (Supplementary Movie 1). (Right) Motor cortex was subsequently lesioned in this rat. Paw trajectories for five consecutive trials from the last training session before and the first training session after the lesion. (e) Fraction of trials with IPIs within 20% of the target (‘successful’ trials) for different experimental conditions (n=5 rats). Lesion data from Kawai et al. shown for comparison (light grey bar). (f) Wireless optogenetic stimulation. (g) Paw trajectories for five consecutive trials for an example rat with and without optogenetic stimulation of motor cortex. h, Same as e, but with and without optogenetic stimulation of motor cortex (n=5 rats). Pooled data in this and subsequent figures shown as mean ± SEM. For tests of significance see Methods. Starring convention: *p<0.05, **p<0.01, ***p<0.001.
dynamics (Roberts et al., 2012). As with transient inactivations, sudden activation can also plausibly affect dynamics in downstream circuits. To probe the effect of transient motor cortex stimulation on skill execution, we used optogenetics, a widely adopted method for manipulating neurons in temporally specific ways (Zhang et al., 2006).

We expressed the optogenetic activator Chrime (Klapoetke et al., 2014) in motor cortex (n=5 rats; Figure B.1a,b; Methods), and stimulated the hemisphere contralateral to the dominant paw after animals had reached asymptotic performance on the task (Figure 4.1f, Figure B.1c). Neither brief (50 ms) nor sustained (1 second) optogenetic stimulation evoked visible motor responses during rest, suggesting that they were sub-threshold for movement initiation. However, both brief and sustained stimulation, triggered on the first lever-press in a trial, interfered with task performance and associated kinematics (Figures 4.1g,h, Figure B.2). Thus, similar to transient inactivations, disrupting normal activity patterns in motor cortex by optogenetic stimulation compromises the animals’ capacity to execute skills robust to permanent lesions.

Effects Nif lesions and inactivation differ

Our results suggested that behavioral effects of transient perturbations may overestimate the essential functions of targeted circuits. To examine whether this caveat should be considered more broadly, we similarly probed whether transient inactivations of sensorimotor nucleus Nif in zebra finches affect their courtship songs. While the song survives Nif lesions (Cardin, 2004), Nif sends excitatory projections to HVC, an essential part of the song control circuit believed to generate the temporal pattern for learned vocalizations through intrinsic network dynamics (Long and Fee, 2008) (Figure 4.2a, Figure B.3a).

We first confirmed the findings of previous lesion studies (Cardin, 2004) by injecting 27-36 nL of NMDA, an excitotoxin, bilaterally into Nif (n=5 birds) (Figure 4.2a, Figure B.3b). When Nif lesioned birds resumed singing two days after surgery, their songs were similar to pre-lesion (Figures 4.2b,c),
consistent with prior studies. Because birds did not sing within the first day of lesions, this result does not preclude short-term effects of Nif silencing (Cardin, 2004). To probe such acute effects, we injected 27 nL muscimol (50 mM) bilaterally into Nif of awake head-restrained adult birds (n=5 birds; Figures 4.2d,e, Figure B.4a; Methods).
Birds typically sang within 20 minutes of the injections, but their songs were severely degraded and reminiscent of subsong (Figures 4.2f,g), highly variable and unstructured utterances normally produced by juvenile birds at the start of vocal learning or by adult birds after bilateral HVC lesions (Aronov et al., 2008). The syllable duration distributions were similar to those reported in HVC-lesioned birds (Aronov et al., 2008) (Figure 4.2h), suggesting that Nif inactivation degrades song by indirectly affecting HVC dynamics.

To exclude the possibility that the behavioral effects were caused by diffusion of muscimol into HVC, we injected the same dose 300-500 µm dorsal to Nif but closer to HVC (n=5 birds), as well as smaller volumes (9 nL) into Nif (n=2 birds). The control injections above Nif did not affect song (Figures 4.2f,g), while the smaller Nif injections evoked effects similar to the larger dose (Figure B.4b).

**HVC dynamics recover after Nif lesion**

Transient circuit manipulations performed in both rats and songbirds revealed strong effects on skill execution not seen after permanent lesions (Figures 4.1, 4.2). The discrepancy could not be explained by experience-dependent relearning in lesioned animals because the skills recovered their idiosyncratic pre-lesion form without any intervening practice (Figures 4.1d,e, 4.2b,c) (Kawai et al., 2015). However, the acute behavioral deficits were consistent with activity manipulations in motor cortex and Nif indirectly affecting the independent functions of downstream circuits. These initially affected circuits, however, seemingly regained their capacity to execute the learned behaviors after permanent lesions (Cardin, 2004; Kawai et al., 2015).

Unlike in rats, where the neural circuits underlying the skills we assay have yet to be characterized, the vocal control circuits in zebra finches have been well delineated (Fee and Scharff, 2010), making it feasible to investigate downstream effects of local circuit manipulations. Based on our results and known anatomy (Figure 4.2), we hypothesized that Nif inactivations perturb vocal output by
removing excitatory input from HVC, thus compromising the function of this song-specialized premotor
Most lesion protocols require surgery, which suppresses singing for a day or two, i.e. exactly the time-frame during which we hypothesize that recovery in HVC function occurs. To monitor neural dynamics in HVC in the immediate aftermath of Nif lesions and to compare it to pre-lesion dynamics, we lesioned Nif in freely behaving birds while recording multi-unit neural activity in HVC (Ali et al., 2013). Stimulation electrodes targeted to Nif were implanted together with recording probes in ipsilateral HVC (Figure 4.3a; Methods). Nif was lesioned unilaterally by injecting 50 µA of current for 30-40 seconds (n=11 birds; Figure 4.3b). Nif was successfully ablated (>80% lesioned; Methods) in 4 out of 11 birds, and subsequent analysis was done on this cohort unless otherwise noted.

Spontaneous (i.e. non-vocal) HVC activity was dramatically reduced immediately following the lesions, consistent with a sudden loss of excitatory input from Nif (Hahnloser and Fee, 2007), but recovered in the ensuing hours (Figure 4.3c). Singing, which was invariably interrupted by the electrical stimulation, resumed after $1.3 \pm 0.9$ hours (Figures 4.3c-e). A fraction of the initial post-lesion vocalizations were severely degraded and did not resemble pre-lesion song (Cardin, 2004) (Figures 4.3d,e, Figure B.5). The effects were less severe than during bilateral Nif inactivation (Figure 4.2f,g), likely reflecting bilateral control of zebra finch song (Schmidt et al., 2004).

For vocalizations that resembled pre-lesion song, neural activity was aligned to a common song template (Methods). While song-aligned activity patterns were similar across renditions in the hours following Nif lesions, they were strikingly different from pre-lesion dynamics (Figures 4.3f-h).
Despite the initial degradation of song and associated HVC dynamics, both gradually recovered (Figures 4.3e-h). By the second day, the song was reliably back to pre-lesion form (Figure 4.3e). Remarkably, the average song-aligned activity patterns in HVC also recovered their pre-lesion structure (Figures 4.3f-h). By the third day, the residual difference was consistent with normal drift in the recordings. Interestingly, the temporal structure of the song-aligned HVC activity recovered
predominantly during the night, while song-related HVC power recovered during the day (Figure 4.3i; Methods).

To assess the extent to which acute post-lesion changes in HVC dynamics were caused by removal of Nif input versus non-specific effects of the current injections, we quantified changes in song-related HVC dynamics as a function of Nif lesion size. We found the extent of Nif damage to be strongly correlated with changes in song-related HVC activity following lesions (Figure 4.3j), consistent with the acute degradation of song and associated HVC activity being due to removal of Nif input to HVC.

**Activity homeostasis explains functional recovery**

The spontaneous and gradual recovery of HVC activity after Nif lesions (Figures 4.3c,h) was suggestive of homeostatic regulation of neural activity (Keck et al., 2013; Marder and Goaillard, 2006; Turrigiano, 1999). To probe whether this could explain the observed song recovery, we modeled HVC as a synaptically connected chain of neurons (a ‘synfire chain’) that receives time-varying excitatory input from Nif (Jin et al., 2007; Li and Greenside, 2006; Long et al., 2010; McCasland, 1987) (Figure 4.4a; Methods). The network generated stable propagation of synchronous spiking activity, much like what is assumed for HVC during singing (Hahnloser et al., 2002). Acute removal of Nif input prevented many neurons in the chain from reaching spiking threshold, causing activity propagation to slow and often stop prematurely (Figures 4.4c,d). Homeostatic regulation of neural activity in the HVC network was implemented by adaptively adjusting either spiking threshold (Watt and Desai, 2010) (Figure 4.4b), input resistance (Welie et al., 2004), or strength of synaptic inputs (Turrigiano et al., 1998) (Figure B.6) of individual HVC neurons (Methods). These mechanisms all had similar effects, increasing probability of HVC spiking while speeding up chain propagation and decreasing the likelihood of early ‘song’ terminations (Figures 4.4b-d, Figure B.6).
Given that HVC is known to control song timing (Long and Fee, 2008), our simulations generated predictions for the temporal structure of song following Nif lesions (Figure 4.4d). In agreement with these predictions, we found a transient increase in premature song terminations (Figures 4.4e,f, Figure B.5). Upon closer inspection of post-lesion song, we also found that the song slowed down (Cardin, 2004), only to recover in the ensuing days (Figure 4.4f), again consistent with the qualitative predictions of our model. Though we cannot exclude other mechanisms, our network simulations are consistent with functional recovery after Nif lesions being due to homeostatic regulation of neural activity in HVC.

