Neural Mechanisms of Trial and Error Learning:

a Study from Bird Song

Alison Guidry Duffy

A dissertation

submitted in partial fulfillment of the

requirements for a degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Adrienne Fairhall, Chair

Marcel den Nijs

David Perkel

Program Authorized to Offer Degree:

Physics

University of Washington

Abstract

Neural Mechanisms of Trial and Error Learning: a Study from Bird Song

Alison Guidry Duffy

Chair of Supervisory Committee:

Adrienne L. Fairhall

Physiology and Biophysics

In this dissertation I examine how variation is generated, shaped and controlled in the brain during trial-and-error learning. The first part of this dissertation concerns the mathematical modeling of networks of neurons that constitute part of the neural architecture of song bird learning. Chapter 1 reviews the theories of learning in neuroscience that shape our subsequent modelling and analysis results as well as background on the song bird neural system and behavior. Chapter 2 examines the modulation of a microcircuit within the basal ganglia and proposes a mechanistic way in which dopamine could modulate the signaling properties of the nucleus to influence behavioral changes in song variability. Chapter 3 addresses the ability of a motor system to maintain stereotyped behavior over long periods of time, through adaptation to changes such as injury or aging. In the context of a model of bird song learning, ongoing instability in neural representation of stable behavior allows a system to more readily adapt and maintain performance with minimal cost. In this framework, behaviors

are made more robust to environmental change by continually seeking new ways of performing the same task. Chapter 4 examines the way that exploration in trial and error learning is shaped by network properties. A reinforcement learning system, inspired by bird song architecture, is able to successfully learn when exploration is driven by variable network dynamics. Further, learning is made more successful when the exploratory dynamics from which variations are selected partially align with the elementary components of the desired behavior.

The second part of this dissertation develops a method of analysis to compare song behavior to patterns of neural activity and suggests an interpretation of the covarying neural-behavioral activity that operates within the theoretical framework of reinforcement learning. Chapter 5 presents an analysis of the covariations between song and neural activity in the ventral tegmental area (abr. VTA) of the zebra finch. This analysis provides evidence that dopamine neurons in VTA encode representations of song error during natural behavior. Additionally, diverse components useful for error calculation exist locally within VTA, which could contribute to the final error signal. Lastly, novel, long-timescale state variations in song and cell activity are present in a subpopulation of VTA; I propose an interpretation in which the error calculation incorporates a normalizing operation relative to internal states.

ACKNOWLEDGEMENTS

First, thank you to my advisor and mentor, Adrienne Fairhall. Adrienne, you have been a great source of inspiration and guidance throughout my PhD. Thanks for your encouragement and friendship throughout this process, the wide-ranging thought and depth of inquiry you bring to our research endeavors, and the enrichment you bring to the neuroscience community. It is all truly appreciated.

Thank you to David Perkel. David, you have been an important additional mentor for me throughout my PhD. I have learned much from our discussions and collaborations. Thank you.

Thank you to the other members of my committee, Marcel den Nijs, Eric Shea-Brown, Paul Wiggins, Andreas Karch, Sara Goering for your time and insightful comments on my work.

Thank you to the members, past and present, of our lab for your support and enriching discussions: Rich Pang, Sharri Zamore, Heather Barnett, Kenneth Latimer, Guillaume Lajoie, Hengji Wang, Yoni Browning, Ben Lansdell, Stephano Recanatesi and Argha Mondal.

Thank you to all of my other collaborators: Agata Budzillo, David Perkel, Elliott Abe, Guillaume Lajoie, Kenneth Latimer, Vikram Gadagkar and Jesse Goldberg. The work in this dissertation is the result of combined research with all of you. I am grateful for your effort and insight in our work.

Thank you to the faculty and staff of both the Physics Department and the Physiology and Biophysics department for creating an environment and community in which students and scientific research can thrive.

To Megha. To my parents, Michael and Jaqueline. To my sister, Anne. Lastly, to Rob and Frieda.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction to:

Neural Mechanisms of Trial and Error Learning: a Study from Bird Song

By

Alison Guidry Duffy

How to read this thesis: Each chapter represents a stand-alone project. Introductions within each chapter will provide the reader with the necessary background to approach each chapter topic. The introductory chapter frames the research questions in a broader context and provides additional background to the entire work.

In this dissertation, I study how variation is generated, shaped and refined in the brain during trial-and-error learning. The first part of this dissertation uses mathematical modeling to study networks of neurons that constitute part of the neural architecture involved in song bird learning. The second part of this dissertation analyzes singing in relation to neural activity in the ventral tegmental area of the brain. In neuroscience, bird song learning is a paradigmatic model

of how trial and error learning is carried out in neural systems. It is a canonical model for two main reasons. One, the bird song neuroanatomical structure has been extensively studied and neural components of the learning and song production process have been distinguished and localized much more precisely than in other systems. This allows behavioral, computational and algorithmic accounts to be mapped to a mechanistic account of neural interactions. Two, the behavior is complex enough to serve as an informative analogue to more complicated animals and behaviors but simple enough to leave hope that a rigorous biophysical understanding of the process can be achieved. Learning algorithms and circuit properties discovered here will likely inform understanding of learning in other species and in higher order behaviors.

Introduction to theories of trial and error learning in neuroscience

Trial and error learning is a broad class of associative learning wherein the learner tries, either purposefully or randomly, different means of achieving a learning outcome until success is achieved. There are many instances of trial and error learning in the animal world. The trial and error learning process was first described rigorously in an animal behavior context by Edward Thorndike in the late 19th century. He described the core processes of trial and error learning as: one, motivation, the basic desire to learn; two, an obstacle—the problem to be solved; three, random activities—out of ignorance, the learning organism tries random acts towards the sought-after solution; four, accidental success—by chance the organism stumbles upon the correct solution; five, an increased likelihood of selecting the correct response—no longer purely accidental; six, mastery of the right response—the organism is now able to

immediately select the correct response when presented with the problem (1)[1]. Notably, trial and error learning does not require the presence of a teacher nor explicit instruction to progress. All one needs is the desire and ability to both explore varied means of executing the task and shift behavior towards more favorable experimental outcomes.

Reinforcement learning is a theoretical approach to trial-and-error learning through interactions with the environment (2). It provides a mathematical framework for many of the ideas expressed by Thorndike and others. Ideas from animal behavior and neuroscience as well as optimal control theory and artificial intelligence have shaped the theory's development (3-5). Reinforcement learning is a powerful tool to understand trial and error learning in neuroscience because it provides a normative scaffolding upon which to form and test hypotheses about the roles of neural activity during learning. It is normative in the sense that it does not describe how animals behave within a learning task, but how they ought to behave to learn optimally (6). The core components of the reinforcement learning scenario are a learning agent and an environment in which it acts. Within this framework, there are four basic elements of reinforcement learning: the policy, the reward function, the value function and a model of the environment.

[1] Thorndike's original paper describes trial and error learning in precise detail and is worth reading in its original form: "*The experiments were upon the intelligent acts of a considerable number of dogs, cats and chicks. The method was to put the animals when hungry in enclosures from which they could escape (and so obtain food) by operating some simple mechanism, e.g., by turning a wooden button that held the door, pulling a loop attached to the bolt, or pressing down a lever. Thus one readily sees what sort of things the animals can learn to do and just how they learn to do them…. The first time that a cat is put into such an enclosure, some minutes generally elapse before its instinctive struggles hit upon the proper movement, while after enough trials it will make the right movement immediately upon being put in the box… The starting point for the formation of any association is the fund of instinctive reactions. Whether or not in any case the necessary act will be learned depends on the possibility that in the course of these reactions the animal will accidentally perform it. The progress from accidental performance to regular, immediate, habitual performance depends on the inhibiting power of effort without pleasure and the strengthening by pleasure of any impulse that leads to it."* 1. Thorndike E (1898) Some Experiments on Animal Intelligence. *Science* 7(181):818-824.

The policy, $\pi(S, a)$, defines how the agent chooses to act at a given time; it maps current environmental states, S, to the agent's actions, a. The agent learns by changing their policy. In general, policies might be either deterministic or stochastic and can be written as:

$$\pi(S, a) = P(a|S).$$

The reward function defines the goal in a learning problem; it maps each perceived state (or state-action pair) of the environment to a single number, the reward, that indicates the desirability of the state. The objective of the agent is to maximize their total cumulative reward. The reward function can be used by the agent to change their policy but is itself unchanged by the agent. Thus, it is a part of the environment in which the agent acts. Another defining feature of the reward function's relationship to the agent is that it is immediate: in biological systems the reward function might be approximately compared to pain and pleasure.

The value function defines what is good in the long run, as opposed to the reward function, which defines an immediate good. Value functions define the desirability of a given state taking into account the cumulative consequences of that state's influence on all future states. Thus, current optimal actions may yield low immediate reward but lead to high-reward future states: 'no pain, no gain' succinctly expresses the idea of using something akin to a value function to determine behavior. A value function for a particular environmental state, may be written as:

$$V(t) = E[r(t) + \gamma r(t+1) + \gamma^2 r(t+2) + \cdots],$$

where r(t) is the reward at a given time step, t, γ is a discounting factor between 0 and 1, which weakens the value of rewards further in the future, and the expectation is taken over all possible actions and states reachable from the current state (2, 7). Action choices are made based on the policy, which can be shaped by the value function. The primary task of reinforcement learning is to accurately estimate the value function. Note that this is much harder than estimating the reward since the value function is determined by not only the current state, but all possible future states that will follow.

Lastly, a model of the environment mimics the behavior of the environment. For example, given a current state and action, a model could predict the next state and likely reward. Models are used for planning: explicitly using possible future outcomes to decide current actions. Model-based learning is more abstract than straightforward trial-and-error learning, which is explicitly unplanned. However, many reinforcement learning algorithms use both: they simultaneously learn by trial-and-error, learn a model of the environment, and use the model for planning actions within the environment (2).

Temporal difference (abr. TD) learning solves the central problem in reinforcement learning, estimating the value function (8), and has been linked to neural correlates in many systems (7, 9, 10). The idea uses the implicit formulation of the value function to iteratively refine an estimate, $\widehat{V}(t)$, of $V(t)$. The definition of $V(t)$ allows for a recursive expression of the current value:

$$V(t) = E[r(t) + \gamma V(t+1)].$$

This implicit expression can be used to compute the temporal difference prediction error, $\delta(t)$, of the estimated value, using only locally available information, instead of waiting to collect all future rewards:

$$\delta(t) = r(t) + \gamma \hat{V}(t+1) - \hat{V}(t).$$

The TD prediction error is an expression of *unexpected,* rewarding stimulus in the agent's environment: it is the discrepancy between the expected outcome and the actual outcome and is positive when a current reward is larger than predicted by the current value estimate and negative when the current reward is smaller than predicted. This expression for error can be used to update the estimate of the local value function:

$$\hat{V}_{new}(t) = \hat{V}_{old}(t) + \eta * \delta(t),$$

where η is the learning rate and determines how much the estimate is changed at each step. Through repeated interactions with the environment, the agent is able to iteratively improve their estimate of the value function. Notably, through experience, the TD prediction error update can drive increases in the value of environmental states that are not themselves intrinsically rewarding but are correlated with future rewarding states. Variants of this temporal difference prediction error can also be used to learn improvements to the agent's policy, as in Actor-Critic models of learning wherein a critic learns updates to the value function and then

instructs the actor in changes to the policy (11), or improvements to state-action pairs, such as in Q-learning (12). The temporal difference prediction error aligns closely with activity in dopaminergic neurons (7, 13, 14), and this correspondence led to Montague et al. (13) to propose the reward prediction hypothesis of dopamine wherein the function of phasic dopamine signaling was to transmit a TD prediction error to drive reward based learning in the basal ganglia. Subsequent studies have found many detailed correspondences between dopaminergic signaling and TD prediction errors in a variety of reinforcement learning algorithms.

Different reinforcement learning strategies combine in the brain with many other strategies for learning, of which trial-and-error learning plays one role. Supervised and unsupervised learning are two such examples (15). The goal of supervised learning is to construct an input-output map that predicts the output for a given input based on a set of explicit training examples wherein the correct input-output mapping is provided. Learning updates to the input-output map depends on the correlation between the output error and the inputs. Importantly, in supervised learning, error is assigned locally, for example, to specific synapses, which requires a high degree of knowledge about the plastic architecture as well as the ultimate learning target. The cerebellum is thought to contribute to this type of error-based learning (15, 16). The goal of unsupervised learning is to construct an input-output map such that the output characterizes the statistical properties of the input. In unsupervised learning, there is no corrective signal, only a form of relaxation dynamics in which local plasticity rules, such as Hebbian plasticity, drive synaptic changes, sometimes in the presence of an additional regularization such as a sparseness constraint. Reinforcement learning bridges these two forms

of learning: there is an evaluation signal in the form of a reward which is then used to compute a reward prediction error. However, the reward is a single scalar value and does not explicitly assign credit for error to specific parts of the learning system (15).

Introduction to vocal learning and song production

> *The sounds uttered by birds offer in several respects the nearest analogy to language, for all the members of the same species utter the same instinctive cries expressive of their emotions; and all the kinds which sing exert their powers instinctively; but the actual song, and even the call-notes, are learned from their parents or foster-parents. These sounds, as Daines Barrington has proved, "are no more innate than language is in man." The first attempts to sing "may be compared to the imperfect endeavor in a child to babble." The young males continue practicing, or as the bird-catchers say, "recording" for ten or eleven months. Their first essays show hardly a rudiment of the future song; but as they grow older we can perceive what they are aiming at; and at last they are said "to sing their song round."*
>
> –Charles Darwin, 1874 (17)

Besides humans, there are only a very few species that we believe come to utter vocalizations through imitation rather than instinct: bats, elephants, cetaceans (whales, dolphins and porpoises) and three related groups of birds: parrots, hummingbirds and song birds (18). Of the bird groups, the species of songbirds, the oscine passerines, are the most numerous and include all of the most complex singers: warblers, larks, wrens, thrushes, finches. Songs are most commonly sung by males to defend territory and attract a mate. Some song birds, like the Australian zebra finch or the American white crowned sparrow, have only a single song which they learn as juveniles and produce in a reliable manner throughout their life.

Others, like the mockingbird in North America, the lyrebirds in Australia or the nightingale in Europe learn hundreds or even thousands of songs (19). Male, European sedge warblers develop a unique repertoire of syllables that they then recombine in inventive combinations as part of their courtship ritual. Female sedge warblers preferentially select mates with more complex and varied repertoires (20).

Bird song and human language share parallels at the behavioral, neurological and genetic level. The Fox-P2 gene is expressed in the basal ganglia of humans and song birds (21). Mutations in this gene leads to speech deficits in humans, particularly relating to sequencing (22), while knockdown of Fox-P2 in song bird striatum produces deficits in song learning (23). The cortical and basal ganglia neural circuits involved in sensory processing, feedback and song learning in song birds have direct mammalian analogues which likely generalize not only to language but other forms of motor and sequence learning in mammals (24). There are striking developmental similarities in the way that humans and birds first learn to produce phenomes (consonant-vowel combinations such a ‘ma’) and to string these phenomes together. Both babble in infancy: human infants produce repetitive sequences of single phenomes that vary exponentially in length (25). Transitions between these single syllables (such as ‘ma-ba’) are acquired in a stepwise fashion in human infants and song birds suggesting similar neural mechanisms might underpin the connecting of phenome units (26). They both require auditory feedback, vocal examples and social interaction to learn properly (27).

While some birds as mentioned above are able to sing enormously complex variations of song, the meaning of those songs appears to be concrete and limited to mating or defense contexts. The combination of vocal communication and the capacity for flexible associations,

high order recursion, and abstract meaning seems to be uniquely human, though to what degree these elements represent novel traits versus more complex versions of an animal continuum is unknown and an area of active debate in disciplines concerned with the evolution of language and cognition (28). Regardless, the most apt comparisons between bird song and human speech are at the level of phonology (the pronunciation of specific sounds), concrete rules for ordering simple sounds, and prosody (patterns of stress and intonation with regards to their relationship to pitch, frequency and amplitude) (27).

The research in this dissertation draw mainly from experiments conducted in the zebra finch, which is the most commonly studied bird in neuroscience. Many of the neural structures in this species extend to other song birds. I present a brief introduction to the development of song and the neural circuity underpinning learning and production of song in zebra finch below.

Song learning in the zebra finch

Behavior

Zebra finches hatch without song[2]. Shortly after hatching, juvenile, male zebra finches begin to produce highly variable vocal babblings while they listen to the songs of adult males, called 'tutor songs' (27). Over several weeks of practice, the juvenile song crystallizes, acquiring more temporal structure and gradually resembling a highly stereotyped copy of the tutor's song. The learning process requires auditory feedback: if the bird is deafened during this period the

[2] In neuroscience, the zebra finch is most commonly studied of the song birds and is best understood.

song develops with abnormal acoustics and increased variability (29) (30). However if initially exposed to sufficient examples of a tutor's song during the critical period, a zebra finch can then successfully improve its song in total isolation (31). This suggests that the zebra finch memorizes a tutor's song and then uses an internal template to evaluate its own performance (32).

As the zebra finch reaches adulthood, song variability gradually diminishes but does not disappear and is modulated by social context: when a male zebra finch sings while courting a female ("directed song"), the variability is much lower than when he sings alone ("undirected song") (33). This remaining adult variability is thought to be important for continued tuning of the birds' song. Furthermore, birds are able to learn in adulthood. In operant conditioning experiments (34, 35), when natural variations in pitch cross a pre-determined threshold at a chosen point in the song trajectory, a brief noise burst is triggered which distorts the bird's perception of their own song. Within hours of this conditioning, the bird learns to adjust the song to avoid the aversive noise.

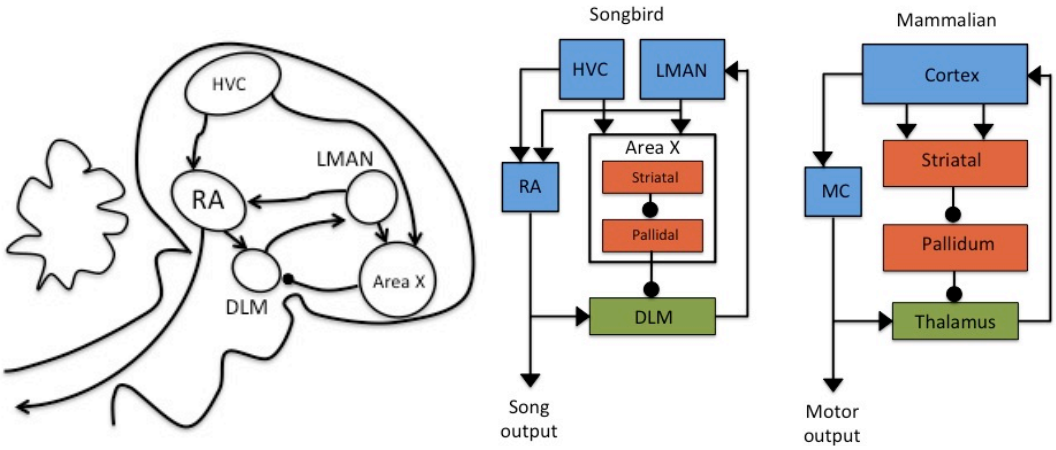

Figure 1. The birdsong system. A. Schematic depiction of brain regions involved in birdsong production and learning. B. Comparison of homologous regions in the avian and mammalian systems. The birdsong system is a useful model for understanding motor learning and variability generation because the circuit functions and connections are well-defined yet analogous to mammalian systems in which a comparable understanding of the connectivity, function and neuronal types are absent.

Neuroanatomy

Birdsong is a powerful system in which to explore trial and error learning because much of the brain circuitry underlying the learning process is well located and studied, (Fig. 1). The forebrain nucleus, RA, (robust nucleus of the arcopallium) projects topographically to the primary motor neurons in the brainstem and controls vocal muscles during song (36). The premotor cortical nucleus HVC (formerly 'high vocal center', now used as a proper name) is a major input to RA. HVC activity is periodic during early development and then evolves into a precise, temporally sparse pattern, wherein every RA-projecting neuron within HVC bursts once with millisecond precision during song (37). HVC has been characterized as a 'clock,' which keeps time in the song motor sequence (38). Lesions in either RA or HVC immediately halt the song (39).

The variability necessary for trial-and-error learning is introduced into this circuit through RA via a secondary circuit called the anterior forebrain pathway (abr. AFP), a specialized basal ganglia thalamocortical circuit needed for learning and plasticity but not for the performance of learned song (31, 40, 41). Basal ganglia thalamocortical loops such as these are evolutionarily conserved in vertebrates and play a role in motor learning and control across many species (42). The frontal cortical-like nucleus, the lateral magnocellular nucleus of the anterior nidopallium (abv. LMAN) projects to RA and is the output of the AFP circuit, which is comprised of LMAN, area X (a part of the basal ganglia) and the medial portion of the dorsolateral nucleus of the thalamus (abv. DLM). The AFP circuit is organized 'myotopically': looping projections from area X to DLM to LMAN and back to area X conserve the topography of RA motor drive from LMAN. This circuit configuration would allow for a variable signal to be evaluated and manipulated within the AFP network itself, a necessary condition for topographically precise refinement and manipulation of the LMAN outputs within the AFP circuit (43).

Although it has been established that the AFP circuit is the source of variability of moment-to-moment fluctuations in song, where and how variability emerges in this circuit is not well understood. Several results implicate LMAN. LMAN projects directly to the primary motor pathway, and while lesions to LMAN have little effect on the crystallized adult song, they halt variability in juveniles (44). In adults, lesions to LMAN prevent learning during distorted auditory feedback, suggesting that LMAN input is responsible for these rapid, time-specific shifts in song (34) (45). Electrical stimulation of LMAN during song causes immediate perturbations in song output (46). Furthermore, as a highly recurrent cortical circuit, LMAN may resemble the structure

of balanced excitatory-inhibitory networks that have been shown to exhibit chaos and therefore could generate its own variable dynamics (47).

During early syllable formation in juvenile birds, input from LMAN imposes a wide distribution of syllable durations. When LMAN is cooled, the durations of these random syllable lengths increase (48, 49). Over the course of early development, the control of rhythm and timing shifts to HVC, and syllable durations are gradually honed to 2-7 stereotyped lengths over the course of learning. In adult birds, LMAN produces different firing patterns depending on social context. In the presence of a female, during stereotyped directed song, the average firing rate of LMAN projection neurons decrease and the spike trains become more stereotyped and time-locked to the song (50). During more variable undirected song, LMAN firing rates increase; there is less song-locked patterning; and burst-like spike patterns emerge (50).

DLM makes excitatory projections to the cortical LMAN and receives inputs from RA and area X. DLM neurons exhibit sharp increase in firing rate prior to song onset; the high firing rate is maintained throughout the course of the song and is augmented by additional spikes in firing rate prior to syllable onset (51). Lesions in DLM eliminate variability in young songbirds (52). DLM may play a critical role in driving the bursting-like spike trains observed in cortical LMAN, particularly since DLM neurons produce intrinsic bursting behavior while cortical neuron types typically do not (51, 53-55). DLM receives excitatory inputs from the motor cortex, RA, and inhibitory input from area X. Excitatory input to DLM from RA has been characterized both as a potential efferent copy of the vocalizations and as a cue to the AFP circuit of the motor activity state and drives song-locked rate modulations in DLM firing patterns during song. However, the

role of this input is not well understood (56, 57). At the same time, single area X inputs to single DLM cells drive input-locked firing events in DLM with millisecond precision (56).

There is evidence that area X, the basal ganglia homologue of the AFP circuit, modulates and contributes to variability downstream in LMAN. Area X receives (1) recurrent inputs from LMAN, (2) precise song-locked inputs from HVC which resemble HVC projections to RA, and (3) tonic and phasic inputs from a dopaminergic midbrain region, the ventral tegmental area (abr. VTA). VTA receives broad inputs from many brain regions including auditory circuits. The pathway from auditory circuits, through VTA, to the basal ganglia is thought to carry out computations that compare the bird's own song to his memory of the tutor song (the 'tutor template') and thereby send evaluation signals to the AFP circuit via area X. Dopamine infusions to area X reduce response to direct excitatory input and decrease firing variability in the area X pallidal neurons (the output neurons of area X) (58). Area X output neurons show highly variable firing patterns during singing, which might then contribute to variability downstream in the rest of the AFP circuit (58). Area X, DLM and LMAN all exhibit intrinsically variable responses to single shock impulses from HVC *in vivo* (59). Finally, in-vivo recordings from area X when LMAN inputs are ablated show that pallidal output neurons undergo context-dependent shifts in firing variability, whereas upstream medial spiny neurons (abr. MSN) do not, implying that some degree of variability emerges from within the x nucleus and is not inherited from LMAN (60). Lesions of area X in juveniles do not affect exploratory babbling but do lead to protracted variability in adult song, suggesting that the changes are never consolidated (52). However, lesions in area X in adults remove local fluctuations in song that are characteristic of the undirected song performance (61).

Theories of learning applied to bird song

Many previous theoretical learning models have used reinforcement learning paradigms to relate components of the birdsong learning circuit to RL algorithmic roles (32, 43, 62, 63).

In the first application of RL theory to the birdsong system, Doya and Sejnowski (64, 65) adopted the classical actor and critic characterization and then added an external "experimenter" that injects randomness into each trial to produce exploration of the performance space. This separation of the actor and the experimenter into two roles was motivated by the structure of the song bird circuitry: RA was cast as the actor carrying out the policy, and LMAN was cast as the experimenter injecting chance into the policy decisions. Using an RL algorithm, synaptic weights from HVC to RA are perturbed randomly at the onset of each trial. Perturbations that lead to improved performance are incorporated into the weight structure, and an additional bias in the direction of that synaptic perturbation is introduced into the new perturbation of the subsequent trial. This is repeated until the model reaches an acceptable level of performance. They proposed that LMAN plays the role of the experimenter and area X that of the critic and the biased random walk that drives synaptic plasticity is a combination of these two roles. However, their model does not specify how the circuitry of the AFP circuit might execute the algorithm and is unrealistic to the observed impact of LMAN on song activity: transient inputs from LMAN lead to transient, sub-syllabic perturbations in song (46), not the song-spanning shifts that should result from static perturbations of the synapses.

Fiete et al. (32) built upon this model by using gradient estimation based on dynamic perturbation of neural conductances. In their model, each LMAN projection neuron forms a glutamatergic synapse onto one RA neuron and delivers Poisson variability to the firing patterns of RA during song. RL via gradient ascent causes changes in the connection strengths of HVC to RA, depending on coincident activation of an HVC→ RA synapse and LMAN→ RA followed by a global positive reinforcement signal, until the song suitably matches the template song. Even though the reinforcement signal is binarized, delayed and temporally imprecise, the model learns in a number of trials comparable to actual repetitions of juvenile song production. Farries and Fairhall (62) took the RL algorithm, applied to two-layer feed-forward networks, closer to biology by using performance-modulated spike-timing dependent plasticity rules to reproduce selected spike trains and population responses. Both of these models dropped the bias in perturbations towards directions of previous improvement, which is likely present in the LMAN signal (43).

In all of these models, the explicit conduit of the reinforcement was unclear. Performance modulated plasticity assumes a reward signal is delivered directly to the primary motor pathway, which then interacts with spike-timing dependent plasticity to alter the connectivity structure of the primary motor pathway. This assumption is notably lacking a known circuit component in an otherwise well-mapped system. Addressing this absence, Fee and Goldberg (43) proposed a hypothesis of learning that explicitly located the reward signal in the dopamine projections from VTA to area X and identified a mechanism through which the reward signal influences future song renditions (58, 66). They proposed that the convergence in area X of the timing information from HVC, the exploratory perturbations from LMAN, the efference copy of RA activity from DLM via LMAN and the dopaminergic evaluation signal from VTA would provide area X the necessary

information to then bias the variable LMAN inputs towards more successful perturbations. This hypothesis shifted the reinforcement learning problem into the AFP pathway: in this theory, changes within area X learned via a reward signal drive LMAN to enact changes in the primary motor pathway simply through repetition of the LMAN signal and spike timing plasticity mechanisms between HVC, RA and LMAN activity. An input-timing dependent plasticity rule has been found at the HVC-RA-LMAN nexus that depends on ordered inputs from HVC and LMAN onto RA so this theory is biologically plausible (67). Tesileanu et al. (2017) used this formulation to construct a 2-stage learning algorithm in which topographic projections from area X through LMAN to RA generate an error-based corrective signal that alters the synaptic weights from HVC to RA (63). In this model, a corrective bias in the LMAN signal is reintroduced, which then drives plasticity in the HVC to RA synapses via a local plasticity rule. Even though this model abstracts the AFP loop to a single layer with time varying Poisson firing rate statistics, it captures the hypothesized, two-stage nature of the bird song circuit in which RL learning is first carried out by the area X 'critic'.

All of these theories of reinforcement learning in song bird address a specific, adolescent and adult stage of song development. Several other learning processes must first happen before the structure is in place for reinforcement learning to proceed (68). One, a tutor memory must form, likely in the auditory cortex. This allows for a comparison between the correct version of the song and the bird's actual performance. Two, sparse, temporally structured activity must form within the HVC nucleus. This builds a lower-dimensional, latent state space upon which reinforcement learning takes place and addresses the 'curse of dimensionality' that reinforcement learning algorithms suffer in the absence of an efficient way to reduce the

dimensionality of the learning task. Three, connections must form between the motor pathway and the auditory cortex so that the memory of the tutor song can be properly aligned with auditory feedback. After these components are in place, reinforcement learning can proceed as hypothesized in the multiple studies above.

Bibliography

1. Thorndike E (1898) Some Experiments on Animal Intelligence. *Science* 7(181):818-824.
2. Sutton RS & Barto AG (1998) *Reinforcement learning : an introduction* (MIT Press, Cambridge, Mass.) pp xviii, 322 p.
3. Bertsekas DP, & Tsitsiklis, J. N. (1996) Neuro-dynamic programming. *Athena Sc*.
4. Rescorla RA, & Wagner, A. R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *In A. H. Black, & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory*:64–99.
5. Bush RR, & Mosteller, F. (1951) A mathematical model for simple learning. *Psychological Review* 58:313–323.
6. Niv Y (2009) Reinforcement learning in the brain. *Journal of Mathematical Psychology* 53:139-154.
7. Schultz W, Dayan P, & Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306):1593-1599.
8. Sutton RS, & Barto, A. G. (1990) Time-derivative models of Pavlovian reinforcement. In M. Gabriel, & J. Moore (Eds.), Learning and computational neuroscience: Foundations of adaptive networks:497–537.
9. Fiorillo CD, Tobler PN, & Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299(5614):1898-1902.
10. Bayer HM & Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47(1):129-141.
11. Barto AG, Sutton, R. S., & Anderson, C. W. (1983) Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man and Cybernetics* 13:834-846.
12. Watkins CJ (1989) Learning with delayed rewards. *Unpublished doctoral dissertation, Cambridge University, Cambridge, UK.*
13. Montague PR, Dayan P, & Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 16(5):1936-1947.
14. Gadagkar V*, et al.* (2016) Dopamine neurons encode performance error in singing birds. *Science* 354(6317):1278-1282.
15. Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural networks : the official journal of the International Neural Network Society* 12(7-8):961-974.
16. Uehara S, Mawase F, & Celnik P (2018) Learning Similar Actions by Reinforcement or Sensory-Prediction Errors Rely on Distinct Physiological Mechanisms. *Cereb Cortex* 28(10):3478-3490.
17. Darwin C (1874) *The descent of man, and selection in relation to sex* (Hurst and company, New York,) New Ed p 705 p.
18. Jarvis ED (2004) Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Sciences* 1016:749-777.

19. Slater P (2011) Bird song and language. in *Oxford Handbooks Online*, ed Tallerman KGaM (Oxford University Press).
20. Buchanan KL & Catchpole CK (2000) Song as an indicator of male parental effort in the sedge warbler. *Proc Biol Sci* 267(1441):321-326.
21. Teramitsu I, Kudo LC, London SE, Geschwind DH, & White SA (2004) Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 24(13):3152-3163.
22. Marcus GF & Fisher SE (2003) FOXP2 in focus: what can genes tell us about speech and language? *Trends in cognitive sciences* 7(6):257-262.
23. Haesler S*, et al.* (2007) Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLoS biology* 5(12):e321.
24. Doupe AJ, Perkel DJ, Reiner A, & Stern EA (2005) Birdbrains could teach basal ganglia research a new song. *Trends in neurosciences* 28(7):353-363.
25. Darshan R, Wood WE, Peters S, Leblois A, & Hansel D (2017) A canonical neural mechanism for behavioral variability. *Nature communications* 8:15415.
26. Lipkind D*, et al.* (2013) Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498(7452):104-108.
27. Doupe AJ & Kuhl PK (1999) Birdsong and human speech: common themes and mechanisms. *Annual review of neuroscience* 22:567-631.
28. Hauser MD, Chomsky N, & Fitch WT (2002) The faculty of language: what is it, who has it, and how did it evolve? *Science* 298(5598):1569-1579.
29. Konishi M (1965) The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift fur Tierpsychologie* 22(7):770-783.
30. Marler P & Tamura M (1964) Culturally Transmitted Patterns of Vocal Behavior in Sparrows. *Science* 146(3650):1483-1486.
31. Brainard MS & Doupe AJ (2002) What songbirds teach us about learning. *Nature* 417(6886):351-358.
32. Fiete IR, Fee MS, & Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *Journal of neurophysiology* 98(4):2038-2057.
33. Kao AC (2005) Learning to talk and listen. *The virtual mentor : VM* 7(8).
34. Andalman AS & Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proceedings of the National Academy of Sciences of the United States of America* 106(30):12518-12523.
35. Tumer EC & Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* 450(7173):1240-1244.
36. Wild JM (1993) Descending projections of the songbird nucleus robustus archistriatalis. *The Journal of comparative neurology* 338(2):225-241.
37. Long MA, Jin DZ, & Fee MS (2010) Support for a synaptic chain model of neuronal sequence generation. *Nature* 468(7322):394-399.
38. Long MA & Fee MS (2008) Using temperature to analyse temporal dynamics in the songbird motor pathway. *Nature* 456(7219):189-194.

39. Nottebohm F, Stokes TM, & Leonard CM (1976) Central control of song in the canary, Serinus canarius. *The Journal of comparative neurology* 165(4):457-486.
40. Nordeen KW & Nordeen EJ (1997) Anatomical and synaptic substrates for avian song learning. *Journal of neurobiology* 33(5):532-548.
41. Farries MA & Perkel DJ (2002) A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22(9):3776-3787.
42. Graybiel AM (2005) The basal ganglia: learning new tricks and loving it. *Current opinion in neurobiology* 15(6):638-644.
43. Fee MS & Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* 198:152-170.
44. Scharff C & Nottebohm F (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 11(9):2896-2913.
45. Warren WC*, et al.* (2010) The genome of a songbird. *Nature* 464(7289):757-762.
46. Kao MH, Doupe AJ, & Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433(7026):638-643.
47. van Vreeswijk C & Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274(5293):1724-1726.
48. Aronov D & Fee MS (2011) Analyzing the dynamics of brain circuits with temperature: design and implementation of a miniature thermoelectric device. *Journal of neuroscience methods* 197(1):32-47.
49. Aronov D, Veit L, Goldberg JH, & Fee MS (2011) Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31(45):16353-16368.
50. Hessler NA & Doupe AJ (1999) Social context modulates singing-related neural activity in the songbird forebrain. *Nature neuroscience* 2(3):209-211.
51. Goldberg JH, Farries MA, & Fee MS (2012) Integration of cortical and pallidal inputs in the basal ganglia-recipient thalamus of singing birds. *Journal of neurophysiology* 108(5):1403-1429.
52. Goldberg JH & Fee MS (2011) Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. *Journal of neurophysiology* 105(6):2729-2739.
53. Person AL & Perkel DJ (2005) Unitary IPSPs drive precise thalamic spiking in a circuit required for learning. *Neuron* 46(1):129-140.
54. Person AL & Perkel DJ (2007) Pallidal neuron activity increases during sensory relay through thalamus in a songbird circuit essential for learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27(32):8687-8698.
55. Luo M & Perkel DJ (1999) Long-range GABAergic projection in a circuit essential for vocal learning. *The Journal of comparative neurology* 403(1):68-84.

56. Goldberg JH & Fee MS (2012) A cortical motor nucleus drives the basal ganglia-recipient thalamus in singing birds. *Nature neuroscience* 15(4):620-627.
57. Goldberg JH, Farries MA, & Fee MS (2013) Basal ganglia output to the thalamus: still a paradox. *Trends in neurosciences* 36(12):695-705.
58. Leblois A, Wendel BJ, & Perkel DJ (2010) Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(16):5730-5743.
59. Leblois A, Bodor AL, Person AL, & Perkel DJ (2009) Millisecond timescale disinhibition mediates fast information transmission through an avian basal ganglia loop. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29(49):15420-15433.
60. Woolley SC, Rajan R, Joshua M, & Doupe AJ (2014) Emergence of context-dependent variability across a basal ganglia network. *Neuron* 82(1):208-223.
61. Kojima S, Kao MH, Doupe AJ, & Brainard MS (2018) The avian basal ganglia are a source of rapid behavioral variation that enables vocal motor exploration. *The Journal of neuroscience : the official journal of the Society for Neuroscience*.
62. Farries MA & Fairhall AL (2007) Reinforcement learning with modulated spike timing dependent synaptic plasticity. *Journal of neurophysiology* 98(6):3648-3665.
63. Tesileanu T, Olveczky B, & Balasubramanian V (2017) Rules and mechanisms for efficient two-stage learning in neural circuits. *Elife* 6.
64. Doya K & Sejnowski TJ (1995) A novel reinforcement model of birdsong vocalization learning. *Advances in Neural Information Processing Systems*
7: 101-108.
65. Doya K & Sejnowski TJ (1998) A computation model of birdsong learning by auditory experience and auditory feedback. *Central Auditory Processing and Neural Modeling*:77-88.
66. Leblois A & Perkel DJ (2012) Striatal dopamine modulates song spectral but not temporal features through D1 receptors. *The European journal of neuroscience* 35(11):1771-1781.
67. Mehaffey WH & Doupe AJ (2015) Naturalistic stimulation drives opposing heterosynaptic plasticity at two inputs to songbird cortex. *Nature neuroscience* 18(9):1272-1280.
68. Mackevicius EL & Fee MS (2018) Building a state space for song learning. *Current opinion in neurobiology* 49:59-68.

Chapter 2

A Microcircuit Mechanism for Dynamic Modulation of Order and Disorder in the Basal Ganglia

This chapter contains work done in collaboration with Dr. Agata Budzillo, Ms. Kimberly Miller, Dr. Adrienne Fairhall and Dr. David Perkel. Dr. Budzillo, Ms. Miller and Dr. Perkel conducted all experiments for this work. I contributed to the design and analysis of the model interpretation of these experimental results in collaboration with Adrienne Fairhall, David Perkel and Agata Budzillo. Throughout this chapter I note where I have explicitly used aspects of our published manuscript, but all of this work was the result of our collaborative efforts. Our published work from this project is: *Dopaminergic modulation of basal ganglia output through coupled excitation-inhibition*; Budzillo, Agata; Duffy, Alison; Miller, Kimberly E; Fairhall, Adrienne L; Perkel, David J; Proceedings of the National Academy of Sciences of the United States of America, 30 May 2017, Vol.114(22), pp.5713-5718

Abstract (quoted from (1))

Learning and maintenance of skilled movements require exploration of motor space and selection of appropriate actions. Vocal learning and social context-dependent plasticity in songbirds depend on a basal ganglia circuit, which actively generates vocal variability. Dopamine in the basal ganglia reduces trial-to-trial neural variability when the bird engages in courtship song. Here, we present evidence for a unique, tonically active, excitatory interneuron in the songbird basal ganglia that makes strong synaptic connections onto output pallidal neurons, often linked in time with inhibitory events. Dopamine receptor activity modulates the coupling of these excitatory and inhibitory events in vitro, which results in a dynamic change in the synchrony of a modeled population of basal ganglia output neurons receiving excitatory and inhibitory inputs. The excitatory interneuron thus serves as one biophysical mechanism for the introduction or modulation of neural variability in this circuit.

Introduction:

The basal ganglia are involved in motor learning and action selection across many species (2). Striatal dopamine critically shapes these roles (3, 4), but little is known about how dopamine modulates microcircuit activity within the basal ganglia. In disorders that effect movement, such as Parkinson's disease, an abnormal synchrony is observed in regions of the basal ganglia (5, 6). In normal functioning, the rhythms of the oscillatory neuron types within the nucleus combine to produce precise, selective motor gestures.

In the song bird, the basal ganglia region that contributes to song learning and song variation is called area X. Area X encompasses several basal ganglia structures and is part of a cortico-basal ganglia thalamic loop called the anterior forebrain pathway (abr. AFP) (Fig. 2.1a) (7, 8). Area X is composed of a large number of spiny-type neurons and many fewer oscillatory, pallidal-type projection neurons (Fig. 2.1b) (9). The AFP loop both receives inputs from the primary motor pathway and projects onto it. The AFP loop is required for learning and changes to song (10). While the critical learning period is during development, ongoing plasticity continues in adulthood (11). Adult males increase the variability in their songs when they are singing alone, in the undirected state (12, 13), and are also able to learn hearing driven changes to song (Fig 2.1c) (11, 14). In experiments that remove AFP activity, the song takes on an immediate and abnormal stereotypy (12, 15, 16).