**Discussion**

While efforts to understand the brain must necessarily rely on reductionist approaches, the simplifications and assumptions made in this pursuit must be scrutinized to prevent misleading conclusions. Given the increased reliance on transient circuit manipulations (e.g. optogenetics, pharmacology, pharmacogenetics, cooling, TMS, etc.) for localizing brain function, we tested whether behavioral effects induced by sudden activity perturbations reliably reflect the computations carried out in the targeted areas. In two different systems, the deficits induced by transient manipulations seemingly overestimated the role of the examined circuits (Figures 4.1, 4.2). This could be explained by the manipulations affecting independent functions of downstream circuits. We found that such off-target effects can resolve after the targeted area is permanently lesioned (Figure 4.3). Importantly, post-lesion recovery did not require any renewed experience with the task, and was consistent with homeostatic regulation of neural activity (Figure 4.4).

Discrepancies between acute and chronic behavioral effects of targeted inactivations/lesions have been recognized in other contexts (Bender and Baizer, 1990; Newsome and Pare, 1988; Talwar et al., 2001; Wilke et al., 2010). Acute effects are almost invariably more severe, a discrepancy typically explained by the brain adaptively compensating for lost function after lesions (Maldonado et al., 2008).
By not allowing time for experience-dependent compensation, transient circuit manipulations are seen as overcoming this ‘caveat’ of lesions. However, if the goal is to assign computations and memory functions to specific brain areas, our results suggest that transient circuit manipulations may have their own interpretive difficulties that stem from acute effects on the independent functions of remote circuits.

That the function of a circuit can be sensitive to sudden perturbations in chronically non-essential inputs is not surprising. The brain - a finely tuned, complex, and heavily interconnected dynamical system – operates in a fairly limited dynamic regime (van Vreeswijk and Sompolinsky, 1998), making it plausible that local perturbations to normal dynamics could interfere with the independent functions of remote circuits (London et al., 2010). For example, sudden removal of permissive inputs could tilt a network’s excitatory-inhibitory balance (Shu et al., 2003), thus compromising its function (Feeney and Baron, 1986). This is seemingly what happens to HVC after Nif is silenced. Loss of excitatory input from Nif causes an acute decrease in the activity of HVC neurons (Hahnloser and Fee, 2007), rendering the network incapable of producing its normal output (Figures 4.3f,g, 4.4).

The intricacies of dissecting interconnected biological networks and assigning functions to discrete nodes in those networks have been recognized in other contexts, including genetic and molecular networks (Phillips, 2008). In such studies, the distinction between permissive and instructive functions is routinely made (Miyashita et al., 2008; Shobe, 2002). Our results suggest that a similar distinction should be considered when interrogating the role of neural circuits in behavior (Taha and Fields, 2006), with a circuit being classified as ‘permissive’ if its activity is acutely required for the expression of a behavior without providing essential information for any of the underlying computations or memories. In contrast, a brain area should be considered ‘instructive’ if it contributes essential information or computation not otherwise available to the system implementing the behavior.
While the behavioral effects of sudden activity perturbations may not reliably reflect the steady-state function(s) of a circuit, lesions can, in certain cases at least, contribute additional insight. Permanent silencing of Nif and motor cortex suggested that the capacity of these brain areas to influence the respective skills we study - evident from transient manipulations - is not exercised under normal conditions, consistent with permissive roles. That Nif and motor cortex have access to the essential control circuits likely reflects instructive roles for these brain areas in behavioral processes that we did not test. Nif, for example, provides early auditory priming of HVC essential for imitative song learning (Roberts et al., 2012), while motor cortical input to subcortical motor circuits is required for the initial acquisition of the skills we train (Kawai et al., 2015) and for modulating other low-level motor behaviors (Stoltz et al., 1999).

Importantly, neural circuit function acutely compromised by sudden changes in permissive input can recover after those inputs are permanently silenced. Both skills we studied recovered after lesions without any task-specific practice, suggesting largely spontaneous recovery processes. While the mechanisms that underlie such recovery will need to be further examined, our results are consistent with a role for homeostatic regulation of neural activity (Keck et al., 2013; Marder and Goaillard, 2006; Turrigiano, 1999) (Figure 4.4, Figure B.6). A similar recovery to the one we observed in HVC of songbirds (Figure 4.3) has been described for the network underlying the pyloric rhythm in crustaceans (Thoby-Brisson and Simmers, 2002), where homeostatic regulation of neuronal dynamics is thought to underlie the recovery of circuit function after removal of permissive or modulatory input (Golowasch et al., 1999).

Though our results are consistent with a relatively simple form of homeostatic regulation (i.e., regulation of neuronal activity), we cannot exclude the involvement of other mechanisms capable of restoring network function. For example, offline replay of activity (Buhry et al., 2011; Pavlides and Winson, 1989), which has been observed in songbirds (Dave and Margoliash, 2000; Hahnloser and Fee,
could serve as a mechanism for probing the state of affected circuits and restoring pre-lesion neuronal dynamics through something like spike time dependent plasticity. Alternatively, it is plausible that one or more of the several nuclei upstream of HVC (Akutagawa and Konishi, 2010; Hamaguchi and Mooney, 2012; Mooney, 2005; Nottebohm et al., 1982; Scharff and Nottebohm, 1991) could alter its dynamics and/or contributions to HVC to compensate for the loss of structured input from NIf (Lewandowski and Schmidt, 2011). Irrespective of the specific mechanism employed or circuit regime regulated, however, our results demonstrate that recovery must occur spontaneously (Figure 4.1d,e and 4.2b,c) – that is, in the absence of new task experience. So long as the recovery of downstream neural dynamics is spontaneous, we argue that it may reasonably be described as self-regulating (i.e., ‘homeostatic’) no matter which regime (i.e., neuronal excitability, structured input, feed-forward connectivity, etc.) is being regulated.

Interestingly, we found that the structure of song-aligned HVC activity recovered predominantly overnight, while overall HVC power recovered during the day (Figure 4.3i). This dissociation is consistent with the synaptic homeostasis hypothesis of sleep (Tononi and Cirelli, 2006) that posits synaptic potentiation during wakefulness and synaptic rescaling and memory consolidation during sleep. Our results suggest that sleep not only consolidates activity patterns associated with recent experiences (Walker and Stickgold, 2004), but may help restore previously established circuit dynamics, and could hence promote functional recovery after brain lesions (Siccoli et al., 2008).

As in our experimental animals, patients with lesions to motor-related brain areas have motor deficits that resolve in the days and weeks following the injury (Levin and Grafman, 2000). Aspects of this recovery are thought to be independent of rehabilitation (Peppen et al., 2004), suggesting spontaneous processes at work. Diaschisis is a broad clinical term referring to the temporary effects of focal brain lesions on remote brain areas (Carrera and Tononi, 2014), yet the underlying mechanisms remain poorly understood (Feeney and Baron, 1986). Our results suggest that focal brain lesions can
affect neural dynamics and function in remote brain areas, and that homeostatic regulation of neuronal
dynamics may help resolve such acute effects, thus contributing to functional recovery after brain injury.

**Methods**

**Animals**

The care and experimental manipulation of all animals were reviewed and approved by the Harvard
Institutional Animal Care and Use Committee. Experimental subjects were female Long Evans rats 3-8
months old at start of training (n=10, Charles River) and adult male zebra finches between 92-205 days
post-hatch (n=27). Because the behavioral effects of our circuit manipulations could not be pre-specified
before the experiments, we chose sample sizes that would allow for identification of outliers and for
validation of experimental reproducibility. No animals were excluded from experiments post-hoc.

**Rat experiments**

*Behavioral training in rats*

Ten rats were trained in the lever-pressing task as previously described (Kawai et al., 2015). Water-
restricted animals were rewarded with water for pressing a lever twice with a prescribed interval
between the presses (700 ms for 9 of the rats, 600 ms for one). All animals were trained using our fully
automated home-cage training system (Poddar et al., 2013). Kinematic tracking of forepaw movements
(Figures 4.1d,g) was done as in (Kawai et al., 2015).

*Motor cortex inactivations*

In rats (n=5) that had reached asymptotic performance in our task (Kawai et al., 2015), a craniotomy was
made to access the caudal forelimb area of primary motor cortex (CFA) in the hemisphere contralateral
to the paw most involved in the lever-press sequence. The center of the CFA was estimated from
stereotactic coordinates (+1.0 mm anterior, +3.00 mm lateral, with respect to Bregma (Neafsey et al., 1986). Kwik-Kast sealant (WPI) was applied to cover the exposed dura. In addition, a protective acrylic cap covering the craniotomy was attached with screws to three nuts secured to the skull with Metabond (Parkell). After recovering from surgery (10 days), animals were trained for at least one additional week to ensure that they were at their asymptotic performance levels.

On injection days, rats were lightly anesthetized with 0.5-1.5% isoflurane and placed in a stereotax. Motor cortex was accessed by removing the custom-made protective cap and the Kwik-Kast plug covering the craniotomy. Muscimol (or PBS for control) was injected at the estimated center of the CFA (Neafsey et al., 1986), 1.5 mm deep, in 9.2 nL increments every 10 seconds using a Nanoject (WPI). The craniotomy was resealed with Kwik-Cast and the protective cap reattached. The whole procedure took 15-30 minutes, and rats resumed normal behavior a few minutes later. Training sessions started 1.5 hours after the injections. ‘Baseline’ performance included sessions after the craniotomies but prior to injection. Experimental days alternated between saline and muscimol injections. To prevent any behavioral compensation in response to muscimol-induced performance deficits, injected animals were tested for only 10 minutes.