How does the AFP loop generate the variability it injects into song? One source of variability emerges from the cortical-like structure, LMAN, which is the output of the circuit and projects directly to the primary motor pathway (10, 13). However, variability emerges independently from Area X as well (17, 18). Concentrations of D1-type dopamine increase in Area X during directed song (19). When a D1 agonist is injected into Area X during undirected

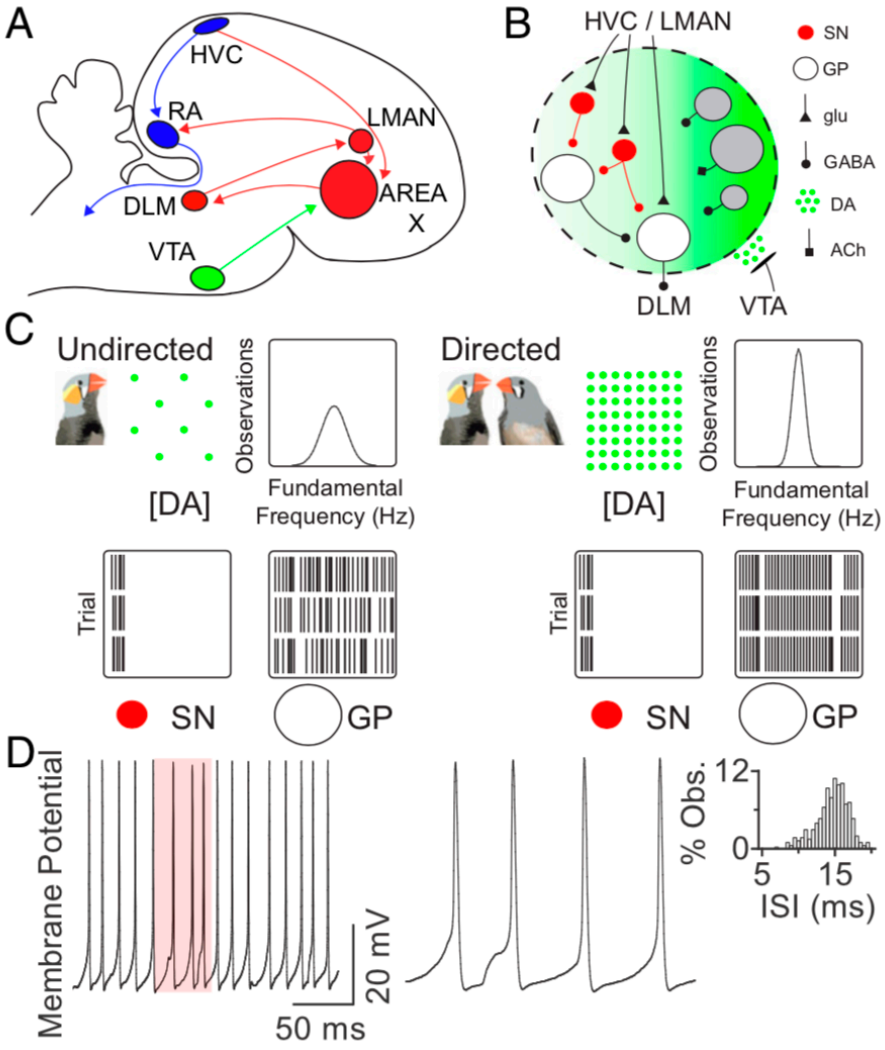

Figure 2.1. (figure and figure legend from Budzillo et al, 2017). Effects of social context and dopamine on area X neuron firing. **A.** Diagram of the songbird brain. Blue, motor pathway; red, learning pathway; green, midbrain dopamine input. **B.** Circuitry within area X. Red, spiny neurons (SN); gray; local interneurons; white, pallidal neurons (GP). ACh, acetylcholine; DA, dopamine; glu, glutamate. **C.** Schematic of social context-dependent changes in behavior and neural activity in area X (after refs. 20 and 23). During courtship, area X DA rises, narrowing the distribution of fundamental frequency across song trials. Simultaneously, area X GP output neuron firing becomes less variable. Input SNs maintain precise firing. **D.** Regular pallidal neuron firing. (Left) Example pallidal neuron re- cording in current-clamp configuration with no current injection. (Center) Magnification of shaded region at Left illustrates underlying synaptic potentials. (Right) Interspike interval (ISI) distribution for this neuron.

singing, the song becomes more precise and resembles directed song (19). Furthermore, dopaminergic projections from the ventral tegmental area (abr. VTA) send time-step specific signals that resemble a reward prediction error from reinforcement learning theory and likely modulate the output activity of area X in a trial-to-trial manner (20) (additionally see Chapter 5 of current work). Together this suggests that dopamine concentrations in area X play an important role in the context dependent variability of motor output.

What is the source of variation within the area X nucleus? HVC projections to spiny neurons during song are precise and punctate (21). The spiny neurons inherit the precision of the HVC inputs and do not become more variable during undirected song (18). However, the output activity of the area X nucleus, the pallidal neurons, is more variable during undirected song, even in the absence of the variable LMAN inputs (17). How then does the modulation of variability emerge from within the area X circuitry and what is the role of dopamine in shaping this process?

To answer this question, we recorded intracellularly from pallidal neurons in brain slices and studied their synaptic inputs (Fig 2.1d). We found a unique, spontaneously active, local glutamatergic neuron type which is the first excitatory cell type to be found in this previously entirely inhibitory nucleus. This excitatory input contributes to the variability of the pallidal firing properties. A simple model of the pallidal cell and the mixed, song-independent inhibitory and excitatory inputs onto it suggests a potential mechanism for dopaminergic modulation. We propose a microcircuit switch that could allow dopamine to control the variability and synchrony of pallidal subpopulations and in turn shape motor outputs depending on social context and song state.

Experimental Results:

Here I present a brief summary of the experimental findings which motivated our modelling choices. For a complete description of the experimental portion of this project, readers are referred to our published work (1) and to Dr. Agata Budzillo's dissertation (22). The central experimental finding is the existence of a novel excitatory neuron type within area X that fires regularly with an in vitro average frequency of approximately 20 Hz (Fig. 2.2a,b). The new excitatory neuron type makes strong connections onto multiple pallidal output neurons. Paired recordings of pallidal cells in voltage clamp showed time-locked excitatory synaptic inputs to both cells, suggesting that the excitatory neurons potentially project broadly to local pallidal populations (Fig. 2.2c). Correlations have been observed in paired recordings of pallidal output cells (unpublished data from Perkel lab). The excitatory inputs are frequently followed by time-locked inhibitory inputs. Application of a dopamine agonist (D1 receptor agonist SKF-38393) increased the likelihood of the coupled excitatory-inhibitory synaptic events without changing the overall rate of the excitatory synaptic inputs (control, 26.7 ± 7.51 ms, vs. D1R agonist, 37.1 ± 7.46 ms; P = 0.003; mean of differences, −10.4; 95% CI, −16.3 to −4.48) (Fig. 2.2d). Glutamate blocks in slice increase the variance of the pallidal single cell ISI distribution. This suggests that the excitatory inputs play a role in regulating the variability of the pallidal firing activity.

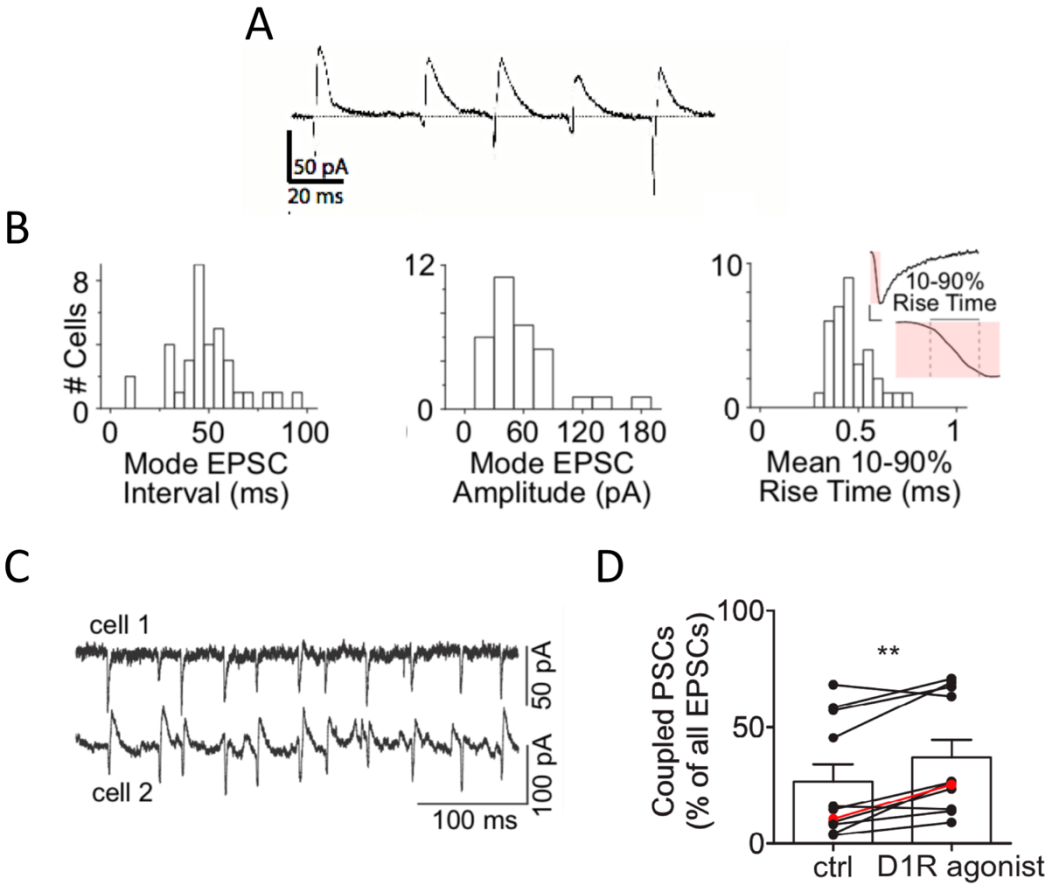

Figure 2.2. Summary of experimental findings. Large, regular, unitary glutamatergic synaptic events impinge on area X pallidal neurons. Figure panels and legends B-D from Budzillo et al 2016 and panel A from Perkel Lab. **A.** Example of voltage clamp recording showing large, regular EPSCs linked with IPSCs. **B.** Summary of EPSC properties. (Left) Inter-EPSC interval mode across 36 neurons, indicating a rate of ∼20 Hz. (Middle) EPSC amplitude mode across 32 neurons, indicating strong excitatory inputs to pallidal cells. (Right) EPSCs had fast rise times (mean rise time, 0.47 ± 0.10 ms; n = 35 neurons), consistent with a unitary origin. **C.** Example paired recording showing that two pallidal cells receive simultaneous EPSCs. The most likely explanation is that they arise from a single presynaptic excitatory neuron. **D.** D1R agonist significantly increased the percentage of all EPSCs that led an IPSC by at most 4 ms (control, 26.7 ± 7.51 ms, vs. D1R agonist, 37.1 ± 7.46 ms; P = 0.003; mean of differences, −10.4; 95% CI, −16.3 to −4.48).

Modeling Results:

We consider two feed forward internal microcircuits that drive an output pallidal population. The first microcircuit input (E) is simply a periodic excitatory synaptic input, putatively from a single, glutamatergic, intrinsically oscillatory interneuron. The second microcircuit input (EI) is a periodic, excitatory synaptic input followed after a brief pause (3-5 ms) by an inhibitory synaptic input. The synaptic inputs sum in our model. We do not distinguish in the model whether the EI microcircuit is from one excitatory cell driving an excitable, inhibitory cell or whether a more exotic form of excitatory-inhibitory transmission exists in a single interneuron.

Both excitatory and inhibitory interactions amongst oscillatory cells can be either synchronizing or desynchronizing depending on many factors, such as the strength of the stimulus, the nature of the cells' response and the relative frequencies of the cells' intrinsic oscillations. We explore several system parameters for the two possible internal microcircuit states: the strength of the microcircuit's synaptic input onto the pallidal population, the relative frequency of the microcircuit drive to the pallidal frequency, the lag time in the coupling of the inhibitory and excitatory inputs and the likelihood that one microcircuit is recruited during any one synaptic event.

In the following sections we first develop a simplified mathematical model of the effects of the putative microcircuits on the pallidal population of cells in the presence and absence of dopamine. We then introduce an entropy metric to quantify the degree to which the pallidal population is ordered or disordered by the microcircuit drive. Lastly, we introduce sources of

noise to the original model and consider how robust our characterizations are under these noise sources.

Phase Response Curves and Firing Maps

Synaptic or injected current into an intrinsically active cell changes when the next spike occurs (Fig 1.3 a). How much the timing of the next spike is shifted depends on when the current perturbation (synaptic or injected) occurs in the oscillatory cell's intrinsic cycle. A phase response curve (abr. PRC) function is the change in phase as a function of the phase at which a perturbation occurs (23). Phase response curves can be computed analytically for various model neurons and can also be measured experimentally (24). Because they are a one-dimensional characterization of a cell's response they present a particularly tractable means of understanding cellular interactions (25). We take the analytic form of our PRC from fits of experimental measurements of the pallidal cell itself (Fig 1.3 b,c).

Going from the infinitesimal PRC of a cell to the PRC of specific inputs

The PRC computed from infinitesimal perturbations in the cell's activity is sometimes called the infinitesimal phase response curve (abr. iPRC) (26). If inputs are sufficiently weak, responses to infinitesimal inputs, the iPRC, can be used as an impulse response function. Convolving the iPRC with the waveform of the input generates a microcircuit phase response curve: the pallidal cell's response to the particular form of the microcircuit's input at different

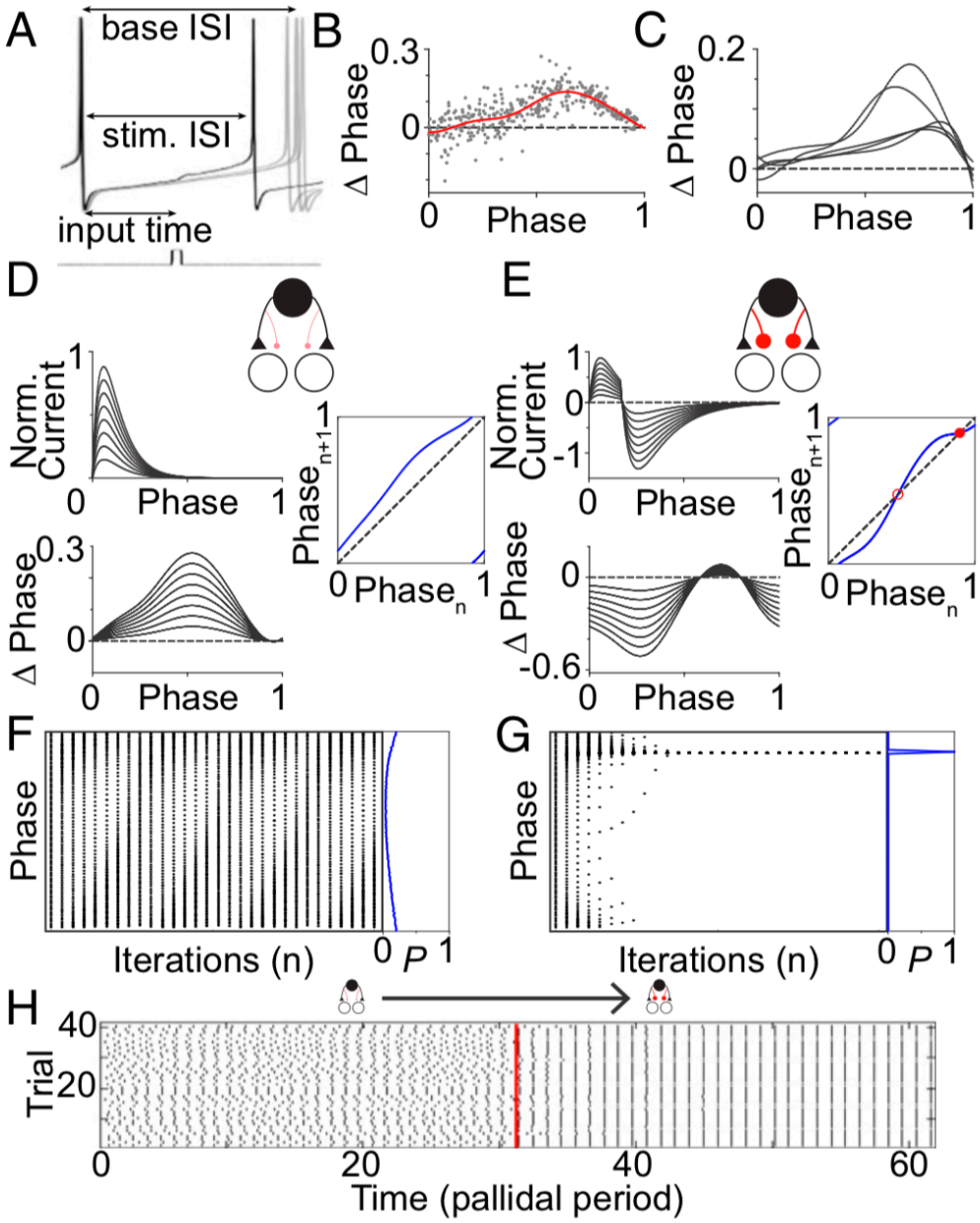
A
base ISI
stim. ISI
input time
B
Δ Phase
0.3
0
0
Phase
1
C
Δ Phase
0.2
0
0
Phase
1
D
Norm. Current
1
0
Phase
1
Phase$_{n+1}$
1
0 Phase$_n$ 1
Δ Phase
0.3
0
0
Phase
1
E
Norm. Current
1
0
-1
0
Phase
1
Phase$_{n+1}$
1
0 Phase$_n$ 1
Δ Phase
0
-0.6
0
Phase
1
F
Phase
Iterations (n)
0 P 1
G
Phase
Iterations (n)
0 P 1
H
Trial
40
20
0
20
40
60
Time (pallidal period)

Figure 2.3 (from Budzillo et al., 2017) **A.** Experimentally measured pallidal iPRC constrains simple model of how DA affects the area X microcircuit. (A) Example of pallidal phase shifts caused by small current pulse (50 pA, 2 ms). **B.** Phase shifts caused by single current pulses in a pallidal neuron. Red curve represents analytic fit to points (R^2 = 0.52). **C.** Individual fits to five pallidal neurons show qualitative similarity. **D.** We convolved the normalized EPSC (Upper Left) with the parameterized iPRC (C) to obtain the microcircuit PRC (Lower Left). Multiple synaptic strengths are shown. (Right) Firing map iteratively relating the phase of the pallidal neuron at the onset of one input to its phase at the time of the next input. **E.** Same as in D but for linked excitatory–inhibitory (EI) synaptic events. Filled red circle in firing map indicates a stable fixed point; open red circle indicates unstable fixed point. **F.** Trajectory of the firing phases of pallidal neuron ensemble relative to excitatory neuron un- der excitatory (E) microcircuit drive. (Left) Pallidal phase ensemble evolution under firing map drive across multiple initial conditions (Fig. 1D). (Right) Blue line plots the resulting phase probability distribution. Note lack of convergence to a single phase (high entropy, low synchrony). **G.** Same as in F for EI microcircuit drive. Note convergence of pallidal ensemble to a single phase (low entropy, high synchrony). **H.** Change in firing of pallidal ensemble over time as microcircuit shifts from excitation only to mixed excitation and inhibition. Each dot represents a pallidal neuron firing event, and each row indicates the progression of a single trial with a different, randomly selected initial phase. Vertical red line indicates the time when the microcircuit switched.

points in phase (26). Using the iPRC we can create microcircuit phase response curves for the effects of the synaptic inputs from both internal circuit motifs (Fig 1.3 d,e) (26, 27).

Definition of a firing map

From the PRC we create discrete firing maps, which map the point in the pallidal phase of the arrival of one microcircuit input to the onset of the next microcircuit input and thereby calculate the trajectory of the interactions of the pallidal cell with this internal circuit:

$$\phi_{n+1} = \{\phi_n + PRC_{syn}(\phi_n) + T_{mc}\}_{mod\ T_p}.$$

T_{mc} is the period of the microcircuit inputs, and T_p is the period of the pallidal cell (27). ϕ_n is the pallidal phase at which the 'nth' synaptic input arrives. The firing map computes the pallidal phase, ϕ_{n+1}, of the arrival of the next synaptic event onto the pallidal cell (Fig 1.3 d,e).

Characterizing the effects of dopamine.

Our experimental findings show that the switch between the E and EI microcircuits is partially mediated the presence of dopamine. The firing maps for the E and EI microcircuits are distinguished by their respective synaptic PRCs: the variation of the waveform of the synaptic input, which is convolved with the iPRC, is what classifies an E firing map from an EI firing map. We therefore must consider the characteristics of firing maps in isolation and as parts of a

coupled E and EI pair. The difference between the E and EI microcircuit effects model the impact of the presence or absence of dopamine.

Definition of population entropy

We calculate the entropy of a population of pallidal cells when begun at random initial points in their phases to quantify both the degree of synchrony in the pallidal population and the trial-to-trial variability of a single cell under the effects of the populations of E and EI microcircuit drives (27). We calculate the entropy of the population as:

$$S = -\sum_{i=1}^{M} p(\varphi_i)\ln\left(p(\varphi_i)\right),$$

where $p(\varphi_i)$ is the probability of finding a cell's phase in bin 'i', and M = 100. The upper bound of this entropy metric is 4.61. The probability is estimated at each time point by evolving a population of firing maps started at random, uniformly distributed phases and normalizing the population distribution as an estimate of the probability density function. We allow the map to evolve 200 iterations and then take the distribution of the last 10 consecutive iterates as an estimate for a steady state distribution. From this we calculate the population entropy. We explore the population entropy over varying amplitudes of synaptic inputs and ratios of pallidal to microcircuit firing frequencies. At the highest intensity of synaptic input used to compute the entropy parameter scape, the greatest magnitude phase shift caused by the EI microcircuit is 0.50 (in units of pallidal phase) and the greatest magnitude phase shift caused by the E

microcircuit is 0.37 (in units of pallidal phase). These ranges were chosen by our fits to the experimental measurements.

When we calculate the entropy over a range of synaptic intensities and ratios of pallidal to microcircuit firing frequencies, we find that, for large regimes within the intensity-frequency parameter space, the two internal microcircuits have complimentary effects: regimes where the single EPSC microcircuit synchronizes its driven population, the linked EPSC-IPSC microcircuit is desynchronizing and vice versa (Fig 2.3 f-h; Fig 2.4 a,b). This suggests an interpretation for the role of varied levels of dopamine in area X: dopamine levels switch the microcircuit drive from being a synchronizing, regularizing drive to a desynchronizing, scrambling drive.

Note that we will use both interpretations of an ensemble estimate of entropy: we are considering the effects on a true population of pallidal cells, and we also use this population to represent the distribution of a single cell's activity to multiple repeated trials of the same microcircuit stimulus. These two interpretations speak to two different experimental observations. One, the putative internal microcircuits project broadly to at least local subpopulations of the pallidal group. Two, single cell recordings of pallidal cells during performance have greater ISI variance during undirected song (17); while in slice, single cell recordings of pallidal neurons in the presence of glutamate blockers have less ISI variance.

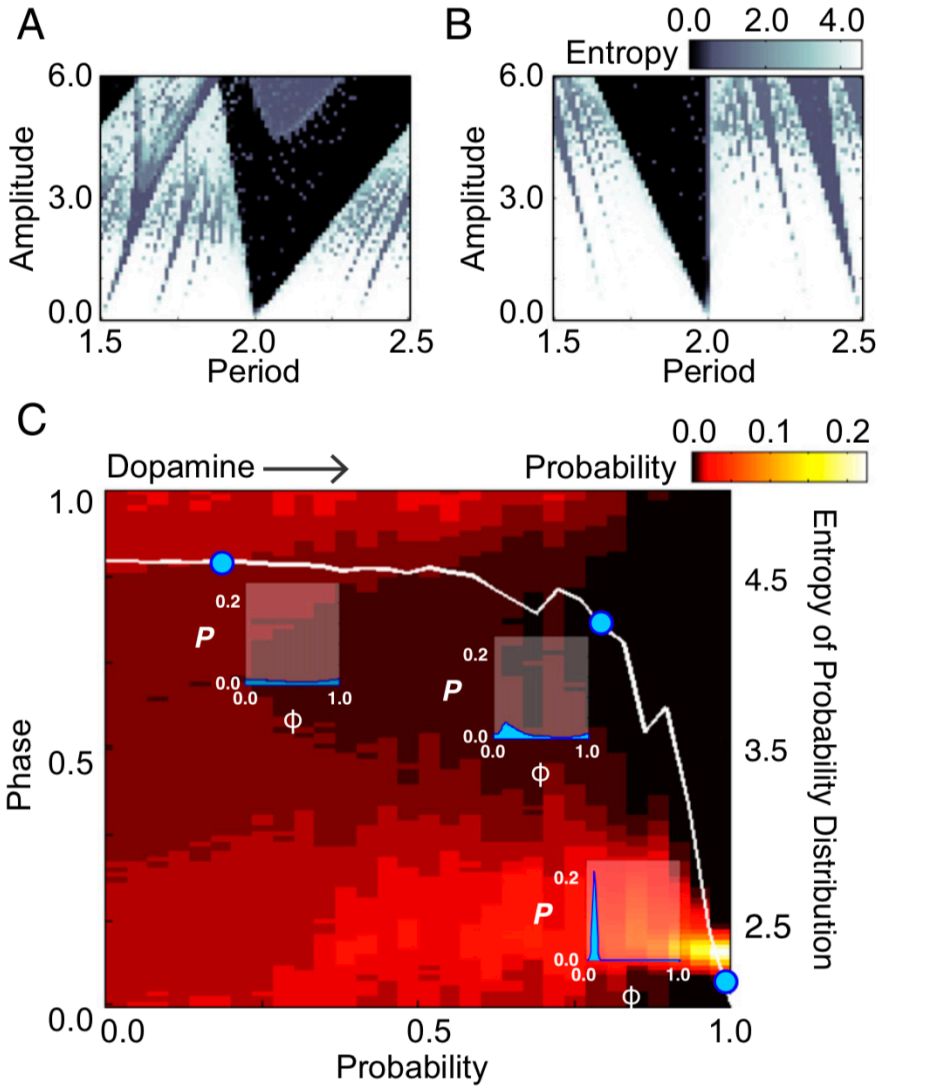

Figure 2.4. (from Budzillo et al. 2017). Neural firing entropy from a simple model of the area X microcircuit in different conditions. **A.** The entropy of the distribution of pallidal firing phase varies with synaptic amplitude and period of E microcircuit drive relative to pallidal period. **B.** Same as A, but for EI microcircuit. **C.** Effects of probabilistic inclusion of the inhibitory element on firing-phase distribution. Pallidal neuron intrinsic firing had low variability (CV = 0.05). Heat map shows the effect on the pallidal phase–probability distribution as the probability of selecting the EI microcircuit varied from 0 to 1 (abscissa). For each EI probability, the resulting phase (left ordinate) probability distribution is plotted as a column of heat values. Entropy from each column is plotted (right ordinate) as a white line. Insets show the probability distribution at three example EI probability values, corresponding to blue circles.

Probabilistic participation of microcircuit states

Experimental observation of these internal microcircuits shows that there is flipping between the two microcircuit states: the coupled EI microcircuit will be disrupted by variable lapses where only the E microcircuit occurs. Indeed, the D1 agonist only increases the fractional participation of the EI microcircuit—it does not create a binary change from one microcircuit to the other (Fig 2.2d). How robust is the synchronization under one microcircuit when it is randomly "perturbed" by the other microcircuit?

We model two aspects of noise: η represents variability in the pallidal ISI and is modeled as a Gaussian random variable with zero mean and variance, σ=0.05, in units of pallidal phase (Fig 2.5 a-c). We model probabilistic jumps between microcircuit states as binomial draws of firing maps f and g:

$$\phi_{n+1} = f(\phi_n);$$

$$\phi_{n+1} = g(\phi_n);$$

$$f(\phi_n) = \{\phi_n + PRC_{EI}(\phi_n) + T_{mc} + \eta\}_{mod\ T_p};$$

$$g(\phi_n) = \{\phi_n + PRC_{E}(\phi_n) + T_{mc} + \eta\}_{mod\ T_p};$$

$$\mathrm{P}\{\phi_{n+1} = f(\phi_n)\} = 1 - \mathrm{P}\{\phi_{n+1} = g(\phi_n)\}\,.$$

Excitatory events couple to inhibitory events in a probabilistic manner. Dopamine shifts the likelihood of linked events. Figure 2.4 panel c shows the time-averaged distribution of a population of initial conditions under a family of random iterated firing maps over which the Bernoulli probability of switching from the E firing map to the EI firing map varies from zero to one. To compute each probability distribution, an initial ensemble of 1000 phases were drawn from a uniform distribution between zero and one and then allowed to iterate 500 times. The time-average of the distribution was calculated by combining the phase locations of the cell ensemble over the last 400 iterations. Small shifts in probability mass can create dramatic differences in the synchrony of the total population.

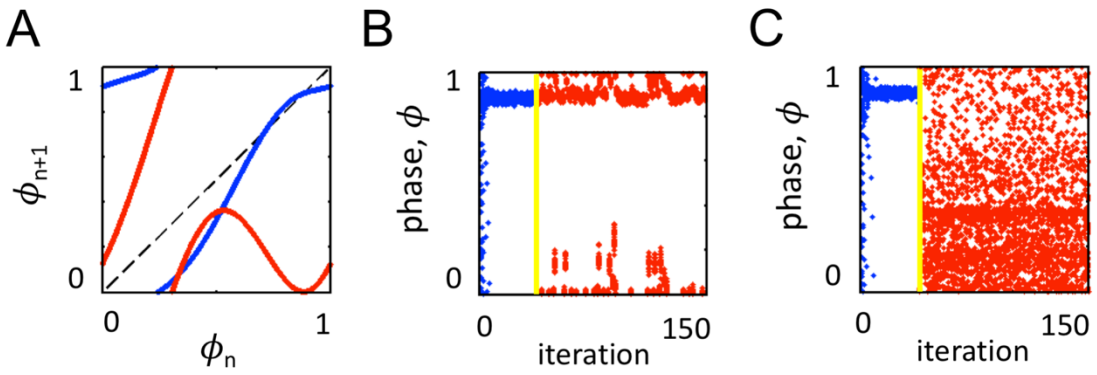

Figure 2.5. Population of model cells driven under the random iterated map model. **A.** The firing maps for example E and EI microcircuits. **B.** Evolution of the phases of 100 pallidal cells initially uniformly randomly distributed over [0,1). The microcircuit model now has added Gaussian noise, (mean,η = 0; variance, σ = 0.05). This leads to jitter about the underlying stable fixed point of the blue firing map. At the yellow line the red map probabilistically replaces the blue firing map with a Bernoulli probability of 0.25. Switching the probability of flipping between and E and EI microcircuits changes the degree of synchronization in the population. Note that the red map begins to desynchronize the phases, but they remain relatively close to the initial fixed point. **C.** Same as in (b) but at the yellow line the red map completely replaces the blue map with Bernoulli switch probability of 1.

Geometric explanation of the switch mechanism

Zero entropy in the deterministic firing maps exists when the firing map has a stable fixed point, which corresponds to the microcircuit fully entraining the pallidal population. In some regimes, switching microcircuits corresponds to the disappearance (or appearance) of a stable fixed point. Because the inhibitory synaptic input follows closely after the excitatory synaptic input and has a slower decay time course, the synaptic phase response curve for the EI map switches the synaptic PRC from being strictly positive (in the E microcircuit) to almost or completely negative-- approximately a reflection across the zero-phase shift axis (though the actual wave form changes) (Fig 2.6a). This corresponds in the firing map to a skewed reflection about the line: $\phi_{n+1} = \phi_n + T_{mc}$. This line is parallel to the line of fixed points ($\phi_{n+1} = \phi_n$), which implies that a reflection about the $\phi_{n+1} = \phi_n + T_{mc}$ line will either create or destroy fixed points when the difference between T_{mc} and T_p is within the range of the synaptic PRC (Fig 2.6b).

This geometric explanation for the asymmetric relationship of the stable regimes within the amplitude-relative frequency parameter scape shown in Figure 2.4, panels A and B applies most directly to the primary tongues of low entropy where we would imagine this mechanism would be most robust to slight non-stationarity in parameter values.

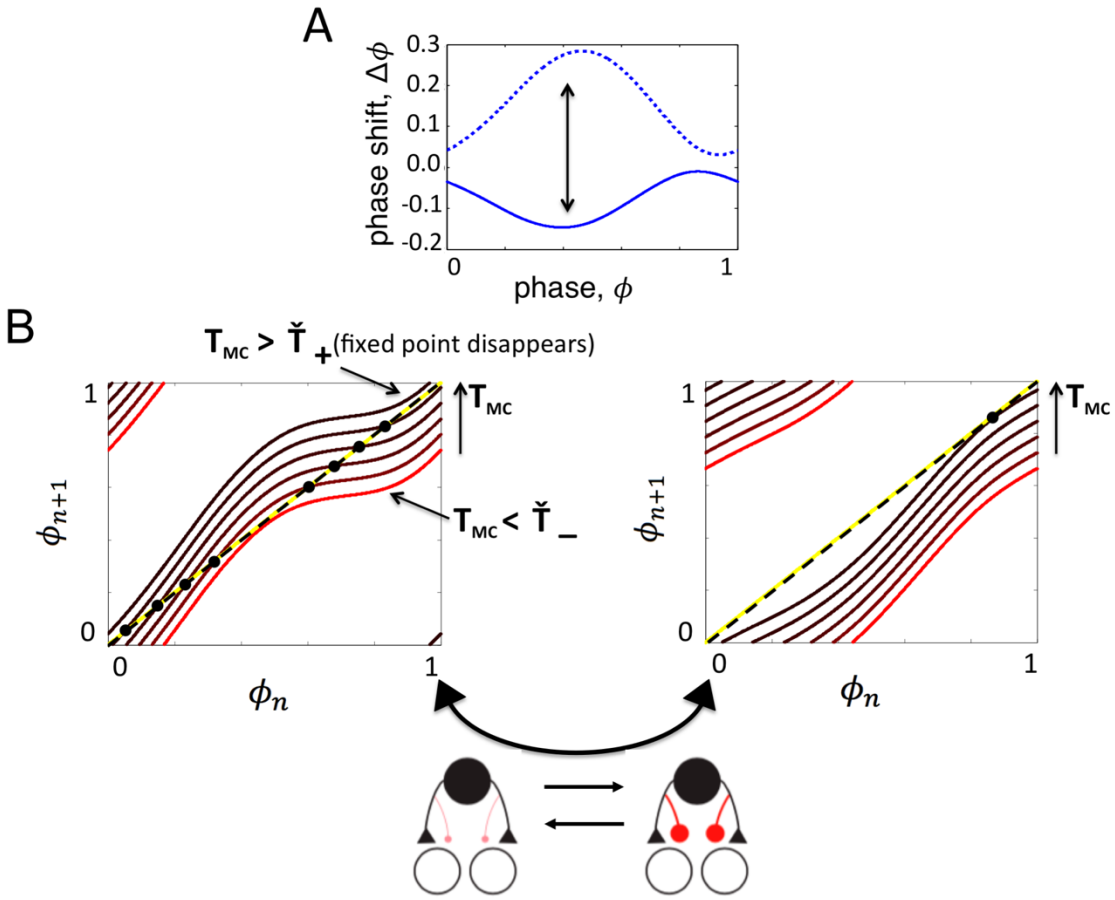

Figure 2.6. Geometric explanation for microcircuit switch mechanism. **A.** An example of the shift in the microcircuit phase response curve between the E and EI microcircuits. The positive PRC is the E microcircuit and the negative PRC is the EI microcircuit. This example is typical of the transition that happens during the E → EI flip: the PRC switches from causing only phase advances to causing phase retreats. This is an approximate reflection about the zero-phase shift axis. **B.** Geometric depiction of a family of firing maps flipping from the E to EI microcircuit configuration. Colors match pairs of firing maps with the same periodic relationship between the microcircuit and the pallidal cells for the left and right panels. Firing maps which cross the $\varphi_n = \varphi_{n+1}$ line and have a stable fixed point under the E microcircuit lose their fixed points upon switching to the EI microcircuit. This translates in the firing map to a shift about the line: $\varphi_n = \varphi_{n+1} + T_{MC}$. In the region, $T_- < T_{MC} < T_+$, switching microcircuits causes the disappearance of a fixed point in the firing map. T_{MC} is period of the driving microcircuit in units of pallidal phase. [T_- , T_+] is the range in which a stable fixed point exists for the E firing map and will vary as a function of the synaptic input strengths and the form of the iPRC.

Discussion:

In this modelling work we explore the possibility that a newly discovered glutamatergic cell type might allow dopamine concentrations both to modulate trial-to-trial variability and to rapidly create and dissolve temporary cell assemblies within the output of the basal ganglia like region, area X, in bird song.

Modulation of the tight coupling of excitatory and inhibitory inputs to regularly firing neurons provides a novel mechanism for controlling both population synchrony and single cell variability. Our model results inform both the single cell and the population level. In the single cell interpretation, changes to the microcircuit input state modulate the variability of the ISI distribution of single pallidal neurons. This change in ISI variance of pallidal firing patterns has been seen experimentally both in vitro over varying concentrations of D1 agonist and in vivo across varying behavioral contexts (17, 28). Our model suggests changes in the coupled excitatory and inhibitory internal circuit as a mechanism. If the periodic synaptic inputs are in certain intensity and frequency regimes, the variability of the ISI distribution will increase, and spike timing dependent precision of the area X signal in relation to particular moments in song will diminish.

At the population level, firing patterns in our model becomes more or less synchronous depending on the driving microcircuit state. There are several reasons the bird song system may take advantage of temporary synchrony in subpopulations of pallidal neurons. There are approximately 400,000 spiny neurons in X, 3,000 pallidal output neurons in X, 10,000 LMAN neurons, 1,000 primary motor neurons and eight muscles participating in song (29). Temporary

synchrony of subpopulations of neurons during task specific network activity could be one means of dimensionality reduction of the neural populations to the number of muscles innervated during specific tasks.

While this work shows a way in which variability could be generated from within area X, the primary source of variable activity from within the AFP circuit is thought to come from the cortical-like region, LMAN (12, 16, 30, 31). LMAN receives inputs from the X-pallidal population via pathways through DLM (12, 31). Pallidal projections to DLM are thought to be roughly one-to-one, but multiple DLM inputs converge onto single LMAN neurons (32, 33). X projections onto DLM control the timing of DLM spiking with sub-millisecond precision and, therefore, exert a strong, temporally precise influence on LMAN (28, 34-36). If area X sometimes entrains LMAN's intrinsically variable output, the momentary coherence of multiple X output neurons could enhance the strength of particular signals downstream to LMAN thus modulating the signal-to-noise ratio of LMAN's output. Previous theoretical studies have shown that recurrent networks can undergo stimulus dependent suppression of their intrinsic, chaotic activity (37). In particular, the stronger the external stimulus, the easier it is to entrain a chaotic network's response to a specific stimulus (37). Temporarily synchronizing subpopulations of X outputs could be one way to strengthen a task-specific signal. Wang et al. 2010 found in a model of feedforward thalamic inputs to primary visual cortex, that the reliability of spike transmission increased sharply at around 20-40 synchronous thalamic inputs within a 5-millisecond time window. From this study they concluded that ensuring reliable spike transmission through the synchrony of small subpopulations of neurons may be a general principle of cortical processing (38).

From this perspective, changes in dopamine levels modulate the degree to which the outputs from Area X play an active control role, wherein a time-varying, modulated signal is transmitted from Area X through DLM and LMAN to drive directed changes in RA activity. In this picture, LMAN and Area X switch back and forth as the primary signaling nucleus of the output from the AFP loop. When Area X is transmitting a decorrelated, low content signal, variations in output from LMAN emerge primarily from within the intrinsic dynamics of the LMAN nucleus and play an exploratory, stochastic role in learning. When area X is transmitting highly correlated subpopulations of spiking activity, the tightly controlled coupling with downstream thalamic recipients results in highly synchronized inputs to the cortical LMAN, which acts as an effective control of LMAN's outputs. Thus, dopaminergic modulation of synchrony in pallidal sub-populations could not only influence shift from directed to undirected song state but also exploration and exploitation patterns within undirected song.

Future work.

One experimental prediction from this work is that there should be observable changes in levels of synchrony across subpopulations within area X over the course of learning and in the state transitions between directed and undirected song. Because our theory proposes that subpopulations of pallidal output cells should selectively synchronize in response to environmental state changes and learning, local field potential recordings may be able to register these types of shifts in subpopulation synchrony.

The current model addresses a discrete microcircuit within area X and its effects on the pallidal cell population. However, there are at least three interacting inputs to the output layer of area X that are modulated during song: the area X inhibitory interneuron layer which is composed of both cholinergic and GABAergic interneurons, the excitatory glutamatergic input from HVC, and the glutamatergic input from the putative excitatory interneuron within the area X structure (9, 39). Further modeling and anatomical investigations are necessary to fully understand how all of these network interactions combine.

The next step in developing the theoretical ideas proposed here would be to construct a larger scale model of the interactions within area X which account for both song specific and state specific inputs to the pallidal population in order to further explore how changes in firing probability due to neuromodulation by dopamine could lead to a modulation of the signal strength in the nucleus' output.

Extended Methods:

The formulation of the firing map model is fully described in the Results section of the text. Below are the methods for the measurement and analytic fits of the iPRC and the synaptic wave forms. This section is quoted from our published work, Budzillo et al 2017 (1).

Calculation of the iPRC

The measurement of the iPRC was made as follows. iPRC experiments were conducted in 5 pallidal neurons in the presence of gabazine and NBQX (10 μM) (24). 2 ms current pulses were injected at a frequency of 2 Hz, with 4 stimulus presentations per sweep, and repeated at different amplitudes (±50/100/250 pA). Each stimulated ISI was associated with a pool of spontaneous ISIs occurring within 30 seconds of the ISI in question. The baseline ISI for that stimulation was defined as the mean of ISIs in this pool longer than the stimulated ISI. If there were fewer than 10 ISIs from which to make this calculation the associated pool of spontaneous ISIs was approximated by a Gaussian distribution with mean and standard deviation equal to that of the spontaneous ISIs occurring within 30 seconds of the stimulation. Input phase for a given stimulus was determined by dividing its input time by the mean baseline ISI, where the input time was the time between the peak of the preceding spike and midpoint of the current pulse. Phase change was the difference between the baseline ISI and the stimulated ISI (the time between the peak of the preceding spike and the peak of the spike following current injection) divided by the mean baseline ISI.