The dosing of muscimol was based on two criteria. (i) To allow comparisons with our lesion study (Kawai et al., 2015), the direct effect of muscimol injections should be restricted to a volume of motor cortex equal to or smaller than what we lesioned in (Kawai et al., 2015) (Figure 4.1c). (ii) To quantify effects on kinematics and performance, drug dosing should not abolish task engagement. We injected increasing concentrations and volumes of muscimol in two rats, and found that relatively larger doses (200 – 400 nL of 25 mM muscimol) degraded performance to the point where animals quit the task.

We converged on a dose of 100 nL of 25 mM muscimol because it generally did not prevent engagement with the task and because we estimated that it affects a volume significantly smaller than
what we lesioned in (Kawai et al., 2015). Our estimate of muscimol spread is bounded by previous studies that injected larger doses of muscimol (Allen et al., 2008; Martin, 1991) (1 µL of 2 and 9 mM respectively) and showed affected volumes of ~4-14 mm³. In comparison, our motor cortex lesions were larger than 23 mm³ (Kawai et al., 2015). We also injected 100 nL of 25 mM fluorescein in one animal using the same protocol as for the experimental animals, sacrificing it 1.5 hour after the injection. Its brain was later sectioned and the approximate spread of the dye visualized using fluorescence microscopy (Figure 4.1c). One of the experimental animals had very severe performance deficits at our chosen dosing, preventing us from characterizing the behavioral effect (i.e. very few successful trials). In this animal we reduced the concentration to 1 mM, at which the task engagement was robust but the performance still affected (Figure 4.1d).

**Optogenetic stimulation**

*Viral injections.* Adeno-associated virus (AAV2/8-hSyn-FLEX-ChrimsonR-tdTomato, UNC vector core (Klapoetke et al., 2014); titer: 5x10¹² vg/mL) was injected into the forelimb motor cortex of isoflurane anesthetized rats (n=5) through multiple small craniotomies (A/P, M/L: +1, ±2; +1, ±4; +1.5, ±2.75; +2.25, ±2.5; +3, ±2, coordinates relative to Bregma). Injections were done in 9.2 nL increments while slowly moving the injection-pipette (Nanoject) from a depth of 1.5 mm to 0.7 mm for a total volume of 0.4 µL per site and 2 µL per hemisphere (Figure B.1a). Animals were allowed to recover for 5 days before starting behavioral training.

*LED implant and stimulation.* Once animals reached asymptotic performance on our task (Kawai et al., 2015), they underwent a second surgery to implant a custom-built device for optogenetic stimulation (Figure B.1c). The custom-built device consisted of a red LED (λ = 615 nm, 110 mW output power, XLAMP XPC LED RD-ORANGE, Cree) on a printed circuit board, powered by two coin-cell batteries (CR2032). An infrared (IR)-sensitive photodiode was used to wirelessly control the LED. After device
implantation and recovery, animals resumed behavioral training. An IR light-source placed on top of the training cage was activated to trigger the LED for the duration of the optogenetic stimulation (1 second or 50 ms, continuous light). Stimulation trials were at least 10 seconds apart to allow the batteries of the LED to recover. Between stimulation trials, rats performed a varying number of non-stimulated trials (range: 0-5), resulting in ~30% of the trials being ‘stimulated’. Optogenetic stimulation was repeated for several sessions (5-12). Batteries were changed daily.

Functional verification. To characterize the effects of optogenetic stimulation on motor cortex activity, acute electrophysiological recordings were done after termination of the behavioral experiments in two of the rats. The animals were anesthetized and fixed in a stereotactic frame. The implanted LED device was carefully removed to expose the previous craniotomy. Using a custom-built recording setup and silicon probes (Buzsaki-64, Neuronexus), we recorded single-unit activity in motor cortex below the craniotomy. The removed LED device was placed next to the silicon probe above the craniotomy. Once stable units were detected, we triggered the LED and illuminated motor cortex for 1 second (30 trials). Recordings were performed at multiple depths (0.1 mm to 2.5 mm). Units were classified as light responsive if at least two consecutive bins of 5 ms during the first 200 ms of illumination had a significant z-score (compared to 1 second of baseline before light onset). These included units with long onset latencies (>10 ms), consistent with indirect activation (Figure B.1b). The relatively high number of light responsive units (69 ± 3%), compared to the number of cells counted as infected by immunohistochemistry (31 ± 2%; see below), is likely due to such indirect effects. Moreover, many of the recorded light-responsive cells were only identified during stimulation, further biasing our results to responsive cells.

Histological verification. At the end of the experiments, animals were transcardially perfused with PBS and subsequently fixed with 4% paraformaldehyde (PFA) in PBS. Brains were removed and post-fixed for at least 24 hrs. Brains were sliced coronally (thickness: 80 µm) and immunohistochemistry performed to
determine the AAV injection site and extent of the transfection (Figure B.1a). Slices were blocked (1% BSA, 0.3% Triton in PBS) at room temperature and incubated with anti-RFP (Chicken, 1:1000, Millipore, AB3528) and anti-NeuN (Mouse, Millipore, MAB377) primary antibodies in blocking buffer for 48 hrs at 4°C. After washing, slices were incubated with anti-Chicken-Alexa 568 (goat, 1:1000, Life Technologies, A-11041) and anti-Mouse-Alexa 647 (goat, 1:1000, Life Technologies, A-31625) overnight at 4°C. Slices were mounted and imaged using a Zeiss Axio Scan Z1 Slide Scanner for overview images and an Olympus FluoView FV1000 confocal microscope for high-resolution images. We determined the spread of the AAV injection based on the fluorescent signal (8.2 mm³ ± 1.3 mm³). In addition we chose 4 regions of interest (size 635 x 635 µm) and counted the number of infected cells relative to the number of neurons (NeuN⁺ cell) to determine the fraction of infected cells (31 ± 2%; n = 2 rats ). Histology was done blind to the outcome of the experiment.

**Data Analysis**

To assess the behavioral effects of the different injections, we measured performance relative to ‘baseline’ training sessions after the craniotomies but prior to any injections. To standardize analysis across experimental conditions in the acute inactivation experiments (muscimol, PBS, or baseline), we only included data from the first 10 minutes of each session, matching the duration of the muscimol sessions. For the optogenetic stimulation experiments, ‘baseline’ was defined as the non-stimulated trials in the same sessions. The number of sessions for each condition ranged from 1-3 (injections) and 5-12 (optogenetic stimulations). Data from training sessions of a given condition were pooled for each animal. To quantify behavioral performance, the fraction of trials with an inter-press interval (IPI) within 20% of the target IPI was calculated. Data on motor cortex lesioned animals presented for comparisons in Figure 4.1e comes from previously published experiments (Kawai et al., 2015), and includes sessions from the second week of post-lesion training.
Zebra finch experiments

All birds were obtained from the Harvard University zebra finch breeding facility and housed on a 13:11 hr light/dark cycle in acoustic isolation with food and water provided ad libitum.

Pharmacological lesions

Birds (n=5) were anesthetized with isoflurane. Nif was localized antidromically by electrical stimulation in HVC (Ali et al., 2013). Bilateral Nif lesions were made by injecting the excitotoxin N-methyl-DL-aspartic acid (NMDA, 4%) into each hemisphere using a Nanoject (WPI). In initial experiments, a single 27 nL bolus of NMDA was injected into the center of Nif. Though this volume produced complete bilateral Nif lesions in one animal, we found that complete lesions were more reliably produced by injecting two boluses of 18 nL (for a total of 36 nL) 200 µm apart along the anterior-posterior axis. We report on the five animals that had 100% bilateral Nif lesions, determined by post-hoc histological inspection (see below, Figure B.3b). One of these received 27 nL and 4 received 36 nL injections.

Reversible inactivations

Birds (n=5) were anesthetized and Nif identified as described above. Craniotomies over Nif were covered with artificial dura (Body Double Fast; Smooth-On, Inc.) and head screws were attached to the skull with dental cement as previously described (Ölveczky et al., 2005). Following post-surgical recovery, awake birds were placed in a foam restraint and head-fixed to a stereotax for ~10 min each morning for 10-14 days to desensitize them to handling and restraint. Following the desensitization training, all birds reliably sang within 30 minutes of the restraint. In the morning of experimental days, muscimol (27 nL, 50 mM) or PBS (27 nL) was injected bilaterally as described in the text. In two birds that routinely sang within 10 minutes of drug administration, we also injected a smaller dose (9 nL) of muscimol into Nif to additionally verify that song degradation was due to direct inactivation of Nif (Figure B.4b).
We note that a previous study aimed at reversibly inactivating Nif in adult songbirds failed to show any obvious effect on song structure (Naie and Hahnloser, 2011), but conflicting results from experiments in juvenile birds and methodological uncertainties regarding drug injection volumes in adult birds make its conclusions tentative. This previous report notwithstanding, all our muscimol injections into Nif produced similarly severe song degradation (Figure 4.2f, Figure B.4b).

**Implantation of recording and stimulation device**

Zebra finches (n=11) were anesthetized with isoflurane and placed in a stereotax. HVC was identified by antidromic stimulation from Area X as previously described (Ali et al., 2013). Nif was similarly identified by stimulating in HVC. For birds targeted for electrolytic Nif lesions, we placed either a monopolar stimulating electrode at the dorsal-posterior edge of Nif (n=8 birds) or a bipolar stimulating electrode straddling Nif in the medial-lateral plane (n=3 birds). A custom recording array (3 channels; 100 kΩ) was implanted in the hemisphere ipsilateral to the Nif-stimulating electrode and within the identified boundaries of HVC as previously described (Ali et al., 2013). All birds exhibited normal song output within 7 days of surgery. Following completion of the experiment, animals were sacrificed, their brains harvested, and the placement of recording electrodes and extent of lesions confirmed histologically.

**Neural and behavioral recordings**

Sound and neural activity were recorded using a custom LabVIEW application (National Instruments) as previously described (Ali et al., 2013). Multi-unit neural activity was recorded from up to three sites in HVC (~250 µm spacing) for three to four weeks per bird. Because stability of the neural recordings is crucial for estimating recovery in HVC dynamics, analysis was done on data collected at the most stable recording site in each bird (determined pre-lesion), though we note that the trends were similar across all channels.
Electrolytic lesions

Electrolytic lesions of Nif were made in the right hemisphere by passing 50 µA of monophasic current through the stimulating electrode for 30-40 seconds. Current injections started while birds were singing, and in all cases immediately terminated song output. Lesion extent was estimated post-hoc as described below.

Histological verification of lesions and inactivations

At the end of the experiments, birds were anesthetized with natriumpentobarbital (Nembutal, IM) and transcardially perfused with PBS, followed by fixation with 4% PFA in PBS. Brains were removed and post-fixed in 4% PFA overnight. Parasagittal sections (75 µm) were cut on a Vibratome (Leica), mounted, and stained with cresyl violet to reconstruct the location of implanted electrodes and lesions (ImageJ). Identification of the injection sites for the muscimol inactivations and circuit tracings were done in alternate brain slices by fluorescence microscopy (Figure 4.2e, Figure B.4a). Histology was done blind to the identity of the animals. Nif was identified based on regions of stronger staining and higher cell density than surrounding areas and were additionally guided by proximate anatomical landmarks (e.g., HVC, the lamina mesopallialis and the lamina pallio-subpallialis).

Lesions. Location and size of the lesions were determined by estimating the extent of necrotic tissue (i.e., loss of neurons and gliosis) in photomicrographs of cresyl violet stained sections as previously described (Roberts et al., 2012). Lesion size was expressed as a percentage of estimated Nif size, measured in intact controls (0.035 ± 0.001 mm³, n=4 birds). In pharmacologically lesioned birds, 100% of Nif was lesioned (Figure B.3b). In electrolytically lesioned birds, 0 - 100% of Nif was lesioned (Figure 4.3j).
**Inactivations.** Fluorescent dye-conjugated dextran (0.5 mg/mL Alexa 594; Invitrogen) were co-injected with the final injection of muscimol for post-hoc verification of the injection site (Figure 4.2e, Figure B.4a). Fluorescence images of the sections were superimposed on those of their adjacent cresyl violet sections (Adobe Photoshop) to determine locations of fluorescence in relation to Nif. All injection sites were found to be within the target nucleus (Figure B.4a).

**Neural circuit tracing**

To visualize Nif (Figure 4.2e, Figure B.3a), fluorescent dye-conjugated cholera toxin subunit B (1 mg/mL, Alexa 488; Invitrogen) was injected into HVC in 2 birds (81 nL per hemisphere). Twenty-one days after surgery, the animals were sacrificed, perfused, and their brains fixed, sectioned, and mounted. Photomicrographs of fluorescent sections were overlaid on those of adjacent cresyl violet sections (Adobe Photoshop) to determine location of fluorescence in relation to anatomical landmarks and density of cell bodies.

**Data analysis - song**

**Syllable segmentation and annotation.** Raw audio recordings were segmented into syllables as previously described (Ali et al., 2013). Spectrograms were calculated for all prospective syllables, and a neural network (5000 inputs, 100 hidden layers, 3-10 output neurons) was trained to identify syllable types using a test dataset created manually by visual inspection of song spectrograms. Accuracy of the automated annotation was verified by visual inspection of a subset of syllable spectrograms.

**Syllable feature quantification.** All non-call vocalizations were characterized by their duration and mean Wiener entropy – both robust acoustic features that are tightly controlled in adult zebra finch song (Ravbar et al., 2012). Syllable durations were estimated from threshold crossings of the acoustic power as previously described (Ali et al., 2013). Wiener entropy, a measure of acoustic randomness, was
calculated using Sound Analysis for MATLAB (Tchernichovski et al., 2000) for 10 ms time windows, advancing in steps of 1 ms, such that entropy was computed for every millisecond. The entropy measurements were averaged across the syllable and log-transformed. On this scale, the Wiener entropies of white noise and of a pure tone are zero and minus infinity, respectively.

Duration probability distributions. Histograms (1.25 ms bins) of syllable durations produced within 1 hour of muscimol/PBS injections were generated for each experiment, normalized by total sample counts, averaged across 2-4 experiments within a bird, and then averaged across birds. Data from HVC-lesioned birds, provided by the authors of Aronov et al. (Aronov et al., 2008), was recorded on the first day of singing after lesion (2-7 days after surgery) and analyzed similarly. Mean duration distributions for all conditions were smoothed with a sliding boxcar window (7-bin width, 1-bin advance).

Entropy-duration joint probability distributions. Two-dimensional histograms, showing the joint distributions of syllable duration and Wiener entropy, were created with bins of width 1.25 ms (duration axis; range: 0-300 ms) and 0.025 (log Wiener entropy axis; range: -4-0). The histogram was normalized by total sample counts to construct an empirical probability distribution. Because these empirical distributions were sparsely sampled, we estimated the true probability distribution by smoothing the empirical distribution with a point-spread function (2-D Gaussian; width: 7 bins; sigma: 3 bins). Distributions were calculated for vocalizations produced during the following time windows: bilateral lesion experiments – the first 2 hours of singing each day; inactivation experiments – 2 hours before (pre), 1 hour after (post), and 6-8 hours after (washout) injection; unilateral lesion – the first 2 hours of singing each day, the first hour of post-lesion singing, and the last 4 hours of singing on the day of lesion.

Distribution similarity measurement. To quantify changes in song elements, we calculated the 1st Wasserstein distance, a common metric of the difference in probability distributions, between syllable entropy-duration distributions for songs produced at different time points or under various experimental conditions (see text). We used an algorithmic implementation in MATLAB and C available
at (http://www.ariel.ac.il/sites/ofirpele/FastEMD/). Distances between bins were Euclidean. Calculations were based on 50,000 samples drawn from the entropy-duration probability distributions and reported in figures as the mean distance per sample.

Motif completion rate. For each bird, a 3-5 syllable dominant song motif was identified by visual inspection of spectrograms. Motif completion rates (MCR) were calculated as:

$$MCR = \frac{\text{# of utterances of complete motifs}}{\text{# of utterances of the first syllable in the motif}}$$

For all birds, motif completion rates were calculated for the first two hours of singing per day; for unilaterally lesioned birds, rates were also calculated for the first hour of singing following lesion. ‘Intact’ motif completion rates (Figure 4.4f) were based on a subset of the lesioned birds (four from the ‘Bilateral’; three from ‘Unilateral’ group) but collected 1-2 weeks prior to the Nif lesions. Data from each bird was normalized to pre-lesion motif completion rates for comparison across animals. See Figure B.5 for examples of truncated motifs.

Motif duration stretch. The durations of the dominant song motifs were calculated as previously described (Ali et al., 2013) for interval durations. For all birds, the mean motif duration was calculated for 100 consecutive renditions, taken at the same time each day (~1 hour after lights on in the morning). For unilaterally lesioned birds, the mean duration was also calculated for the first 100 identifiable motifs produced immediately after lesion. As noted above, ‘Intact’ data were collected from birds that were later lesioned. Motif durations were normalized to pre-lesion values for comparison across animals. See Figure B.5 for examples of aligned and excluded vocalizations.

Data analysis – neural recordings

Spontaneous activity. To record spontaneous HVC activity, minute long recordings were made every 15 minutes. These recordings were bandpass filtered (1-5 kHz; 2-pole Butterworth; zero-phase) and segments within 500 ms of vocalization-related activity were marked for exclusion from subsequent
Individual spikes were detected by an amplitude threshold set to 3-8 standard deviations of the estimated noise in the recordings. For each bird, the spontaneous firing rates were normalized to the mean firing rate in the two hours prior to lesion. Shown in Figure 4.3c is the across-bird mean and standard error, smoothed with a sliding boxcar window (5 bin width, 1 bin advance).

Alignment of the neural recordings to song. A dynamic time warping (DTW) algorithm was used to align individual song motifs to a common template as previously described (Ali et al., 2013). The warping path derived from this alignment was then applied to the corresponding HVC recordings with a premotor lead of 35 ms (Ali et al., 2013). The aligned neural traces were squared (to calculate signal power) and smoothed (5 ms boxcar window, 1 ms advance).

HVC activity correlation. The recovery of temporal dynamics in HVC was calculated as the Pearson’s correlation between the song-aligned neural power (averaged over 25 consecutive motifs) immediately before lesion and the same at different times after lesion. The pre-lesion data point in Figure 4.3g represents the correlation between the mean power envelopes for two consecutive blocks of 25 motifs recorded immediately prior to lesion. Normal drift in the song-related HVC signal ('control') was calculated similarly. The running correlation in Figure 4.3f shows Pearson’s correlation between the mean song-aligned activity pattern of pre-lesion songs on the day of lesion and the mean activity patterns in a sliding window of 25 song motif.