Analytical form of the experimental iPRC

The experimental iPRC was fit to an analytical form of sum of sines and cosines:

$$iPRC(\phi) = a_0 + \sum_{p=1}^{p=3} a_p \sin(p\omega\phi) + b_p \cos(p\omega\phi)$$

All cells measured showed qualitative similarity in their iPRC: almost entirely phase advancing with the biggest phase advance occurring approximately two thirds of the way into the oscillation cycle. Because of this similarity, we chose an analytic iPRC fit to a single

representative cell. For this cell, the R-squared value of the fit parameters is 0.52. The parameters of the fit are provided below.

Table 2.1 Fit parameters for the iPRC.

Parameter	Value	Confidence interval (95%)
a_0	0.055601369321404;	(0.04054, 0.07066)
a_1	-0.061151688548318	(-0.09477, -0.02753)
a_2	-0.006863029049976	(-0.02946, 0.01574)
a_3	-0.005532332972041	(-0.01231, 0.001244)
b_1	-0.033316012078074	(-0.07761, 0.01098)
b_2	0.015454546227122	(-0.0007489, 0.03166)
b_3	-0.001488447238460	(-0.0116, 0.008619)
w	5.848805665250545	(4.72, 6.978)

Calculation of synaptic waveforms. PRCs were generated by convolving a representative iPRC at +50 pA current injection with two classes of synaptic input observed in our data: an excitatory synaptic input and a coupled excitatory-inhibitory input. Excitatory synaptic input was modeled by a difference of exponentials:

$$Syn_E(t) = A_E * (e^{-\frac{t}{\tau_{E1}}} - e^{-\frac{t}{\tau_{E2}}}) + I.$$

Coupled excitatory/inhibitory input was modeled by summing two such differences of exponentials, one representing the excitatory component and the other representing the inhibitory component of the event:

$$Syn_{EI}(t) = Syn_E(t) + Syn_I(t) - I.$$

Coupled EPSCs from 6 pallidal neurons were collected and fit with the above equation in order to directly draw the distribution of parameters for events from the data. The following tau parameters yielded the best fit after examining both individual and coupled synaptic events: τ_{E1} = 1.2 ms; τ_{E2} = 0.7 ms; τ_{I1} = 2.5ms; τ_{I1} = 0.7 ms. These parameters were inserted as constants back into the above equations. Excitatory components of coupled events observed in our dataset precede inhibitory components. We found that the best fits to our data occurred when activation time courses $\tau_{E2}$and τ_{I1} were equivalent, but a short delay was imposed on the inhibitory component (mean/SD delay = 2.25 ± 0.61 ms; n= 231). This step activation was approximated by a hyperbolic tangent function so that *x* in both rising and falling phases of the inhibitory component was replaced with $(t-d)*\frac{1}{2}(\tanh(c*(t-d))+1)$, where *d* = delay and *c* = 1000. Allowing the inhibitory component of the equation to shift in time resulted in robust synaptic event fits (mean R-squared = 0.96 ± 0.03; n = 231 coupled EPSCs).

Bibliography

1. Budzillo A, Duffy A, Miller KE, Fairhall AL, & Perkel DJ (2017) Dopaminergic modulation of basal ganglia output through coupled excitation-inhibition. *Proceedings of the National Academy of Sciences of the United States of America* 114(22):5713-5718.
2. Graybiel AM (2005) The basal ganglia: learning new tricks and loving it. *Current opinion in neurobiology* 15(6):638-644.

3. Bromberg-Martin ES, Matsumoto M, & Hikosaka O (2010) Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron* 68(5):815-834.
4. Wise RA (2009) Roles for nigrostriatal--not just mesocorticolimbic--dopamine in reward and addiction. *Trends in neurosciences* 32(10):517-524.
5. Wichmann T, DeLong MR, Guridi J, & Obeso JA (2011) Milestones in research on the pathophysiology of Parkinson's disease. *Movement disorders : official journal of the Movement Disorder Society* 26(6):1032-1041.
6. Gatev P, Darbin O, & Wichmann T (2006) Oscillations in the basal ganglia under normal conditions and in movement disorders. *Movement disorders : official journal of the Movement Disorder Society* 21(10):1566-1577.
7. Bottjer SW, Miesner EA, & Arnold AP (1984) Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* 224(4651):901-903.
8. Scharff C & Nottebohm F (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 11(9):2896-2913.
9. Farries MA & Perkel DJ (2002) A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 22(9):3776-3787.
10. Olveczky BP, Andalman AS, & Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS biology* 3(5):e153.
11. Tumer EC & Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* 450(7173):1240-1244.
12. Kao MH, Doupe AJ, & Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433(7026):638-643.
13. Kao MH & Brainard MS (2006) Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. *Journal of neurophysiology* 96(3):1441-1455.
14. Charlesworth JD, Tumer EC, Warren TL, & Brainard MS (2011) Learning the microstructure of successful behavior. *Nature neuroscience* 14(3):373-380.
15. Brainard MS & Doupe AJ (2000) Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* 404(6779):762-766.
16. Goldberg JH & Fee MS (2011) Vocal babbling in songbirds requires the basal ganglia-recipient motor thalamus but not the basal ganglia. *Journal of neurophysiology* 105(6):2729-2739.
17. Woolley SC & Kao MH (2014) Variability in action: Contributions of a songbird cortical-basal ganglia circuit to vocal motor learning and control. *Neuroscience*.
18. Woolley SC, Rajan R, Joshua M, & Doupe AJ (2014) Emergence of context-dependent variability across a basal ganglia network. *Neuron* 82(1):208-223.
19. Leblois A, Wendel BJ, & Perkel DJ (2010) Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(16):5730-5743.
20. Gadagkar V*, et al.* (2016) Dopamine neurons encode performance error in singing birds. *Science* 354(6317):1278-1282.

21. Hahnloser RH, Kozhevnikov AA, & Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419(6902):65-70.
22. Budzillo A (2015) Microcircuitry of the songbird basal ganglia nucleus area X. Doctor of Philosophy Dissertation (University of Washington).
23. Winfree AT (2001) *The geometry of biological time* (Springer, New York) 2nd Ed pp xxvi, 777 p.
24. Farries MA & Wilson CJ (2012) Phase response curves of subthalamic neurons measured with synaptic input and current injection. *Journal of neurophysiology* 108(7):1822-1837.
25. Ermentrout GB & Koppell N (1986) Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM J. Appl. Math* 46:233–253.
26. T. NTSML (2012) Chapter: Experimentally estimating phase response curves of neurons: Theoretical and practical issues. Phase Response Curves in Neuroscience: Theory, Experiment, and Analysis. Springer Series in Computational Neuroscience, eds Schultheiss NW, Prinz AA, Butera RJ. 6:95-129.
27. Wilson CJ, Beverlin B, 2nd, & Netoff T (2011) Chaotic desynchronization as the therapeutic mechanism of deep brain stimulation. *Frontiers in systems neuroscience* 5:50.
28. Leblois A, Bodor AL, Person AL, & Perkel DJ (2009) Millisecond timescale disinhibition mediates fast information transmission through an avian basal ganglia loop. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 29(49):15420-15433.
29. Fee MS & Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* 198:152-170.
30. Ali F*, et al.* (2013) The basal ganglia is necessary for learning spectral, but not temporal, features of birdsong. *Neuron* 80(2):494-506.
31. Aronov D, Veit L, Goldberg JH, & Fee MS (2011) Two distinct modes of forebrain circuit dynamics underlie temporal patterning in the vocalizations of young songbirds. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31(45):16353-16368.
32. Boettiger CA & Doupe AJ (1998) Intrinsic and thalamic excitatory inputs onto songbird LMAN neurons differ in their pharmacological and temporal properties. *Journal of neurophysiology* 79(5):2615-2628.
33. Bottjer SW, Brady JD, & Walsh JP (1998) Intrinsic and synaptic properties of neurons in the vocal-control nucleus IMAN from in vitro slice preparations of juvenile and adult zebra finches. *Journal of neurobiology* 37(4):642-658.
34. Goldberg JH, Farries MA, & Fee MS (2012) Integration of cortical and pallidal inputs in the basal ganglia-recipient thalamus of singing birds. *Journal of neurophysiology* 108(5):1403-1429.
35. Person AL & Perkel DJ (2005) Unitary IPSPs drive precise thalamic spiking in a circuit required for learning. *Neuron* 46(1):129-140.
36. Person AL & Perkel DJ (2007) Pallidal neuron activity increases during sensory relay through thalamus in a songbird circuit essential for learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27(32):8687-8698.

37. Rajan K, Abbott LF, & Sompolinsky H (2010) Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* 82(1 Pt 1):011903.
38. Wang HP, Spencer D, Fellous JM, & Sejnowski TJ (2010) Synchrony of thalamocortical inputs maximizes cortical reliability. *Science* 328(5974):106-109.
39. Goldberg JH & Fee MS (2010) Singing-related neural activity distinguishes four classes of putative striatal neurons in the songbird basal ganglia. *Journal of neurophysiology* 103(4):2002-2014.

Chapter 3

Variation in Sequence Dynamics Improves Maintenance of Stereotyped Behavior: an Example from Bird Song

This chapter reports on research done in collaboration with Elliott Abe, Dr. David Perkel and Dr. Adrienne Fairhall. We have submitted the following text for publication and it is currently under review. I wrote the following text, and all of my collaborators contributed editing. In the course of this project, I contributed to the model and research design, simulation implementation, analysis and interpretation of the results.

Alison Duffy*[1], Elliott Abe*[1,2], David J. Perkel[4], Adrienne Fairhall[1,3]

[1]Dept. Physics, University of Washington
[2]Dept. Biology, Institute of Neuroscience, University of Oregon
[3]Dept. Physiology & Biophysics, University of Washington
[4]Depts. Biology & Otolaryngology, University of Washington
*A. Duffy and E. Abe share first authorship

Abstract:

Performing a stereotyped behavior successfully over time requires both maintaining performance quality and adapting efficiently to environmental or physical changes affecting performance. The bird song system is a paradigmatic example of learning a stereotyped behavior and therefore a good place to study the interaction of these two goals. Through a model of bird song learning we show how instability in neural representation of stable behavior confers advantages for adaptation and maintenance with minimal cost to performance quality. A precise, temporally sparse sequence from the premotor nucleus HVC is crucial to the performance of song in songbirds. We find that learning in the presence of sequence variations facilitates rapid relearning after shifts in the target song or muscle structure and results in decreased error with neuron loss. This robustness is due to the prevention of the buildup of correlations in the learned connectivity. In the absence of sequence variations these correlations

grow due to the relatively low dimensionality of the exploratory variation in comparison to the number of plastic synapses. Our results suggest one would expect to see variability in neural systems executing stereotyped behaviors, and this variability is an advantageous feature rather than a challenge to overcome.

Significance Statement:

In this work we show a novel way by which the nervous system maintains precise, stereotyped behavior in the face of environmental and neural changes. Through a model of bird song learning, we show how instability in neural representation of stable behavior can allow a system to more readily adapt and maintain performance with minimal cost. In this perspective, behaviors are made more robust to environmental change by continually seeking subtly new ways of performing the same task. Thus, one should expect to find variability in neural systems executing stereotyped behaviors, and this variability can serve a constructive role in maintaining skilled behavior.

Introduction

When we first learn to play the piano, ride a bicycle, or use a spoon, the learning trajectory is similar: initial attempts at the task are clumsy and erratic, one quite different from the next. But over time we slowly improve and eventually are able to complete these tasks in a highly skilled and reliable manner. At this point we usually think of learning as complete: we can do the desired task, we do it well, and we do it the same way each time. This trial-and-error learning process is common to many stereotyped tasks learned in development and performed in a seemingly automatic manner in adulthood. However, there is a need for ongoing plasticity in order to perform in a stereotyped manner despite changes over time, such as new

environmental conditions or physiological growth or injury. Active maintenance must therefore be an important element of the neural pathways that carry out these repetitive tasks.

A well-characterized biological example of a learned, stereotyped behavior is the courtship song of songbirds. We use this as an example in which we might expect such maintenance to occur. Juvenile male songbirds learn to sing from a male tutor in a trial-and error-manner. Once the bird reaches adulthood, the song becomes stereotyped with millisecond precision. Despite the adult song stability, there is ample evidence that song plasticity is maintained in adulthood (1-6). Adult songbirds are able to shift the pitch of individual syllables in response to white noise stimuli or shifted auditory feedback (1, 2). In deafened birds, song eventually degrades, implying ongoing plasticity (3, 4, 6).

One hypothesis about how stereotyped motor output is generated is that it emerges from precisely controlled and sequenced neural activity. Is stable behavior therefore underpinned by long term, stable representations at the population and single neuron level? This question applies to the bird song system, where singing is governed by precisely timed and sequenced firing in the premotor nucleus HVC (proper name) (Fig. 3.1a). HVC neurons project to RA (the robust nucleus of the arcopallium), and, in the adult, fire in a rapid burst exactly once during the song (7). Temporally precise, though less sparse, firing patterns in RA drive downstream motor neurons, which then drive the vocal muscles during song (8-11). The HVC projection neurons' burst-onset times collectively tile the duration of the song (7). In most mechanistic descriptions of the bird song system, the synaptic weights of the HVC projections onto RA are understood to encode the form of the song at each moment in time (12-14).

The long-term precision in song has been assumed to rely upon the scaffolding provided by long-term precision in HVC firing. Song maintenance has been thought to occur through the retuning of the motor output relative to this precise timing. However, HVC neurons undergo cell death and replacement by neurogenesis, both continuously and in a seasonal manner (15). In addition, recent experimental results have revealed some degree of longer-term changes in single, pre-motor neuron activity patterns during singing (16, 17). These variations in the sequence dynamics raise intriguing questions. How can a static, stereotyped behavior survive variable premotor firing patterns? What advantage might be gained by instabilities in the neural representation of the song?

We investigate these questions through a simple computational model of the bird song learning system. Previous theoretical work has used a reinforcement learning (RL) framework to model song learning in the projections structure from HVC to RA (12-14). In RL theory, an actor performs a task and varies its performances in a trial-and-error manner. A critic evaluates each performance and provides feedback. The actor adjusts its actions to improve performance. We adopt an RL model based on Fiete et al. (12) (Fig. 3.1b) and assume HVC's sparse firing pattern has emerged earlier in development. The learning process trains RA to drive motor outputs (here, m_1 and m_2) to reproduce the target song based on variable inputs from LMAN (Fig. 1b). Via gradient descent, reinforcement learning changes the connection strengths from HVC to RA depending on the levels of coincident activation of an HVC-to-RA synapse and a LMAN-to-RA synapse, followed by a global reinforcement signal.

We explore three different approaches to perturbing premotor firing patterns in a subset of HVC neurons: pausing activity while imposing synaptic decay, only pausing activity,

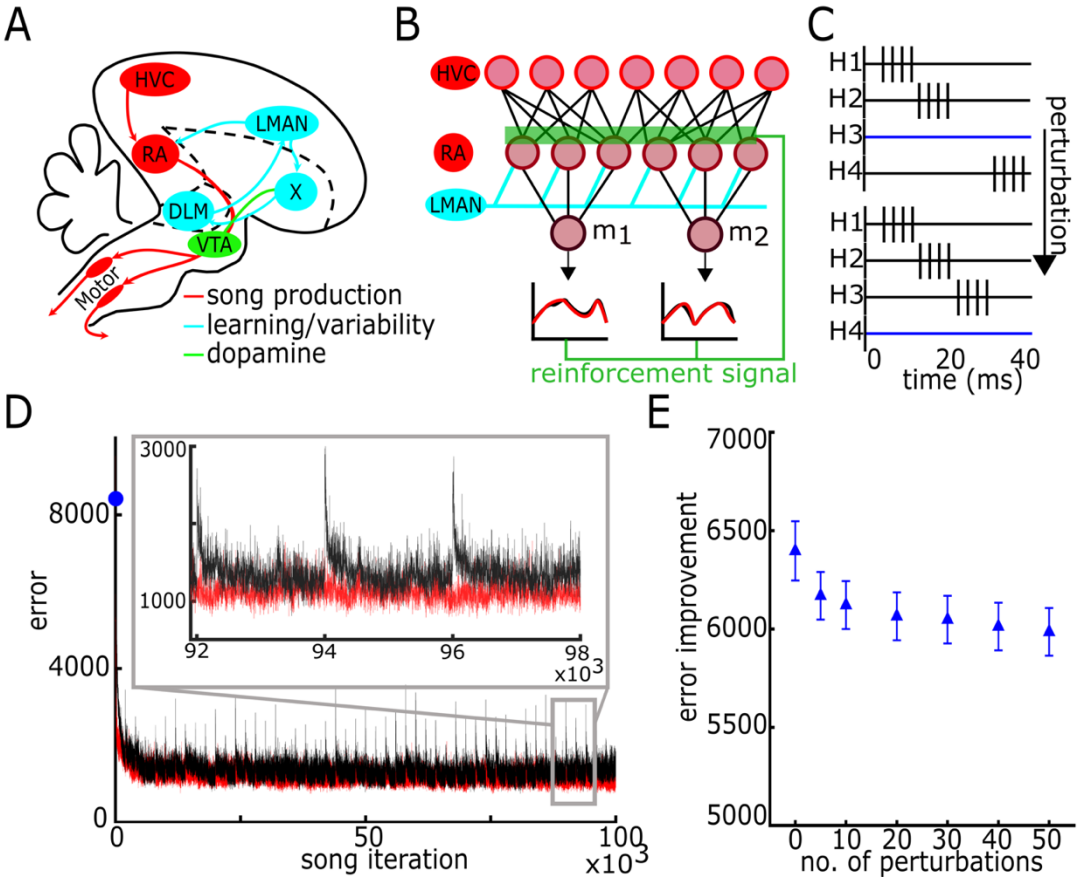

Figure 3.1. Model of bird song learning. **A**. Schematic of the avian song system. **B**. Schematic of learning model. Learning occurs on the synaptic weights from HVC to RA. **C**. Schematic of HVC firing patterns and perturbation events. **D**. Learning trajectory. Error is defined as the absolute difference between the target m_1 and m_2 and the model output m_1 and m_2. Red trace shows unperturbed trajectory. Black trace shows trajectory with 50 HVC perturbation events. Blue dot on y-axis indicates the error before learning. Inset shows three perturbation events. **E**. The difference between initial and final error in learning trajectory as a function the number of HVC perturbation events per 10^5 iterations averaged over 50 trials. Error improvement is defined as difference between the first iteration and average of the last 500 iterations for each trial. HVC perturbations slightly decrease error improvement.

and shifting firing timing. We then test the effects of these perturbations on network robustness in three different ways: performance error, robustness to RA cell loss, and speed of relearning an altered motor output. Next, we explore the underlying mechanisms for these effects. Finally, we make a quantitative comparison of the three different approaches. The three perturbation methods produce qualitatively similar results.

We find that varying HVC activity patterns balances two goals of the system: maintaining quality in song performance and adapting efficiently to environmental or physical changes that affect performance. Instability in HVC firing activity slightly degrades the performance quality. In exchange, however, the system is able to learn changes in muscle activity faster, and better distributes song encoding over the downstream RA network. Our results also suggest a possible mechanism underlying this effect. Variability in neural representation of stereotyped tasks may thus confer robustness and facilitate active maintenance of motor performance.

Results.

Basic learning framework

Figure 1d shows the learning trajectory, defined as the total error between the m_1 and m_2 templates and the produced versions. Although learning converges, error continues to fluctuate and does not go to zero, in part because gradient descent converges to a local, not a global, minimum. Continued error fluctuations are due to the ongoing variable inputs from

LMAN to RA, which drive both trial-to-trial variability in the RA firing patterns and changes in the HVC-to-RA connection strengths. After the initial convergence, the average error is stable over the approximately 100,000 subsequent iterations of the simulation (Fig. 3.1d, red trace) and is consistent with the stable form of adult birdsong (18).

Introducing changes in HVC activity

We next examine how random changes in HVC firing activity affect circuit activity and plasticity once the song has been learned. We assume the basic temporal structure of HVC inputs and the song dynamics have been learned previously during development, but synaptic plasticity and learning continue in adulthood (1-6). In our first perturbation scheme, "paused with synaptic weakening", we halt activity in 6% of active HVC projection cells while simultaneously activating the same number of previously paused cells in discrete offline episodes (see Methods) (Fig. 3.1d, black trace). The timing of the activity patterns of the paused and activated cells is independent and random. The synaptic projections of paused cells undergo synaptic weakening.

As expected, changing a subset of HVC firing activity increases song error (Fig. 3.1d,e). However, asymptotic error is only slightly increased by increasing the frequency of HVC disruption events (Fig. 3.1d,e). This daily error trajectory is consistent with the behavioral observation that song is more variable in the morning than in the evening (Fig. 3.1d) (19).

Impacts of HVC perturbations on network robustness

Although introducing pauses in a subset of HVC firing increases the overall error in song performance, advantages are gained. A likely change during aging is cell loss. We first ask how robust the song system is to partial loss of the RA network, a test which replicates experimental ablations or cell death. We next consider changes in the muscle transformation of the RA output to song or in the target song itself, both of which require a relearning of the upstream HVC-to-RA connection structure. Although adult zebra finch song is a highly stereotyped behavior, there are several reasons an adult bird might alter the network structure generating song. Injury or simply aging in the vocal muscles leads to the same neuronal drive generating a different song output: to keep the song stable, the bird must learn to control the altered muscles in a new manner.

Adaptations to partial loss of network

After 10^5 iterations of song we randomly remove a subset (8%) of RA neurons and then measure the resulting increase in error when the song is performed using the partial RA network (Fig. 3.2a). Removing a subset of the RA neurons increases error; however, the magnitude of the induced error decreases with the introduction of HVC perturbation episodes (Fig. 3.2b). Perturbing HVC firing activity distributes the song representation over more of the RA network and allows subsets of the RA network to continue to represent the song more accurately.

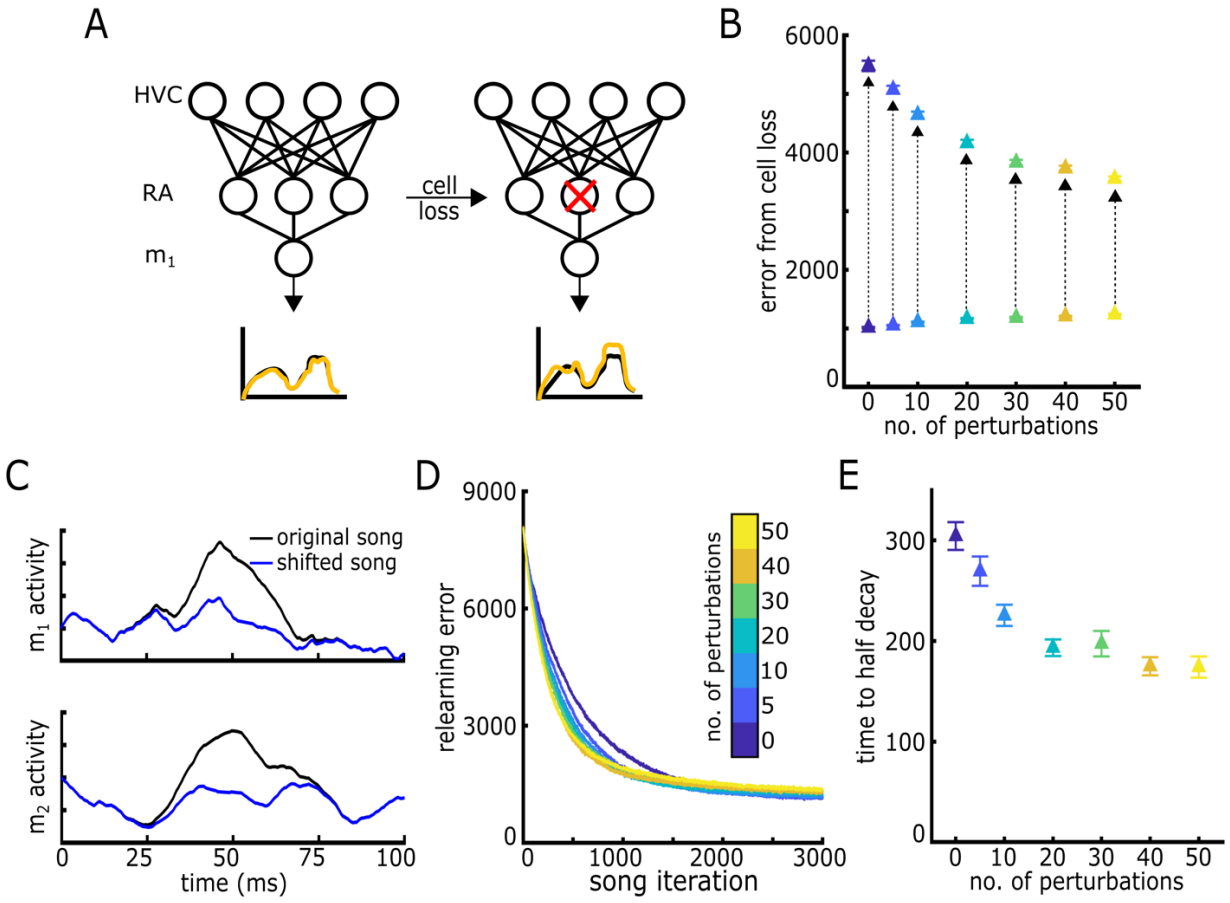

Figure 3.2. Tests of robustness. **A**. Schematic RA cell loss. After a subset of RA is removed, the song is performed, and the new error is computed. **B**. Error in song as a function of number of HVC perturbations per 10^5 iterations for the full RA network and for subpopulations of RA. Arrows indicate error from full RA network to error from RA network after cell loss. Error is averaged over randomly drawn subsets of RA and over trials. HVC perturbations increase the ability of a subpopulation of the RA network to represent song. **C**. Target song shift. After 10^5 iterations the target song is altered and relearning proceeds for 3,000 iterations. **D**. Re-learning trajectory after target song changes. **E**. Number of iterations to half decay of learning trajectory. Half-decay is defined relative to initial error at relearning onset and final error at 3,000 iterations after onset. HVC perturbations significantly speed up the adaptation process with minor penalty in final error.

Adaptation to environmental and physical changes

We model environmental and physical changes in song context as shifts in the motor target trajectory that the network is trying to produce: $m_1 \rightarrow m_1'$ and $m_2 \rightarrow m_2'$ (Fig. 3.2c). We then ask how quickly and successfully the shifted song target is learned by the network as a function of the number of HVC perturbation episodes occurring in the 10^5 iteration maintenance protocol. Periodic changes in HVC firing activity during initial learning increase the speed with which the network is able to adapt to alterations to the target template. Speed of re-learning increases with the frequency of HVC perturbation episodes (Fig. 3.2d,e). However, a penalty is paid for the increased adaptation speed with an increase in final error after 3,000 iterations of relearning (Fig 3.2d, Appendix 3.S1a,b).

Origins of increased robustness

What changes occur in the network structure due to perturbations in HVC activity? The learned components of our model are the synaptic strengths of projections, W, of which each entry, W_{ij}, represents the connection strength from HVC neuron 'j' to RA neuron 'i'. The distributions of synaptic weights after 10^5 song iterations change very little due to HVC perturbations (Appendix 3.S2a-c). However, when we order the HVC cells according to the time in the song when the neuron fires, a clear change in structure in W emerges due to the HVC perturbations (Fig. 3a). Without variations in HVC activity, projection patterns from HVC neurons that fire at similar times become highly correlated (Fig. 3b,c). The HVC perturbations

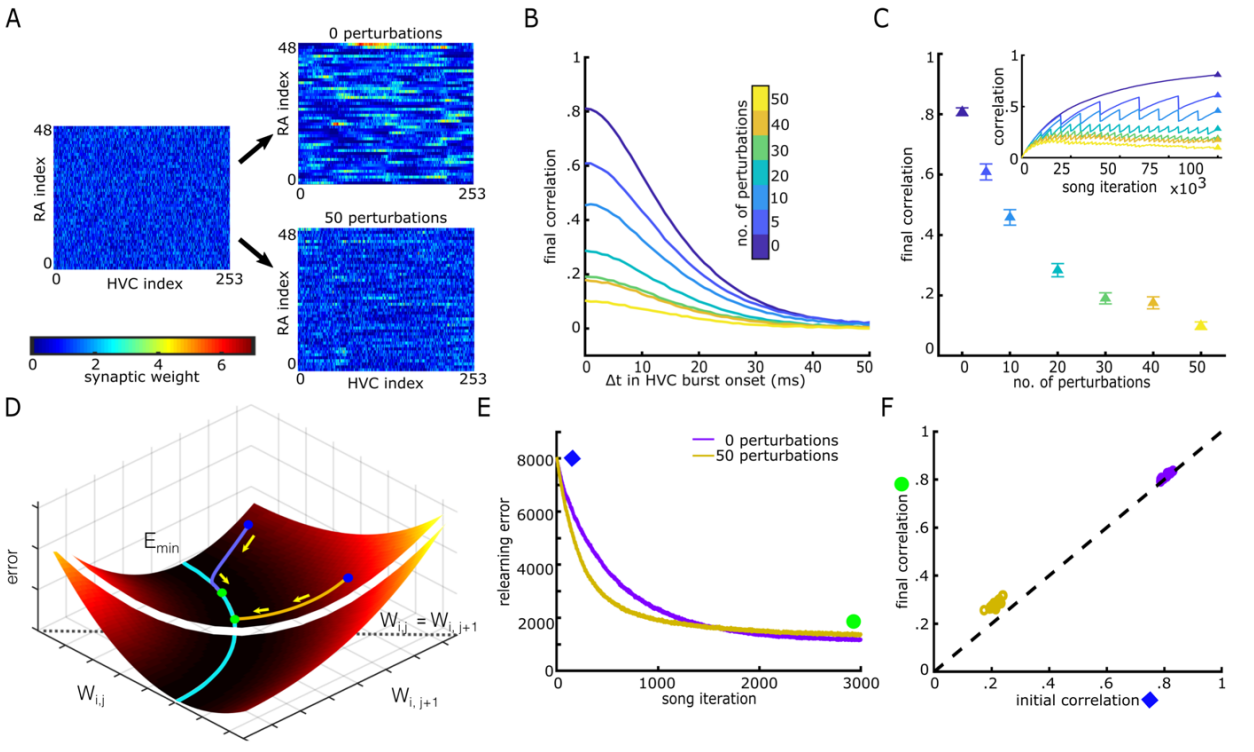

Figure 3.3. Mechanisms underlying robustness. **A**. Example initial and final W matrices after 10^5 iterations of learning with HVC index ordered by when the HVC neuron bursts. Left: initial W for 0 and 50 HVC perturbations; top and bottom right: final W for 0 and 50 HVC perturbation events. Perturbations prevent temporal correlations in HVC projection structure. **B**. Average pairwise correlations after 10^5 iterations of learning between HVC neurons' synaptic projections to RA (columns of W) as a function of the time difference in firing onset. Averages are taken over HVC neurons and learning trials. Correlations decrease with more HVC perturbation events. **C**. Maximum pairwise correlations between HVC neurons' synaptic projections to RA as a function of the number of HVC perturbation events. Inset: trajectory of maximum correlations over learning iterations. Drops in correlation occur at HVC perturbation events. **D**. Schematic of the progression of W over the course of learning. Error first proceeds quickly to a minimum. Correlations in the LMAN exploratory drive then slowly push the solution towards higher correlations if learning continues without perturbations. The dotted grey line shows W space where pairwise correlations = 1. Solid blue line shows region of approximately equal error. Purple trajectory shows W matrix undergoing learning from an initial position of low pairwise correlation. Gold trajectory shows W matrix undergoing learning from an initial position of high pairwise correlation. Blue and green dots show initial and final positions. The error landscape is not depicted near $W_{i,j} = W_{i,j+1}$ where error increases. **E**. Average re-learning trajectory (over 50 trials) for altered song target. **F**. Maximum HVC pairwise correlation in W at the beginning and end of 3,000 song iterations for representative trials from E. The correlation strength of the initial weight matrix before learning strongly influences the correlation strength of the final weight matrix after learning.

decorrelate nearby HVC projections as a monotonic function of the frequency of HVC perturbations. This is quantified by computing the average pairwise correlation values of individual HVC cells' projection structures to RA as a function of the time difference of the HVC neurons' burst onset (Fig 3.3b,c).

The timescale on which these correlations grow is much longer than the learning trajectory (Fig. 3.3c, inset). In the absence of any HVC activity changes, error decreases to within 5% of its total descent within 2000 iterations whereas correlations in the weight matrix reach 50% of their final values at 20,000 iterations and reach 95% of their final values only at 75,000 iterations (Fig. 3.3c, inset). This suggests that some aspect of the continued plasticity causes drift in the weight matrix structure along a valley of solutions with approximately equal error but increasing correlation strength (Fig. 3.3d).

What causes this drift? HVC synapses made on to a single RA neuron share variability from LMAN. Because of the extended synaptic time course of inputs from LMAN, a single LMAN firing event affects plasticity in synapses from multiple HVC cells that fire at similar times. One would expect that correlated variability from LMAN biases the reachable search space towards solutions which themselves are correlated. We confirm that LMAN's synaptic timescales drive the buildup of correlations in additional simulations where we vary the synaptic time course of the inputs from LMAN. We observe a monotonic dependency of the correlations that emerged in W on the time course of individual LMAN inputs (Fig. Appendix 3.S3).

Disrupting the activity of HVC cells slows this growth in correlation. Temporarily silencing HVC cells while weakening synapses introduces a random change into the local projection structure from HVC to RA at that moment in song. This lowers the total exposure

sequentially firing HVC cells have to correlated noise inputs and therefore allows the cells' synaptic strengths to remain more independent.

Because HVC and RA have many more degrees of freedom than the downstream motor pools, there are redundancies in the possible firing patterns in RA: multiple firing patterns will produce the same m_1 and m_2 output. Within these firing patterns with identical motor pool output, the correlation structure of W can vary substantially. These redundancies give rise to a manifold of equivalent local minima on the error function.

Randomly varying HVC activity pushes the synaptic weight structure into a region in W space with higher error but lower pairwise correlation. Learning that begins in a less correlated state first approaches the asymptotic error value in a less correlated solution state, even though the value of the local error minimum is the essentially the same (Fig. 3.3e,f) (20). This daily repositioning of the projection structure on the error landscape leads to a slower accumulation of passive correlations due to the correlated LMAN variability. The flexibility and generality of this repositioning is possible because of the multiplicity of solutions with comparable error.

In additional simulations we tested the hypothesis that correlations across cells' synapses slow learning by comparing the learning trajectories from sets of initial random weight matrices wherein each HVC cell's initial projection weights are drawn from identical Gaussian distributions, but the correlations between HVC cells' synapses vary (See Extended Methods and Fig. S4 a,b, c). In this way we isolate correlation level in W while holding all other statistics of W constant. We find that increasing correlations in the initial W structure dramatically slows learning here as well (Fig. S4 d,e). This result further suggests that

correlations in synaptic strengths is the network feature that leads to variable re-learning speeds.

Other forms of HVC plasticity

Lastly, we compare our initial form of HVC plasticity with two other perturbation schemes (Fig. 3.4 a-c). The second scheme, ("perturbation: paused"), is the same as the original scheme but with no synaptic decay in paused cells (Fig. 3.4b). In our third scheme ("time shift"), instead of temporarily silencing subsets of HVC cells, we shift the timing of 6% of HVC projection cells such that they fire at new, randomly-selected times in the song (Fig. 3.4c). This third scheme represents an extreme version of HVC plasticity in our hypothesized role decorrelating synaptic structure.

All of the tested perturbations lead to qualitatively similar increases in network robustness (Fig. 3.4 d-f). The perturbation scheme we show in detail (paused with synaptic weakening) performs best under cell loss and motor re-learning. However, this form of perturbation is the only scheme that leads to slightly higher error in motor relearning (Fig. 3.2 d, Appendix 3.S1a). From these comparisons, we predict this robustness advantage is a general property of varying upstream HVC firing activity and does not rely on the particular form of variations we chose.

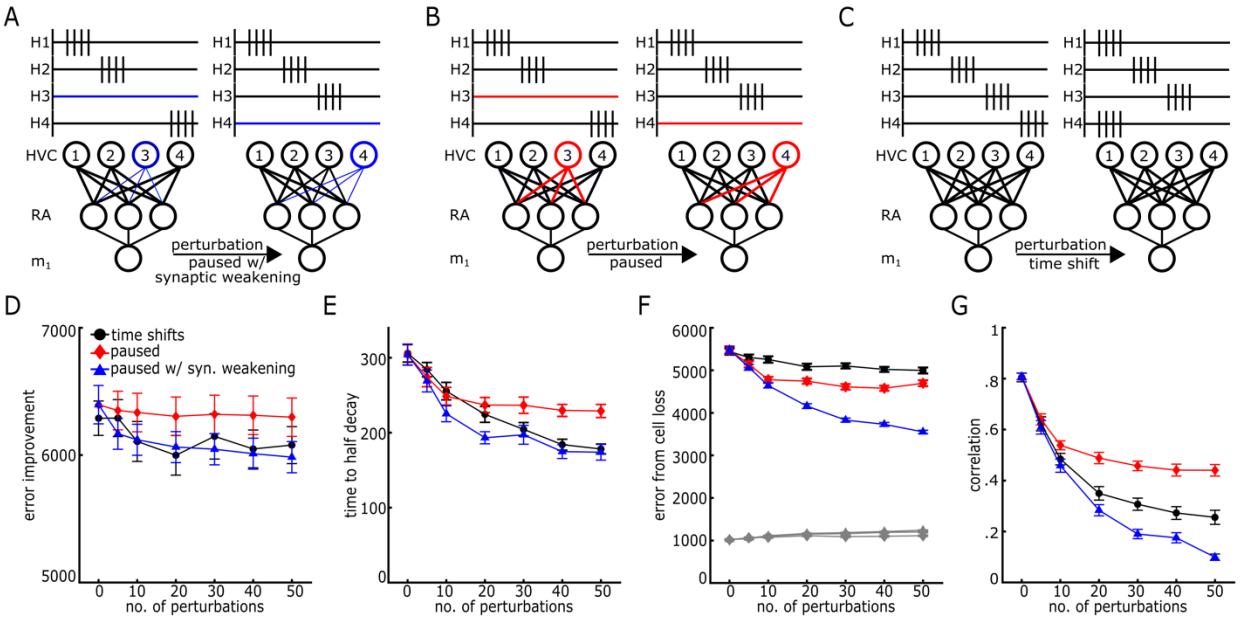

Figure 3.4. Model comparisons of perturbing HVC activity. **A**. HVC perturbation scheme wherein a subset of HVC cells are silenced and re-activated at each perturbation event. Synapses weaken in silent cells. **B**. The same scheme as in 'A', but without synaptic weakening. **C**. HVC perturbation scheme wherein a subset of HVC cells' firing times randomly shift at each perturbation event. **D**. (D, E, F and G compare performance and robustness metrics across perturbation schemes and frequency.) Error improvement over 10^5 iterations. Minor cost is incurred for increased perturbation frequency. **E**. Number of iterations to half decay of re-learning trajectory. **F**. Increase in error due to RA cell loss. Grey indicates performance error with full RA network. Other colors indicate performance error with partial RA network. **G**. Maximum pairwise correlations between HVC neurons' synaptic projections to RA.

Discussion:

The main finding of this work is that repeated variation in the activity of HVC neurons, while slightly elevating overall error in song, increases robustness to physical and environmental changes. Specifically, such perturbations result in faster adaptation to changes in muscles or target song, and decreased error caused by cell loss. When HVC activity patterns remain constant, the strength of synapses from HVC neurons that fire at similar times to target RA neurons become highly correlated due to an overlapping exploratory signal from LMAN. In contrast, when HVC activity patterns vary, synaptic weights remain decorrelated, allowing more rapid adaptation to a changed song template and greater robustness to the loss of RA neurons. These benefits could result in an increased ability to maintain high quality song performance over the lifetime of the bird. This work explains a recent unexpected observation and provides a method for potentially increasing robustness in neural networks more generally.

Broadly, this work identifies a way in which ongoing plasticity can lead to deleterious correlations when a portion of the system is static. Here, keeping the firing activity of HVC neurons fixed while allowing learning on the HVC to RA synapses leads to a buildup of correlations and weakens the system's response to stressors that the continued plasticity presumably exists to address. This work suggests that repetitive behaviors should be performed in a dynamic manner that pushes the system to retain access to many degrees of freedom. Repetitive behaviors are made more robust to environmental change by continually seeking new ways of performing the same behavior.

The extent to which stable behavior is underpinned by stable representation at the population and single neuron level varies across systems and contexts and is a question of active debate. A common assumption is that, once mastered, stereotyped tasks are represented by stable neural activity and that plasticity occurs due to learning. However, Rokni et al. found that when monkeys perform a familiar reaching task, the tuning curves of neurons in the supplementary motor area undergo slow, random drift (21). They attributed this drift to noisy synaptic plasticity and found that behavior remains stable despite this noise, presumably due to the high redundancy of the active neural networks. This type of random drift is exactly what one would expect from the ongoing plasticity we consider in this work and we predict that similar shifts in RA firing properties should be observable over long timescales. We theorize that this random drift, though still allowing stable representation of an otherwise static system, will pose a challenge to the long-term maintenance of the task unless accompanied by other types of variation in the network dynamics.

The presence of correlations in exploratory inputs across plastic synapses is critical to our findings and based on neuroanatomical observations in the song system. Specifically, there is high divergence and convergence in the projections from HVC to RA, but low convergence and divergence in the projections from LMAN to RA (22). While using a single Poisson varying LMAN input for each RA neuron is a modelling simplification, we expect that the true connectivity structure still gives rise to high correlations. There are approximately 50 synaptic connections from LMAN onto a single RA neuron, possibly originating from a small number of co-localized LMAN neurons (23-25). There are approximately 1000 HVC synaptic projections onto one RA neuron originating from approximately 200 HVC neurons (26). Therefore, if all HVC

synapses are affected by some portion of LMAN activity, correlations in LMAN activity across HVC inputs must exist.

Furthermore, correlations in an exploratory drive are likely for any system where synaptic strengths are learned individually: if the smallest unit of independent variation in a network is a single cell, then the number of independent exploratory processes will be approximately the number of independent cells whereas the number of synaptic connections to be learned will be the number of synapses per cell times the total number of cells. The finite dimensional space of the exploratory inputs will limit the amount of independent variation the learning synapses can undergo.

How much variation is enough for adaptation advantages to be relevant? Fully answering this question depends on the rate of ongoing plasticity as well as the degree to which adaptation is needed and is likely to depend critically on the specific system. However, we saw a significant robustness advantage even at our lowest frequency of perturbation (once per 20,000 song iterations). Note that the two experimental results from zebra finch differ substantially in the amount of variability reported in the HVC nucleus: Liberti et al. report that approximately 40% of HVC projection cells change activity patterns over a period of 5 days, whereas Katlowitz et al. report 3.6% of projection cells change activity patterns over timescales ranging from 3-56 days, as well as ongoing, random jitter in the burst-onset timing (16, 17). In other songbird species, HVC cell death and neurogenesis are seasonally regulated, with an almost doubling of HVC projection cells to RA during breeding season (27). Our model considers a range of variation frequencies and approximately spans the frequency regimes of these experimental results. We predict that our results would be qualitatively the same were we to

allow changes to happen continuously and perhaps result in even better performance, since punctate changes to network structure would not exists.

In addition to HVC perturbations, other methods could potentially accomplish the same decorrelation of HVC projections. Plasticity in the synapses from LMAN to RA in such a way that synapses from different HVC neurons were exposed to a changing set of shared and unshared LMAN inputs would present another method to reduce the buildup of correlations in the HVC to RA synapses. Recent experimental work has identified synaptic plasticity in the LMAN to RA synapses, but it is not known whether this form of plasticity redistributes LMAN-HVC coincident activity onto individual RA neurons (28). In addition, recurrent connections among RA neurons show synaptic plasticity (29); their functional role in a learning model has not yet been explored, but they could contribute to the type of decorrelation needed for robust learning. Also, note that the temporally sparse firing pattern in HVC already contributes to reducing the amount of harmful correlation in LMAN inputs across HVC synapses by limiting the active learning time window for each synapse to a region around the single HVC burst event for each neuron (assuming some local form of HVC spike timing dependent plasticity as we do for this work). Earlier theoretical work has shown how temporally sparse HVC firing patterns could increase initial song learning speeds by reducing the amount of interference in weight updates (30) (I.R. Fiete, R.H. Hahnloser, M.S. Fee, H.S. Seung, Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong, Journal of neurophysiology, 92 (2004) 2274-2282). Our work identifies another way in which the sparse firing patterns from HVC could be well suited to rapid learning by decreasing correlations

amongst HVC projections to RA. In other systems, a variety of decorrelation mechanisms could achieve the same effect that we observed with HVC perturbations.

This biological learning strategy can be seen in the context of machine-learning techniques that introduce a stochastic element in the forward pass of the network, such as "drop-out". Methods using drop-out are similar to the bird song strategy presented here, wherein a subset of neurons and the connections to and from the subset are probabilistically removed during portions of training to avoid over-fitting (31). In the context of artificial neural networks, overfitting is generally taken to mean that a network has been too closely tuned to the training set of inputs and outputs. The result is that when a new input is introduced from the same statistical class, the network has learned the vagaries of the training set rather than the statistical properties of the full set of possible inputs. Again, there is an interesting parallel to the bird song strategy: maintenance of song through the trials of normal life may require adapting to an altered version of the target song or muscle program. Varying HVC inputs prevents the system from over-fitting to a single target. The addition of synaptic weakening during the equivalent 'drop-out' periods in our model presents a potentially beneficial feature in artificial learning systems as well.

However, there are important ways in which this finding is uniquely biological. The issue of maintenance plasticity is currently specific to biological systems; it is not a feature of most machine learning algorithms because it is not needed. Once an artificial RL system reaches an asymptotic final error, changes to the weight structure of the system are halted. However, ongoing plasticity is an inherent feature of biological systems. It is in the context of this ongoing plasticity that correlations grow around static components of the system. While currently not

within the scope of most machine learning tasks, it is possible that machine-learning algorithms could benefit from ongoing plasticity mechanisms to provide robustness to changing constraints.

In this work we find that perturbing HVC firing activity balances two goals of the system: maintaining quality in song performance and adapting efficiently to environmental or physical changes that affect performance. This result is to an extent anti-optimal: in our model, the bird is not learning a single stereotyped behavior to the greatest possible precision but retains the ability to adapt to environmental and physical changes. In this contingent picture, optimality must be broadly interpreted as maximizing performance quality across competing goals. It is notable that this contingent strategy is possibly present even in the zebra finch song system, which has traditionally been thought of as an extreme example of learning a single, highly stereotyped behavior. For biological systems, learning always takes place in a fluctuating environment that may at any moment change the conditions of performance. This requires a contingent relationship to optimal behaviors and to the notion of optimality itself.

Methods:

Base model. The base model comprises three feed-forward layers, HVC (N=500), RA (N=48) and Motor Pools (N=2), with an independent, Poisson firing process synapsing onto each RA cell from LMAN and was modified from Fiete et al. (12). HVC and RA layers are modeled as a conductance-based leaky integrate-and-fire neurons. Song production is modeled by non-

spiking motor pool output units that receive input from RA. The goal of learning in the model is for the two motor pools to reproduce a target motor trajectory. ***Learning Parameters.*** Only HVC projections to RA are plastic and are determined by $dW_{ij}/dt = \eta\, R(t)\, e_{ij}(t)$,where $W_{ij}$is the synaptic strength from the jth HVC cell to the ith RA cell, R(t) is the reinforcement signal, $\eta$is the learning rate and $e_{ij}(t)$ is the eligibility trace over which weight changes occur, defined as $e_{ij}(t) = \int_0^t dt'\, \mathrm{G}(t-t')\,(s_i{}^{LMAN}(t') - < s_i{}^{LMAN} >)\, s_{ij}{}^{HVC}(t')$, where $G(t) = t^n e^{t/\tau_e}$, $s_i{}^{LMAN}(t)$ is the LMAN input to the ith RA cell, $s_{ij}{}^{HVC}(t)$ is the synaptic input from the jth HVC cell to the ith RA cell. The reinforcement signal, R(t), is defined as, $R(t) = 2 * \Theta[D(t) - \bar{D}(t)] - 1$, where D(t) is the mean squared error of the current motor output and $\bar{D}(t)$is the average mean squared error of the previous 5 trials. ***Perturbations to HVC.*** HVC perturbations events occur offline at regular intervals. We vary the number of perturbation events from 0 to 50 over 10^5 song iterations. We model three perturbation schemes. (1) 'Paused with synaptic weakening: 500 cells are active and 200 cells are silent at any one time. At each perturbation event 30 active cells go silent and 30 silent cells become active. Synaptic decay occurs while cells are silent. (2) 'Paused': this scheme is the same as (1) except synapses of silent cells are frozen. (3) 'Time-shift': 500 cells are utilized in the HVC layer. At each perturbation event 5% of cells randomly shift burst-onset times within the song. We simulate each perturbation frequency and scheme 50 times. ***Tests of Network Robustness.*** *Shift in motor targets*. We add a gaussian form centered at the midpoint of the target motor output and measure the speed (time to half decay of error) and the final error after 3000 iterations of re-learning. *Cell Loss in RA.* We silence activity in 1/12 of the RA network, and measure performance. We average the resulting error over 500 randomly drawn subpopulations. ***Pairwise Correlation of HVC Synaptic***

Structure. The pairwise correlation between HVC cells' synaptic projections is calculated as a function of the difference in timing between the HVC cells' burst onsets. We denote the outgoing weights for HVC neuron *p* at time *t* as ${W^t}_p \equiv {W^t}_{(:,p)}$. For all pairs of HVC projection vectors, ${W^t}_p$ and ${W^{t\prime}}_q$ for which the HVC neurons' burst onset times *t* and *t'* are within $\tau \pm \Delta\tau$, ($\Delta\tau$ = 0.5 ms) we compute the average pairwise correlation at time separation, τ, as:

$$C_\tau({W^t}_p, {W^{t\prime}}_q) = \langle \frac{({W^t}_p - \overline{{W^t}_p})({W^{t\prime}}_q - \overline{{W^{t\prime}}_q})}{\sqrt{({W^t}_p - \overline{{W^t}_p})^2({W^{t\prime}}_q - \overline{{W^{t\prime}}_q})^2}} \rangle_{all\ p,q:\ (t-t\prime)\subseteq\tau\pm\Delta\tau}.$$

Acknowledgements: This work was funded by the National Science Foundation, the Washington Research Foundation and the National Institute for Health.

Bibliography

1. Andalman AS & Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proceedings of the National Academy of Sciences of the United States of America* 106(30):12518-12523.
2. Sober SJ & Brainard MS (2009) Adult birdsong is actively maintained by error correction. *Nature Neuroscience* 12(7):927-931.
3. Nordeen KW & Nordeen EJ (1992) Auditory feedback is necessary for the maintenance of stereotyped song in adult zebra finches. *Behavioral and Neural Biology* 57(1):58-66.
4. Nordeen KW & Nordeen EJ (2010) Deafening-induced vocal deterioration in adult songbirds is reversed by disrupting a basal ganglia-forebrain circuit. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience* 30(21):7392-7400.
5. Sossinka R, and Böhner, J. (1980) Song Types in the Zebra Finch *Poephila guttata castanotis1. Zeitschrift für Tierpsychologie* (53):123-132.
6. Okanoya K & Yamaguchi A (1997) Adult Bengalese finches (Lonchura striata var. domestica) require real-time auditory feedback to produce normal song syntax. *Journal of Neurobiology* 33(4):343-356.
7. Hahnloser RH, Kozhevnikov AA, & Fee MS (2002) An ultra-sparse code underlies the generation of neural sequences in a songbird. *Nature* 419(6902):65-70.
8. Suthers RA & Margoliash D (2002) Motor control of birdsong. *Current Opinion in Neurobiology* 12(6):684-690.
9. Wild JM (1993) Descending projections of the songbird nucleus robustus archistriatalis. *The Journal of Comparative Neurology* 338(2):225-241.
10. Yu AC & Margoliash D (1996) Temporal hierarchical control of singing in birds. *Science* 273(5283):1871-1875.
11. Wild JM (2004) Functional neuroanatomy of the sensorimotor control of singing. *Annals of the New York Academy of Sciences* 1016:438-462.

12. Fiete IR, Fee MS, & Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *Journal of Neurophysiology* 98(4):2038-2057.
13. Farries MA & Fairhall AL (2007) Reinforcement learning with modulated spike timing dependent synaptic plasticity. *Journal of Neurophysiology* 98(6):3648-3665.
14. Doya K & Sejnowski TJ (1998) A computation model of birdsong learning by auditory experience and auditory feedback. *Central Auditory Processing and Neural Modeling*:77-88.
15. Barnea A & Pravosudov V (2011) Birds as a model to study adult neurogenesis: bridging evolutionary, comparative and neuroethological approaches. *The European Journal of Neuroscience* 34(6):884-907.
16. Liberti WA, 3rd*, et al.* (2016) Unstable Neurons Underlie a Stable Learned Behavior. *Nature Neuroscience* 19(12):1665-1671.
17. Katlowitz KA, Picardo MA, & Long MA (2018) Stable Sequential Activity Underlying the Maintenance of a Precisely Executed Skilled Behavior. *Neuron* 98(6):1133-1140 e1133.
18. Nordby JCC, S. Elizabeth; Beecher, Michale D. (2002) Adult Song Sparrows Do Not Alter Their Song Repertoires *Ethnology* 108:39-50.
19. Deregnaucourt S, Mitra PP, Feher O, Pytte C, & Tchernichovski O (2005) How sleep affects the developmental learning of bird song. *Nature* 433(7027):710-716.
20. Polycarpou MM & Ioannou PA (1992) Learning and convergence analysis of neural-type structured networks. *IEEE transactions on neural networks* 3(1):39-50.
21. Rokni U, Richardson AG, Bizzi E, & Seung HS (2007) Motor learning with unstable neural representations. *Neuron* 54(4):653-666.
22. Luo M, Ding L, & Perkel DJ (2001) An avian basal ganglia pathway essential for vocal learning forms a closed topographic loop. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 21(17):6836-6845.
23. Canady RA, Burd GD, DeVoogd TJ, & Nottebohm F (1988) Effect of testosterone on input received by an identified neuron type of the canary song system: a Golgi/electron

microscopy/degeneration study. *The Journal of Neuroscience : the official journal of the Society for Neuroscience* 8(10):3770-3784.

24. Herrmann K & Arnold AP (1991) The development of afferent projections to the robust archistriatal nucleus in male zebra finches: a quantitative electron microscopic study. *The Journal of Neuroscience : the official journal of the Society for Neuroscience* 11(7):2063-2074.
25. Gurney ME (1981) Hormonal control of cell form and number in the zebra finch song system. *The Journal of Neuroscience : the official journal of the Society for Neuroscience* 1(6):658-673.
26. Kittelberger JM & Mooney R (1999) Lesions of an avian forebrain nucleus that disrupt song development alter synaptic connectivity and transmission in the vocal premotor pathway. *The Journal of Neuroscience : the official journal of the Society for Neuroscience* 19(21):9385-9398.
27. Brenowitz EA (2004) Plasticity of the adult avian song control system. *Annals of the New York Academy of Sciences* 1016:560-585.
28. Mehaffey WH & Doupe AJ (2015) Naturalistic stimulation drives opposing heterosynaptic plasticity at two inputs to songbird cortex. *Nature Neuroscience* 18(9):1272-1280.
29. Sizemore M & Perkel DJ (2011) Premotor synaptic plasticity limited to the critical period for song learning. *Proceedings of the National Academy of Sciences of the United States of America* 108(42):17492-17497.
30. I.R. Fiete, R.H. Hahnloser, M.S. Fee, H.S. Seung (2004) Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong, Journal of neurophysiology, 922274-2282
31. Nitish Srivastava GEH, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov (2014) Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research* 15(1):1929-1958.

Appendix:

Figures:

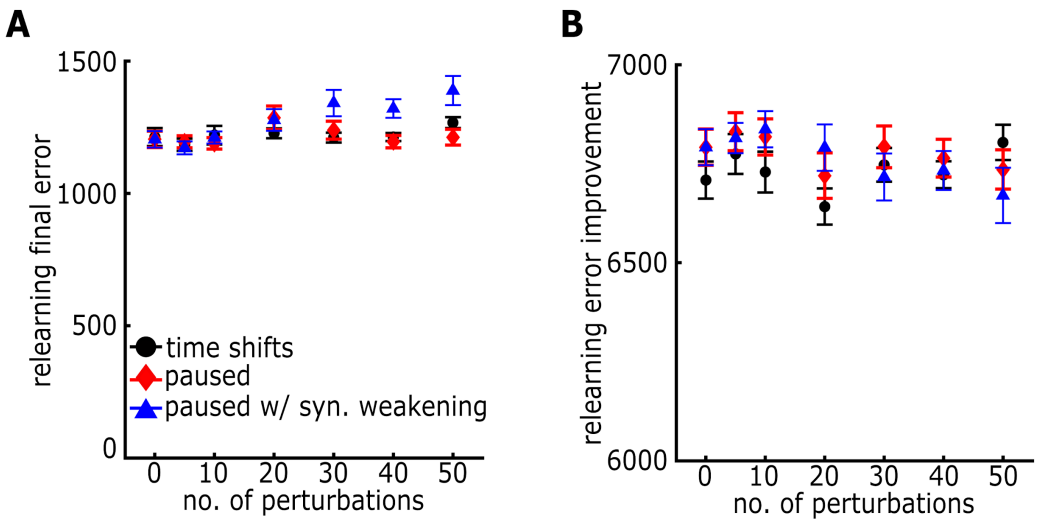

Figure S1. Comparisons of final error. **A.** Comparison of final error after song re-learning. All perturbation schemes re-learn the shifted target song with equal accuracy across frequencies of perturbation except for the perturbation: 'paused with synaptic weakening,' which resulted in slightly less improvement in song at higher frequencies of HVC perturbations. Color schemes same as in Fig. 4. **B.** Comparison of error improvement after song re-learning. Error improvement is not significantly different across all perturbation schemes and frequency.

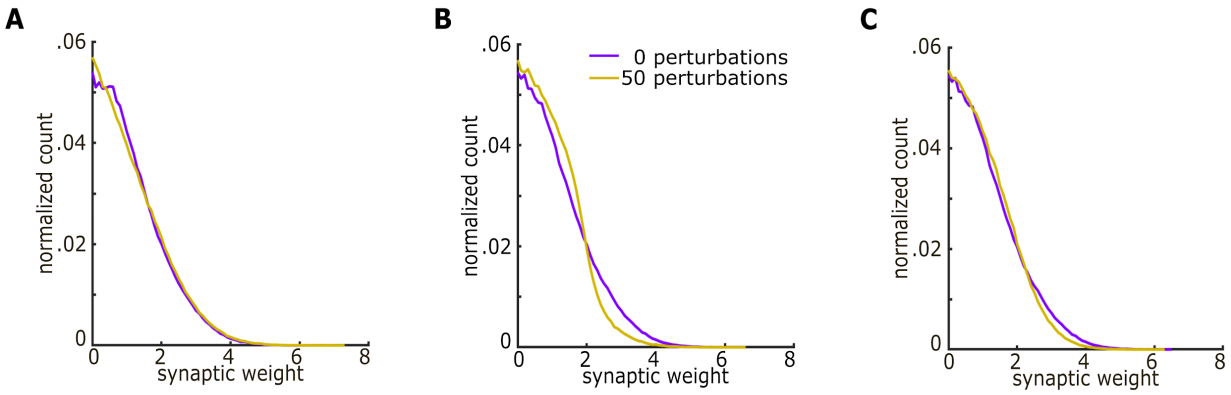

Figure S2. Synaptic weight distributions. **A**. Distributions of synaptic weights after 10^5 maintenance trials for the paused perturbation scheme compared to the no perturbation control distribution. **B**. Distributions of synaptic weights after 10^5 maintenance trials for the paused with synaptic weakening perturbation scheme compared to the no perturbation control distribution. **C**. Distributions of synaptic weights after 10^5 maintenance trials for the time shifted perturbation scheme compared to the no perturbation control distribution.

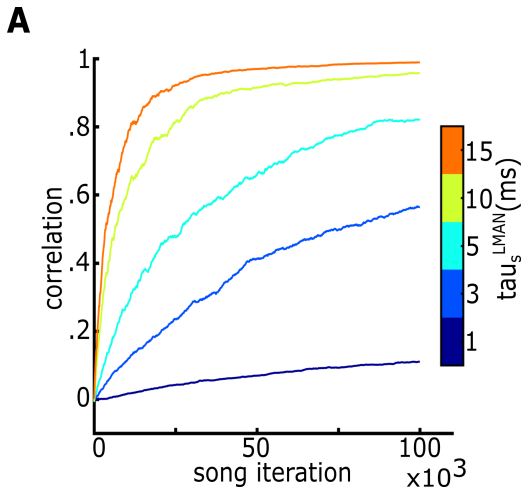

Figure S3. Synaptic weight correlation dependence on LMAN time constants. **A**. Example of the accumulation of average pairwise correlations over the course of 10^5 maintenance trials for varying time courses of LMAN inputs. As inputs from LMAN become more punctate in time, correlations across HVC projection strengths decrease. This shows that the build-up of correlations in the weight matrix is due to shared LMAN inputs. In the actual bird song system, LMAN synaptic inputs to RA are NMDA receptor mediated with long time courses (approximately 70-75 milliseconds in adults) (1)

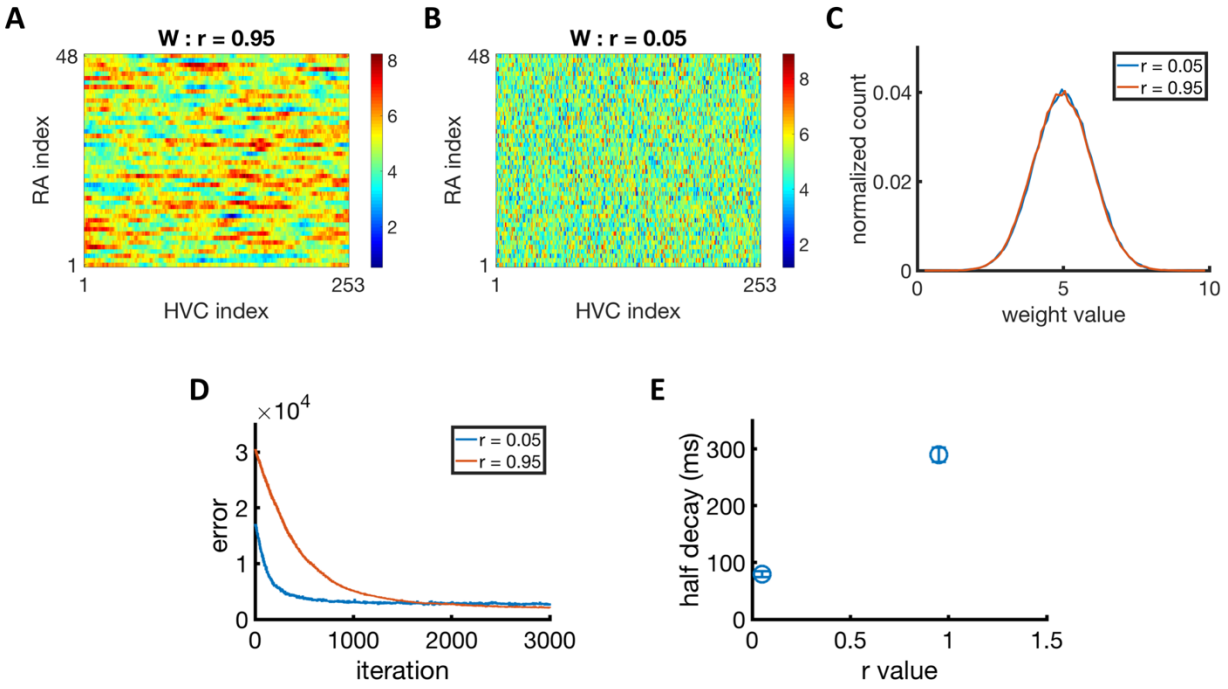

Figure S4. Learning speed depends on pairwise correlations of HVC synaptic weights in random Gaussian weight matrices. **A.** Example weight matrix with correlation between nearest neighbor HVC synaptic projection strengths, r = 0.95. **B.** Example weight matrix with correlation between nearest neighbor HVC synaptic projection strengths, r = 0.05. **C.** Comparison of synaptic weight distributions in weight matrices with r = 0.05 and r = 0.95. The overall distributions are identical and Gaussian distributed. It is only the correlations between columns that differ. **D.** Average learning trajectory for each correlation level in the initial, random weight matrices before learning begins (over 25 trials per correlation level). **E.** Time to half decay of the traces in panel D. If the initial weight matrix has highly correlated columns, learning proceeds much more slowly.

Statistical Significance:

To test whether the number of perturbations of each type affected the final error, the half time for relearning, the error from RA cell loss and the final weight correlations, we carried out one-way analysis of variance (ANOVA) using Prism (GraphPad Software). In all cases, there was a significant main effect of the number of perturbations ($p < 0.0001$ in each case).

Extended Methods*:

*The text for the following extended methods section was written primarily by Elliott Abe with editing and additional writing by myself, and editing by David Perkel and Adrienne Fairhall.
In this section, we lay out the basic framework for the network and then describe the three different approaches utilized in the perturbations to HVC activity patterns. Following this, we detail how we measured network robustness. Lastly, we quantify changes to the network structure.

A. Base Model and Network Architecture

In this initial approach, the base model comprised three layers with feed-forward connectivity, representing the premotor network of HVC, RA and motor pools, with empiric synapses from LMAN driving variability (Fig. 1a,b).

1. Neuronal Parameters

Our base model is derived from that of Fiete et al., (2007) (2). We assume a model structure shown in Fig. 1b, with 500 active HVC neurons projecting to 48 RA neurons. Each RA neuron is represented as a single-compartment conductance-based model. Each RA neuron projects to one of two motor pools, which are low-dimensional representations of song features, such as fundamental frequency or amplitude. The LMAN input to each RA neuron is taken to be an independent Poisson process.

Each neuron in HVC and RA layers is modeled as a conductance-based leaky integrate-and-fire neuron:

$$C_m\,(dV_i/dt) \;=\; -g_L(V_i - V_L) \;- g_{E,i}(V_i - V_E) \;- g_{I,i}(V_i - V_I) \qquad (1)$$

where $g_{L,i}, g_{E,i}, g_{I,i}$ represent the leak, excitatory, and inhibitory conductance, respectively for the i[th] neuron. Within this model, a spike is generated when V_i crosses the threshold voltage V_θ, and is reset to V_{reset}.

a. HVC Layer

The onset times for HVC bursts were drawn from a uniform random distribution over the time-course of the motif. $g_{I,j}(t) = 0$ for all neurons. $g_{E,j}(t) = 0$ for all neurons at all times in the song motif, except for one 6-ms excitatory pulse with magnitude 0.13 mS/cm^2 to drive a single burst.

b. RA Layer with LMAN Input

RA neurons receive excitatory input from HVC, and both excitatory and inhibitory input from LMAN. Biological LMAN neurons are glutamatergic and excitatory; we modeled this excitatory connection and also introduced an inhibitory component representing di-synaptic inhibition from LMAN to RA via RA interneurons (3). This balanced synaptic input allows stable song over long time scales. In addition, there is weak recurrent inhibitory activity in RA.

In RA, the excitatory synaptic conductances are described by:

$$g_{E,i}{}^{RA}(t) = 0.0024 \sum_j [W_{ij}\, s_j{}^{HVC}(t) + (s_i{}^{LMAN+}(t) + s_i{}^{LMAN-}(t))], \qquad (2)$$

where W_{ij} represents the synaptic weight from the j[th] HVC neuron to the i[th] RA neuron, $s_j{}^{HVC}$ represents the synaptic activation level from the jth HVC neuron, and $s_i{}^{LMAN+}$ represents the excitatory component of the synaptic activation from the ith LMAN input, the one projecting to the ith RA neuron, and $s_i{}^{LMAN-}$ represents the inhibitory input. The recurrent inhibitory synaptic conductances in RA are described by:

$$g_{I,i}{}^{RA}(t) = (0.2/N_{RA}) \sum_i s_i{}^{RA}(t), \qquad (3)$$

where N_{RA} is the number of RA neurons. Following an action potential in neuron i, the synaptic activation $s_i(t)$ is increased by one, and decays exponentially with time constant τ_s. The synaptic time course is described by:

$$ds_i(t)/dt = -s_i(t)/\tau_s. \qquad (4)$$

Throughout the song, each LMAN input is modeled as an independent Poisson process, with a constant mean firing rate $\lambda = 80Hz$.

c. The Motor Pools

Song production is modeled by two non-spiking motor-pool output units that receive input from RA. These motor pools represent premotor nuclei controlling song features such as fundamental frequency and amplitude, and are defined by:

$$\tau_m \, dm_k(t)/dt + m_k(t) = \sum_i A_{ki} s_i^{RA}(t) + b_k, \qquad (5)$$

where each motor pool has time constant τ_m and tonic activation b_k. Each motor pool sums activity from RA weighted by a fixed set of output weights A. Half of the RA neurons project to motor pool 1 (m_1) and half to motor pool 2 (m_2). Half of m_1 RA neurons are excitatory ($A_{ki} > 0$) and half are inhibitory ($A_{ki} < 0$).

The goal of learning in the model is for the two motor pools to reproduce a target motor trajectory. To generate this trajectory, the weight matrix was assigned random values on the interval [0, 4]; this led to a particular trajectory of motor-pool activity and was used as the target motor trajectory for subsequent learning throughout our study.

2. Learning Parameters

In our model, only HVC projections to RA are plastic. These changes are determined by

$$dW_{ij}/dt \;=\; \eta\, R(t)\, e_{ij}(t), \quad (6)$$

where R(t) is the reinforcement signal at every time point t in the song motif, and $\eta$is the learning rate. ηdetermines the size of the synaptic changes after each trial, and was empirically determined for the longer simulations and optimized to have a stable, decreasing error. The eligibility for plasticity at each synapse, at each time point, $e_{ij}\,(t)$, is defined by:

$$e_{ij}(t) \;=\; \int_0^t dt'\, \mathrm{G}(t-t')\, (s_i{}^{LMAN}(t') - < s_i{}^{LMAN} >)\, s_{ij}{}^{HVC}(t'), \qquad (7)$$

where $G(t) \;=\; t^n e^{t/\tau_e}$, n=5, and time constant $\tau_e = 5ms$. For learning to occur, coincident activity from HVC and LMAN must occur at a single RA site.

Calculating eligibility is computationally intensive. To speed up simulation time, we compute the convolution using Fourier transforms.

a. Learning Dynamics

The reinforcement signal is created by comparing the motor pool outputs with the target motor trajectory and is defined as:

$$R(t) \;=\; 2 * \Theta[D(t) \;-\; \bar{D}(t)] - 1, \quad (8)$$

where Θ is the Heaviside function, $D(t)$ is the time delayed activity for a trial and $\bar{D}(t)$is the adaptive threshold calculated by averaging the past five trials of $D(t)$. $D(t)$is defined as:

$$D(t + T_{delay}) \;=\; -([\bar{m}_1(t) \;- m_1(t)]^2 + [\bar{m}_2(t) \;- m_2(t)]^2). \qquad (9)$$

Thus, when performance is better than the average of the previous five trials, the reinforcement signal is +1 and when it is worse than the average of the previous five trials, reinforcement is -1. Error shown in figures is the absolute difference between the target and actual motor-pool activity summed across both motor pools.

When the network first starts a trial, there must be enough HVC cells active to drive the rest of the network. To prevent such edge effects from entering our calculations, we confined our analyses of error to within a conservative window, defined as the middle 100 ms of the song.

B. Perturbations to HVC sequencing

In the following sections, we describe the network configurations used to perturb HVC and the implementation of the tests of robustness. To model changes in HVC firing, we define an "epoch" in the learning process by a given number of iterations of the song: N_{stop}. After each epoch a perturbation event occurs. We vary the number of perturbations from zero to fifty, and extend the simulation as needed beyond 100,000 iterations to provide a complete unperturbed final epoch of song. For each number of perturbations, we simulate 50 trials with different randomly generated initial weight matrices, by selecting the random seed.

1. HVC Perturbations: Pausing with Synaptic Weakening

500 HVC neurons were active on each song iteration; 200 additional HVC neurons were paused. At the end of each epoch, we perturbed HVC by randomly sampling 30 neurons from the conservative window of the "active pool" of HVC cell to be placed in this "paused pool" and sample 30 neurons from the "paused pool" to enter the active pool. While neurons are in this paused pool, their synapses undergo synaptic weakening via the following equation:

$$\frac{dW_{ij}^{paused}}{dt} = -(W_{ij}{}^{paused} - W_{ij}{}^{initial})/\tau_{LTD}, \quad (10)$$

where W_{ij}^{paused} is the current synaptic weight of the i[th] synapse of the paused j[th] HVC neuron and $W_{ij}^{initial}$ is the initial paused synaptic weight which we randomly set at the beginning of the simulation. In this way, the synaptic weights of paused neurons decay back to their initial, low values.

2. HVC Perturbations: Pausing

This framework utilizes the same network configurations as above without synaptic weakening. While neurons are in the paused pool, all of the associated synapses are frozen until placed back into the active pool.

3. HVC Perturbations: Time-Shifts

In these simulations there are 500 active neurons in the HVC layer. At the end of each epoch, we perturb HVC by randomly choosing 5% of the cells and shift their burst-onset times randomly within the song (Fig. 3.2). The new times are chosen using the equation:

$$t_{new} = \Delta t + t_{old}, \quad (11)$$

where Δt is drawn from a uniform distribution between $[-t_{old}, T - t_{old}]$, where T = length of the song.

C. Exploring the Impact of Perturbations on Network Robustness

To test the robustness of the network, we developed three measures. First, we quantified the ability of the network to learn the target activity under different numbers of perturbations. Next, we changed the target song and investigated the speed and quality of relearning. Finally, we simulated RA neuron loss and measured the error introduced in the song.

1. Error Improvement

To quantify how well the network learned the template we defined the measure of "error improvement" as the difference between the initial error and the average of the last 500 iterations.

2. Cell Loss in RA

We halted activity in subpopulations (1/12 of the network) of RA, and analyzed how this affected performance error. The loss of neurons was restricted so that losses were equal for both the m_1 and m_2 motor pools. We repeated this test over 500 randomly drawn subpopulations and report the mean resulting error.

3. Shifts in Song and Relearning

To shift the target song, we added a gaussian waveform, with temporal width $\sigma = 10\,ms$, centered at the midpoint of the song to the original template motor-pool activity (Fig. 2c). To calculate the speed of relearning, we measured the time to half decay of the error.

<u>Origins of Robustness</u>

1. Pairwise Correlation of HVC Firing Times

The pairwise correlation between individual HVC neurons' synaptic weights was calculated as a function of the difference in timing between the HVC neurons' burst onsets. We denote the outgoing weights for HVC neuron *p* at time *t* as ${W^t}_p \equiv \boldsymbol{W^t}_{(:,\boldsymbol{p})}$. For all pairs of HVC projection vectors, ${W^t}_p$ and ${W^{t\prime}}_q$ for which the HVC neurons' burst-onset times *t* and *t'* are within $\tau \pm \Delta\tau$, we compute the average pairwise correlation at time separation τ, as:

$$C_\tau({W^t}_p, {W^{t\prime}}_q) = \left\langle \frac{({W^t}_p - \overline{{W^t}_p})({W^{t\prime}}_q - \overline{{W^{t\prime}}_q})}{\sqrt{({W^t}_p - \overline{{W^t}_p})^2 ({W^{t\prime}}_q - \overline{{W^{t\prime}}_q})^2}} \right\rangle_{all\ p,q:\,(t-t\prime)\subseteq\tau\pm\Delta\tau} . \qquad (10)$$

We compute C_t for all timing intervals τbetween 0 and 50 ms; we take $\Delta\tau$ = 0.5 ms. To track the evolution of the synaptic weight structure, these correlations were computed every 200 iterations throughout learning.

2. Generation of Correlated, Gaussian-Distributed Random Weight Matrices

To test our hypothesis that increasing correlations in the synaptic projections between HVC cells that fired at nearby times slows re-learning, we generated random weight matrices such that,

$$W^t{}_p = rW^{t+\delta t}{}_p + (1 - r^2) * \vec{X} \qquad (11)$$

where $\vec{X}$ is a vector of independent gaussian random variables with $\mu = 5$ and $\sigma^2 = 1$, $\vec{X} \sim \mathcal{N}(5 * \vec{1}, \mathbb{I})$, $W^{t+\delta t}{}_p$ are the synaptic weights from the HVC cell which fires at the smallest latency after the HVC cell that fires at time t with synaptic weights, $W^t{}_p$, and r is the correlation strength between these two HVC projection vectors. See figure S4 for results.

Table 3.1 Parameters

Conductance model	Description	Parameter name	Value
	Membrane capacitance	Cm	$1\ \mu F/cm^2$
	Leak equilibrium potential	V_L	-60 mV
	HVC leak conductance	g_L	0.3 mS/cm^2
	RA leak conductance	g_L	0.44 mS/cm^2
	Action potential threshold	V_θ	-50 mV
	Reset potential	V_{reset}	-55 mV

Connectivity	Weight Matrix	W	
	Paused Steady State Weights	$W^{initial}$	
	Paused Pool Weights	W_{Paused}	
	RA Neuron Count	N_{RA}	48
	HVC Active Neuron Count	N_{HVC}	500
	HVC Paused Neuron Count	$N_{HVC\ Paused}$	200
	LMAN Neuron Count	N_{LMAN}	48
Synaptic parameters	Synaptic Time Constant	τ_s	5 ms
Motor Pool	Motor Pool Time Constant	τ_m	5 ms
	Tonic Activation 1	b_1	60
	Tonic Activation 2	b_2	40
	RA Positive Output Weights 1	A_{ij}	$440/N_{RA}$
	RA Negative Output Weights 1	A_{ij}	$-440/N_{RA}$
	RA Positive Output Weights 2	A_{ij}	$660/N_{RA}$
	RA Negative Output Weights 2	A_{ij}	$-660/N_{RA}$
Learning	Learning Rate	η	$5 * 10^{-5}$
	Eligibility Time Constant	τ_e	5 ms
	Iterations in a Epoch	N_{stop}	50,000-2,000

	Number of Epochs	N_{Epoch}	1-50
	LTD Time COnstant	τ_{LTD}	.001
	Number of neurons place in paused pool	N_{change}	30 Each

Appendix References:

1. Stark, L. L. and D. J. Perkel (1999). "Two-stage, input-specific synaptic maturation in a nucleus essential for vocal production in the zebra finch." J Neurosci **19**(20): 9107-9116.
2. Fiete IR, Fee MS, Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. J Neurophysiol 98(4):2038–2057.
3. Spiro JE, Dalva MB, Mooney R (1999) Long-range inhibition within the zebra finch song nucleus RA can coordinate the firing of multiple projection neurons. J Neurophysiol 81(6):3007–20.

Chapter 4

Learning from Disorder in Neural Networks

This chapter discusses the role of variation in trial and error learning and summarizes work-in-progress of an ongoing project that implements reinforcement learning via stochastic gradient descent from a chaotic reservoir. This project is a collaboration with Dr. Guillaume Lajoie and Dr. Adrienne Fairhall.

"The starting point for the formation of any association is the fund of instinctive reactions. Whether or not in any case the necessary act will be learned depends on the possibility that in the course of these reactions the animal will accidentally perform it."

Edward Thorndike, 1898, from *Some Experiments on Animal Intelligence*

"The ground beneath my feet is nothing but an enormous unfolded newspaper. Sometimes a photograph comes by; it is a nondescript curiosity, and from the flowers there uniformly rises the smell, the good smell, of printer's ink. I heard it said in my youth that the smell of hot bread is intolerable to sick people, but I repeat that the flowers smell of printer's ink…"

Andre Breton, 1924, 'automatic writing' from the surrealist novel, *Soluble Fish*

Exploration is a fundamental behavior across orders and complexities of organisms. While noise in nervous systems is often a hinderance, it can also play an important role. How the nervous system casts the dice is a relevant question whenever a degree of randomness is required. General categories of behaviors where chance plays a constructive role include pursuing a goal in an unknown environment or with unknown tools, acting with an underspecified goal, decision making with incomplete information or creating something new.

Motor learning is one example of pursuing a goal with unknown or partially unknown tools: the process of learning a new motor task requires exploring motor space to find appropriate movements and repeat them. In trial and error learning random gestures are explored; good ones are repeated, and bad ones are abandoned (1). The theory of reinforcement learning (RL) describes an explicit trade-off between exploration and exploitation: as an agent explores a new environment they must decide whether to behave in a way that capitalizes on their current knowledge of the path to the maximal reward (exploitation), or in a way that seeks out unknown aspects of the environment in a hopes of finding an even larger reward (exploration) (2). Unlike some other forms of learning, reinforcement learning requires some degree of stochastic exploration to discover correct forms of behavior (3).

There are several results from motor learning in humans that suggest increasing variability during motor learning improves final performance (4). In one study, when humans were asked to learn a computer-based motor skill, the learning rate was doubled when experimenters forced one group to introduce more variability into the learning process (5). In another study participants were trained in a reinforcement learning paradigm to execute a stereotyped reaching motion (6). They did not know the correct trajectory but were only given a numerical reward at the end of each trial indicating how they had performed. Learning speed was highly correlated with task-relevant motor variability: participants whose initial motor variations were more aligned with the hidden trajectory learned faster. Furthermore, training aligned the temporal structure of the motor variability more with the relevant task dimensions.

From these studies, it seems that variability in motor learning tasks can be deployed for efficient exploration and shaped relative to the demands of the search.

In songbirds a specialized circuit injects variation into the song learning process. The cortical-like output (LMAN) of the anterior forebrain pathway (AFP) projects to the primary motor pathway with which adult song birds generate the motor patterns necessary for song (7). Pharmacologically silencing LMAN in juvenile birds results in immediate and abnormal stereotypy in song (8). In adulthood, the song is stereotyped and consistent from rendition to rendition but small variations persist. However, birds retain the ability to change their song. It is easy to imagine why adult song plasticity might be useful: changes in environment and physical changes due to aging or injury would necessitate that adult birds maintain the ability to adjust their song structure. Indeed, adult birds exhibit two distinct modes of performance in laboratory contexts: a practice mode (undirected song) which they sing in isolation, and a performance mode (directed song) which they sing in presence of female birds. The undirected song is significantly more variable than the directed song, and a corresponding increase in variability during undirected song is observed in recordings from LMAN activity (9). Lesioning LMAN removes the increased variation in undirected song (9). Furthermore, adult birds make use of specific variations at local points of their song for ongoing learning. Adult male Bengalese finches as well as zebra finches can shift the pitch of specific portions of their song to avoid an aversive noise stimulus that is conditioned on the variations in pitch rendition-to-rendition (10, 11). Together, this shows that adult birds are able to modulate the degree of variability and exploration in their song performances, and that the residual variations in adult song serve a useful role in ongoing learning.

Even though the early babbling stage of song learning is highly variable and unstructured, there are constraints on the type of vocalizations that are eventually learned. Zebra finches raised in isolation without ever hearing an adult bird's song still learn an abnormal version of the enculturated zebra finch song (12). This suggests genetic priors such as muscle constraints and network organization that shape variations and underpin the learned versions of song.