HVC mean power. The mean HVC power was calculated per motif and averaged over the 25-motif windows as described above for the correlation. For analyses pooled across birds, mean HVC power was normalized to the pre-lesion value.

Day versus night recovery. Recovery of HVC activity in the first 60 hours following lesion, during which most of the post-lesion recovery occurred, was parsed into 3 daytime and 2 nighttime intervals. Daytime recovery was calculated as the change in correlation to pre-lesion activity (or normalized mean power) between the first 25 motifs in the morning (or immediately following lesion) and the last 25 motifs that
evening; nighttime recovery is the change between the last 25 motifs of the day and the first 25 of the subsequent morning.

**Modeling**

*Network architecture.* Based on previous experimental findings (Hahnloser et al., 2002; Long et al., 2010), we modeled the HVC network as a synfire-chain of bursting neurons. The model consisted of 1200 integrate-and-burst neurons organized into 80 nodes. Each of the 15 neurons in a node projected to all neurons in the next node, forming a chain topology. The subthreshold membrane potential of the \(i^{th}\) neuron, \(V_i\), obeys:

\[
C \frac{dV_i}{dt} = -g_L(V_i - V_L) + I_{syn,i} + I_{Nif,i} + \sqrt{\tau_\eta} \sigma \eta_i(t),
\]

where \(C = 1 \mu F/cm^2\) is the membrane capacitance, \(g_L = 0.1 mS/cm^2\) is the leak conductance, \(V_L = -60 mV\) is the leak potential, \(I_{syn,i}\) is the synaptic input, \(I_{Nif,i}\) represents external input to the HVC neurons from Nif, \(\eta_i(t)\) is a zero-mean Gaussian white noise with covariance \(\langle \eta_i(t) \eta_i(t') \rangle = \delta(t-t')\), \(\tau_\eta = 10 ms\) and \(\sigma = 200 nA/cm^2\). The synaptic input is given by \(I_{syn,i}(t) = W \sum_j M_{ij} \sum_{k} \epsilon(t - t_{jk})\), where \(t_{jk}\) denotes the \(k^{th}\) spike of \(j^{th}\) neuron, \(\epsilon(t) = \Theta(t)e^{-t/\tau_s}\) with \(\tau_s = 5 ms\), \(W = 87 nA/cm^2\) and \(M_{ij}\) is 1 for synapses from a neuron to the neurons in the next node and 0 otherwise. The Nif input is a different waveform for each HVC neuron and does not change across simulations. The waveforms were randomly generated by simulating an Ornstein-Uhlenbeck process with an autocorrelation time scale of 50 ms, starting from a random initial point. Noise and drift were chosen such that the resulting waveforms had a mean of 97 nA/cm\(^2\) and standard deviation of 53 nA/cm\(^2\). When the membrane potential of the integrate-and-burst neuron reaches threshold, \(V_{th} = -50 mV\), the neuron emits 4 spikes with 2 ms intervals, modeling the bursts generated by calcium spikes in RA-projecting HVC neurons (Fiete et al., 2010; Long et al., 2010), and the membrane potential...
is reset to \( V_R = -55 \text{ mV} \) after a refractory period of 4 ms. Chain propagation was started by a 5 ms pulse input with magnitude 6.7 \( \mu \text{A/cm}^2 \) to the neurons in the first node. The parameters of the model were chosen to approximate the results of our experiments. Some of these parameters were subject to change as explained below.

*Homeostatic regulation of neural activity.* We implemented three different homeostatic plasticity rules, each of which can adaptively modify the excitability of HVC neurons.

Rule 1) If during a simulated chain propagation a neuron did not spike, its spiking threshold decreased by 1 \( \mu \text{V} \). If the neuron produced more than 8 spikes, or 2 bursts, the threshold increased by 1 \( \mu \text{V} \). This rule is used in Figure 4.4 and Figure B.6a. Such homeostatic changes in spiking thresholds have been observed in experiments (Watt and Desai, 2010).

Rule 2) If during a simulated chain propagation a neuron did not spike, the leak conductance of the neuron decreased by 0.1 \( \mu \text{S/cm}^2 \). If the neuron produced more than 8 spikes, or 2 bursts, the leak conductance increased by 0.1 \( \mu \text{S/cm}^2 \). This rule amounts to changing the neuron’s input resistance, defined as the change in membrane potential in response to injected current, divided by the current. This rule is used in Figure B.6b. Homeostatic changes to input resistance have also been observed in experiments (Welie et al., 2004).

Rule 3) If during a simulated chain propagation a neuron did not spike, all synaptic weights to that neuron increased by 6.7 \( \text{pA/cm}^2 \). If the neuron produced more than 8 spikes, or 2 bursts, the synaptic weights decreased by 6.7 \( \text{pA/cm}^2 \). This rule is used in Figure B.6c. Activity-dependent homeostatic changes to a neuron’s synaptic inputs have been observed in experiments, e.g. in cortical neurons (Turrigiano et al., 1998).
In Figure 4.4 and Figure B.6, a ‘motif’ was considered complete if at least one neuron in each of the 80 nodes produced a spike. Motif duration was calculated as the average time from the propagation initiation until the neurons in the last node produced spikes. We ran simulations with modified parameters to verify that our results presented in Figure 4.4 were qualitatively robust.

**Statistical Analysis**

All statistics on data pooled across animals is reported in the main text as mean ± SD and depicted in figure error bars as mean ± SEM. Where appropriate, distributions passed tests for normality (Kolmogorov-Smirnov), equal variance (Levene), and/or sphericity (Mauchly), unless otherwise noted. Multiple comparison corrected tests were used where justified. Statistical tests for specific experiments were performed as described below.

**Figure 4.1e:** Comparison of fraction of trials with IPIs within 20% of the target for different experimental treatments (n=5 rats). Mauchly’s test indicated a violation of sphericity (W=0.134, p=0.049), and a Huynh-Feldt degrees of freedom correction was applied. Subsequent repeated-measures ANOVA revealed significant differences between the treatments \((F_{(1.17,4.59)}=35.7, p=0.002)\). Post-hoc comparisons using Dunnett’s test showed significant differences between PBS (control) and muscimol injections \((p = 0.0002)\), but not between PBS and baseline \((p=0.99)\).

**Figure 4.1h:** Effect of optogenetic stimulation of motor cortex on task performance. A two-tailed, paired \(t\)-test revealed significant differences in performance in the light off and light on conditions (n=5 rats; \(p=3\times10^{-5}\)).

**Figure 4.2c:** Comparison of Wasserstein distances between joint entropy-duration distributions before and after bilateral Nif lesions (n=5 birds). Repeated-measures ANOVA showed no significant difference on any day \((F_{(3,12)}=2.21, p=0.14)\).
Figure 4.2g: Same as 2c, but comparing pre-injection songs to songs after muscimol/PBS injections (n=5 birds). Mauchly’s test indicated a violation of sphericity (W=1.9x10^{-4}, p=0.014), and a Huynh-Feldt degree of freedom correction was applied. Subsequent repeated-measures ANOVA revealed significant differences between the treatments (F_{1.36,5.43}=19.7, p=0.004). Post-hoc comparisons using Dunnett’s test showed significant differences between PBS (control) and muscimol injections (p=1x10^{-5}); no other condition significantly differed from PBS (p>0.92).

Figure 4.2h: Comparison of the syllable duration distributions following HVC lesions (n=5 birds) and Nif inactivations (n=5 birds). A Kolmogorov-Smirnov test on the mean distribution across animals showed no significant differences (p=0.24).

Figure 4.3e: Comparison of Wasserstein distances between joint entropy-duration distributions before and after unilateral lesions to Nif (n=4 birds). Repeated-measures ANOVA revealed that lesions produced significant differences in song structure (F_{5,15}=17.7, p=8x10^{-6}). Post-hoc comparisons using Dunnett’s test showed significant differences from baseline until the second day after lesion (Post and 8h: p<0.001; 1d: p=0.002; p>0.05 thereafter).

Figure 4.3g: Comparisons of HVC dynamics in intact controls and following Nif lesions. A two-tailed, paired t-test revealed significant differences in correlation immediately before and after lesion (n=4 birds; p=0.003). In addition, two-tailed unpaired t-tests showed significant differences between lesion and control conditions at matched time points until the third day post-lesion (p<0.03 before, p=0.1 at the start of the third day).

Figure 4.3h: Comparisons of normalized HVC activity in intact controls and following Nif lesion. A two-tailed, paired t-test revealed significant differences in activity immediately before and after lesion (n=4 birds; p=0.002). In addition, two-tailed unpaired t-tests showed significant differences between lesion and control conditions at matched time points until the third day post-lesion (p<0.03 before, p=0.29 at the end of the third day).
Figure 4.3i: (Top) Comparison of recovery of correlation to pre-lesion HVC dynamics during day and night (n=4 birds). Two-tailed one-sample t-tests revealed significant recovery overnight but not during the day (test against mean zero; p=0.01 and p=0.053, respectively). (Bottom) Comparison of recovery of HVC activity to pre-lesion levels during day and night (n=4 birds). Two-tailed one-sample t-tests revealed significant recovery during the day but not overnight (test against mean zero; p=0.007 and p=0.48, respectively).

Figure 4.3j: (Top) Correlation to pre-lesion HVC dynamics immediately following Nif lesions as a function of the fraction of Nif lesioned (n=11 birds). A two-tailed t-test revealed the Pearson’s linear correlation coefficient, R=-0.91, to be significantly different from zero (p=1x10^-4). (Bottom) Normalized HVC activity immediately following Nif lesions as a function of the fraction of Nif lesioned (n=11 birds). A two-tailed t-test revealed the Pearson’s linear correlation coefficient, R=-0.87, to be significantly different from zero (p=5x10^-4).

Figure 4.4f: (Top) Comparison of post-lesion motif completion rates to pre-lesion baseline for unilateral (n=4 birds) and bilateral (n=5 birds) Nif lesions. Repeated measures ANOVA revealed that lesions resulted in significant reductions of completion rates in unilateral (F[8,16]=7.0, p=5x10^-4), but not bilateral (F[6,18]=4.1, p=0.07), lesions. Post-hoc analysis of the unilateral lesion data using Dunnett’s test showed motif completion rates to be significantly different from pre-lesion on the day of lesion (p=5x10^-4), but not thereafter (p>0.11). (Bottom) Comparison of post-lesion motif tempo to pre-lesion baseline for birds with unilateral (n=4 birds) and bilateral (n=5 birds) Nif lesions. Repeated measures ANOVA revealed that Nif lesions had a significant effect on motif tempo in both unilateral (F[8,16]=10.4, p=4.6x10^-5) and bilateral (F[6,18]=17.5, p=1.3x10^-6) conditions. Post-hoc analysis using Dunnett’s test showed motif tempo was slowed down for both unilateral and bilateral lesions: the effects remained significant (p<0.05) throughout the 7 days in the bilaterally lesioned birds, and through the first 4 days in unilaterally lesioned birds.
Acknowledgements. We thank Ed Soucy and Joel Greenwood for technical assistance. We are grateful to Markus Meister, Joshua Sanes, Naoshige Uchida, Kenneth Blum, Ashesh Dhawale, Max Josch, Adam Kampff, and Evan Feinberg for their feedback on our manuscript. This work was supported by a McKnight Scholars Award to BPÖ, HSFP and EMBO fellowships to SBEW, an NRSA fellowship to RK, and a Rubicon fellowship from the Netherlands Organization for Scientific Research to SMHG.
Appendix A:

Supplemental Material for Chapter 2
Figure A.1: Conceptual models for how temporal structure of birdsong can be modified, involving plasticity in RA and HVC (or upstream of HVC) respectively.

A (top) Presumed functional organization of the motor pathway underlying song. A synaptic chain network in HVC serves as a generic time-keeper. The unary time code produced by RA-projecting HVC neurons is transformed into a specific motor program through connections to neurons in motor cortex analogue RA, which control vocal musculature. (bottom) Within this framework, changes to temporal structure (e.g. lengthening of a song segment) can be implemented by reorganizing the connections between time-keeper neurons in HVC and muscle-related neurons in RA. Additional plasticity within RA could also serve to transform the motor program in RA. After learning each time-keeper neuron would drive a different set of RA neurons than before. 

B (top) Alternatively, HVC neurons could encode specific motor (i.e. song) elements, the specific timings of which are amenable to change. (bottom) Within this framework, the duration of a song segment can be altered by changing the speed of propagation in the part of the HVC network encoding the segment, leaving the connection between HVC neurons and RA neurons intact. Plasticity underlying this form of learning could plausibly be implemented within the HVC network or at the level of its inputs.
Figure A.2: The slower learning rate for syllables relative to inter-syllable gaps in our tCAF paradigm can be explained by the syllables being temporally further removed from the reinforcer.

(A) Example of a ‘syllable + gap’ segment that stretched after 4 days of tCAF. Though gaps change more than syllables, note the significant change also in syllable duration, marked by white lines before the stretch and green lines after (see also Figure S3C). (B) Syllable and gap duration distributions before and after tCAF for the example in A. The mean stretch in the syllable and gap over the 4 days was 9.8 ms ($p = 10^{-28}$) and 20.5 ms ($p = 10^{-62}$) respectively. (C) Example spectrograms showing no delay in the noise-feedback (N) (top) and 50 ms delay (D+N) relative to the end of the targeted gap (G). (D) Summary of the delayed noise-feedback experiments showing a 79.7% reduction in learning rates of the targeted gaps in the 50ms delay condition ($n = 3$ birds). (E) Gap and syllable learning rates across 21 birds. After correcting for the difference in the reinforcement delay between syllables and gaps (see Supplemental Experimental Procedures for details), the capacity for syllables to change was not significantly different from that of gaps (Gap vs. Syllable ‘delay corrected’, $p = 0.35$).
Figure A.3: Rendition-by-rendition estimates of pitch and duration for a targeted syllable during normal singing and during CAF.

(A) Song spectrogram from a bird that underwent pCAF and tCAF targeting the same syllable (S) (tCAF target also included the ensuing gap [G]). (B) Rendition-by-rendition estimates of syllable pitch for a day of pCAF. Difference in mean syllable pitch over the course of the day was 32.9 Hz ($p = 10^{-71}$; using first and last 100 songs of the day for comparison). (C) Rendition-by-rendition estimates of syllable duration during a day of tCAF in the same bird. Increase in mean syllable duration over the course of the day was 1.6 ms ($p = 10^{-10}$). Dashed horizontal lines in (B) and (C) indicate the thresholds for white noise (for tCAF, threshold is estimated as actual online threshold was applied to $S+G$). (D) Scatterplot of duration vs. pitch for the targeted syllable in (C) ($r = 0.004$, $p = 0.88$). (E) Summary statistics for birds that did both tCAF and pCAF ($n = 5$ birds) showing no significant correlation between pitch and duration for targeted syllables either in baseline (no CAF) or during CAF ($p = 0.89$ and 0.09 respectively). Summary statistics computed only on catch trials without white noise interference (see Methods).
Figure A.4: Acute and persistent effects of Area X lesions on spectral and temporal variability. (A) Variability in the pitch of harmonic stack syllables showed a small, but significant decrease immediately following Area X lesions (1 - 3 days post-lesion), but recovered to pre-lesion levels after 3 days. The coefficient of variation (CV) for pitch went from 2.4 ± 1.7 % before lesion to 1.8 ± 1.3 % immediately after (n = 7 birds, average reduction of 22.1%; p = 0.001), consistent with ref. (Kojima et al., 2013). Pitch CV measured 4 or more days after lesion, however, was 2.5 ± 2.1% (p = 0.90 compared to before lesion), indicating that the effect of Area X lesions on pitch variability is transient. (B) Variability in the duration of syllables and gaps, however, remained unchanged after Area X lesions. CV of interval durations was 3.0 ± 0.8% before lesions and 3.3 ± 1.8% (p = 0.65 compared to pre-lesion) in the first three days after lesions and 2.9 ± 0.6 (p = 0.91 compared to pre-lesion) 4 or more days after lesion.
**Figure A.5: Histology of Area X, LMAN, and MMAN lesions.**

**A-B** Sagittal brain slices showing chemical lesions targeting Area X (A) and LMAN (B) respectively. Lesion boundaries demarcated by dashed red lines and estimated from close inspection of viable cell bodies; estimated locations of the targeted nuclei is demarcated by black lines and based on the corresponding (medial-lateral) slices in zebra finch atlases. **C** Coronal brain slice showing lesions targeting MMAN. Same convention as in A and B.
Figure A.6: Two independent measures of the temporal relationship between HVC activity and vocal output shows neural activity leading sound by ~35 ms.

(A) Normalized cross-correlation function between HVC activity and sound amplitude averaged across 7 birds (solid line: mean, dashed line: s.e.m.; peak time = 35 ms; data from each bird is an average over 4 days (100 renditions/day)). The mean lag at peak correlation, across all recorded birds, was 34.9 ± 4.4 ms. (B) Estimated covariance between the temporal variability in HVC activity and song over a range of HVC signal lead times (see Supplemental Experimental Procedures). Red squares show the mean covariance z-scores, across n = 7 birds, each averaged over 4 days (~100 renditions/day), at the specified HVC signal lead times; error bars report s.e.m. A Gaussian fit to this data (µ = 34.7 ms, σ = 33.3 ms, R² = 0.97) is shown in black.
Table A.1: Coordinates and pharmacological doses for Area X, LMAN, and MMAN lesions.

All coordinates are relative to the bifurcation of the midsagittal sinus. Head angle is relative to beak horizontal. Injections of 4% NMA at each site were in steps of 9.2 nL using a Nanoject II (Drummond Scientific) every 10 sec. In two birds, LMAN was lesioned electrolytically (200 μA, 60 sec at each of the 4 sites below).