Noise in a motor task such as bird song can arise in the nervous system from many sources and at many scales (13). Stochastic vesicle release in synapses (14, 15), the probabilistic gating of voltage-dependent ion channels (16) contribute to cellular noise. The noise from these sources can then be amplified or suppressed by the larger scale connectivity structure of networks of neurons. Classical balanced network architectures which have been postulated to exist in cortex can give rise to chaotic dynamics that amplify small differences in individual cellular activity (17). Peripheral variability also contributes to variability in motor tasks: force-dependent scaling of trial-to-trial variations in performance is seen as a basic property of muscles (4). However, this type of peripheral system noise may not be useful to motor learning because it is not controllable (4).

What sources of variation are useful for exploration in trial and error learning tasks? First, the direction and dimensionality of variations should be relevant to the task: exploration that wanders too far from or never accesses relevant regions of action will fail to reach an optimal solution. Second, the scale of variations should be relevant to the task: if variations are too small, learning will be too slow; if variations are too large, the exploratory process will jump over optimal solutions. Lastly, in trial and error learning, successful, random acts are repeated

with higher probability. Therefore, the probability distribution from which random acts are drawn must be changeable. Variability in network-level dynamics may be particularly useful to tasks such as song learning in birds because it is well established that this type of variability can be modulated on fast timescales by external network inputs or neuromodulator concentrations and on slower timescales through changes to connectivity via synaptic plasticity (18, 19).

In ongoing work, we study the impact of using chaotic dynamics from a neural network as the source of exploration in a reinforcement learning problem inspired by bird song architecture. We use numerical simulations to study the effects on learning of varying the degree of intrinsic variability in network dynamics. We find that a reinforcement learning system that implements stochastic gradient descent is able to learn even with a relatively high degree of structure in the exploratory dynamics. However, performance divergences when the exploratory dynamics have insufficient trial-to-trial variation. Increased variation leads to better performance. Chaotic dynamics outperform a Poisson process when the network dynamics are more variable than the stochastic process. Lastly, when the target of learning is partially aligned to fall within the product of the strange attractor on which the chaotic dynamics vary, performance is further improved. This could represent a beneficial interaction between an evolved, genetic prior and trial-and-error learning wherein learning takes place embedded within a pre-existing, structured architecture.

In the next sections I summarize completed work and discuss future directions.

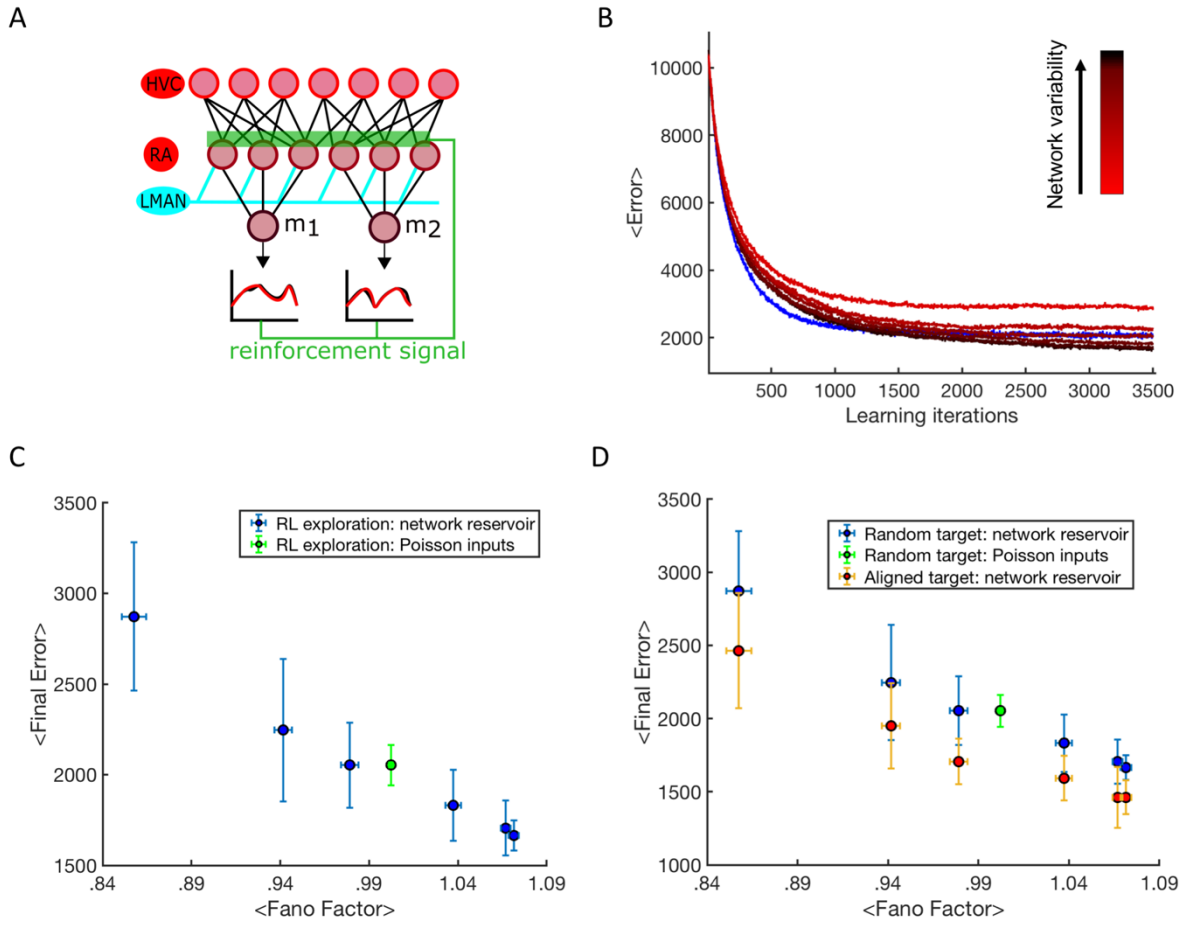

Figure 4.1. RL learning from chaotic dynamics. **A.** Model of bird song-inspired network architecture (This figure is taken from chapter 3 and was made in collaboration with Elliott Abe for our manuscript). HVC projects in a feed-forward manner to RA, which projects to downstream motor neurons that then drive song production. LMAN projects to RA and drives trial-to-trial variations in RA activity. Synapses from HVC to RA are changed based on the evaluation of the motor output from m1 and m2. For a more complete description of this RL learning model see Chapter 3. LMAN is modelled as a recurrent network driven by a static white noise realization. **B.** Learning trajectories driven by varying degrees of chaotic activity from the LMAN network. Trajectories are averaged over 25 learning trials. The blue trace is the average of learning trials which use independent Poisson variation instead of pooled LMAN network activity. Error is defined as the absolute difference between the target m_1 and m_2 and the model output m_1 and m_2. **C.** Final error averaged over the last 50 iterations of learning as a function of variability in the LMAN drive. The Fano Factor (FF) is defined as the variance divided by the mean of spike counts in a given time window. The average of the FF is taken across song and over learning trials. Vertical and horizontal error bars represent the standard error in mean Final Error and mean FF respectively. More variable inputs from LMAN lead to improved performance. **D.** Final error plotted as in panel C with the addition of trials wherein the target of learning is chosen to partially align with the natural product of the reachable phase space in the LMAN dynamics. Aligning the search space with the task improves performance.

Results:

Our learning task is inspired by the song motor pathway in song bird. The model consists of three feed-forward layers, HVC, RA and motor pools and was adapted from Fiete et al. (Fig 4.1a) (20)[1]. RA is driven by both HVC activity and inputs from LMAN. HVC and RA layers are modeled as a conductance-based leaky integrate-and-fire neurons. Song production is modeled by two non-spiking motor pool output units that receive input from RA. The goal of learning in the model is for the two motor pools to reproduce a target motor trajectory.

Learning Parameters

Changes in the system's structure happen through synaptic plasticity in the HVC to RA synapses. Only HVC projections to RA are plastic and are determined by:

$$dW_{ij}/dt \; = \; \eta \, R(t) \, e_{ij}(t) \, ,$$

where W_{ij} is the synaptic strength from the jth HVC cell to the ith RA cell, R(t) is the reinforcement signal, η is the learning rate, and $e_{ij}\,(t)$ is the eligibility trace over which weight changes occur. The eligibility trace is defined as:

[1] The feed-forward RL model studied in this section is an adapted from the model as used in Chapter 3. Here we explore the impact of changing the nature of the variable inputs from LMAN. For more details on the feed forward portion of this model, see the Extended Methods section of Chapter 3.

$$e_{ij}(t) = \int_0^t dt' \, \mathrm{G}(t-t') \, (s_i^{LMAN}(t') - < s_i^{LMAN} >) \, s_{ij}^{HVC}(t'),$$

where $G(t) = t^n e^{t/\tau_e}$, $s_i^{LMAN}(t)$ is the LMAN input to the ith RA cell, and $s_{ij}^{HVC}(t)$ is the synaptic input from the jth HVC cell to the ith RA cell. The reinforcement signal, R(t), is defined as: $R(t) = 2 * \Theta[D(t) - \bar{D}(t)] - 1$, where D(t) is the mean squared error of the current motor output and $\bar{D}(t)$is the average mean squared error of the previous 5 trials.

Variable inputs from LMAN

RA receives inputs from LMAN, which we model as convergent inputs from a recurrent network. We adjust the parameters of the network to modulate the trial-to-trial variability of the network and thus modulate the degree of variability LMAN injects into the learning process. We model individual units in the recurrent network as theta neurons, a mapping of a quadratic, integrate-and-fire neuron onto the unit circle (18). The level of variability is controlled in this network by σ, the variance of a static, white noise drive to each unit in this network and η, the intrinsic excitability of individual units (Fig. 4.2). We numerically solve a mean-field approximation of our networks' firing rates for σ and η to guide our selection of σ and η such that the average firing rates of the different network realizations are the same. For each level of variability in the LMAN network, we simulate 25 learning trajectories and compare the average error trajectories over the 25 trials across levels of variability in LMAN activity (Fig 4.1b). For each iteration within an individual learning trajectory, the LMAN network is driven with the same white noise realization; it is the intrinsically chaotic dynamics of the network which leads

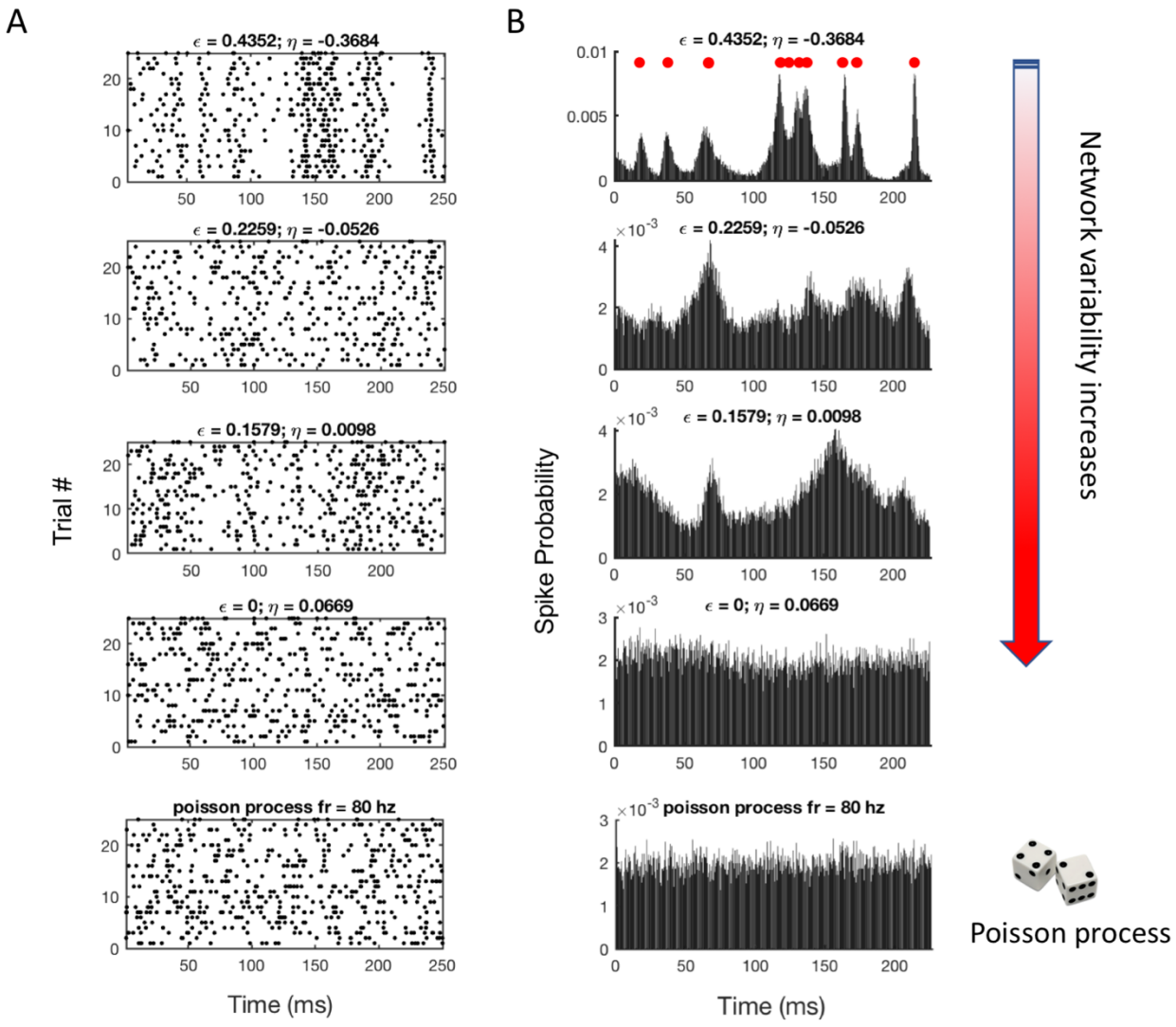

Figure 4.2. LMAN network activity patterns as degree of network chaos varies. **A.** Example raster plots from LMAN network activity. 25 trials are plotted for each network realization. In each figure all trials receive the same white noise realization as the network drive: differences in initial conditions and intrinsic chaotic dynamics drive trial-to-trial variability. Note, the white noise realization does change as network variability increases. **B.** Corresponding normalized PSTH of the spiking activity over 3,000 trials for the example cells on the left over the time course of the learned song trajectory. This is the probability density function of spiking activity over the course of song for a single input. The top panel shows an example of generating LMAN activity patterns from the probability density function to generate an attractor aligned song target. Each red point represents an independent random spike time drawn from the probability density function.

to iteration-to-iteration variations (Fig 4.2). Within the bird song framework, this can be seen as approximating a regular excitatory drive from DLM to LMAN. To compare this input to a more traditional method of driving exploration in a stochastic gradient descent task, we model a control, exploratory LMAN input as independent, homogeneous Poisson spiking.

Figure 4.1b shows the learning trajectories over degrees of LMAN network variability and compares these trajectories to a Poisson drive. As variability increases in the LMAN network, learning improves. Figure 4.1c shows the average error in the last 50 iterations (3450 to 3500) of the learning trajectory as a function of the Fano Factor (FF) of LMAN inputs. The FF is an experimentally accessible measure of spiking variability and quantifies the variability of each LMAN network realization. The FF is defined as the ratio of the variance of spike counts to the mean of spike counts over repeated trials within a fixed time window:

$$FF = \frac{\sigma^2{}_{spks}}{\langle N_{spks} \rangle}.$$

We compute the average FF in 30 ms, fixed windows across the song trajectory and average these values across all inputs to RA. The true value of the FF for a Poisson process is 1. We use this theoretical value to linearly rescale the numerical computation of the FF for the Poisson process and the network realizations (true numerical calculation of the Poisson process = 0.97122) due to numerical error. More variable inputs from LMAN lead to improved learning performance. The learning trajectory is able to converge even for lower levels of variation in the exploratory drive. When the variability of the network dynamics exceeds FF=1, the chaotic dynamics are more successful at driving learning than a purely stochastic Poisson process (Fig 4.1c).

Partially supervised learning: aligning a target trajectory with the chaotic attractor

In the previous section, the success of the learning depends simply on the fact that chaotic network dynamics approach or exceed the variability of a stochastic process. This variability allows the stochastic gradient descent algorithm that is being implemented in this task to sufficiently explore the space of possible solutions and converge to a satisfactory approximation of the target. However, this does not address the how network structure might shape a learning process in ways that stochastic inputs do not. Here we treat the structure within the chaotic activity as the 'prior' upon which the trial and error learning is carried out. We suggest that this prior could be built either genetically or during the earliest passive listening period of juvenile bird development. This direction investigates why learning built on a chaotic attractor might not be simply sufficient but advantageous in comparison to purely stochastic forms of exploration.

Approximating the chaotic attractor

We approximate phase space region of the chaotic attractor by building up a probability density function of the recurrent network output over the course of the song time interval. To build the probability density function we normalize the peristimulus time histogram (PSTH) of inputs to each RA unit independently over repeated trials. Trials are defined by network trajectories under the same substantiation of white noise drive. The PSTH is formed from 3,000 network trajectories for each level of network variability. These trajectories are used only to build the probability density function and not for learning. Figure 4.2 panel B shows the

probability density function for LMAN inputs to a single RA unit over the course of the song for varying levels of variability.

Adding partial supervision: an attractor-aligned target

We then generate an 'attractor-aligned' target song by drawing spikes independently from this probability distribution and driving RA using this randomly selected activity pattern from LMAN to create a target song output in the motor pool (Fig. 4.2b top panel). Note that this output is not an actual chaotic trajectory: the dynamics of the chaotic network are not mimicked, but the spike patterns are drawn from locations in phase space such that they are likely to overlap with the chaotic attractor. In this way the alignment is crude: dynamics correlated in time or space will not appear in this alignment. Furthermore, the song generating process filters the output of the LMAN network through the non-linear input-output relationship of the conductance-based, leaky integrate and fire models in the RA network and the exponential filter of the synaptic input from LMAN. Therefore, while the task is now correlated with the chaotic attractor, it is removed in several respects from the actual network dynamics. In this way, we call the task 'partially supervised': the exploratory drive contains partial information about the desired learning target.

Figure 4.1 panel D compares the final error after learning when the task is partially supervised to when it is fully unsupervised. At each level of variation in the network dynamics, partial supervision leads to a decrease in final error. This is true even in the highly variable networks where the phase space of the attractor is large and unstructured. Thus, when the phase space of the network dynamics aligns with the learning task, learning outcomes improve.

Discussion

In this work we have shown that a reinforcement learning task implemented in a feedforward neural network via stochastic gradient descent can learn using the variable dynamics from a recurrent neural network as the source of exploratory activity and that when the network dynamics share partial alignment with the task goal, learning improves.

Stochastic gradient descent assumes that an agent can take unbiased samples of their environment. In this way they develop an unbiased estimate of the direction of greatest gradient and are able to proceed along that route toward a minimum error solution. Using a recurrent network to explore the error landscape may violate this assumption of an unbiased sampling: although the network output is chaotic, it still exists in a confined region of phase space: correlations across individual units and in time makes it such that sampling from structured network dynamics is not an unbiased sampling of the local region of phase space. Indeed, we found that if the network becomes too regular, the learning process diverges (figure not shown). However, even in networks with lower variability than a stochastic Poisson process, learning is able to converge. Future work is needed to explain the path to divergence.

In our modeling process we reverse engineer an artificial alignment between the network structure generating task variations and the task itself. In the bird, we imagine that this correspondence exists due to genetic predispositions in neural development that interact with the form of the enculturated song. Birds raised in isolation without any exposure to tutor songs still develop an abnormal version of the enculturated song. Furthermore, when colonies are bred from initially isolated birds, an enculturated form of the song emerges in 3-4 generations

implying that enculturated song is partially genetically encoded (12). The early learning process also likely generates this type of network-task correspondence. Selectivity to the bird's own song as well as to the tutor song emerges in LMAN neurons during the course of development, which suggests that early auditory and motor experience also shapes the structure of this network (21).

Shaping the LMAN source of exploration to partially align with the task goal may be another way to address the 'curse of dimensionality' that exists in reinforcement learning, wherein the dimensionality of the task becomes intractably large. Refining and shrinking the task and task-exploration space narrows the search. It has been suggested in other work that one role of the highly stereotyped, sequenced activity in HVC, which emerges prior to the learning of specific song elements, is to narrow the latent space upon which RL operates and speed up learning (22, 23). Searches confined to priors embedded in LMAN structure would add another element of structure to the RL task.

Future work

We intend to pursue two future directions in this project. One, we will exchange our current reward-based plasticity rule for a local spike-timing dependent plasticity rule and instead vary the drive to the LMAN network to modulate the reliability of the network dynamics in a time-dependent, error-dependent manner. This will more fully align with the actual learning process likely present in song bird (24). It will also take advantage of another salient aspect of variability arising from chaotic network dynamics: variation can be modulated rapidly in a time-dependent manner using network inputs (19).

Two, we intent to expand the idea of 'partial alignment' of the target trajectory with the network structure and explore how introducing elements of temporal and spatial correlations in the task alignment influences learning. This will make use of the dynamics of the recurrent network and follows notions of elemental gestures proposed to be present in song production (25).

Bibliography

1. Thorndike E (1898) Some Experiments on Animal Intelligence. *Science* 7(181):818-824.
2. Sutton RS & Barto AG (1998) *Reinforcement learning : an introduction* (MIT Press, Cambridge, Mass.) pp xviii, 322 p.
3. Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? *Neural networks : the official journal of the International Neural Network Society* 12(7-8):961-974.
4. Dhawale AK, Smith MA, & Olveczky BP (2017) The Role of Variability in Motor Learning. *Annual review of neuroscience* 40:479-498.
5. Wymbs NF, Bastian AJ, & Celnik PA (2016) Motor Skills Are Strengthened through Reconsolidation. *Current biology : CB* 26(3):338-343.
6. Wu HG, Miyamoto YR, Gonzalez Castro LN, Olveczky BP, & Smith MA (2014) Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nature neuroscience* 17(2):312-321.
7. Perkel DJ (2004) Origin of the anterior forebrain pathway. *Annals of the New York Academy of Sciences* 1016:736-748.
8. Olveczky BP, Andalman AS, & Fee MS (2005) Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS biology* 3(5):e153.
9. Kao MH, Wright BD, & Doupe AJ (2008) Neurons in a forebrain nucleus required for vocal plasticity rapidly switch between precise firing and variable bursting depending on social context. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28(49):13232-13247.
10. Tumer EC & Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* 450(7173):1240-1244.
11. Gadagkar V*, et al.* (2016) Dopamine neurons encode performance error in singing birds. *Science* 354(6317):1278-1282.
12. Feher O, Wang H, Saar S, Mitra PP, & Tchernichovski O (2009) De novo establishment of wild-type song culture in the zebra finch. *Nature* 459(7246):564-568.
13. Renart A & Machens CK (2014) Variability in neural activity and behavior. *Current opinion in neurobiology* 25:211-220.

14. Calvin WH & Stevens CF (1968) Synaptic noise and other sources of randomness in motoneuron interspike intervals. *Journal of neurophysiology* 31(4):574-587.
15. Katz B & Miledi R (1970) Membrane noise produced by acetylcholine. *Nature* 226(5249):962-963.
16. White JA, Rubinstein JT, & Kay AR (2000) Channel noise in neurons. *Trends in neurosciences* 23(3):131-137.
17. Sompolinsky H, Crisanti A, & Sommers HJ (1988) Chaos in random neural networks. *Physical review letters* 61(3):259-262.
18. Lajoie G, Lin KK, & Shea-Brown E (2013) Chaos and reliability in balanced spiking networks with temporal drive. *Physical review. E, Statistical, nonlinear, and soft matter physics* 87(5):052901.
19. Rajan K, Abbott LF, & Sompolinsky H (2010) Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical review. E, Statistical, nonlinear, and soft matter physics* 82(1 Pt 1):011903.
20. Fiete IR, Fee MS, & Seung HS (2007) Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *Journal of neurophysiology* 98(4):2038-2057.
21. Solis MM & Doupe AJ (1999) Contributions of tutor and bird's own song experience to neural selectivity in the songbird anterior forebrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 19(11):4559-4584.
22. Mackevicius EL & Fee MS (2018) Building a state space for song learning. *Current opinion in neurobiology* 49:59-68.
23. Fiete IR, Hahnloser RH, Fee MS, & Seung HS (2004) Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *Journal of neurophysiology* 92(4):2274-2282.
24. Fee MS & Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* 198:152-170.
25. Amador A, Perl YS, Mindlin GB, & Margoliash D (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495(7439):59-64.

Chapter 5

Analysis of VTA's Relationship to Natural Fluctuations in Song: Dopaminergic Evaluation of Trial-to-trial Variations in Performance and Its Relationship to a Reinforcement Learning Framework

The research presented in this chapter is the result of a collaboration with Dr. Vikram Gadagkar, Dr. Kenneth Latimer, Dr. Jesse Goldberg and Dr. Adrienne Fairhall. Vikram Gadagkar and Jesse Goldberg collected the data and performed the original experiments. Kenneth Latimer, Adrienne Fairhall and myself designed the statistical modelling approach and significance analysis. Kenneth Latimer implemented the Gaussian Process model. I performed the data analysis and led the analysis design. All collaborators contributed to the design and interpretation of the analysis results. Two manuscripts are in preparation from this research: one focused on the role VTA in encoding a reward prediction error and a second focused on the global state hypothesis.

Mental facts cannot properly be studied apart from the physical environment of which they take cognizance*.... Mind and world in short have evolved together, and in consequence are something of a mutual fit.*

--William James, pp. 2-3 *Psychology (Briefer Course)*

Introduction:

Value judgements are contextual. If you are learning a song, say Claude Debussy's Syrinx, L. 129 for solo flute, and you play an A sharp, how do you know you played the right note? It depends on where you are in the song: at one point in the song an A sharp is correct; at another point in the song, it is an error. Furthermore, how do you decide if you played the whole song well? This depends on your manner of judgement: you likely did not play it as well

as the solo flutist of the New York Philharmonic; however, perhaps you played it better than you did when you were first learning the song, or even better than you played it yesterday. And then, there are styles of playing that don't differentiate along an error axis at all. Some music historians believe Debussy originally wrote Syrinx, L. 129 without bar lines or breath marks, so as to leave the performer wide room for interpretation and emotional expression. A forlorn rendition is no more or less correct than a lighthearted one. A quiet performance in a small room is no more or less correct than a loud performance in a concert hall.

Evaluating your own performance requires a comparison of what you heard, your sensory feedback, to your internal benchmarks of success. The internal benchmarks are specific to each moment in song and to your current skill level. They are likely also influenced by things like your mood, your level of alertness and the environment in which you are performing. Value judgements are necessary for learning: without them we would have no compass by which to change our actions.

How are internal value judgements arrived at, and how do we use them to learn? This process has been understood in many contexts through a theoretical framework called Reinforcement Learning (RL) (1). In an RL account, learning progresses in a trial and error manner; good trials are reinforced while poor trials are not, and through this process, the student improves. The student evaluates trials based on pre-existing expectations of performance quality. Trials that are better than expected are rewarded whereas trials that are worse than expected are not. This is not an absolute definition of error: as expectations change, so too will value judgements. Expectations are updated as skill level progresses and, importantly for memory and computational capacity, do not rely on the full history of practice.

A student need only make differential changes to an expectation cumulatively based on the previous trials in order to learn in this framework. In RL theory, the difference between expected and actual performance quality is called the reward prediction error (RPE).

Where in the brain do these value judgements form? In humans, errors in musical performance (2, 3) and speech (4) relate to changes in electroencephalogram readings that may originate from dopamine neurons in the ventral tegmental area (VTA). In rodents and primates, dopamine neurons in VTA encode reward-like signals in response to the modulation of external rewards such as juice or food as well as to the expectation of future reward (5-7). In electrical self-stimulation experiments, rodents will choose to electrically stimulate dopamine neurons over food or sex, further suggesting these neurons' relationship to aspects of reward and reinforcement (8).

VTA dopamine neurons are also part of the song learning circuits in song birds. Song birds learn to sing in a trial and error manner. They use auditory feedback to improve performance and have VTA projections into the basal ganglia portion of a basal ganglia-thalamic-cortical loop that is necessary for song learning (Fig. 5.1a,b) (9-11). The most widely studied song bird, the zebra finch, learns a single mating song during development and then sings it in a stereotyped manner throughout adulthood.

In previous work, our collaborators showed that in the zebra finch, dopamine neurons in VTA encode signals resembling reward prediction errors in response to experimentally-manipulated, internal benchmarks of song performance (12). Gadagkar et al. controlled the birds' perceived song quality with distorted auditory feedback. Days before neural recordings, a

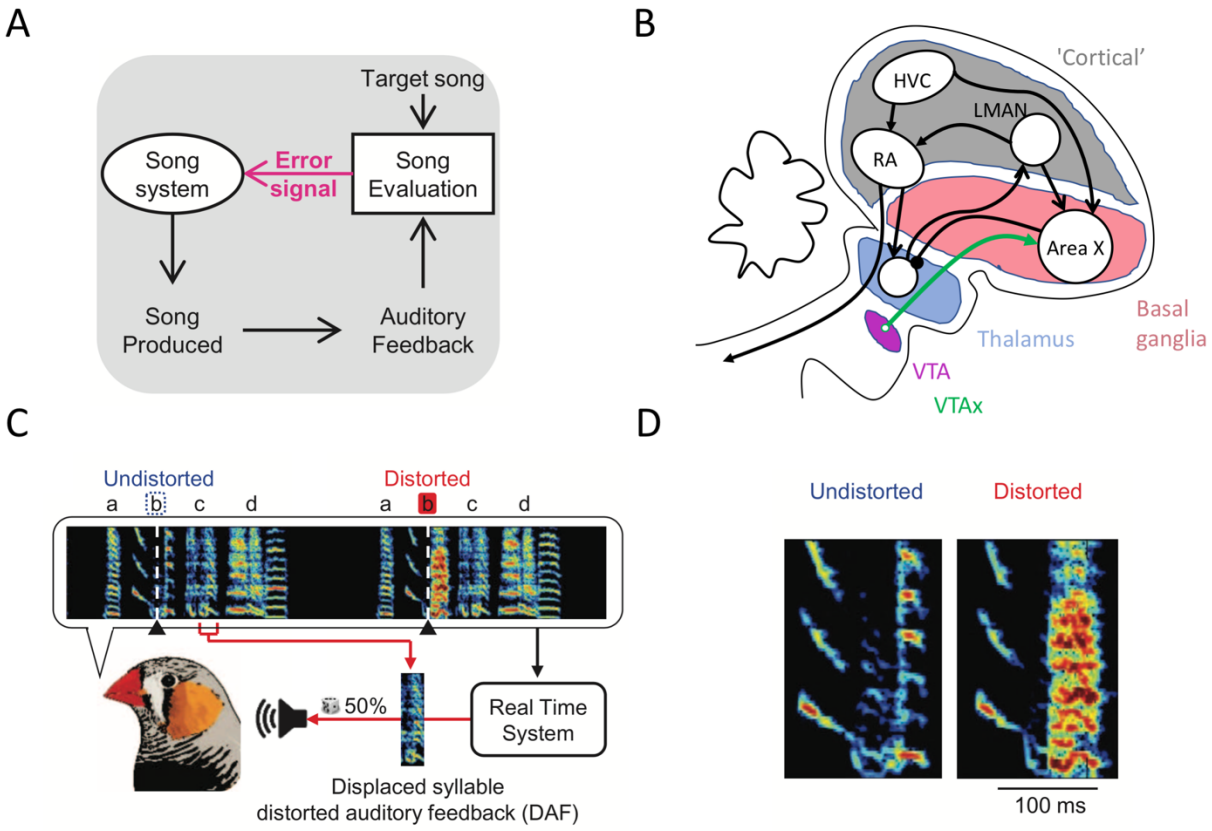

Figure 5.1. Schematics of song bird learning circuit and experimental paradigm from Gadagkar et al. 2016. Panels a,c and d are from Gadagkar et al. 2016. Panel b is adapted from Gadagkar et al. 2016. **A.** Schematic of learning system. Evaluation of auditory feedback is thought to be used in evaluating song and driving learning through an error signal. **B.** Schematic of the song bird learning circuitry. VTA is hypothesized to send an error signal to the basal ganglia region of a learning loop that involves the basal ganglia, thalamus and a cortical-like region. **C.** Schematic of experimental distortion paradigm. At a chosen target time, a portion a displaced syllable was randomly played during production of the target syllable. Distortions were randomly interleaved with undisturbed renditions. **D.** Close-up view of target syllable during an undistorted and distorted rendition.

portion of a specific syllable was either distorted with distorted auditory feedback (DAF) or left undistorted in randomly interspersed renditions[1] (Fig. 5.1 c,d). In this way, they induced a shifted expectation of song quality at the point of distortion. When they recorded from neurons in VTA they observed two distinct groups of responses. A subset of neurons responded with a significant error-like signal, defined as the z-scored difference in average firing rate 50 to 125 milliseconds after distortion onset between renditions with and without the distortion. They labelled these neurons as "VTA-error" neurons (n = 17 neurons; error response = 3.3 +- 0.5). They labelled the neurons which did not respond with an error signal as "VTA-other" neurons (n=108; error response = 0.1 +- 0.9). All VTA-error neurons responded to the auditory distortion with phasic suppression of their firing rates and to the undistorted syllable with phasic activation of their firing rates. This response resembles an RPE signal from RL theory: the undistorted syllables are evaluated as better than expected based on the recent history of randomly-inflicted distortion at that point in song and result in an elevated firing rate. Thirteen of the seventeen VTA-error neurons were antidromically identified as projecting to area X. 95% of X-projecting VTA neurons are dopaminergic. This projection structure and cell type agree with previous findings and hypotheses that describe dopaminergic VTA projections to area X as an evaluation signal which induces synaptic plasticity within area X (13, 14).

These results extended the connection between the RL theoretical framework and experiment to a behavior for which, in a natural context, there is no external reward, only internal evaluation based on internal benchmarks. However, the experiment created an

[1] The distortions were constructed either from a portion of another syllable from the same bird, shifted in time, or synthesized to resemble the acoustic structure of the broadband portion of the bird's song.

artificial and exaggerated binary for the bird to respond to: the song was either permitted to proceed in a natural manner, or a portion of the song was distorted artificially such that the bird's auditory feedback was dramatically different from the natural song. This exaggerated distortion led to an exaggerated neural response that exceeded the cells' natural firing rates by several standard deviations and permitted an unambiguous reading of the difference in responses to the natural song and the distorted auditory feedback. This experimental paradigm also offered a reasonable way to guess at the bird's internal and subjective value judgement of their own song within the experimental context: it is reasonable to assume that the bird prefers their own song to a distorted version of it.

Although this experiment intervened in a natural behavior, it was not a natural context by design. In addition to the artificiality of the distortion, the bird had no advanced cue of whether a random distortion would happen in a given rendition, which may or may not be true when the bird is varying their own song. If the bird has an internal model of song production, it is quite possible that an intention to sing the song in a particular way would proceed a fluctuation in song. In RL theory, this could change the timing relationship of the RPE signal to the song variation it evaluates (1). Additionally, while one interpretation of the spiking response to the distortion events is "better or worse than expected" another clear valence of the experimental context is "hearing one's own song versus hearing an externally imposed sound". To what degree this experimental paradigm exaggerates the experience of singing good and bad versions of the song versus creating a new context is unknown.

This experiment creates an unusual opportunity to study in a natural context specific neural signals that have been designated in RL theory and through this experiment as reward

prediction errors. The experiment and theory together generate a strong hypothesis about the VTA's role in natural song: VTA activity should encode fluctuations in natural song activity and the VTA-error neurons' activity should resemble a reward prediction error. Further, the experiment allows us to analyze VTA activity during natural song with an empirically-derived division between the role of the VTA-error and VTA-other cells. From this, we hypothesize that the VTA-error cell population should resemble an RPE signal while the VTA-other cell population should not. Thus, we ask in this analysis, do VTA neurons' activity patterns relate to natural fluctuations in song? If so, what is the structure of these relationships and do they relate to an RPE framework?

To answer these questions, we parameterized the song into a low dimensional set of time-varying song features (Fig. 5.2a). We agnostically fit the relationship between rendition-to-rendition variations in song features and spike counts at local time steps in song (Fig. 5.2b) across many spike-song segment latencies and looked for when song feature variations predicted spike counts (Fig. 5.3 a,b). We characterized the timing of these predictive fits as well as the shapes (i.e. the tuning curves). We assessed the significance of finding predictive fits using a population method of bootstrapping that retained the underlying temporal structure of song and spiking.

We find that the activity of the VTA-error neural population correlates to fluctuations in natural song in a manner consistent with an RPE signal. Both the timing and nature of the activity patterns fit what would be expected from an RL theory. Furthermore, a subset of the VTA-error cells' activity correlates with what we term macroscopic changes in song: the number of syllables sung and the number of repetitions of single syllables.

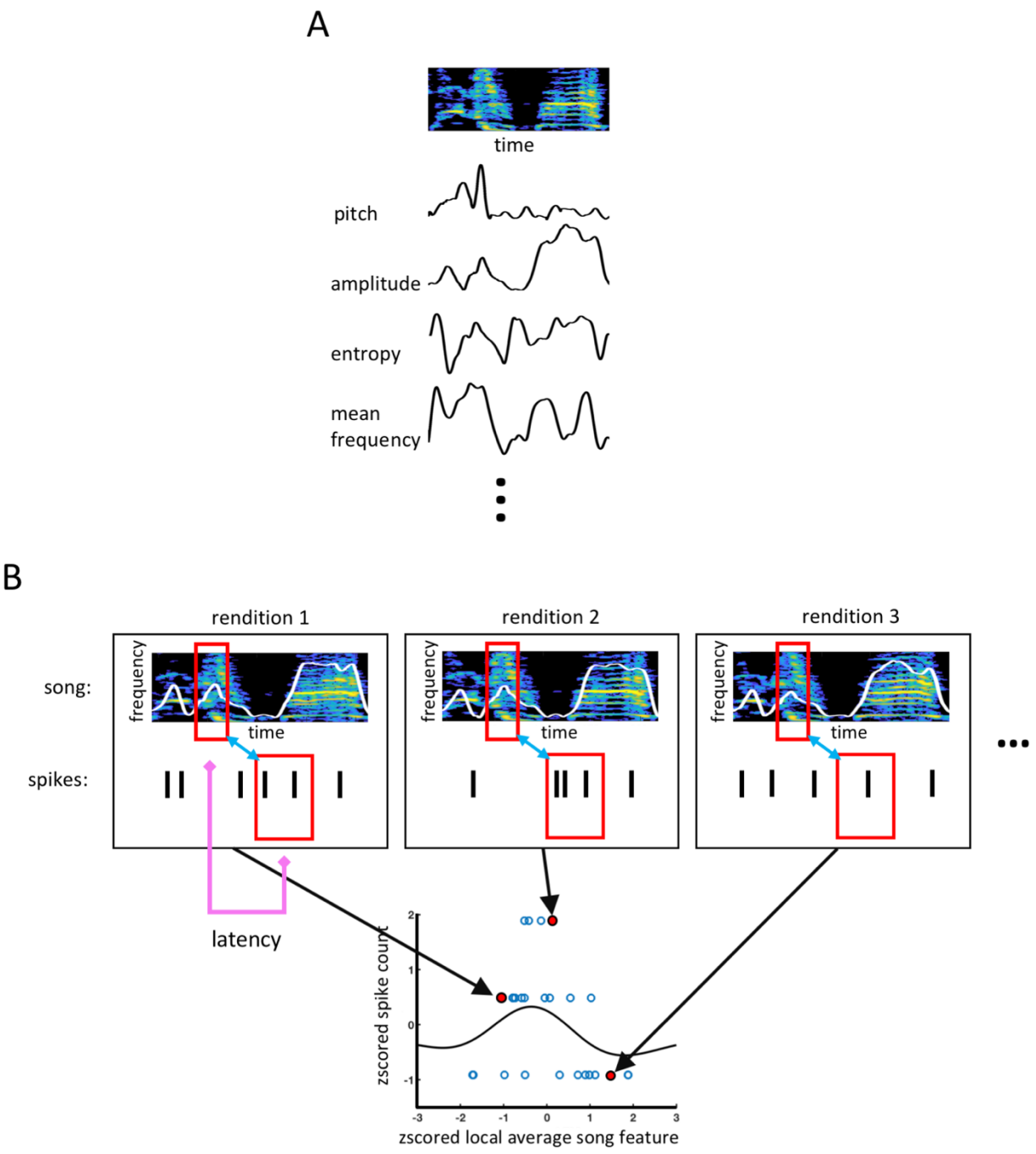

Figure 5.2. Schematics of analysis scheme. **A.** Spectrogram of song along with examples of extracted song features. In total we use 8 feature parameterizations of song: amplitude, pitch, entropy, goodness of pitch, mean frequency, AM, FM, and aperiodicity. **B.** Schematic of fitting song fluctuations to spike counts within specific time windows. Local feature averages are used to predict local spike counts. A Gaussian process model is used to fit the relationship between local feature averages and spike counts. This example shows the process for a single song feature for visualization purposes; however, we use all eight song features at each point in song to construct the GP model.