<table>
<thead>
<tr>
<th>Target</th>
<th>Amount of NMA (nL)</th>
<th>Anterior (mm)</th>
<th>Lateral (mm)</th>
<th>Depth (mm)</th>
<th>Head angle (degrees)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>83</td>
<td>5.5</td>
<td>1.5</td>
<td>2.9</td>
<td>0</td>
</tr>
<tr>
<td>Site 2</td>
<td>83</td>
<td>5.5</td>
<td>2.0</td>
<td>2.9</td>
<td>0</td>
</tr>
<tr>
<td>Site 3</td>
<td>83</td>
<td>6.1</td>
<td>1.5</td>
<td>2.9</td>
<td>0</td>
</tr>
<tr>
<td>Site 4</td>
<td>83</td>
<td>6.1</td>
<td>2.0</td>
<td>2.9</td>
<td>0</td>
</tr>
<tr>
<td>LMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>74</td>
<td>4.7</td>
<td>1.75</td>
<td>1.8</td>
<td>20</td>
</tr>
<tr>
<td>Site 2</td>
<td>74</td>
<td>5.3</td>
<td>1.75</td>
<td>1.8</td>
<td>20</td>
</tr>
<tr>
<td>Site 3*</td>
<td>-</td>
<td>5.3</td>
<td>1.2</td>
<td>1.8</td>
<td>20</td>
</tr>
<tr>
<td>Site 4*</td>
<td>-</td>
<td>5.3</td>
<td>2.2</td>
<td>1.8</td>
<td>20</td>
</tr>
<tr>
<td>MMAN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 1</td>
<td>37</td>
<td>5.3</td>
<td>0.3</td>
<td>1.8</td>
<td>20</td>
</tr>
<tr>
<td>Site 2</td>
<td>37</td>
<td>5.3</td>
<td>0.7</td>
<td>1.8</td>
<td>20</td>
</tr>
<tr>
<td>Site 3</td>
<td>28</td>
<td>4.9</td>
<td>0.3</td>
<td>1.8</td>
<td>20</td>
</tr>
<tr>
<td>Site 4</td>
<td>28</td>
<td>4.9</td>
<td>0.7</td>
<td>1.8</td>
<td>20</td>
</tr>
</tbody>
</table>

* Only for the 2 birds lesioned electrolytically
Table A.2: Extent of Area X, LMAN, and MMAN lesions.
Estimate of the lesions relative to age-matched intact brains. See Supplemental Experimental Procedures for detailed description of quantification methods used.

<table>
<thead>
<tr>
<th>Area X lesions</th>
<th>Bird 1&lt;sup&gt;i&lt;/sup&gt;</th>
<th>Bird 2&lt;sup&gt;ii,iv&lt;/sup&gt;</th>
<th>Bird 3&lt;sup&gt;iii&lt;/sup&gt;</th>
<th>Bird 4&lt;sup&gt;iii,iv&lt;/sup&gt;</th>
<th>Bird 5&lt;sup&gt;iii&lt;/sup&gt;</th>
<th>Bird 6&lt;sup&gt;ii&lt;/sup&gt;</th>
<th>Bird 7&lt;sup&gt;ii&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>% lesioned</td>
<td>98</td>
<td>97</td>
<td>96</td>
<td>96</td>
<td>94</td>
<td>94</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>Bird 8&lt;sup&gt;iii,iv&lt;/sup&gt;</td>
<td>Bird 9&lt;sup&gt;ii&lt;/sup&gt;</td>
<td>Bird 10&lt;sup&gt;iii,iv&lt;/sup&gt;</td>
<td>Bird 11&lt;sup&gt;iii&lt;/sup&gt;</td>
<td>Bird 12&lt;sup&gt;ii&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>84</td>
<td>83</td>
<td>83</td>
<td>82</td>
<td>72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LMAN lesions</th>
<th>Bird 13</th>
<th>Bird 14&lt;sup&gt;ii&lt;/sup&gt;</th>
<th>Bird 15&lt;sup&gt;ii&lt;/sup&gt;</th>
<th>Bird 16</th>
<th>Bird 17&lt;sup&gt;i&lt;/sup&gt;</th>
<th>Bird 18&lt;sup&gt;i&lt;/sup&gt;</th>
<th>Bird 19&lt;sup&gt;ii&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>% lesioned</td>
<td>100</td>
<td>100</td>
<td>95</td>
<td>93</td>
<td>90</td>
<td>89</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>Bird 20&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Bird 21&lt;sup&gt;i,ii&lt;/sup&gt;</td>
<td>Bird 22&lt;sup&gt;ii&lt;/sup&gt;</td>
<td>Bird 23&lt;sup&gt;ii&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>82</td>
<td>81</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MMAN lesions</th>
<th>Bird 23&lt;sup&gt;ii&lt;/sup&gt;</th>
<th>Bird 24&lt;sup&gt;ii&lt;/sup&gt;</th>
<th>Bird 25&lt;sup&gt;ii&lt;/sup&gt;</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>% lesioned</td>
<td>100</td>
<td>100</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>i</sup> pCAF  
<sup>ii</sup> tCAF  
<sup>iii</sup> Return to baseline after pCAF  
<sup>iv</sup> Return to baseline after tCAF
Supplemental Experimental Procedures

Lesion quantification. We measured the volumes of intact LMAN and Area X across three adult control birds by tracing the boundaries of Area X and LMAN in 100 µm thick Nissl-stained sagittal brain slices. LMAN and Area X appear as regions with a higher density of Nissl-stained cell bodies or stronger staining than surrounding areas. In identifying Area X and LMAN, we were also guided by published brain atlases of the zebra finch brain that provide well-established landmarks relative to where LMAN and Area X can be located (e.g. lamina pallio-subpallialis and lamina mesopallialis) (Nixdorf-Bergweiler and Bischof, 2007). The areas traced were then summed and multiplied by 0.1mm (section thickness) to obtain volume. The average volumes we obtained for LMAN (0.22 mm³) and Area X (2.28 mm³) in control birds were very similar to those reported in published studies (Airey et al., 2000; Thompson et al., 2011). For experimental birds, lesion areas were typically distinct in their lack of neuron cell body staining, the presence of gliosis, and other obvious tissue damage when viewed under a 10x microscope objective. The boundaries of unlesioned portions of LMAN and Area X with healthy-looking cells were traced, summed, multiplied by 0.1 mm and then divided by the average Area X and LMAN volumes in control birds (see Figure A.5 for examples of boundaries, Table A.2 for quantification). Lesions of Area X did not extend to LMAN as verified by careful examination of cell bodies in all slices with LMAN. Moreover, analysis of songs (blind to lesion treatments) showed no permanent effect on vocal variability after Area X lesions (unlike actual LMAN lesions that significantly reduced variability), providing further functional confirmation of the fact that LMAN remained intact in the Area X lesioned birds.

For MMAN, as previously reported (Foster and Bottjer, 2001), the boundaries of the nucleus are not always clear in Nissl stained section, thus a reliable quantification based on tracing intact areas was not possible. We thus followed Foster & Bottjer's (2001) method of lesion assessment. Briefly, in the coronal plane, MMAN's location is well delineated by the two laminae (lamina mesopallialis and lamina pallio-subpallialis) on the dorsal-ventral axis; the lateral ventricle and the medial-most extent of LMAN
identify the medial and lateral extent of MMAN respectively. The rostral-caudal axis of MMAN generally coincides with LMAN's (see (Foster and Bottjer, 2001) for details). These landmarks provide the approximate location of MMAN along the three axes. We thus counted slices with LMAN present (as well as slices immediately before and after the rostral-caudal extent of LMAN) which were not lesioned in the expected MMAN location. Based on this criterion, two birds had complete lesions while the third had 3 partially lesioned slices out of a total of 12 in both hemispheres. We thus gave this bird a conservative lower-bound estimate of 75% lesion. All lesion quantifications were done blind to CAF outcomes.

*Pitch estimation algorithm*. The discrete Fourier transform of a 5 ms signal has a 200 Hz frequency resolution (time-frequency tradeoff) which is too coarse for pitch estimation. However, because we target harmonic stacks, we make the assumption that the data can be well approximated by a fundamental frequency and its harmonics. This allows us to obtain highly precise pitch estimates of even very short (i.e. 5 ms) syllable snippets. The algorithm is as follows: 1) Perform a fast Fourier transform (FFT) on the full 5 ms dataset (220 points given our 44.15 kHz sample rate), which has power at 0, 201, 402, ...Hz, etc (rounded to nearest Hz). 2) Discard the first data point of the dataset to create a new FFT now with power at 0, 202, 404...Hz, etc. 3) Repeat the procedure of discarding the first data point in the diminishing data set to create up to 73 frequency based descriptions of the 5 ms sound, each offset relative to its neighbor by, on average, 2.7 Hz in the target fundamental range (typically estimated at 400 - 600 Hz). 4) Give the algorithm an estimate (accurate to within 100 Hz) of the fundamental frequency of the syllable (this ‘prior’ is updated as the syllable pitch shifts during pCAF). Sum the power in the frequency bins that best corresponds to the fundamental and first 19 harmonics of this ‘prior’ for each of the 73 FFT datasets. The fundamental of the FFT set with the highest sum (normalized to the power across all frequency bins) is used as the estimate of the pitch. When tested using playback of 5
ms sound with known pitch in a typical bird cage, the algorithm was able to accurately estimate pitch down to a 2-3 Hz resolution even with signal-to-noise ratio of 5:1 (proportion of white noise added to signal). Our algorithm gave estimates of pitch that were highly correlated with off-line estimates using a previously published algorithm \((r = 0.92 \pm 0.09 , n = 10 \text{ birds})\) (Andalman and Fee, 2009). Our algorithm was implemented in LabVIEW and estimated the pitch of a 5 ms sound in less than 3 ms.

**Interval duration estimates.** Song recordings were first segmented to contain only target motifs. The total power in each segment was normalized to 1 and converted to spectrograms (5 ms Hamming window with 1 ms advancement) and log-transformed. These processed signals were the ones used for all subsequent analysis. Interval (i.e. syllable or gap) durations were estimated using the following two-step procedure. First, we generated motif templates to which all renditions in a catch trial block were aligned. These templates were obtained as follows: (i) Motif renditions were visually inspected and up to 5 motifs were selected. All renditions were aligned to these motifs using a DTW algorithm (see below) and all the aligned renditions were averaged to obtain an average template for each starting motif. (ii) The average templates were summed across frequency bins to produce “power envelopes”. For each such power envelope, the distribution of power was fitted to a mixture of two Gaussians, with the Gaussian with the lowest mean considered to be noise. (iii) A threshold on each power envelope was then set at 2-3 standard deviations above the mean of the noise Gaussian (kept constant within a bird). Time points where this threshold was crossed were noted and labeled as interval onsets and offsets depending on whether it was an up-cross or down-cross respectively.

In the second step, interval durations (syllables and gaps) were estimated as follows. (i) Each rendition was aligned to the average templates. (ii) Time points in the DTW-aligned renditions that matched the interval onset and offset time points in the average template were used to estimate interval durations. As a result of this procedure, we obtained up to 5 different estimates (one for each
average motif template) for the same interval duration. Statistics of the length of an interval across renditions of a catch trial block were calculated separately for each set of estimates (belonging to the same average template) and then averaged across different average templates to obtain a more robust estimate. We used the same average templates and onset and offset times for drives across multiple days (e.g., tCAF up and down) so as to minimize the differences in estimates due to use of different daily templates.

Delayed noise feedback experiments. When we targeted ‘syllable+gap’ segments the average delay of the white noise feedback relative to the center of the gaps was 15.6 ± 6.1 ms, whereas the average delay to the center of the syllables was 62.8 ± 17.0 ms. Thus syllables were, on average, 47.2 ms further removed from the reinforcer (i.e. white noise). To test the effect of delayed reinforcement on learning in the temporal domain, we compared learning rates in tCAF experiments with no delay between the end of the target gap and the reinforcer ($Gap_{no\, delay}$) to rates after experimentally adding a 50 ms delay ($Gap_{50\, ms\, delay}$) (Figure A.2C). This 50 ms delay reduced the learning rate within a bird by, on average, 79.7 ± 4.1% (n = 3 birds) relative to the no delay condition. To estimate what syllable learning rates would be if the reinforcer was delivered right after the end of the syllable ($Syllable_{delay\, corrected}$), we applied the following formula (see Figure A.2E):

$$Syllable_{delay\, corrected} = Syllable \times \frac{Gap_{no\, delay}}{Gap_{50\, ms\, delay}}$$  \hspace{1cm} (1)

where ‘Syllable’ is the learning rate in tCAF experiments targeting ‘syllable + gap’ segments.

Relationship between HVC activity and vocal output (Cross-Correlation Method). We estimated the time-lag between HVC activity and vocalization by cross-correlating sound amplitude and HVC activity and identifying the lag at which the function peaks. ~100 renditions of song and associated HVC traces from a morning catch-trial block were aligned to an average template (as described in the main text’s
Experimental Procedures). The mean neural traces and sound amplitudes were calculated (as described in the main text), normalized by their maxima, and mean subtracted. The two resultant time series were cross-correlated; the cross-correlation function was then normalized by its peak value. For each bird (n = 7), this procedure was repeated for each of 4 catch trial blocks. The normalized correlation functions were averaged across the 4 blocks for each bird. Over the 7 birds, we found the peak in the cross-correlation function to indicate a lag between HVC activity and sound of 34.9 ± 4.4 ms. In addition, we calculated the mean cross-correlation function across birds by averaging over each bird’s mean cross-correlation function (Figure S6A); this function exhibited a clearly defined peak at 35 ms lag.

*Relationship between HVC activity and vocal output (Covariance Method).* We estimated the covariance between temporal variability in HVC activity and vocal output over a range of time lags. ~100 renditions of the song were aligned to an average template (as described in the main text’s Experimental Procedures). The warping paths from these alignments were then applied to the corresponding neural traces with 26 different time lags (range: -36 ms to +100 ms; negative shifts imply that vocalization temporally precedes HVC activity). For each lag, we calculated the correlation coefficient for all possible neural trace pairs in the block. Because the warping path used to align each neural trace was derived from the corresponding song, the average correlation coefficient reflected the covariance between the temporal variability in the song and the corresponding HVC activity. For each bird (n = 7), the procedure was repeated for 4 catch trial blocks; the mean correlation coefficient calculated at each lag time, averaged across blocks, and then converted to a z-score. We then averaged these z-scores across birds to generate a mean ‘covariance’ profile (Figure A.6B). A Gaussian fit to the data (μ = 34.7 ms, σ = 33.3 ms, R² = 0.97) indicated a 34.7 ms lag between HVC activity and sound.
Appendix B:

Supplemental Material for Chapter 4
Figure B.1: Light stimulation of motor cortical neurons expressing the optogenetic activator Chrimson.
(a) Representative example of AAV-injections into motor cortex, showing Chrimson-tdTomato expression at different magnifications in a coronal brain section (~1.5 mm anterior to bregma). The scheme of the brain (right) is adapted from Paxinos’ rat atlas (Paxinos and Watson, 2007). The estimated spread of the injections was 8.3 ± 1.3 mm³ (mean ± SD), with an average of 31.5 ± 1.9% infected cells (Supplementary Methods). (b) Heatmap showing the instantaneous firing rates of 28 single units recorded in an anesthetized rat in response to a 1 second light pulse, averaged over 30 stimulations (Supplementary Methods). (c) A custom-built battery-operated wireless optogenetic stimulation device, consisting of a printed circuit board with integrated IR sensor and LED (λ = 615 nm). The IR sensor gates the circuit and allows the LED to be triggered by an IR light-source placed on top of the rat’s cage. During surgery, the LED is affixed atop a small craniotomy above motor cortex.
Figure B.2: Both brief and sustained optogenetic stimulation of motor cortex cause significant performance deficits in our task. 
(a) Optogenetic stimulation was triggered on the first lever press in a trial, and lasted for either 50 ms or 1 second. (b) Both sustained (1 sec, left, compare Figure 1h, n = 5 rats, p = 3 x 10^{-5}, paired t-test) and brief (50 ms, right, n = 3 rats, p = 0.01, paired t-test) optogenetic activation of motor cortex disrupted task performance. (c) Comparing the effects of the two stimulation protocols on task performance (ratio light on/light off) shows that sustained stimulation has a significantly larger effect on performance (1 sec: n = 5; 50 ms: n = 3, p = 0.004, unpaired t-test).
Figure B.3: Localization and lesioning of Nif.
(a) Injection of fluorescently labelled cholera toxin subunit B (green) into HVC retrogradely labels Nif and anterogradely labels downstream control nucleus RA. (b) Bilateral injections of the excitotoxin NMDA produced focal lesions of Nif. Shown are Nissl stained sections from both hemispheres in the same example bird. Red arrows indicate estimated location of Nif; dashed green line shows the extent of the lesion.
Figure B.4: Muscimol injections into Nif.
(a) (Top) Nissl-stained parasagittal section of a zebra finch brain. (Middle) Magnified view of the region demarcated with a green square atop. Red arrows (left) indicate the location of Nif; violet overlay (right) shows the spread of fluorescent dye co-injected with muscimol. Orange star indicates estimated center of injection based on brightness of the fluorescence. (Bottom) Estimated injection sites relative to the boundaries of Nif for all muscimol-injected birds. Colors denote different animals. (b) Syllable spectrograms (left) and entropy-duration distributions (right) for a bird injected with different volumes of muscimol in Nif. Example spectrograms for 9 nL and 27 nL injections are from recordings made 3 min and 7 min after the injections, respectively. That song disruption was similarly rapid and severe for both volumes (in conjunction with the lack of effect from injections above Nif) limits the possibility that the effects on song were due to diffusion of the drug into HVC.
Figure B.5: Spectrograms of vocalizations following unilateral Nif lesion for the example bird in Figure 4.3.
All examples were recorded within the first hour of singing after lesion. (Top) Example spectrogram of a motif recorded just prior to lesion. (Left) Example spectrograms of vocalizations in which motif syllables could not be reliably identified and thus were excluded from subsequent analysis. (Center) Example spectrograms of identifiable motifs that were included in the alignment-dependent analysis (Figures 4.3f-j). (Right) Example spectrograms of songs with identifiable syllables, but truncated motifs.
Figure B.6: Different mechanisms for homeostatic regulation of neural activity produce similar effects. (a) (Top) Effect of Nif removal on membrane excitability during simulated songs in a model neuron (from the 40th node), smoothed with a 100-point moving average and averaged over 40 model ‘experiments’. A rule for homeostatic regulation of activity drives a reduction in spiking threshold after Nif removal. (Middle) Fraction of simulations in a 100-point window for which activity in the model HVC network propagated to the end, averaged over 40 model ‘experiments’ (Supplementary Methods). (Bottom) A 100-point moving average over the time to complete a full chain propagation, averaged over 40 model ‘experiments’. Orange triangle denotes time of ‘lesioning’. Same as in Figure 4.4b,d. (b-c) Same as in a, but with homeostatic regulation of membrane leak conductance b and synaptic input strength c (Methods).


Eales, L.A. (1987). Do zebra finch males that have been raised by another species still tend to select a conspecific song tutor? Anim. Behav. 35, 1347–1355.