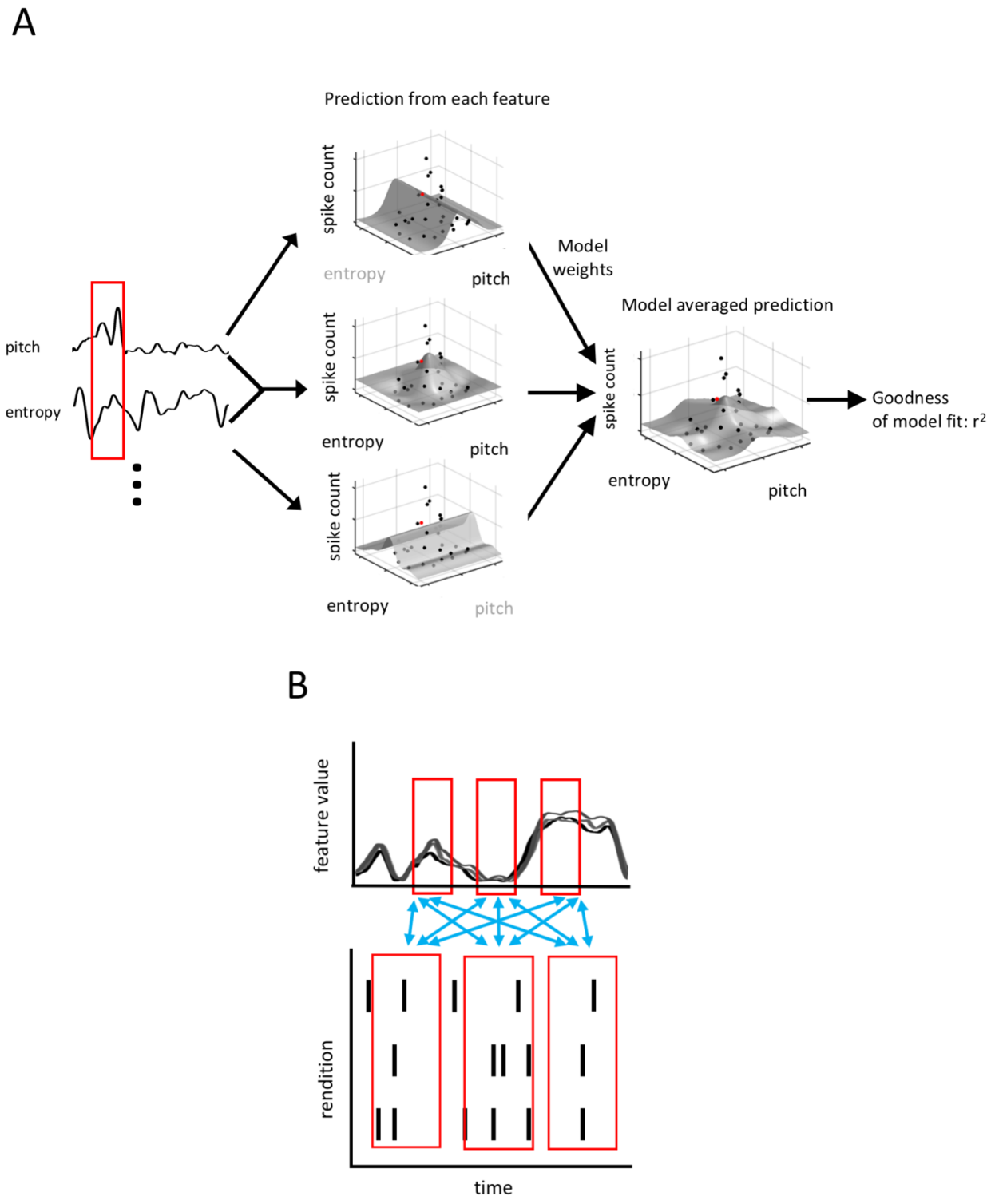

Figure 5.3 legend on following page.

Figure 5.3. Schematic of Gaussian Process model and fitting process. **A.** Schematic of fitting a single, multivariate model using a collection of song features. The multidimensional model takes a weighted average of the model predictions from every combination of our eight song features. The weights are determined by the log-likelihood of each feature combination and a penalty for more features. Here we show a schematic of fitting 2 features. The final model's goodness of fit is quantified by the r^2 value computed using leave-one-out cross validation. (See Methods and Appendix A for more details). **B.** Schematic of extending the song-spike count relationship to all parts of song and spike train latencies. The modeling technique laid out in panel a and Figure 5.2b is extended here across a range of song-spike latencies, building a matrix of r^2 values. The top panel shows a sliding window along the song. One example feature is shown here for reference; however, we use all eight features at each time point in song to build the GP model. The bottom panel shows the aligned spiking response for three example renditions in a raster plot. Spikes counts are binned in 100 ms windows.

These correlated activity patterns are sometimes are predictive instead of responsive– they precede the song event itself. Again, interpreting this cell activity in an RL theoretical framework, this could imply an intent to sing in a particular way that is functioning like an internalized cue, which draws the RPE signal into a predictive timing relationship with the song event it is evaluating. Many of these macroscopic relationships were found in the subset of VTA-error cells which did not project to Area X, suggesting that this level of song variation and evaluation may happen via another circuit.

The activity of the VTA-other neural population also correlates with fluctuations in natural song; however, the timing of this activity does not fit what would be expected from an RPE signal. Instead, the VTA-other population contains heterogeneous information about song and song fluctuations at diverse timing latencies. This diverse population of responses contain necessary components for calculating an RPE value; this suggests that some of the computations towards the hypothetical RPE signal sent to area X could take place within VTA.

Lastly, in the course of our analysis, we discovered novel behavioral fluctuations that are related across the entire song and correspond to long timescale changes in the overall firing rates of one subpopulation of VTA-other cells. Previously, these song-spanning, behavioral state changes have been identified in the transitions between directed and undirected song, when the male is singing in the presence of a female or alone, or across diurnal cycles. However, the song shifts we observed were all within a consistent, undirected song environment over short periods of time (2-15 minutes) and were correlated with the firing rates of the VTA-other subpopulation in a continuous manner (rather than the binary state change that is usually described in directed versus undirected behavior). This is a qualitatively different type of song

variability than described in learning: in most accounts of adult song variation, the song is varied independently syllable-to-syllable or at even finer time steps within song (15-17). We term this neural activity and the correlated behavioral changes, 'global state changes' and hypothesize that they arise from changes in broader state variables in the bird such as intent, arousal and motivation. This type of behavioral-neural global fluctuation suggests a modification to the existing RL theoretical account of the song bird circuit. In the current account, variations are independent and specific to particular time points within the song. This characterization is based on experiments which show that adult birds can learn syllable-specific changes to their song that do not influence other parts of the song (16-19), as well as the ultra-sparse firing patterns in HVC which are ideally suited to characterize song independently at each time step (20).

VTA receives convergent inputs from brain regions which relate to arousal and motivation and is well situated to incorporate this type of global information into the reward prediction error computation. Global state variations change the song output but may not emerge from the learning circuit-specific variations in song nor change the underlying error of the rendition. We hypothesize that the VTA-other global state cells could be enacting a type of gain-scaling or re-normalization on the error signal that accounts for global state variations and removes the confound for learning that they would introduce in the local, sensory signal.

In the following sections we present our analysis methods and results and then discuss interpretations and resulting experimental predictions.

Methods:

What is the general analysis problem we face? We wish to quantify non-stationary spiking responses to a time-varying sensory signal (song). We can't use traditional ideas of spike-triggered averages or spike-triggered covariances because we assume these cells' responses are inherently non-stationary. If the cells are encoding RPE-like responses to song fluctuations, the responses are specific to the contextual time-step in song and should differ according to the timing. In other words, an identical fluctuation at the beginning of the song should elicit a different response than at the middle.

Additionally, the relevant dimensions of the signal space should vary across song time-step contexts as well. Different parameterizations of the song should work better as low dimensional representations of error-relevant song variation at different song points. For example, pitch is often used as a low-dimensional parameterization of song at moments of harmonic stacks: at these moments, there are well-defined, discrete frequency harmonics, and pitch straightforwardly quantifies the frequency power structure. However, in regions of song where the power distribution of frequencies is broader, pitch is not well defined.

Lastly, we expect that an RPE tuning curve should also differ across song time-steps. Perhaps at one point in song, the bird is trying to maintain an existing manner of performance; at this instant, the tuning curve of an RPE-like signal should peak near either the mean or the mode of the song variants and be lowest at the edges of the song variation distribution. At another point in song, the bird could be trying to adjust their performance; at this instant, the

tuning curve of an RPE-like signal should peak at whatever shifted variant the bird aspires to but is not yet consistently producing: perhaps an outlier song variant receives the strongest feedback. Because the RPE signal is specific to time-step, we cannot average responses across a sliding window, only across song renditions.

Our analysis should be flexible enough to account for these time-varying aspects of the song signal and spiking response, but also impose enough constraint to reasonably assess the significance of what we find.

Methods Summary

We designed an analysis that identifies timing and shape signatures of a hypothetical RPE signal and compared these signatures in the VTA-error and VTA-other cell populations. We chose a low dimensional, time-varying representation of song based on established song parameterizations (Fig. 5.2a). We identified syllables across renditions and aligned each rendition at the syllable onset. We linearly warped spike timing and song so that all syllables of the same kind were the same duration. We binned spike counts in a sliding window for each syllable in every cell. We used a multi-dimensional regression model to ask whether local song feature fluctuations predicted spike counts from rendition-to-rendition (Fig. 5.2b and Fig 5.3a,b). We used a leave-one-out cross validation technique to assess how well variations in song features from rendition-to-rendition predicted spike counts in our model. We fit the model to many different song segment-spike bin latencies (Fig 5.3b) and looked at the latency

distribution for song-spike count pairs in which song features predicted spike counts over the population of cell responses.

To assess tuning curve shapes, we used the parameters of variants of a generalized linear model (GLM) fit between song feature fluctuations and spike counts to characterize the shapes of fits. We looked for whether tuning curve shapes had single peaks or were multi-peaked in instances when song features predicted spike counts via our model.

We assessed the significance of the relationship between song and spike count by randomizing the relationship between entire spike trains and song many times and re-fitting our model to the randomized data. In this way we preserved potential temporal relationships in spiking and song and addressed the significance problem of multiple fits. We computed individual p-values of song segment-spike count fits. We also computed population measures of significance across all fits in a syllable and across all fits in the cell population (either VTA-error or VTA-other).

Extended Methods

Parameterizing song

We parameterized song using non-linear features extracted from the song spectrogram. The song spectrogram is the discrete Fourier transform of the song sound wave within sliding temporal windows across song. The song features were computed within each temporal window from the frequency spectrum and vary across song. We used open source MATLAB

software, Sound Analysis Pro 2011 (SAP 2011), to assemble the spectrogram as well as to define and extract song features. SAP 2011 is a customized software package written to analyze animal communication and is originally and most frequently used to study bird song (21). We used an existing SAP feature set for our parameterization because these features have been used in many previous studies to link zebra finch song variations to neural activity or neuromodulator concentrations (22-24), to study variation in song over development (15, 18, 25), and to drive adult learning in DAF paradigms (17, 19, 26). Therefore, we can use this form of dimensionality reduction of song knowing in advance that these dimensions are relevant to song variation in other contexts[2]. The features extracted were Wiener entropy, pitch, goodness of pitch, amplitude, amplitude modulation (AM), frequency modulation (FM), mean frequency and aperiodicity. These features result in an eight-dimensional representation of song at each time-step. We further applied a moving-average filter (35 ms) to smooth the feature signals in time and sampled the smoothed value every 5 ms across song.

Aligning syllables across renditions

To compare song across renditions, syllables were classified using custom LabView code from the Goldberg lab (12). Clusters of unique syllables were labelled alphabetically as 'a', 'b',

[2] It would have been a reasonable alternative to attempt a more generic dimensionality reduction on the song spectrogram, such as principle component analysis (PCA). A PCA decomposition would extract linear combinations of orthogonal features and would not find relevant non-linear variations, which are known to exist and be relevant to learning. The features and feature combinations would be time-varying across song and would be difficult to interpret in relation to existing behavioral findings. However, this type of generic decomposition could in theory more successfully discover song variations across the full song extent by discovering relevant dimensions of variations at each time step from the full high-dimensional spectral structure.

‘c’ etc. depending on order within a rendition. The number of syllables each bird sings varies bird-to-bird from 3-7 syllables. We identified syllable onsets and offsets across renditions for every syllable set in which there were greater than 15 renditions of that syllable using either an amplitude threshold chosen to match the amplitude variance of that syllable or a derivative measure of the amplitude that found the first characteristic maximum or minimum, which identified the syllable border. The choice between using a magnitude threshold or a derivative measure was made based on the slope of the amplitude modulation at syllable onset or offset: for syllables in which there was a fast rise in amplitude at syllable onset or offset, a magnitude threshold was used; for syllables in which there was a gradual rise in amplitude at syllable onset or offset, the derivative measure was used. This decision was made by visual inspection of the syllable. All alignments were further checked by eye. Renditions in which alignment was ambiguous by eye were excluded from analysis.

All syllable types (i.e. ‘a’ or ‘b’ etc.) were isolated and aligned across renditions by syllable onset times. Individual syllable types have a stereotyped, characteristic duration; however, there is some variation of this duration from rendition-to-rendition. Variations in syllable duration across renditions could be part of what an RPE signal encodes; however, these timing misalignments depend on the location of the initial alignment point. Therefore, we restricted our analysis to local differences in feature values, rather than differences in timing. In order to make sure that minor differences in syllable lengths were not misaligning local syllable features at the later parts of the syllable, we linearly time-warped the feature wave forms of each syllable rendition such that they all lasted the median duration of that syllable type (27).

Parameterizing and aligning spiking activity

Spike sorting was performed offline by the Goldberg lab using custom MATLAB software (12). For every syllable, we considered the spike train +- 500 ms around the syllable onset. In order to align spiking activity, we first applied the same linear time-warping map to the spike train that we used to align syllables for each rendition (27). In all cases, we applied this map to the time window in which the syllable took place. When possible, we generated a piece-wise linear time warping map based on syllable boundaries in surrounding syllables across the entire motif. When other syllables were missing from the motif we allowed that region of the spike train to remain un-warped. In the time windows where there was no song with which to build a time warping map we left the spike train un-warped.

We binned spike counts within a sliding window (100 ms) across the 1000 ms length of spike train we considered for each syllable. We chose this spike count window based on the firing rate of the VTA-error neurons we considered (mean firing rate = 13 +- 5 Hz).

We recomputed our results without applying the linear time-warping to the song and spike train and found that this step was not necessary to maintain the qualitative aspects of our population results. This is because differences in timing from rendition to rendition are not large. However, because we were primarily interested in local song variation and not timing variations, we used the linear time warping in our final analysis.

Fitting spikes to song with a Gaussian process regression

We used a regression approach to determine if spike counts are related to the variation in song. The relationship between spike counts and song is likely non-linear and related to a variable number of features depending on the point in song. To address this, we used a non-parametric Gaussian process (GP) regression to fit the relationship between our 8 song features and spike counts within single time windows (Fig. 5.3a) (e.g. a song segment beginning 20 ms after syllable onset and a spike count window beginning 100 ms after syllable onset) (28).

There is a good deal of model uncertainty in this task: it is unclear which features to use at a given point in song and how many should be used. Furthermore, the prediction of the model depends heavily on which features are used. To address this uncertainty, we used a Bayesian model averaging approach to determine the predicted spike count wherein we integrated over all possible values of $\mathcal{M}$ and weighted their predictions according to their posterior probability given the observed spike counts (29). For a derivation and further elaboration of the GP model, see Appendix A.

To assess the success of the model in using song features to predict spike counts, we used a leave-one-out cross validation procedure to estimate the mean squared error for predicting new observations (30). We compared the GP model mean squared error to a null hypothesis in which spike counts were not related to song using an r^2 metric, which quantifies how the predictive song features are of spike counts in our model relative to the mean spike count over all renditions:

$$r^2 = 1 - \frac{MSE_{loo}^{(GP)}}{MSE_{loo}^{(null)}},$$

where the mean squared error, $MSE_{loo}^{(GP)}$, for the model prediction is,

$$MSE_{loo}^{(GP)} = \frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2.$$

The prediction of the i[th] left-out point from the regression using all other T-1 points is $\hat{y}_i$, and the actual value is y_i. The summation, T, is over all renditions. The mean squared error of the null hypothesis is:

$$MSE_{loo}^{(null)} = \frac{1}{T}\sum_{i=1}^{T}\left(y_i - \overline{y_{/i}}\right)^2$$

where,

$$\overline{y_{/i}} = \frac{1}{T-1}\sum_{j\neq i} y_j. \qquad (32)$$

For each cell, in every non-distorted syllable for which there were $N \geq 15$ renditions, we sampled the smoothed song features every 5 ms across the syllable and sampled spike counts in 100 ms windows every 10 ms across $\pm 500\ ms$ of spike train around syllable onset. We fit the multi-dimensional GP model across all spike bin-song segment pairs and generated song-spike relationships at many time latencies. We additionally fit the GP model to every feature individually. For all of these fits we computed the r^2 value.

We implemented this procedure over a limited parameter search of the spike bin window width (50, 100 and 150 ms) and the moving-average, smoothing filter we applied to the song features (25, 35, 45 ms). We found that our results were qualitatively the same except for when we used the 150 ms spike bin window, which was too broad and diminished the relationship between spike count and song fluctuations.

Characterizing tuning curves of cell responses

The GP model is flexible in that it will fit any relationship between the independent and dependent variables, easily scales to multiple dimensions and is computationally efficient. However, from the output of the model we have no easily interpretable means of characterizing the shape of the fit. In order to characterize the form of the spike-count to song relationships across the large number of fits we assessed, we needed an automated way to categorize the shapes of the tuning curves.

To do this, we used a generalized linear model (GLM) and a modification of the GLM called a generalized quadratic model (GQM) that introduces a quadratic transformation of the song features (31). A GLM consists of a linear stimulus filter, an invertible non-linearity (the link function) and a stochastic exponential non-linearity, such as a Poisson process:

$$y|\boldsymbol{x} \sim Poiss\big(f(\boldsymbol{w}^T\boldsymbol{x})\big) \qquad (1)$$

where here, y is the spike count, f is the inverse link function, **w** is the stimulus filter and $\boldsymbol{x}$ is the stimulus. In the GQM extension of the GLM, Eq. 1 becomes:

$$y|\boldsymbol{x} \sim Poiss\big(f(Q(\boldsymbol{x}))\big) \qquad (2)$$

$$Q(\boldsymbol{x}) = \boldsymbol{x}^T A\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c, \qquad (3)$$

where Q is a quadratic function of $\boldsymbol{x}$ with coefficients $A, \boldsymbol{b}, c$. We consider song features individually in this tuning curve analysis, so dim($\boldsymbol{x}$) = 1, and the quadratic coefficients become scalars (a,b,c). We take the link function to be an exponential and the noise process to be Poisson. Thus, we can fit the quadratic coefficients by maximizing the log-likelihood:

$$\log P(Y|X,a,b,c) = \sum_{i=1}^{N}[-\exp\left(Q(x_i)\right) + a * y_i {x_i}^2 + \ b * y_i x_i + c - \log\left(\Gamma(y_i+1)\right)], \qquad (4)$$

$$\Gamma(y) = (y-1)! \qquad (5)$$

We maximized the log-likelihood using the MATLAB function fminunc. The sign of the quadratic coefficient, a, of this model determines whether the data is better fit by an upwards-facing, quadratic basis in which the data is double-peaked, or a downwards-facing quadratic basis in which the data is single-peaked. We compare this model to an alternative model fit where the quadratic term is removed and the model is again of a traditional GLM form:

$$y|x \sim Poiss\big(f(L(x))\big) \qquad (6)$$

$$L(x) = b^T x + c. \qquad (7)$$

The log-likelihood of this model is:

$$\log P(Y|X,a,b,c) = \sum_{i=1}^{N} [-\exp\left(L(x_i)\right) + b\, y_i x_i + c - \log\left(\Gamma(y_i+1)\right)]. \qquad (8)$$

We compare the performance of the two models using the Akaike information criterion (AIC) (32). The AIC metric is defined as:

$$AIC = 2k - 2\ln\hat{\mathcal{L}}, \qquad (9)$$

$$\hat{\mathcal{L}} = argmax_{a,b,c}\ \log P(Y|X,a,b,c), \qquad (10)$$

where $\hat{\mathcal{L}}$ is the maximum of the log-likelihood function for a given model and k is the number of estimated parameters. This metric balances goodness of fit with model complexity. A lower AIC metric indicates better performance. Therefore, the difference in the AIC metrics of two models indicates the relative success of one model over another, taking into account differences in model complexity (32, 33). We can then ask, when the quadratic model is a better fit to the

data than the linear model, is the tuning curve relationship of spike counts to song features single peaked or double peaked? We predict that an RPE-like signal should be single peaked.

We compared the GQM and GLM models on all GP model fits with $r^2 > 0$ for all individual song features which had themselves predictive fits within the multi-dimensional model. We calculated the fraction of fits with the quadratic coefficient, a < 0, as a function of the AIC comparison between the two models:

$$\Delta AIC \equiv 2k_{linear} - 2\ln\hat{\mathcal{L}}_{linear} - 2k_{quad} + 2\ln\hat{\mathcal{L}}_{quad}. \qquad (11)$$

Significance testing

Assessing the significance of the model predictions must be done on a population level for this type of analysis. We generated model fits to thousands of spike count-song segment pairs for each syllable. Simply by chance, a portion of these fits would generate a predictive $r^2 > 0$ value.

Furthermore, spike-song pairs are correlated, not only because of overlapping spike counts and song segment windows, but also because of possible underlying correlations in the song and spike fluctuations across the song. To address this, we randomized the relationship between entire spike trains and song renditions and then re-performed our model fits on the randomized, spike count-song segment pairs across all time steps. By leaving the temporal structure of the song and spiking activity intact and only randomizing the relationship between them, we built a randomized population of fits for each cell-syllable pair, which retained the

unknown, underlying temporal structure possibly present in the spike trains and song (34). We repeated this procedure 500 times for the VTA-error cell population and 100 times for the VTA-other cell population to build a distribution of coherently randomized cell sets[3].

From this distribution of randomized cell sets, we computed not only single-tailed p-values assessments of the r^2 values of the individual spike count- song segment model fits but also on population measures of significance in the VTA-error and VTA-other cell populations independently. The population measures we assessed were:

(1) *The frequency of the predictive signal across the whole cell population.* An $r^2 > 0$ indicates the model predicts the data better than an estimate based solely on the mean spike count across renditions, and we call this a 'predictive signal'. We therefore assessed the significance of the total number of $r^2 > 0$ spike count-song segment fits within the populations of VTA-error and VTA-other syllable-cell pairs respectively with a single-tailed p-value test.

(2) *The spread of the predictive signal across the population.* We asked whether a small number of cells were accounting for the majority of the signal seen in (1) or if the signal appeared across multiple cells and syllables in the population. To answer this, we first labeled each cell-syllable pair as 'significant' if the number of positive r2 values within the RPE latency window (0-150 ms) had a single-tailed p-value < 0.05.

[3] We were unable to perform more bootstrapping trials on the data due to computational constraints because of the large number of fits this analysis required. Generating the current randomized populations for the VTA-error and VTA-other populations required over 2,000,000 individual model fits.

We then calculated the single-tailed p-value for the number of significant cell-syllable pairs across the entire cell population.

(3) *The significance of the magnitude of the peak in signal frequency within the RPE latency window of the latency distribution of all* $r^2 > 0$ *fits across the full cell population.* We compared the variance of the latency distributions of the randomized populations to the variance of the peak we found in the actual data. We computed the single-tailed p-value for the maximum fluctuation of a latency distribution at *any point* in the latency domain. In this way we tested the significance not only of finding a peak in the data at the RPE window but of finding a peak of that size anywhere in the latency distribution.

(4) *The significance of the shapes of tuning curves we find via our GLM parameterization technique.* We computed the single-tailed p-value of the fraction of single peaked tuning curves in the real population relative to the randomized populations.

Note that this population significance strategy allows us to assess the VTA-error cell activity as a population, but not the significance of particular song segment-spike count pairs. More data is needed for this level of confidence.

Removing the influence of correlations across syllables

For each cell recording, the number of renditions of each syllable varied. Some syllables were repeated multiple times within a motif; in other instances the bird truncated their song before reaching later syllables. Therefore, while we performed the coherent bootstrapping

procedure on each syllable-cell pair, we were unable to extend the same coherence across syllables. This left open the possibility for partial correlations across syllables to go unrecognized in our significance analysis. This could create a discrepancy between our true population measures and the randomized population measures.

To address this, we considered the significance of a population of randomly drawn subsets of our data in which we considered only one syllable for each cell (Fig. 5.4a). In this way we generated subpopulations of the cell-syllable pairs that did not have any underlying correlation structure that wouldn't also be present in a randomized population. We sub-sampled 5,000 times from the full population of cell-syllable pairs. For each subsample, we computed two of our population measures of cell response: the total number of positive r2 values within the RPE latency window across all cell-syllable pairs, and the number of cell-syllable pairs that had a significant number of positive r2 values within the RPE latency window (i.e., the total magnitude of the signal and the spread of the signal across cells). We performed the same structured subsampling on our 500 randomized sets to compute p-values for each of the subsample's population measures of signal response. The resulting p-value distribution of the subsamples of the data is shown in Figure 5.4 panels B and C. Because we are subsampling from a population with a moderate signal, some subsamples naturally miss the significant signal in the data altogether, resulting in high p-values. Conversely, some subsamples pick up an unusually high portion of the signal, resulting in very low p-values. What we wish to know is: how shifted is the significance distribution of the subsamples towards low, significant p-values in comparison to chance?

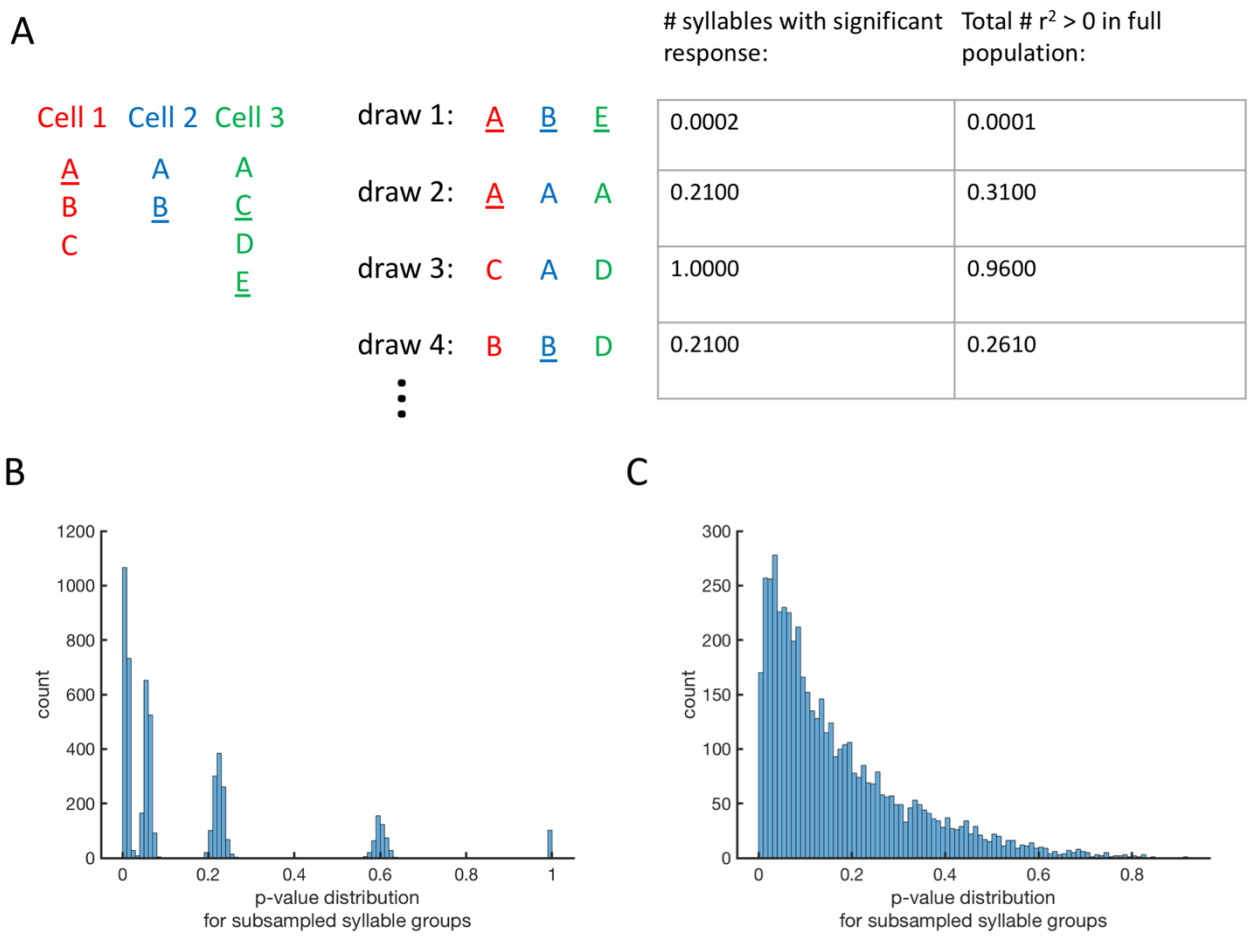

Figure 5.4. Results from sub-sampling cell-syllable pairs within the VTA-error cell population. **A.** Schematic of subsampling scheme (cartoon with 3 cells from 3 different birds). For every cell we draw a single syllable randomly from the motif to include in the sub-population. In the figure underlined syllables have a significant RPE-like signal present. For each drawn subpopulation we generate two population metrics of significance and compute their p-values via a one-tailed p-value test: the number of syllables with a significant RPE-like signal and the total number model fits with latencies within the RPE window with $r^2 > 0$ in the full subpopulation. We draw 5,000 subpopulations. This cartoon example shows 4 draws from 3 different birds' cells and the subsequent p-values of the population measures in the table. **B.** The p-value distribution over 5,000 subpopulations of the VTA-error cell for the number of significant syllables in the subpopulation. The discretized nature of the distribution arises from the discrete numbers of possible significant syllables that can be found via this method. The median p-value is 0.056 and the mean p-value is 0.148. **C.** The p-value distribution over 5,000 subpopulations of the VTA-error cell of the number of fits with $r^2 > 0$ in each subpopulation. The median p-value is 0.1184 and the mean p-value is 0.1688.

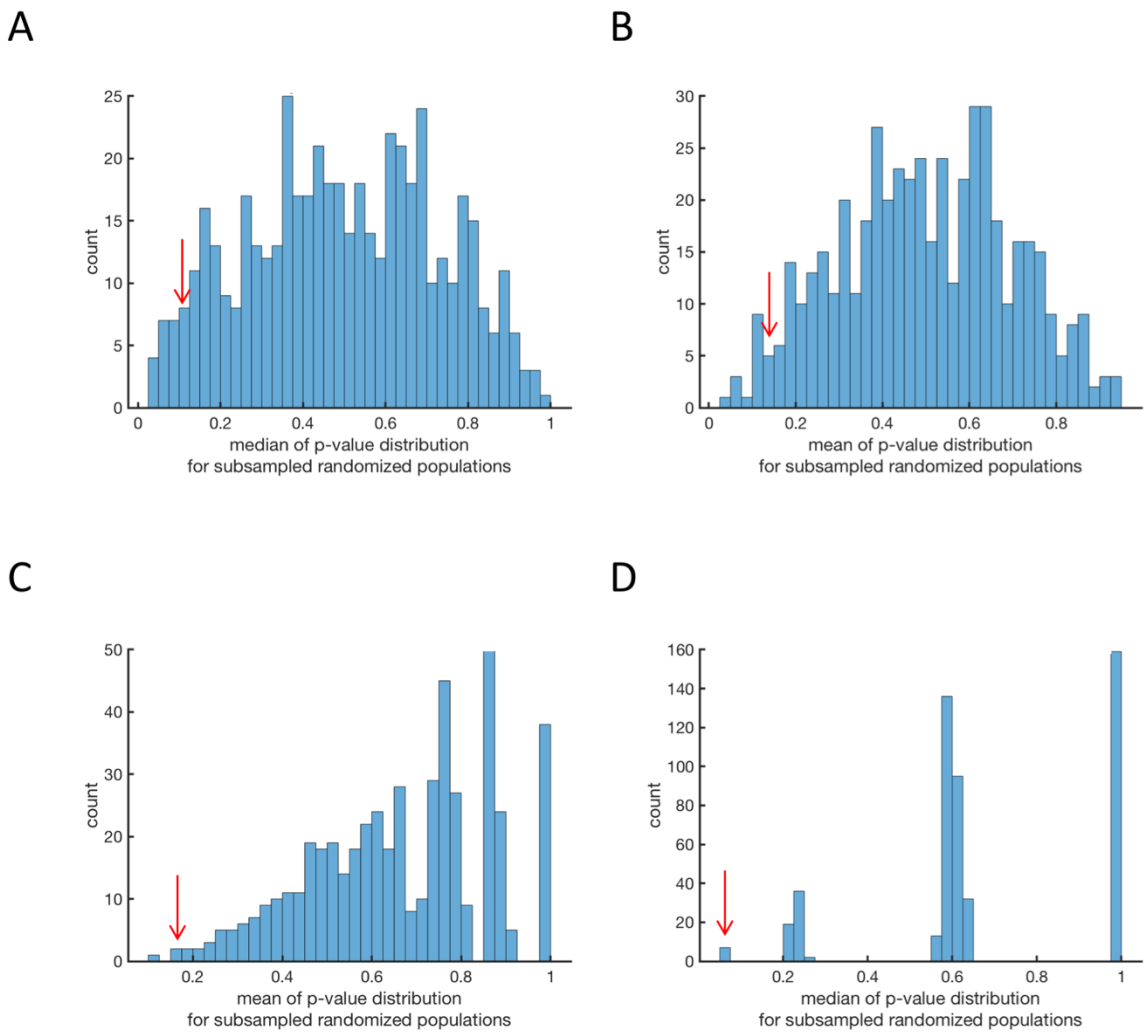

Figure 5.5. Randomized distributions of subpopulation metric statistics. The red arrows show the values in the actual data. **A.** Median p-value distribution of the total number of fits with $r^2>0$ for the subsampled, randomized populations. From this distribution, the one-tailed p-value of the true median is 0.048. **B.** Mean p-value distribution of the total number of fits with $r^2>0$ for the subsampled, randomized populations. From this distribution, the one-tailed p-value of the true mean is 0.046. **C.** Mean p-value distribution of the number of cell-syllable pairs with a significant RPE-like signal for the subsampled, randomized populations. From this distribution, the one-tailed p-value of the true mean is 0.002. **D.** Median p-value distribution of the number of cell-syllable pairs with a significant RPE-like signal for the subsampled, randomized populations. From this distribution, the one-tailed p-value of the true median is 0.002.

To answer this, we compared the mean and median of the p-value distribution in the real data to the means and medians of p-value distributions of the randomized subsample sets (Fig. 5.5). In both population measures, the data retained significance. Therefore, we are confident that the VTA-error cell population has a significant response to song fluctuations.

Results

VTA-error cells: relationship to local, fine-timescale song fluctuations. We find that the VTA-error cell population contains correlations to song fluctuations that display signatures consistent with an RPE-like signal and are significant at the population level. We applied our GP model to song feature-spike count pairs across every naturalistic syllable for each cell at 5 ms sliding increments over song and 10 ms sliding increments over the spike train (song feature moving average filter = 35 ms; spike count binning window = 100 ms).

Figure 5.6 shows one example syllable and cell activity pattern. In Figure 5.6 panel A, the middle heat map, positive r^2 values indicate that song features predict spike counts in the GP model and are coded in color. Negative r^2 values indicate no predictive relationship and are coded in a grey scale. The pink parallelogram indicates the time latency region that is consistent with an RPE response (RPE response latency is defined as a 0: 150 ms latency between spiking activity and song feature fluctuation. This latency range was chosen to cover the range in timing

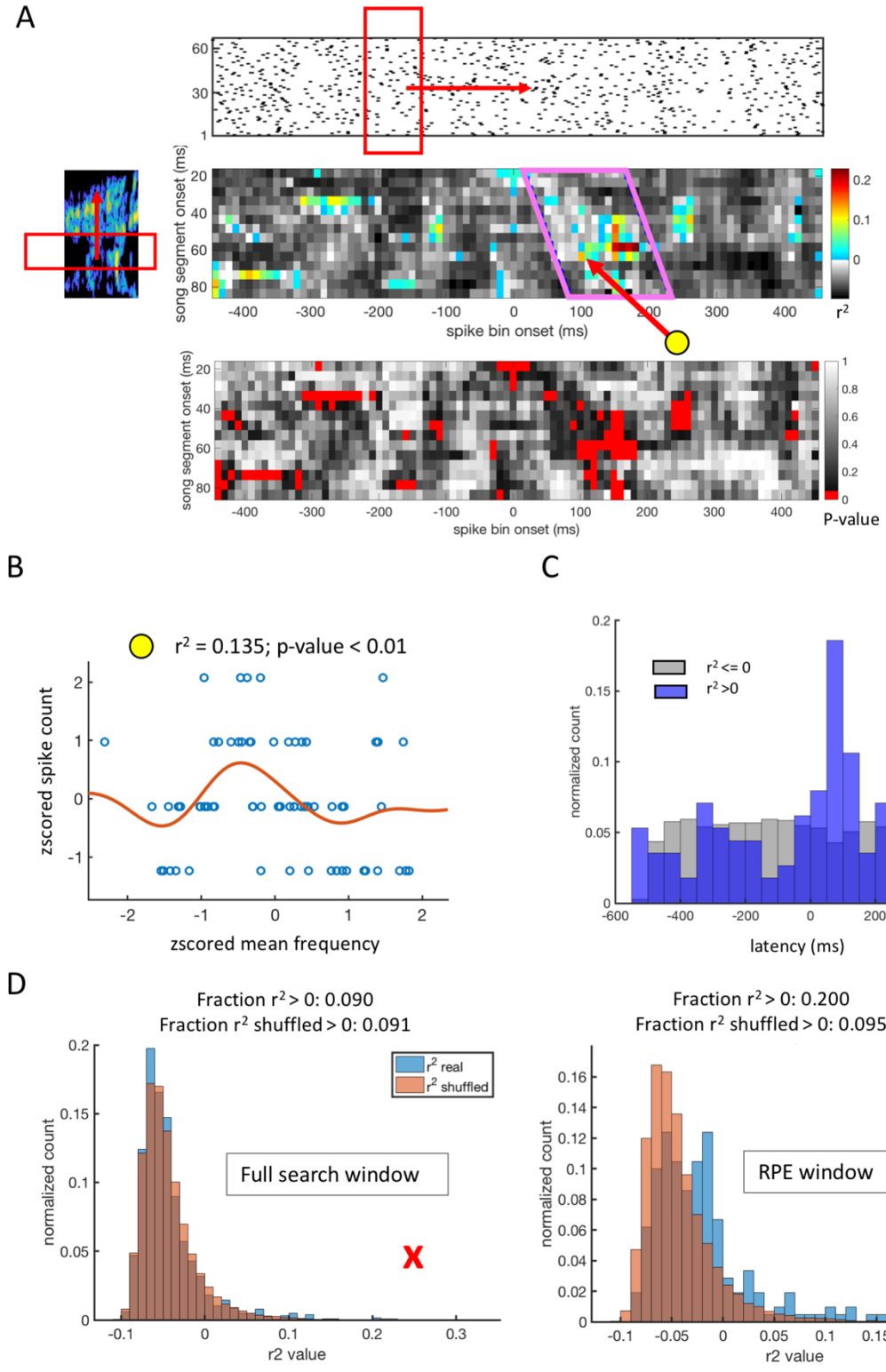

Figure 5.6. See next page for legend entries.

Figure 5.6. Single cell example from the VTA-error cell population. **A**. Top: raster of cell activity for each rendition aligned to syllable onset time. Left: spectrogram of example syllable. Time is aligned to syllable onset as in the raster. Middle heat map: heat map of r^2 values for fitted relationship between binned spike count and local feature averages over renditions across many latencies. $R^2 > 0$ indicates a predictive relationship. The x-axis is the time point of the beginning of the spike count bin and is aligned to the raster above, and the y-axis is the time point of the beginning of the local song segment and is aligned to the song spectrogram to the left. Time t=0 is aligned to the syllable onset on both axes. The pink parallelogram indicates the region where the latency fits the hypothesized response for a reward prediction error (RPE) (0:150 ms). The arrow with the yellow dot points to the spike-song relationship that is illustrated for a single song feature in panel b. The lower heat map reports the p-values for each corresponding pixel in the middle heat map of r^2 values. P-values < 0.05 are colored red. All other p-values are indicated on a grey scale. **B**. Example fit between one local feature average and spike counts for cell example in panel a. Each point represents a single rendition. **C**. Histogram of latencies for predictive fits in panel a. The blue histogram is for latencies with $r^2>0$ and the grey histogram is for fits with $r^2<=0$. Note there are many more predictive relationships within the expected RPE window for this example. **D.** r^2 distributions for randomized and actual data. Left: r^2 distributions for the shuffled and real data compared across all latencies. The real and shuffled distributions appear quite similar. The number of $r^2 > 0$ in the real data is not significantly different from what would be expected by chance. Right: the distribution of r^2 values that fall within the RPE latency window ((0, 150] ms) compared to the randomized distribution from within this same latency range. This distribution is shifted away from the randomized distribution, with many more $r^2 > 0$. This population has a significantly greater number of $r^2 > 0$ than expected by chance (p-value = 0.02).

relationships found in Gadagkar et al., 2016 (12)). Latency is defined as the time difference between the midpoints of the spike bin window and the song segment window[4].

The heat map at the bottom of Figure 5.6 panel A shows the significance level for each r^2 value individually using a single-tailed p-value test. P-values < 0.05 are colored in red. Larger p-values are grey scale. Figure 5.6 panel B shows one example fit to an individual song feature for the song segment-spike count pair indicated by the yellow dot in panel A. Figure 5.6 panel C shows the latency distributions for all spike count-song segment pairs with $r^2 > 0$ and all $r^2 \leq 0$ for the cell-syllable shown in panel A. There is a prominent peak within the RPE latency window in the number of spike count-song segment pairs for the predictive $r^2 > 0$ distribution that does not exist in the $r^2 \leq 0$ distribution.

The left hand plot in Figure 5.6 panel D shows the distributions of r2 values for all the points in the r2 heat map in Figure 5.6 panel A compared to a randomized distribution built from 500 randomized constructions of the r^2 matrix[5]. These distributions are quite similar. However, the right hand plot of Figure 5.6 panel D shows the real and randomized r^2 distributions for all model fits within the RPE latency window. The real data now has a visibly shifted r^2 distribution: there are more positive r^2 values than the chance distribution. We quantify this shift by comparing the total number of positive r^2 in the real and randomized distributions. This population has a significantly greater number of $r^2 > 0$ fits than expected by chance (p-value = 0.02).

[4] $latency = \left[onset^{spike\ bin} + \frac{offset^{spike\ bin} - onset^{spike\ bin}}{2}\right] - \left[onset^{song} + \frac{offset^{song} - onset^{song}}{2}\right]$

[5] P-values in Figure 5.6 panel A were computed using a single-tailed p-value test on this distribution.

Across the entire VTA-error population (N = 18), within the RPE latency window, there is a significant number of song-spike count pairs for which song features predict spike counts in the model ($r^2 > 0$), and the predictive fits are distributed across multiple cells within the population. Across all VTA-error cells, there are 1173 model fits to song-spike count pairs with $r^2 > 0$ (p-value < 0.01) within the RPE latency window (Fig. 5.7a). Within the RPE latency window, a significant number of model fits with $r^2 > 0$ is present in 10 out of the 53 cell-syllable sets that we analyzed (p-value < 0.01), and in at least one syllable in 6 of the 18 VTA-error cells (Fig. 5.7b).

In addition to there being a significant predictive signal between song feature and spike counts, the distribution of the timing of these predictive signals across the VTA-error population matches what we expect from an RPE-like response (Fig. 5.7c). In Figure 5.7 panel C, the blue line is the distribution of latencies across all VTA-error cells of spike count- song segment pairs in which song features predict spike counts ($r^2 > 0$). The black line is the mean of the same latency distributions across all randomized population draws (number of randomized, population draws = 500). The grey band is one standard deviation of these distributions at each discrete time bin across these distributions' domain. The true data has a large peak within the expected RPE latency region (3.74 std from mean). The p-value of this variance in relation to the randomized latency distributions anywhere in the latency domain is < 0.01. From this result, we find that spike counts are most predictive of song feature fluctuations in the latency window 0 to 100 ms after the song fluctuation occurs. This timing is consistent with the timing of an RPE signal.

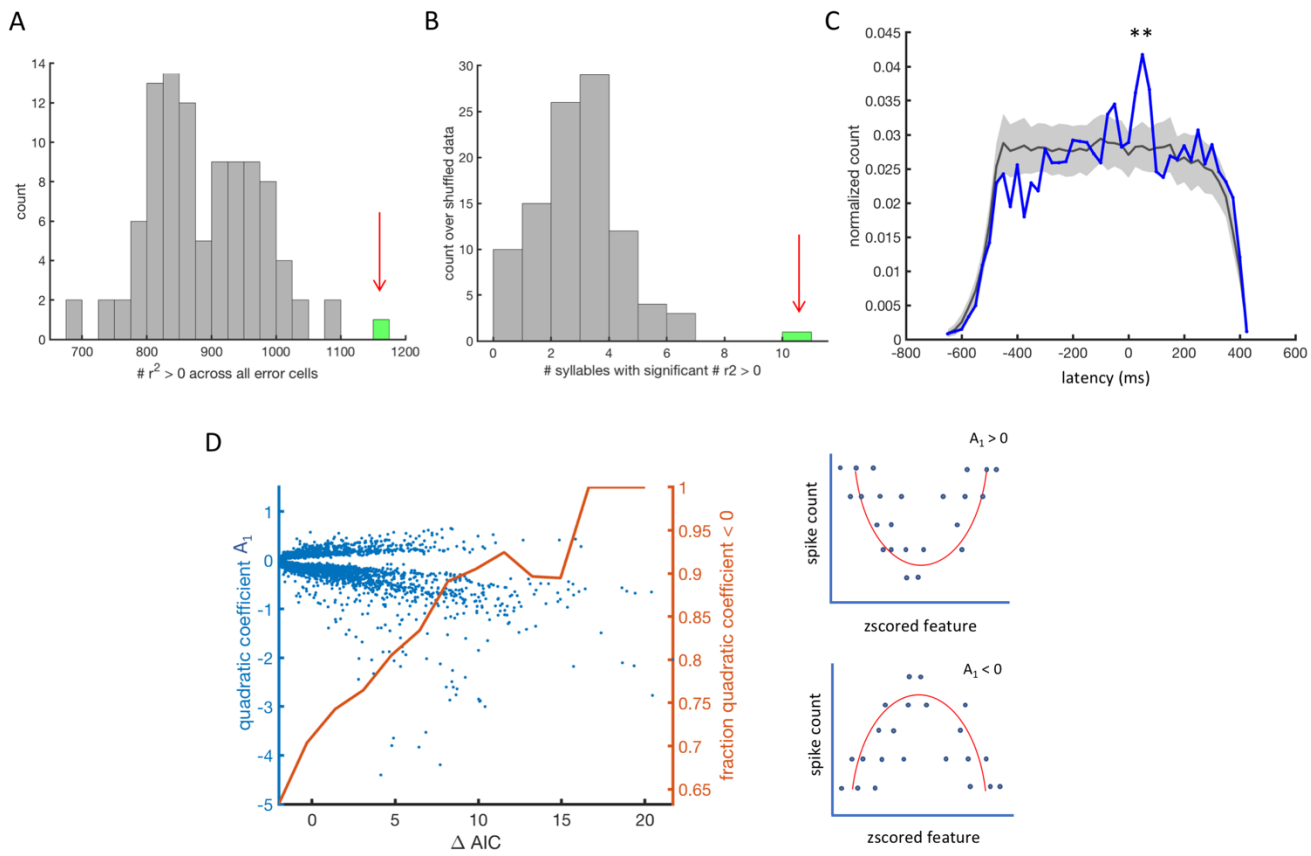

Figure 5.7. Population measures of the VTA-error cell population. **A.** Randomized distribution of total number of model fits to spike count- song segment pairs within the RPE latency window (0: 150 ms) for which there is a predictive relationship ($r^2 > 0$) compared to the actual data. The real population count is indicated in green. The p-value from this one-tailed test is < 0.01. **B.** Randomized distribution of the total number of significant cell-syllables. Cell-syllables are termed significant when there are a significant number of predictive model fits within the RPE latency window. The true syllable count is indicated in green. The p-value from this one-tailed test is < 0.01. **C.** VTA-error population latency distribution. The blue line is the distribution of latencies across all VTA-error cells of spike count- song segment pairs in which song features predict spike counts ($r^2 > 0$). The black line shows the mean of the same latency distributions for across all randomized population draws (number of randomized, population draws = 500). The grey band is one standard deviation of these distributions. The true data has a large peak within the expected RPE latency region. This peak is 3.84 standard deviations from the mean. The ** indicates a p-value of this variance in relation to the maximum variance in randomized latency distributions less than 0.01. The time-bin of this latency distribution is 25 ms. We tested this result over varying time bin widths (50, 75, 100 ms) and confirmed that the significance of

(continued Fig. 5.7) the result does not depend strongly on the choice of time bin. **D.** Quantification of tuning curve shape for cell responses to song fluctuations. Here we re-fit all spike-song fluctuation relationships with a GLM and GQM (see Methods) using up to linear and up to quadratic bases. If the first quadratic coefficient is negative, the fit is downwards facing and has a single peak (schematic bottom-right). If the first coefficient is positive, the fit is upwards facing and has two peaks (schematic top-right). The Δ AIC metric compares the relative success of the linear versus quadratic model. Δ AIC values > 0 indicate the quadratic model performs better than the linear model. Left: each point represents the fit to a single feature with an $r^2>0$ within a multi-dimensional model fit with an $r^2>0$. The predictive fits from the GP model have more single-peaked tuning curves than double peaked when their shape is better characterized as quadratic rather than linear, as is expected for an RPE signal. This fraction of single peaked tuning curves is significant (p-value < 0.01).

Lastly, we quantified the tuning curve shapes of cell activity patterns that were predicted by song feature fluctuations. From our hypothesis that the VTA-error cells are responding to song fluctuations with an RPE-like signal, we can make further predictions about the forms of spike count relationships to song. We expect that RPE tuning curves should vary in shape. Perhaps at one point in song, the bird is trying to maintain an existing manner of performance; at this instant, an RPE-like signal should peak at either the mean or the mode of the song variants' distribution. At another point in song, the bird could be trying to adjust their performance; at this instant, an RPE-like signal should peak at whatever shifted variant the bird aspires to, but is not yet consistently producing: if the goal is at the edge of the distribution of syllables, then the spike count to song relationship would be monotonic; if the attempted song adjustment is more modest, then the peak in spike counts would simply be shifted away from the center of the song variant distribution. However, we do not expect an RPE-like signals to have multiple maxima: we assume there is a single 'best' version of the song at any one point in time[6].

In this tuning curve shape analysis we considered the individual song features within the subset of GP model fits for which $r^2 > 0$, which themselves individually predicted spike counts ($r^2 > 0$ for the 1D feature fit). We chose this select subset of spike count-song feature pairs because we are interested only in the shapes of the tuning curves that we believe might actually carry information about song. We re-fit all spike-song fluctuation relationships with a generalized linear model (GLM) and an extension of the GLM, a generalized quadratic model

[6] This additionally assumes that the song parameterization we've chosen maps in an approximately 1:1 manner to whatever internal representation the bird actually uses to assess their own song.

(GQM) (see Methods) using up to linear and up to quadratic bases. We chose these models for this analysis because the model parameters can be used to quantify aspects of the tuning curve shapes. Within the GQM, if the first quadratic coefficient of the GQM is negative, the model fit is downwards facing and the data have a single peak (Fig. 5.7d lower right). If the first coefficient is positive, the model fit is upwards facing and the data have two peaks (Fig. 5.7d upper right).

The ΔAIC metric compares the relative success of the linear versus quadratic model. ΔAIC values > 0 indicate the quadratic model performs better than the linear model, taking into account both the likelihood of the model fit and the complexity of the model used. In the left panel of Figure 5.7d, each point represents a fit to a single feature with an $r^2>0$ within a multi-dimensional model fit with an $r^2>0$. The orange line plots the fraction of fits in which the quadratic coefficient (A_1) is negative for all fits with an ΔAIC greater than the abscissa value (i.e. a cumulative function where the accumulation is taken for all points greater than, rather than less than the x value). As the quadratic model does increasingly better than the linear model, a greater fraction of fits are single peaked. Thus, the predictive fits from the GP model have more single-peaked tuning curves than double peaked when their shape is better characterized as quadratic rather than linear, as we expect for an RPE signal. This fraction of single peaked tuning curves is significant (2-tailed z-test: p-value < 0.02). This is consistent with our hypothesis that an RPE signal should respond most strongly to a single best performance of song.

The ΔAIC measure also allows us to examine the fraction of tuning curves that are better fit by a linear versus quadratic model. The VTA-error population did not differ from chance in this fraction metric (fraction fits with ΔAIC > 0 = 0.43; 2-tailed z-test: p-value = .34).

This is also consistent with our hypothesis that an RPE signal should have both monotonic responses and single-peaked responses depending on the current level of song error.

VTA-error cells: relationship to macroscopic song fluctuations. In addition to the local, fine time scale song fluctuation correlations reported in the previous section, a subset (N=6/18) of VTA-error cells responded to either the presence or absence of entire syllables in particular renditions or how many times a single syllable was repeated (Fig 5.8a). We term these song fluctuations "macroscopic song fluctuations". Surprisingly, the latencies of these correlated activity patterns were both positive and negative: in three instances, the cell response preceded the song event with which it covaried.

Figure 5.8 panel B shows one example of this type of response: in a spike count window 300 ms before syllable 'd', spike counts predict whether syllable 'd' will be sung on a particular rendition or if the song will stop early after syllable 'c'. Figure 5.8 panel C shows the average spike count response 300 ms before syllable 'd' categorized by whether the song continues after syllable 'c' or ends at syllable 'c'. This cell's response in this early time window is strongly predictive of how long the rendition will last (N = 30 renditions).

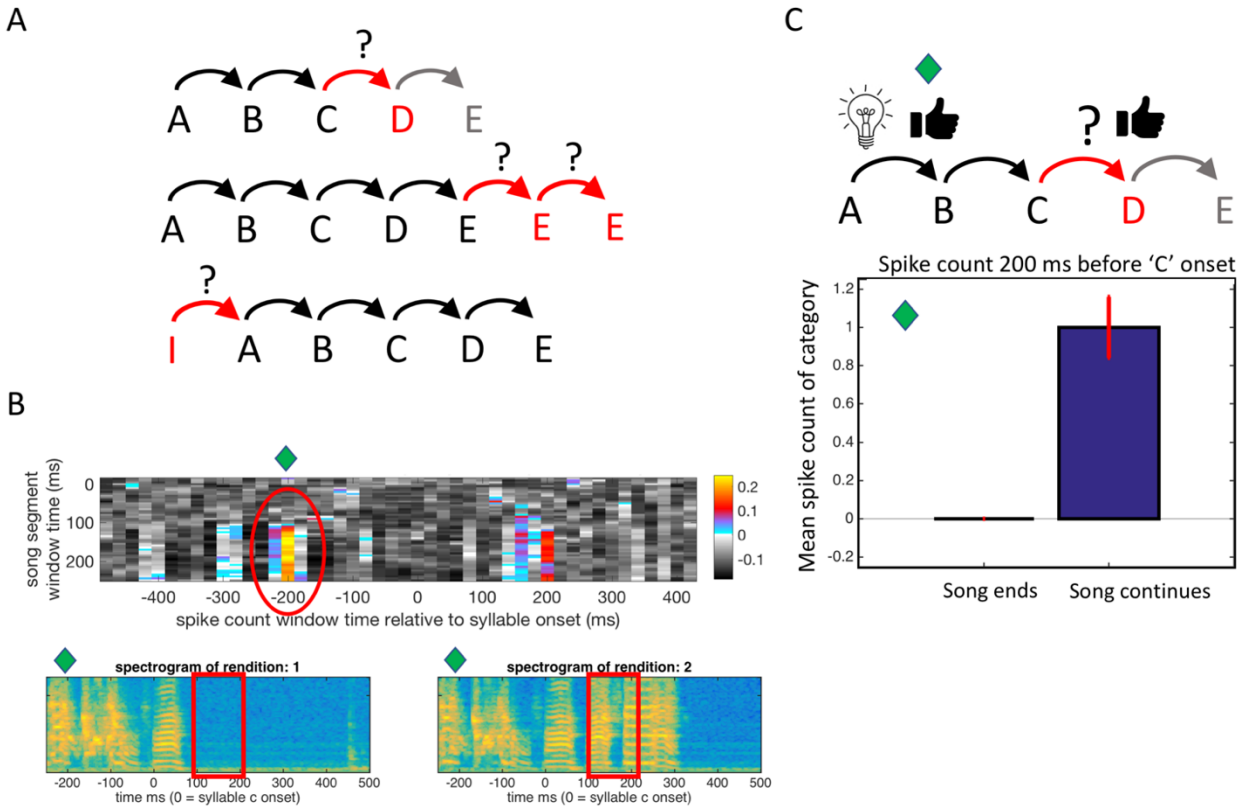

Figure 5.8. Macroscopic song fluctuation relationship. **A.** Schematics of example macroscopic cell responses seen in the VTA-error cell set. The red transitions indicate parts of the song which do not appear in all of the renditions and are coded for by at least one error cell. These types of macro variations and responses are in at least 6 of the 18 error cells analyzed. The letter, 'I' indicates an introductory syllable. **B.** One example of spike counts predicting song activity. Top panel: the r^2 array as in Figure 5.7a with time aligned to syllable 'c' onset. $r^2>0$ are coded in color. The activity highlighted by the red circle shows a strong relationship between spike counts in the -200 ms time bin and activity 100-200 ms after syllable 'c' onset. Bottom panels: two example spectrograms of renditions from this example. In one rendition the song continues after syllable 'c' and in the other rendition the song ends. **C.** Continuation of example in panel b. Spike count in window 200 ms before syllable 'c' strongly predicts whether song will continue after syllable 'c'. Bottom panel: bar plot of average spike count in the -200 ms spike count window partitioned by whether the song continues or ends. Red bars indicate standard deviation of spike count responses in each category. Top panel: schematic of activity within an RPE interpretation. This predictive cell activity suggests the presence of an internal 'cue' that signals the bird's intention to sing a long or short rendition of their song.

VTA-other cells: relationship to local, fine-timescale song fluctuations. We repeated our analysis on the VTA-other cell population. This population does not show a significant peak in activity in the spike-song latency consistent with an RPE signal. However, there are strong, varied types of song-spike relationships within this population that together could contribute to the final error calculation.

Across the VTA-other population (N = 24), there is a significant number of song-spike count pairs for which song features predict spike counts in the model ($r^2 > 0$), and the predictive fits are distributed across multiple cells within the population. Across all VTA-other cells, there are 4523 model fits to song-spike count pairs with $r^2 > 0$ (p-value < 0.01) within the RPE latency window (Fig 5.9a). Within the RPE latency window, a significant number of model fits with $r^2 > 0$ is present in 27 out of the 74 cell-syllable sets that we analyzed (p-value < 0.01) (Fig 5.9b). However, at the population level, spiking responses to song fluctuations in the VTA-other cells are not strongly clustered at an RPE latency (Fig 5.9c). The largest positive deviation from the randomized mean was +2.30 std and was not significant (one-tailed test: p-value = 0.33). Therefore, although there are complex, strong responses in the VTA-other cells, as a population, they do not have the same RPE latency signature that we find in the VTA-error population. This serves as an additional control on the VTA-error population's timing profile.

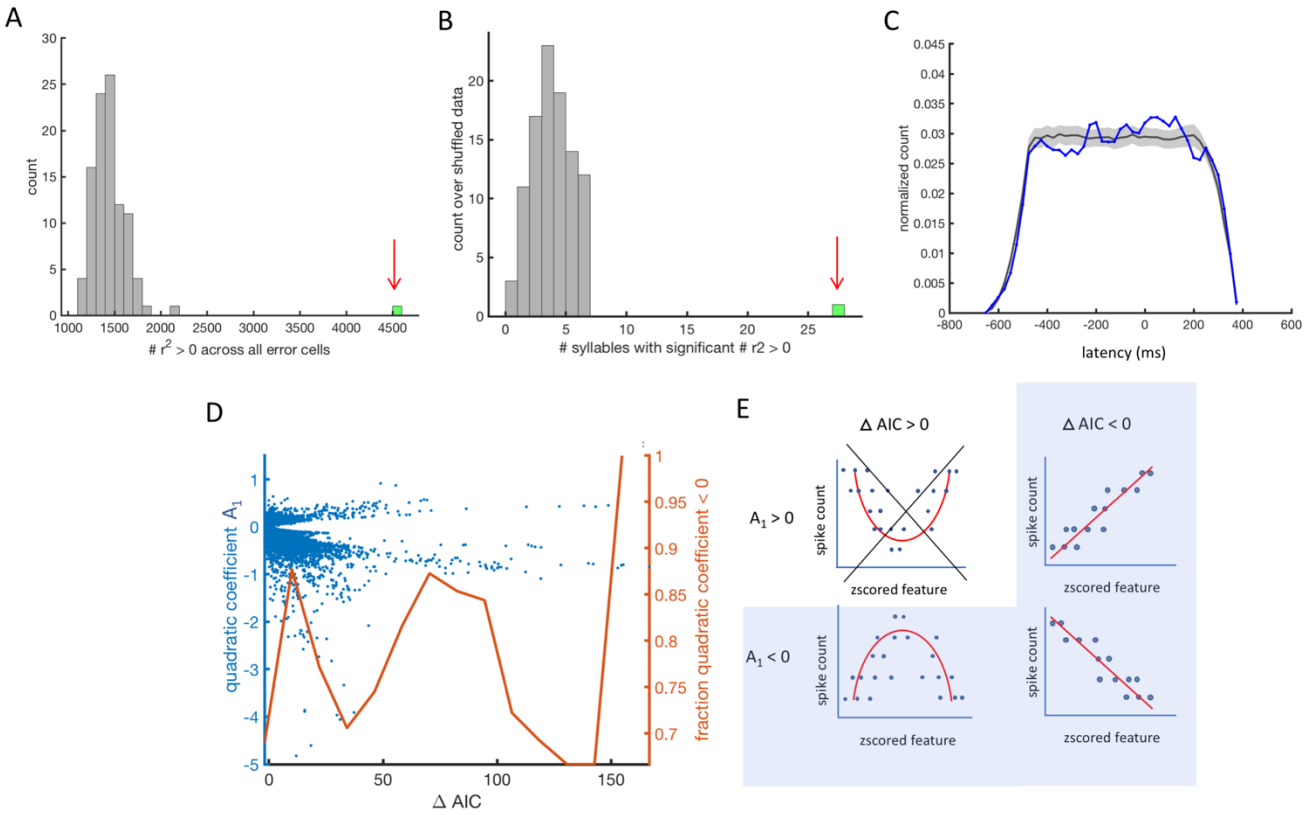

Figure 5.9. Population measures of the VTA-other cell population. **A.** Randomized distribution of total number of model fits to spike count- song segment pairs within the RPE latency window for which there is a predictive relationship ($r^2 > 0$) compared to the actual data. The real population count is indicated in green. The p-value from this one-tailed test is < 0.01. **B.** Randomized distribution of the total number of significant cell-syllables. Cell-syllables are termed significant when there are a significant number of predictive model fits within the RPE latency window. The true syllable count is indicated in green. The p-value from this one-tailed test is < 0.01. We show the RPE window for consistency with the VTA-error population. However, note from panel C, that the RPE latency window does not denote a region of unusual signal strength in this population. **C.** VTA-other population latency distribution. The blue line is the distribution of latencies across all VTA-other cells of spike count- song segment pairs in which song features predict spike counts ($r^2 > 0$). The black line shows the mean of the same latency distributions for across all randomized population draws (number of randomized, population draws = 100). The grey band is one standard deviation of these distributions. This peak is 2.30 standard deviations from the mean. This variance is not significant (one-sided z-test: p-value = 0.21). The time-bin of this latency distribution is 25 ms. We tested this result over varying time bin widths (50, 75, 100 ms) and confirmed that the significance of the result does not depend strongly on the choice of time bin. **D.** Quantification of tuning curve shape for cell responses to song fluctuations. Here we re-fit all spike-song fluctuation relationships with a

(Continued Fig. 5.9) GLM and GQM (see Methods) using up to linear and up to quadratic bases. The Δ AIC metric compares the relative success of the linear versus quadratic model. Δ AIC values > 0 indicate the quadratic model performs better than the linear model. Left: each point represents the fit to a single feature with an $r^2>0$ within a multi-dimensional model fit with an $r^2>0$. The predictive fits from the GP model have more single-peaked tuning curves than double peaked when their shape is better characterized as quadratic rather than linear. This fraction of single peaked tuning curves is significant (two-tailed z-test: p-value < 0.02). Additionally, there is significantly high fraction of fits which are better fit by an up-to linear basis set rather than an up-to quadratic basis set (two-tailed z-test: p-value < 0.02). Right: schematic summarizing types of tuning curves which are over-represented in the VTA-other population: linear fits are more common; when fits are quadratic, they are more likely to represent single peaks rather than double peaks in the data.

The tuning curves of the VTA-other cells similarly were more likely to be single peaked rather than double-peaked when the data were better characterized with the addition of a quadratic basis in the models used to fit the tuning curves of single features (2-tailed z-test: p-value < 0.02) (same procedure as in VTA-error cells; see Methods.) (Fig 5.9d: left plot). In addition, they were more likely to be better fit by the linear basis (GLM) model (fraction fits with ΔAIC > 0 = 0.35; 2-tailed z-test: p-value = 0.03) (schematized in Fig 5.9d: right hand plots). In this respect, they differed from the VTA-error cell population.

Figure 5.10 panel A shows one example syllable and cell relationship from the VTA-other population. The format is the same as in Figure 5.6. Note that in this example song fluctuations predict spike counts across broad regions of the spike train and song (Fig 5.10 a,c). Furthermore, song features very strongly predict spike counts (Fig 5.10 a,b). Figure 5.10 panel B shows one example fit to a single song feature. The strength of this fit is representative of many other song-spike pairs and song feature types in this example. Figure 5.10 panel C plots the latency distribution of the r^2>0 activity: it is broadly distributed across time with no preferred latency between spike counts and song features. This is qualitatively different than the previous VTA-error cell example wherein correlations were more localized in time. Furthermore, from the raster plot in panel A, it is clear that there is a broad state transition in firing rate that corresponds to shifts in song. We label these types of song-cell relationships 'global state' relationships since the long-timescale firing rate of these cells across a broad time window capture variations across the entire song.

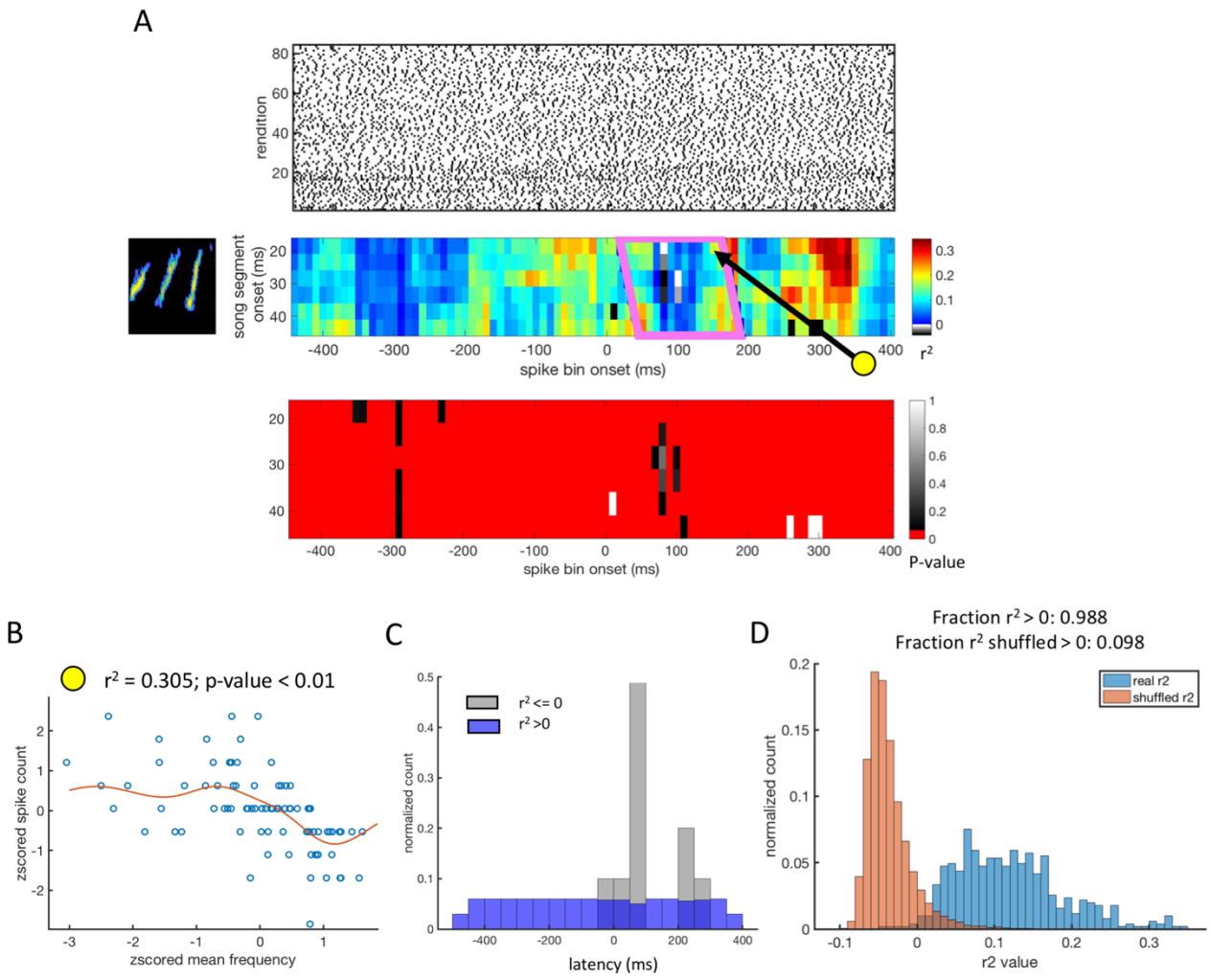

Figure 5.10. VTA-other cell example (1) of a 'global state' cell. **A.** Top: raster of cell activity for each rendition aligned to syllable onset time. Left: spectrogram of example syllable. Time is aligned to syllable onset as in the raster. Middle heat map: heat map of r^2 values for fitted relationship between binned spike count and local feature averages over renditions across many latencies. $R^2 > 0$ indicates a predictive relationship. The x-axis is the time point of the beginning of the spike count bin and is aligned to the raster above, and the y-axis is the time point of the beginning of the local song segment and is aligned to the song spectrogram to the left. Time t=0 is aligned to the syllable onset on both axes. Pink parallelogram indicates the region where the latency fits the hypothesized response for a reward prediction error (RPE) (0:150 ms). (Note that the RPE window clearly does not have significance here; the window indicator is for reference to the VTA-error example). The arrow with the yellow dot points to the spike-song relationship that is illustrated for a single song feature in panel b. The lower heat map reports the p-values for each corresponding pixel in the middle heat map of r^2 values. P-values < 0.05 are colored red. All other p-values are indicated on a grey scale. **B**. Example fit

(continued Fig. 5.10) between one local feature average and spike counts for cell example in panel a. Each point represents a single rendition. **C**. Histogram of latencies for predictive fits in panel a. The blue histogram is for latencies with $r^2>0$ and the grey histogram is for fits with $r^2<=0$. **D.** r^2 distributions for randomized and actual data. r^2 distributions for the shuffled and real data compared across all latencies. This population has a significantly greater number of $r^2 > 0$ than expected by chance (p-value < 0.01).

Figure 5.11 shows another example of a global state response. This cell contains both global state information and timing information: the long-timescale firing rate of the cell strongly predicts variations in song across the entire song (Fig 5.11a,b,c); in addition, there are stereotyped spiking events that happen across all renditions at the same time point in song (raster plot in Fig 5.11a). We find global state responses in 5/24 VTA-other cells.

In addition to the global state relationships, there were other types of song-related, cell activity in the VTA-other population. Figure 5.12 shows a predictive relationship that we label a 'motor plan' because the cell covariation with song precedes the song fluctuation. The latency distribution for this cell strongly peaks within -100 to 0 ms (Fig. 5.12b). Figure 5.13 shows activity that we term a 'clock' because of the precise locking to the time point in song (Fig 5.13a,b). In this example, there are very few spike count-song segment pairs in which spike counts are predicted by song fluctuations; the comparison to the randomized r^2 distribution shows less predictive signal across the syllable than expected by chance (Fig 5.13c). Therefore, although the latency distribution is weighted towards negative latencies (Fig 5.13b), we should not interpret this as implying that this cell anticipates song fluctuations. Instead of differences, this cell encodes similarities across renditions: consistent timing in song.

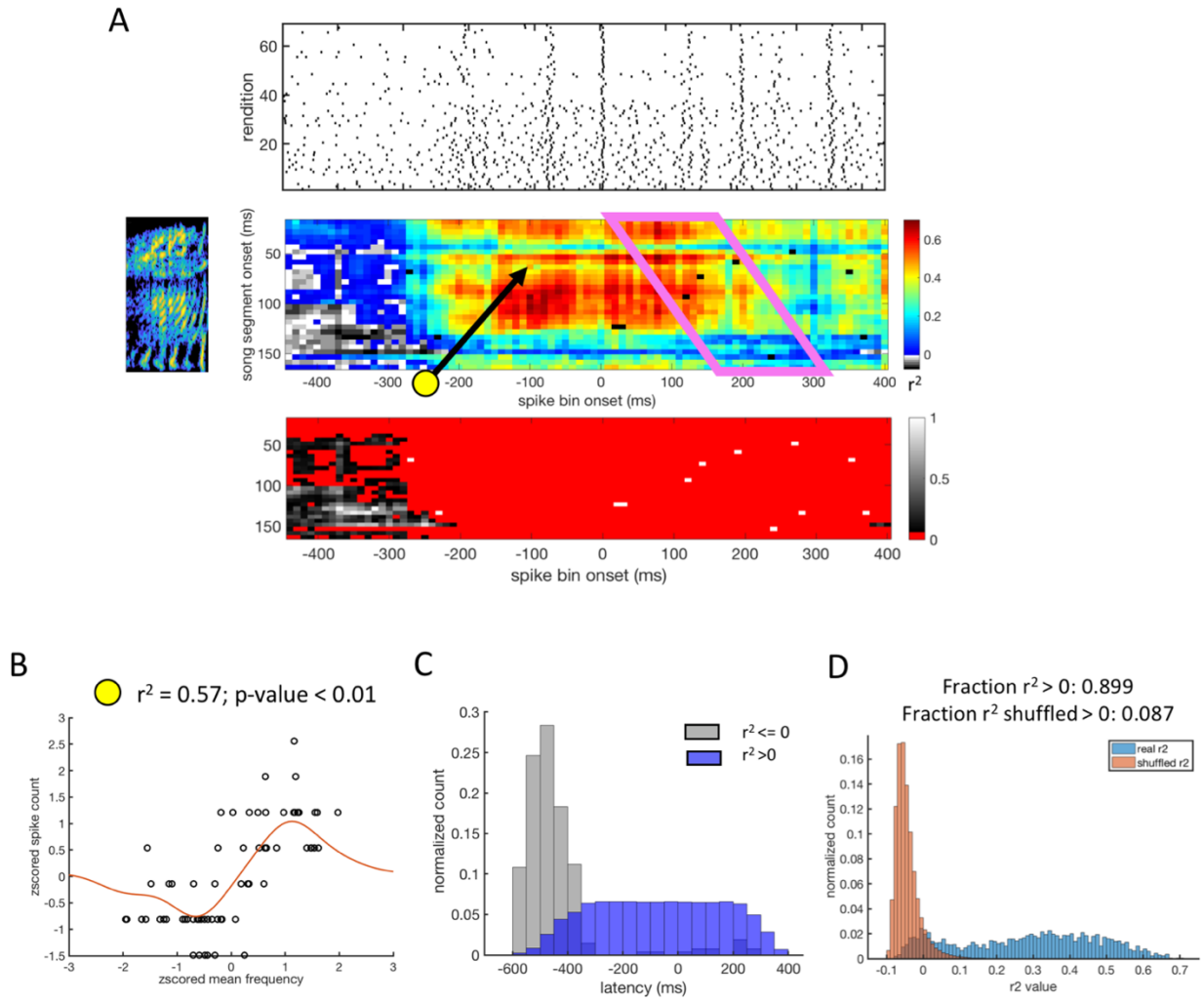

Figure 5.11. VTA-other cell example (2) of a 'global state' cell. **A.** Top: raster of cell activity for each rendition aligned to syllable onset time. Left: spectrogram of example syllable. Time is aligned to syllable onset as in the raster. Middle heat map: heat map of r^2 values for fitted relationship between binned spike count and local feature averages over renditions across many latencies. $R^2 > 0$ indicates a predictive relationship. The x-axis is the time point of the beginning of the spike count bin and is aligned to the raster above, and the y-axis is the time point of the beginning of the local song segment and is aligned to the song spectrogram on the left. Time t=0 is aligned to the syllable onset. Pink parallelogram indicates the region where the latency fits the hypothesized response for a reward prediction error (RPE) (0:150 ms). (Note that the RPE window clearly does not have significance here; the window indicator is for reference to the VTA-error example). The arrow with the yellow dot points to the spike-song relationship that is illustrated for a single song feature in panel b. The lower heat map reports

(continued Fig. 5.11) the p-values for each corresponding pixel in the middle heat map of r^2 values. P-values < 0.05 are colored red. All other p-values are indicated on a grey scale. **B**. Example fit between one local feature average and spike counts for cell example in panel a. Each point represents a single rendition. **C**. Histogram of latencies for predictive fits in panel a. The blue histogram is for latencies with $r^2>0$ and the grey histogram is for fits with $r^2<=0$. **D.** r^2 distributions for randomized and actual data. r^2 distributions for the shuffled and real data compared across all latencies. This population has a significantly greater number of $r^2 > 0$ than expected by chance (p-value < 0.01).

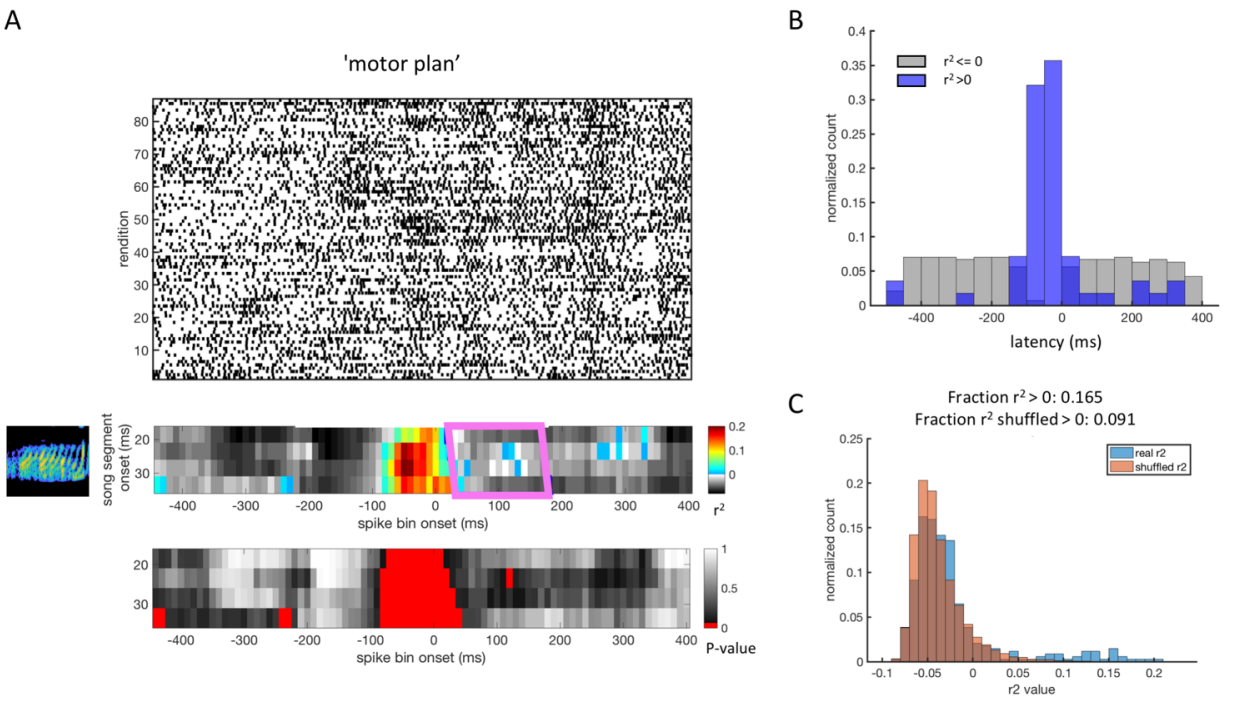

Figure 5.12. VTA-other cell example that anticipates song fluctuations. **A.** Top: raster of cell activity for each rendition aligned to syllable onset time. Left: spectrogram of example syllable. Time is aligned to syllable onset as in the raster. Middle heat map: r^2 values for fitted relationship between binned spike count and local feature averages over renditions across many latencies. $R^2 > 0$ indicates a predictive relationship. The x-axis is the time point of the beginning of the spike count bin and is aligned to the raster above, and the y-axis is the time point of the onset of the time window of the local song segment and is aligned to the song spectrogram to the left. Time t=0 is aligned to the syllable onset. The pink parallelogram indicates the region where the latency fits the hypothesized response for a reward prediction error (RPE) (0:150 ms). (Note that the RPE window clearly does not have significance here; the window indicator is for reference to the VTA-error example). The lower heat map reports the p-values for each corresponding pixel in the middle heat map of r^2 values. P-values < 0.05 are colored red. All other p-values are indicated on a grey scale. **B**. Histogram of latencies for predictive fits in panel a. The blue histogram is for latencies with $r^2>0$ and the grey histogram is for fits with $r^2<=0$. **C.** r^2 distributions for randomized and actual data. r^2 distributions for the shuffled and real data compared across all latencies. This population has a significantly greater number of $r^2 > 0$ than expected by chance (p-value = 0.02).

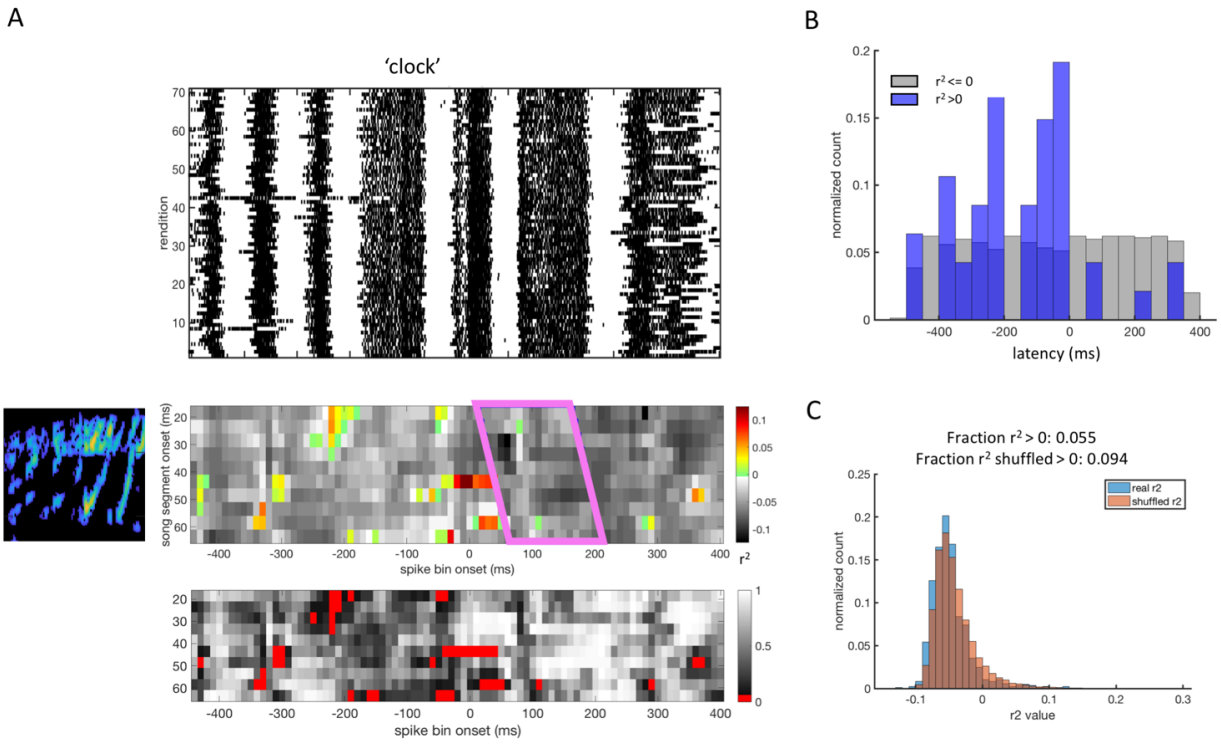

Figure 5.13. VTA-other cell example that tracks song timing. A. Top: raster of cell activity for each rendition aligned to syllable onset time. Left: spectrogram of example syllable. Time is aligned to syllable onset as in the raster. Middle heat map: heat map of r^2 values for fitted relationship between binned spike count and local feature averages over renditions across many latencies. $R^2 > 0$ indicates a predictive relationship. The x-axis is the time point of the beginning of the spike count bin and is aligned to the raster above, and the y-axis is the time point of the beginning of the local song segment and is aligned to the song spectrogram to the left. Time t=0 is aligned to the syllable onset on both axes. Pink parallelogram indicates the region where the latency fits the hypothesized response for a reward prediction error (RPE) (0:150 ms). (Note that the RPE window clearly does not have significance here; the window indicator is for reference to VTA-error example). The lower heat map reports the p-values for each corresponding pixel in the middle heat map of r^2 values. P-values < 0.05 are colored red. All other p-values are indicated on a grey scale. **B**. Histogram of latencies for predictive fits in panel a. The blue histogram is for latencies with $r^2>0$ and the grey histogram is for fits with $r^2<=0$. **C.** r^2 distributions for randomized and actual data. r^2 distributions for the shuffled and real data compared across all latencies. This population does not have a significantly greater number of $r^2 > 0$ than expected by chance (p-value = 0.84).

Discussion:

Value judgements in the brain are necessary to drive appropriate changes in behavior during learning. Using experimentally-constrained tasks with external rewards, previous studies have found that dopamine neurons in VTA encode a key component of value judgement: the mismatch between expected performance and actual performance, the reward prediction error. However, extending these findings to natural behavior and internalized reward has been a challenge. Here, we make use of a novel opportunity to use an experimental context to partition VTA cells into error and non-error cell classes and analyze their responses in a natural behavior context. We examine natural song fluctuations at a local, within-syllable scale and compare these variations to variations in spike counts in VTA cells. Our analysis finds evidence that VTA dopaminergic neurons' activity patterns correlate with variations in natural behavior in a manner consistent with an RPE signal. This finding corroborates and extends complementary discoveries of RPE-like signals emerging from dopamine neurons in VTA in artificial, experimentally-constrained tasks.

There are several ways in which the VTA-error cell class relate to song fluctuations in an RPE-like manner. One, the error cells in the experimental context of Gadagkar et al. (12) show RPE-like signals when the song is artificially distorted in a manner that we assume is 'worse' to the bird's sensibilities; from this we define the error cell class. Two, the natural cell activity correlations to song fluctuations parallel the timing of responses to the experimental paradigm: there is a strong, significant response within the expected RPE latency window. Three, the shapes of the tuning curves of individual cell responses match what would be expected for an

error signal: they are both linear and quadratic, but the quadratic fits are significantly more likely to be single-peaked. This shape profile is consistent with the RPE signal reinforcing a single, optimal song form. Four, the non-error cell set does not show a significant timing peak within the RPE window; this serves as a control for the error-cell timing profile. Five, the non-error cell set shows a different tuning curve profile: while there are still more single peaked quadratic fits, there are significantly more linear tuning curves than quadratic tuning curves. Thus, the presence of RPE signatures in the VTA-error cell activity patterns suggests that a value judgement could be happening in part via VTA dopamine neurons during natural song.

Additionally, we find that a subset of the VTA-error cells relate to 'macroscopic' variations in song: whether particular syllables are present and the number of times an individual syllable is sung in a given motif. These correlated activity patterns sometimes anticipate the song variation itself. The macroscopic example in Figure 5.8 resembles classic RPE activity from earlier experiments with external cues and rewards, but with the stimulus and reward signals internalized (5). In this interpretation, the bird decides early on to sing a long or short version of the song (sing syllable 'D E" or stop after 'C"), which is the 'conditioned stimulus' (CS) equivalent, and that decision is followed by an RPE signal (in cartoon shown as a thumbs up) (Fig 5.8c). Then, when the song is actually extended or not, which is the 'unconditioned stimulus' (US) equivalent, the behavior is followed by another RPE signal (the $r^2>0$ activity in the r2 heat map at 200 ms in Fig 5.8b). A long song is the rewarded action within this interpretation. Intriguingly, this cell is one of the VTA-error cells *not* identified as projecting to area X. It is possible that RPE responses to macroscopic rendition variations are sent to another brain region and that this evaluation is carried out via a secondary circuit.

This is consistent with an RL theory of temporal difference learning in which the RPE signal drifts backwards in time to the point at which the cue predicting a rewarding event is presented (5). An intention to sing a motif in a given manner could serve as an internalized cue signal, and operate in much the same way that light cues or sound cues operate in classical reward-conditioning experiments (5, 35, 36). Longer recordings from error cells during natural song are needed to fully support this interpretation.

Additionally, a straightforward variant of Gadagkar et al.'s previous experimental paradigm could test this idea of an internalized cue signal. In one set of experiments, Gadagkar et al. distorted two separate syllables randomly and independently within a given motif. In this way, they were able to test whether responses to an earlier part of the song affected responses to a later part of the song. They found that when the distortions happened independently of one another, the RPE-like cell responses to the distortions were also independent. We predict from the macroscopic responses we see in some VTA-error cells and from RL theory that introducing correlations between the probability of an early song distortion and a late song distortion should dampen responses to the later distortion. Another variation of this experiment would be to simply introduce another external cue that signals to the bird whether a distortion would happen on a given rendition. We again predict that the RPE signal should weaken after the actual distortion event and appear near the cue.

While, as a population, the VTA-other cell activity did not relate to song fluctuations in a manner consistent with an RPE signal, many cells showed diverse and strong relationships to fluctuations in song. These findings are consistent with previous studies which found that both

dopaminergic and non-dopaminergic cells in VTA contribute to an RPE calculation and that at least some element of the RPE signal is computed locally within VTA (35-38). Some of the VTA-other correlations with song anticipated song fluctuations and could represent elements of intent, which is a critical component of an error calculation. Other cells coded not for song feature fluctuations but for song timing and could provide the timing state context of the RPE signal. Lastly, we found novel fluctuations in spike counts spanning entire song renditions, which coded for song fluctuations across the entire song. We term these cells 'global state cells'. These types of correlated fluctuations do not fit the current picture of how fluctuations in song drive learning in the song motor pathway in which independent, local fluctuations drive element-specific changes in song (39).

What might be the role of the global state cells in VTA? Further experiments are needed to fully elucidate the role of these cells. We suggest a hypothesis in which fluctuations in song are due to both error-relevant and error-irrelevant changes in the song state and that the role of the global state cells in VTA is to enact a gain-scaling that removes error-irrelevant song fluctuations from the error signals projected to Area X. Global state variations change the song output, but may not emerge from the AFP learning circuit nor change the underlying error of the rendition.

A simple example is singing renditions at different volumes. Singing a song loudly or softly causes a shift in song amplitude across the entire song but wouldn't necessarily be an error. However, it would confound decoding local variations and assigning accurate error in song amplitude from auditory input. Figure 5.14 panel A demonstrates this relationship. Song

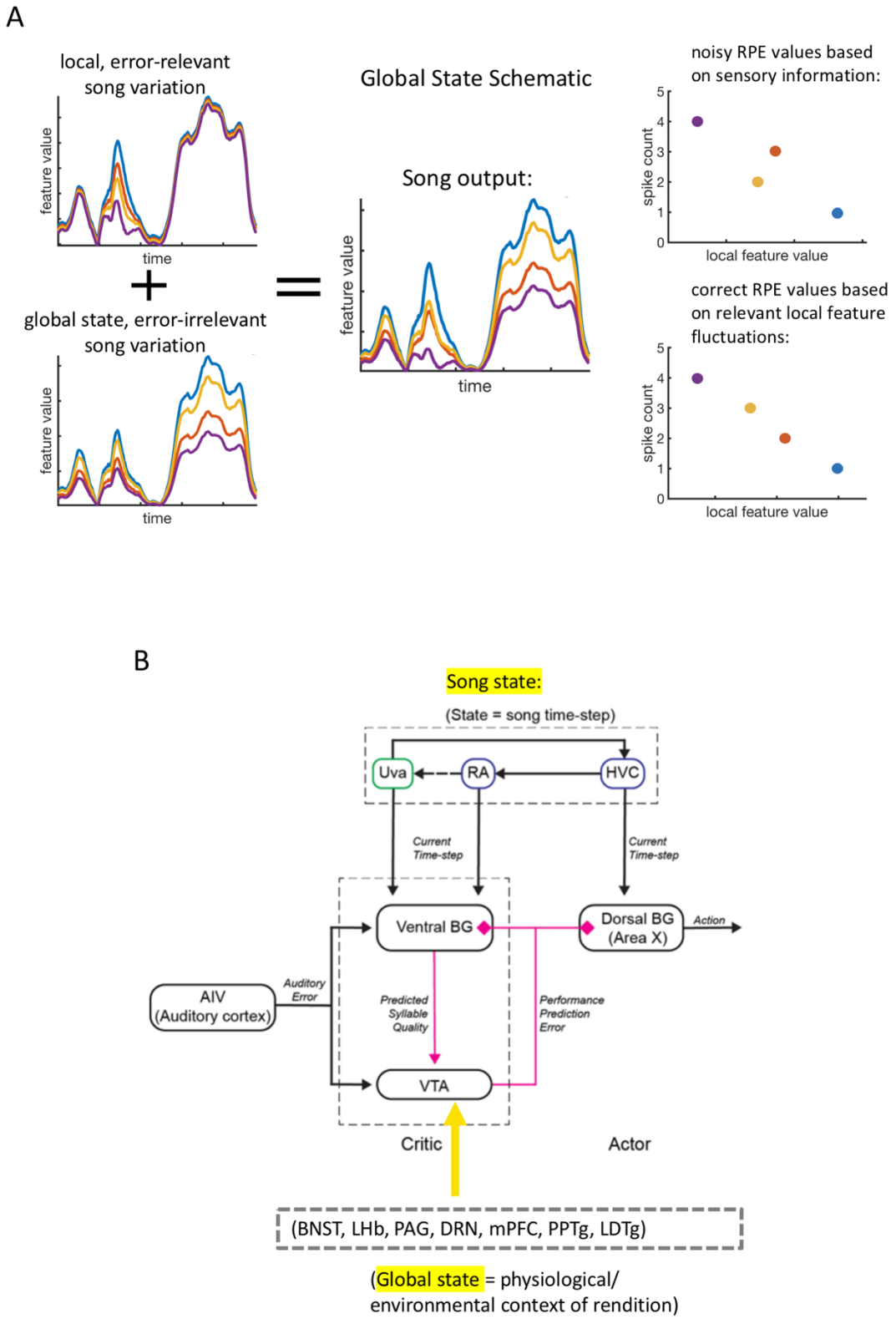

Figure 5.14 (legend on following page.)

Figure 5.14. Schematic of the 'global state' hypothesis. **A.** Schematic of the 'global state' hypothesis. Fluctuations in song are due to both error-relevant and error-irrelevant changes in the song state. The RPE signal is made more robust and accurate by being able to subtract out error-irrelevant song fluctuations caused by changes in the bird's 'global state' (e.g. time of day, arousal, motivation, intent to sing loudly or softly, etc.). A simple example is singing renditions at different volumes. Singing a song loudly or softly would cause a shift in song amplitude across the entire song but wouldn't necessarily be an error. However, it would confound decoding local variations in song amplitude from auditory input. Having access to variations in song which don't induce changes in error, which we call the 'global state fluctuations,' would make the RPE calculation more accurate and robust. **B.** Schematic of reinforcement learning in song (schematic expanded and adapted from Chen et al. 2018). Multiple projections to VTA could contribute aspects of our hypothesized 'global state' variables to the error calculation. In this theory we divide state into 2 components: 'song state', which is the time step in song, and 'global state' which is the physiological and environmental context of the song. Global state creates variation from rendition to rendition but is not necessarily indicative of local differences in error.

output is a noisy representation of local song variations: local, error-relevant song variations are masked by global variations in song. The error signal is made more robust and accurate by the removal of error-irrelevant song fluctuations caused by changes in the bird's 'global state' (Fig. 5.14a). Access to this global state information is crucial for accurate, local error computation.

Even in this simple example, there is strong behavioral evidence in zebra finch that song volume is both a learned, local song element and varied contextually on a global scale rendition-to-rendition. Previous work found that zebra finches sing at different volumes depending on their physical distance from a female, so intentional, global song variation is a practiced skill (40). Furthermore, amplitude of song is a learned trait. Ritschard and Brumm compared the amplitude of young male zebra finch songs to those of their genetic fathers as well as to those of their tutors from whom they had learned their song. They found that the amplitude of specific tutee song elements was strongly correlated with the amplitude of the tutor's song elements, but not with the genetic father's, implying that song amplitude is, in part, a learned trait (41). The work of Kao et al. (2005) provides additional evidence that amplitude fluctuations are learned: electrical micro-stimulation of LMAN sites during singing induced song element-specific variation in amplitude (22). Because LMAN is the output nucleus of the AFP learning circuit onto the primary song motor pathway, this finding again implies that amplitude is, in part, a learned trait which is varied via activity from LMAN.

Amplitude is a particularly simple example, but global state fluctuations could as easily affect different song elements in a non-linear, time step-specific manner and derive from other sources. In male zebra finches, undirected song is used in a variety of contexts. The contexts in

which male zebra finches sing undirected song in the wild are: (1) during nest guarding or bonding—near the nest with a female inside or to induce a female to enter the nest; (2) sexual advertising—near the flock or nesting colony; (3) unmated males in the flock; (4) during visual isolation from the flock (42). Additionally, the frequency of undirected song increases quickly in zebra finch fathers from day 8 to day 24 post-hatch of their offspring, after which singing occurs at sustained, high levels (43); this coincides with the period during which juveniles are thought to memorize the tutor song and could represent an additional teaching context. The prosody of song in these various contexts may differ and would lead to global variants in song which needn't differ in error value. Time of day (44), arousal, emotional state or motivation are additional examples of possible sources of global state variations.

In both mammal and birds, VTA receives convergent inputs from brain regions that relate to arousal and motivation and, therefore, is well situated to incorporate this type of global information into the reward prediction error computation (Fig. 5.14b) (45). Tian et al. (2015) recorded from multiple inputs to RPE-encoding dopaminergic neurons in mice VTA during a classic reward conditioning task. They found that the variables needed to compute an RPE signal were present in multiple projections from the ventral and dorsal striatum, the ventral pallidum, subthalamic nucleus, lateral hypothalamus, rostromedial tegmental nucleus (RMTg) and pedenculo-pontine tegmental nucleus (PPTg) and non-dopaminergic cells within VTA. Individual inputs were often multiplexed: they contained combinations of the RPE calculation (36). Tian et al. concluded that even though the theoretical calculation of an RPE signal is a very simple subtraction, ($Value_{actual} - Value_{predicted} = RPE$), the actual calculation of this operation involves diverse brain regions and non-linear processes that

converge in VTA. This complexity makes more sense and has a clear role if error is computed relative to additional organism states such as arousal or motivation as well as relative to other error-invariant transformations of the task, such as, in our simple example, singing the song at different volumes. The full functionality of this distributed error calculation would be difficult to see in simplified laboratory tasks that are designed to remove these types of contingencies from behavior or average over them.

It is significant that we did not find global song responses in any of the VTA-error cell population. This suggests that although covarying, global fluctuations in song exist during undirected song, the VTA-error cells are only encoding locally independent song fluctuations that presumably are the most relevant for learning. Whatever the global state fluctuations relate to in the brain, they do not hold a strong valence in the observed, putative error signal that projects to Area X in our data. The convergent inputs into VTA from auditory, motor and motivational brain regions make it a reasonable nexus at which the removal of global fluctuations from a local error evaluation could take place. Simultaneous recordings from VTA-x-projecting error cells and global state cells would strengthen this observation.

How might this theory change the current reinforcement learning (RL) picture in song bird? The type of RL currently used to describe song bird learning is particularly simple. The motor nuclei, HVC and RA, generate a song 'policy:' which vocalizations to carry out when. A cortical-thalamic-basal ganglia loop, AFP (anterior forebrain pathway) serves as the 'actor' and injects variability into the vocal output via projections from LMAN to RA. A pathway from auditory centers through VTA evaluates the vocal outputs by potentially comparing them to a memory of the tutor's song and projects this evaluation in the form of an reward-prediction

error signal back to the basal ganglia region of the AFP loop, area X. Area X then biases the variable outputs of LMAN towards better variations, which consolidates an improved song policy. In this picture, the effects of actions taken early in the song sequence do not propagate to later parts of song (46). Sutton and Barto describe this type of RL as being a simple, partial version of 'full RL', that they term 'contextual bandit', because actions only affect current rewards and not future states (1). This learning task is a purely associative search task: a trial-and-error search for the best actions and an association of these actions with the appropriate situation (here: the time step in song).

HVC's sparse patterns of activity support this straightforward, temporally independent, version of RL (20, 39). In adult birds, HVC projection neurons to RA burst exactly once during song in a highly stereotyped manner. Concurrently, HVC projections to area X are almost as sparse. This provides a unique timestamp for each moment in song to both the downstream 'policy' in RA and the 'actor-evaluator' circuit. HVC to RA synapses only need substantiate a local policy in the song sequence: the sparse HVC firing pattern establishes this independence.

From one perspective, it seems like global fluctuations in song could introduce a more complex version of RL into the song bird system: now there is the possibility for actions taken in previous steps in song to influence the value of current actions. To return to the simple example of volume, if the song starts out loud, singing the end softly would be an error: the value of a particular action at a later part of the song now depends on the action at the beginning of song. However, if we add a new global state variable to the current time-step definition of state (Fig 5.14b) in the RL account which can appropriately scale the error signal at each time step, the simple 'contextual bandit' form of RL remains intact in the motor pathway

and correlated fluctuations across song need not introduce a sequentially-dependent form of learning. This structure is a form of hierarchical RL that could retain the simplicity of the 'contextual bandit' RL at a granular, within-rendition scale. Future modeling work is required to fully explore the implications of variants of a single behavior on an RL structure.

What is the nature of the hypothetical neural computation that could remove global state fluctuations from the error signal? In sensory systems, arousal, attention, adaptation and other contextual factors are known to modulate neuronal responses to well-defined tasks (47, 48). Especially noteworthy here, auditory cortex processes simple acoustic features differently depending on both the external environmental context as well as internal states (49). Gain-scaling has been observed in many sensory systems and is a well-established operation in the brain. Normalizing the song-error signal relative to global state changes could be a higher order analog of this basic operation. If this type of error scaling exists, it would present another way in which bird song vocalizations can be understood in relation to human language wherein error invariance across emotional cadences and other contextual variations of a spoken word is highly likely.

This project uses the structure of an experimentally-grounded characterization of individual neurons to analyze the same neurons in natural behavior. The connection to an existing experiment as well as to an RL framework anchors our interpretations of natural behavior in a constrained laboratory paradigm and theory. The unusually high stereotypy of the natural behavior we consider, zebra finch song, allows reasonable inferences to be made both in the experimental and natural context about the intentions of the bird and a reasonable way to segment and align a complex behavior. We found a parallel relationship between the VTA-

error cell activity in experimental and natural contexts that corroborates the experimental finding that VTA-error cells are encoding time-step specific reward prediction errors in song. However, we found additional complexities that exist outside the binary of the experimental distortion paradigm.

The novel interpretations and findings that emerge from examining the variations of natural behavior are: (1) value judgements at multiple scales: error-like signals relating to macroscopic changes in song, such as the number of syllables within a given rendition, (2) value judgments upon intention as well as action: cell activity predictive of macroscopic song variations, (3) evidence for local components of the error calculation: a subset of VTA-other cells that contain diverse aspects of a hypothetical error equation, and (4) global state changes that effect performance and possibly judgement: a subset of VTA-other cells that encode global changes across song within an undirected song context. Together these insights suggest that prediction error is a much more mutable quantity than previously thought in the songbird. In this more complex picture, error is not simply calculated relative to a static memory of the tutor's song, as in current theoretical accounts, but arrived at through a dynamic, contextual process. These additional findings were unexpected in the course of our analysis and require careful follow-up experiments in greater isolation and in more controlled conditions.

However, our examination of natural behavior was necessary to observe even the possibility of these expanded computations. A frequent debate in neuroscience is whether artificial behavioral paradigms serve as true building blocks for understanding neural activity in complex, freely behaving contexts, or whether they represent a different context that will not extrapolate. This experimentally guided study of natural behavior is a fruitful direction that

permits the control of experimental contexts and the complexity of natural contexts to interact and build upon one another.

Appendix A: *Construction of the Gaussian process regression model*

The following derivation of the Gaussian process regression model is based on Kenneth Latimer's derivation in the methods section of our manuscript which is currently in preparation.

We used a regression approach to determine if spike counts are related to the variation in song. The relationship between spike counts and song is likely non-linear and related to a variable number of features depending on the point in song. To address this, we used a non-parametric Gaussian process (GP) regression to fit the relationship between our 8 song features and spike counts across song and spike window pairs (28). We use a Bayesian average approach to combine a weighted average of GP regressions using all non-zero subsets of song features into a single model prediction. We model the relationship between the set of N z-scored song features on a single rendition, i, $\boldsymbol{x}_i$, and the spike counts on that given rendition, y_i, in single time windows (e.g. the song feature values 20 ms after syllable onset and the spike count in a 100 ms window, 75 ms after syllable onset).

We selected a subset of features for a single GP regression, where feature is indexed by $\mathcal{M}$ such that $\mathcal{M} \subseteq \{1,2, \ldots N\}, \mathcal{M} \neq 0$, N = 8. The GP regression model for a single set of $\mathcal{M}$ is:

$$y_i | \, f_{\mathcal{M}} \sim \mathcal{N}\left(f_{\mathcal{M}}\left(\boldsymbol{x}_{i,\mathcal{M}}\right), \sigma^2\right) \qquad (1)$$

$$f_{\mathcal{M}}(\cdot) \sim \mathcal{GP}\left(\mu, \omega^2 \kappa^{(\mathcal{M})}(\boldsymbol{x}, \boldsymbol{x}')\right) \qquad (2)$$

where κ is the covariance function and defines how spike counts will be correlated with one another in feature space. We use the commonly selected kernel function for κ,

$$\kappa^{(\mathcal{M})}(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x}_{\mathcal{M}} - \boldsymbol{x}'_{\mathcal{M}}\|^2}{2l^2}\right). \qquad (3)$$

ω^2 is the GP variance term and specifies how strongly the spike counts vary as a function of the song features. σ^2 is the variance term that captures noise in the spike count (i.e. how much the spike counts vary at a single point in song feature space). The length scale, l, determines how close points must be in feature space to have correlated spike counts. To reduce computational complexity, we set $l = 0.5$, for all model fits and z-scored the individual song features at each time step we considered.

The likelihood of all T renditions of spike count-song segment pairs is:

$$p(y_{1:T}|\boldsymbol{x}_{1:T}, \mathcal{M}, l, \sigma^2, \omega^2, \mu) = \mathcal{N}\left([y_1, y_2, \ldots, y_T]^{\mathrm{T}};\ [\mu, \mu, \ldots, \mu]^{\mathrm{T}},\ \omega^2 \mathrm{K}^{(\mathcal{M})} + \sigma^2 I_T\right); \qquad (4)$$

$$\mathrm{K}^{(\mathcal{M})} = \begin{bmatrix} \kappa^{(\mathcal{M})}(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & \kappa^{(\mathcal{M})}(\boldsymbol{x}_1, \boldsymbol{x}_T) \\ \vdots & \ddots & \vdots \\ \kappa^{(\mathcal{M})}(\boldsymbol{x}_T, \boldsymbol{x}_1) & \cdots & \kappa^{(\mathcal{M})}(\boldsymbol{x}_T, \boldsymbol{x}_T) \end{bmatrix}, \qquad (5)$$

where I_T is the identity matrix of dimension T.

The prediction mean-squared error for the GP model is:

$$MSE_{loo}^{(GP)} = \frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2 \qquad (6)$$

$$\hat{y}_i = \quad \mathbb{E}\left[y_i \mid \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right] \qquad (7)$$

where $\hat{y}_i$ is the predicted spike count from the model for rendition i, and $\boldsymbol{x}_{/i}$ and $\boldsymbol{y}_{/i}$ are the song features and spike counts for all renditions except for the i^{th} rendition.

There is a good deal of model uncertainty in this task: it is unclear which features to use at a given point in song and how many should be used. Furthermore, the prediction of the model depends heavily on which features are used. To address this uncertainty, we used a Bayesian model averaging approach to determine the predicted spike count wherein we integrate over all possible values of $\mathcal{M}$ and weight their predictions according to their posterior probability given the observed spike counts (29):

$$\mathbb{E}\left[y_i \mid \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right] \; = \sum_{\mathcal{M}\subseteq\{1,2,\ldots N\},\mathcal{M}\neq 0} \int_0^\infty \mathbb{E}\left[y_i \mid \mathcal{M}, r, \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right] p\left(\mathcal{M}, r \middle| \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}\right) dr, \quad (8)$$

where $r = \frac{\sigma^2}{\omega^2}$ is the ratio of the GP variance to the observation noise. We integrate over all possible values of $\mathcal{M}$ and weight their predictions according to their posterior probability given the observed spike counts.

$$\mathbb{E}\left[y_i \mid \mu, \mathcal{M}, r, \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right] = \mathrm{K}_{/i,i}^{(\mathcal{M})^{\mathrm{T}}}\left(\mathrm{K}_{/i,/i}^{(\mathcal{M})} + rI_{(T-1)}\right)^{-1}\left(\boldsymbol{y}_{/i} - \mu\right) + \mu\,. \quad (9)$$

Thus, we incorporate all combinations of features into a single model prediction for each song-spike count pair. We re-parameterize (σ^2, ω^2) to (ψ^2, r^2) where ψ^2 is the total variance:

$$\psi^2 = \sigma^2 + \omega^2, \quad (10)$$

$$\sigma^2 = \frac{r}{r+1}\psi^2, \qquad \omega^2 = \frac{\psi^2}{r+1} \quad (11)$$

And evaluate the posterior over model parameters using Bayes' rule:

$$p\left(\mathcal{M}, r|\boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}\right) = \frac{p\left(\boldsymbol{y}_{/i}|r, \mathcal{M}, \boldsymbol{x}_{/i}\right)p(\mathcal{M}, r)}{\sum_{\mathcal{M}^*\subseteq\{1,2,\ldots N\}} p(\mathcal{M}^*)\int_0^\infty p\left(\boldsymbol{y}_{/i}|r^*, \mathcal{M}^*, \boldsymbol{x}_{/i}\right)p(\mathcal{M}^*, r^*)dP(r^*)}\,. \quad (12)$$

We again use Bayes' rule to compute the likelihood term in Eq. 12:

$$p\left(\boldsymbol{y}_{/i}|r, \mathcal{M}, \boldsymbol{x}_{/i}\right) = \frac{p\left(\boldsymbol{y}_{/i}|\mu, \psi^2, r, \mathcal{M}, \boldsymbol{x}_{/i}\right)p(\mu, \psi^2)}{p\left(\mu, \psi^2|r, \mu, \boldsymbol{x}_{/i}, \psi^2, \mathcal{M}\right)}. \quad (13)$$

The likelihood term is computed as in Eq. 4. We again use Bayes' rule to compute the posterior over μ and ψ^2 :

$$p\left(\mu, \psi^2 | r, \mu, \boldsymbol{x}_{/i}, \psi^2, \mathcal{M}\right) \propto p\left(\boldsymbol{y}_{/i} | \mu, \psi^2, r, \mathcal{M}, \boldsymbol{x}_{/i}\right) p(\mu, \psi^2). \quad (14)$$

We then place a conjugate normal-inverse gamma prior over μ and ψ^2 :

$$(\mu, \psi^2) \sim \mathrm{N\_}\Gamma^{-1}(\mu_0, \lambda_0, \alpha_0, \beta_0) \quad (15)$$

where,

$$\mu_0 = 0; \quad \lambda_0 = 1; \quad \alpha_0 = 10; \quad \beta_0 = \alpha_0 + 1. \quad (16)$$

Thus,

$$\left(\mu, \psi^2 | r, \mu, \psi^2, \mathcal{M}, \boldsymbol{x}_{/i}\right) \sim \mathrm{N\_}\Gamma^{-1}\left(\mu_{post}^{(i)}, \lambda_{post}^{(i)}, \alpha_{post}^{(i)}, \beta_{post}^{(i)}\right), \quad (17)$$

where,

$$\mu_{post}^{(i)} = \frac{b^{(i)}}{a^{(i)}}, \quad (18)$$

$$\lambda_{post}^{(i)} = a^{(i)}, \quad (19)$$

$$\alpha_{post}^{(i)} = \alpha_0 + \frac{T-1}{2}, \qquad (20)$$

$$\beta_{post}^{(i)} = \frac{1}{2}\left(c^{(i)} - \frac{b^{(i)}}{a^{(i)}}\right), \qquad (21)$$

$$a^{(i)} = (r+1)\mathbf{1}^{\mathrm{T}}\left(\mathrm{K}_{/i,/i}^{(\mathcal{M})} + rI\right)^{-1}\mathbf{1} + \lambda_0, \qquad (22)$$

$$b^{(i)} = (r+1)\mathbf{1}^{\mathrm{T}}\left(\mathrm{K}_{/i,/i}^{(\mathcal{M})} + rI\right)^{-1}\boldsymbol{y}_{/i}, \qquad (23)$$

$$c^{(i)} = (r+1)\mathbf{1}^{\mathrm{T}}\left(\mathrm{K}_{/i,/i}^{(\mathcal{M})} + rI\right)^{-1}\boldsymbol{y}_{/i} + 2\beta_0, \qquad (23)$$

where **1** is a vector of ones. With this, we can compute all of the terms in Eq. 13.

The integral over Eq. 12 is over one dimension and thus tractable. We chose a discrete distribution for the prior $P(r)$ to increase computation speed:

$$P(r) = Uniform(\{3,4,5.67,9\}), \qquad (24)$$

such that the GP model variance could be 25%, 20%, 15% or 10% of the total variance.

We imposed a truncated binomial prior over $\mathcal{M}$ such that $|\mathcal{M}| \geq 1$, that favored models with fewer features:

$$p(\mathcal{M}) = \frac{1}{1-(1-p)^N}\binom{N}{|\mathcal{M}|}p^{|\mathcal{M}|}(1-p)^{N-|\mathcal{M}|}. \qquad (25)$$

We set p=0.1 so that approximately 2/3 of the prior probability mass rests on single-feature models. Using this normal inverse-gamma description of the posterior, we can compute the prediction of y_i, given $\mathcal{M}$ and r:

$$\mathbb{E}\left[y_i \mid \mathcal{M}, r, \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right] = \mathbb{E}\left[\mathbb{E}\left[y_i \mid \mu, \mathcal{M}, r, \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right] | \mathcal{M}, r, \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right], \qquad (26)$$

$$= \mathrm{K}_{/i,i}^{(\mathcal{M})^{\mathrm{T}}}\left(\mathrm{K}_{/i,/i}^{(\mathcal{M})} + rI_{(T-1)}\right)^{-1}\left(\boldsymbol{y}_{/i} - \mu_{post}^{(i)}\right) + \mu_{post}^{(i)}\,. \qquad (27)$$

We then insert Eq. 27 and Eq. 12 into Eq. 8 to obtain the prediction of y_i.

Evaluating the model performance.

To evaluate model performance we use a leave-one-out cross validation method to estimate the mean-squared prediction error for new observations as in (30):

$$MSE_{loo}^{(GP)} = \frac{1}{T}\sum_{i=1}^{T}(y_i - \hat{y}_i)^2, \qquad (28)$$

$$\hat{y}_i = \mathbb{E}\left[y_i \mid \boldsymbol{x}_{/i}, \boldsymbol{y}_{/i}, \boldsymbol{x}_i\right], \qquad (29)$$

where $\hat{y}_i$ is the model prediction spike count for rendition 'i', and $\boldsymbol{x}_{/i}$ and $\boldsymbol{y}_{/i}$ are the song features and spike counts of all renditions excluding the ith rendition.

We then compare the GP model to a model with constant mean firing rate equal to the mean spike count over all renditions excluding the ith rendition:

$$y_{(i)} \sim \mathcal{N}(\alpha, \tau^2). \quad (30)$$

The predictive mean-squared error of this model is:

$$MSE_{loo}^{(null)} = \frac{1}{T}\sum_{i=1}^{T}\left(y_i - \overline{y_{/i}}\right)^2 \quad (31)$$

where,

$$\overline{y_{/i}} = \frac{1}{T-1}\sum_{j \neq i} y_j. \quad (32)$$

The r^2 value, is:

$$r^2 = 1 - \frac{MSE_{loo}^{(GP)}}{MSE_{loo}^{(null)}}. \quad (33)$$

An r^2 > 0 indicates that the model predicts new observations better than simply using the mean. The maximum theoretical value the r^2 can take is one—this indicates perfect model prediction and, in practice, is never achieved. We use the r^2 value as our measure of model performance.

Bibliography:

1. Sutton RS & Barto AG (1998) *Reinforcement learning : an introduction* (MIT Press, Cambridge, Mass.) pp xviii, 322 p.
2. Maidhof C, Vavatzanidis N, Prinz W, Rieger M, & Koelsch S (2010) Processing expectancy violations during music performance and perception: an ERP study. *J Cogn Neurosci* 22(10):2401-2413.
3. Katahira K, Abla D, Masuda S, & Okanoya K (2008) Feedback-based error monitoring processes during musical performance: an ERP study. *Neurosci Res* 61(1):120-128.
4. Trewartha KM & Phillips NA (2013) Detecting self-produced speech errors before and after articulation: an ERP investigation. *Front Hum Neurosci* 7:763.
5. Schultz W, Dayan P, & Montague PR (1997) A neural substrate of prediction and reward. *Science* 275(5306):1593-1599.
6. Bayer HM & Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47(1):129-141.
7. Fiorillo CD, Tobler PN, & Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299(5614):1898-1902.
8. Wise RA & Rompre PP (1989) Brain dopamine and reward. *Annu Rev Psychol* 40:191-225.
9. Person AL, Gale SD, Farries MA, & Perkel DJ (2008) Organization of the songbird basal ganglia, including area X. *The Journal of comparative neurology* 508(5):840-866.
10. Konishi M (1965) The role of auditory feedback in the control of vocalization in the white-crowned sparrow. *Zeitschrift fur Tierpsychologie* 22(7):770-783.

11. Scharff C & Nottebohm F (1991) A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: implications for vocal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 11(9):2896-2913.
12. Gadagkar V*, et al.* (2016) Dopamine neurons encode performance error in singing birds. *Science* 354(6317):1278-1282.
13. Gale SD, Person AL, & Perkel DJ (2008) A novel basal ganglia pathway forms a loop linking a vocal learning circuit with its dopaminergic input. *The Journal of comparative neurology* 508(5):824-839.
14. Ding L & Perkel DJ (2004) Long-term potentiation in an avian basal ganglia nucleus essential for vocal learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 24(2):488-494.
15. Ravbar P, Lipkind D, Parra LC, & Tchernichovski O (2012) Vocal exploration is locally regulated during song learning. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 32(10):3422-3432.
16. Charlesworth JD, Tumer EC, Warren TL, & Brainard MS (2011) Learning the microstructure of successful behavior. *Nature neuroscience* 14(3):373-380.
17. Tumer EC & Brainard MS (2007) Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. *Nature* 450(7173):1240-1244.
18. Lipkind D & Tchernichovski O (2011) Quantification of developmental birdsong learning from the subsyllabic scale to cultural evolution. *Proceedings of the National Academy of Sciences of the United States of America* 108 Suppl 3:15572-15579.
19. Andalman AS & Fee MS (2009) A basal ganglia-forebrain circuit in the songbird biases motor output to avoid vocal errors. *Proceedings of the National Academy of Sciences of the United States of America* 106(30):12518-12523.
20. Fiete IR, Hahnloser RH, Fee MS, & Seung HS (2004) Temporal sparseness of the premotor drive is important for rapid learning in a neural network model of birdsong. *Journal of neurophysiology* 92(4):2274-2282.
21. Tchernichovski O, Nottebohm F, Ho CE, Pesaran B, & Mitra PP (2000) A procedure for an automated measurement of song similarity. *Anim Behav* 59(6):1167-1176.
22. Kao MH, Doupe AJ, & Brainard MS (2005) Contributions of an avian basal ganglia-forebrain circuit to real-time modulation of song. *Nature* 433(7026):638-643.
23. Woolley SC & Kao MH (2014) Variability in action: Contributions of a songbird cortical-basal ganglia circuit to vocal motor learning and control. *Neuroscience*.
24. Leblois A, Wendel BJ, & Perkel DJ (2010) Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(16):5730-5743.
25. Deregnaucourt S*, et al.* (2004) Song development: in search of the error-signal. *Annals of the New York Academy of Sciences* 1016:364-376.
26. Sober SJ & Brainard MS (2009) Adult birdsong is actively maintained by error correction. *Nature neuroscience* 12(7):927-931.
27. Kao MH, Wright BD, & Doupe AJ (2008) Neurons in a forebrain nucleus required for vocal plasticity rapidly switch between precise firing and variable bursting depending on

social context. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 28(49):13232-13247.
28. Williams CK, & Rasmussen, C. E. (2006) Gaussian processes for machine learning. *the MIT Press* 2(4).
29. Hoeting JA, Madigan, D., Raftery, A. E., & Volinsky, C. T. (1998) Bayesian model averaging *Proceedings of the AAAI Workshop on Integrating Multiple Learned Models* 335:77-83.
30. Vehtari A, Gelman, A., & Gabry, J. (2017) Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and Computing* 27:1413-1432.
31. Il Memmming Park EA, Nicholas Priebe, & Jonathon Pillow (2013) Spectral methods for neural characterization using generalized quadratic models. *Advances in Neural Information Processing Systems* 26:2454-2462.
32. Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6):716 - 723.
33. Raftery REKAE (1995) Bayes Factors. *Journal of American Statsitical Association* 90:773-795.
34. Tusher VG, Tibshirani R, & Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences of the United States of America* 98(9):5116-5121.
35. Wood J, Simon NW, Koerner FS, Kass RE, & Moghaddam B (2017) Networks of VTA Neurons Encode Real-Time Information about Uncertain Numbers of Actions Executed to Earn a Reward. *Front Behav Neurosci* 11:140.
36. Ju Tian RH, Jeremiah Y. Cohen, Fumitaka Osakada, Dmitry Kobak, Christian K. Machens, Edward M. Callaway, Naoshige Uchida, and Mitsuko Watabe-Uchida (2016) Distributed and Mixed Information in Monosynaptic Inputs to Dopamine Neurons. *Neuron* 91:1374-1389.
37. Dobi A, Margolis EB, Wang HL, Harvey BK, & Morales M (2010) Glutamatergic and nonglutamatergic neurons of the ventral tegmental area establish local synaptic contacts with dopaminergic and nondopaminergic neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(1):218-229.
38. Cohen JY, Haesler S, Vong L, Lowell BB, & Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature* 482(7383):85-88.
39. Fee MS & Goldberg JH (2011) A hypothesis for basal ganglia-dependent reinforcement learning in the songbird. *Neuroscience* 198:152-170.
40. Brumm HS, P. (2005) Animals can vary signal amplitude with receiver distance: evidence from zebra finch song. *Anim Behav* 72:699-705.
41. M. Ritschard HB (2011) Effects of vocal learning, phonetics and inheritance on song amplitude in zebra finches. *Anim Behav* 82:1415-1422.
42. Zann R (1996) *The Zebra Finch: a synthesis of field and laboratory studies* (Oxford University Press).
43. ten Cate C (1982) Behavioural differences between zebra finch and Bengalese finch (foster) parents raising zebra finch offspring. *Behaviour* 81:152-172.

44. Glaze CM & Troyer TW (2006) Temporal structure in zebra finch song: implications for motor coding. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26(3):991-1005.
45. Morales M & Margolis EB (2017) Ventral tegmental area: cellular heterogeneity, connectivity and behaviour. *Nature reviews. Neuroscience* 18(2):73-85.
46. Mackevicius EL & Fee MS (2018) Building a state space for song learning. *Current opinion in neurobiology* 49:59-68.
47. Rabinowitz NC, Goris RL, Cohen M, & Simoncelli EP (2015) Attention stabilizes the shared gain of V4 populations. *Elife* 4:e08998.
48. Cohen MR & Maunsell JH (2010) A neuronal population measure of attention predicts behavioral performance on individual trials. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(45):15241-15253.
49. Angeloni C & Geffen MN (2018) Contextual modulation of sound processing in the auditory cortex. *Current opinion in neurobiology* 49:8-15.