Sequence Variability And Learning In Bengalese Finch Song

A thesis submitted

by

Malcolm Kennedy Department of Biology McGill University, Montreal

to McGill University in partial fulfillment of the requirements of the degree of Master of Science.

December 2019

© Malcolm Kennedy 2019

Abstract

Songbirds are one of a small number of vertebrate taxa that learn their vocalizations over development (Jarvis, 2019). Unlike some songbird species (e.g. zebra finch, white-crowned sparrow), Bengalese finches (*Lonchura striata domestica*) produce sequences of acoustic elements (syllables) that are variable from redition to rendition, as well as across individuals (Okanoya, 2004a, 2004b). However, these sequences are not random: both behavioural and neural evidence has supported the encoding of sequential context over more than the preceding syllable during sequence generation in Bengalese finches (Jin & Kozhevnikov, 2011; Warren et al, 2012). I investigate between-bird variance in sequence variability in Bengalese finch song, and assess the role of sequence variability in learning success. I find that sequence variability is constrained by sequential context has been accounted for, than when it has not. In addition, I find that the variability of the sequences in which a syllable is embedded significantly relates to the extent of acoustic and sequence learning for that syllable.

Résumé

Les oiseaux chanteurs (oiseaux appartenants à la clade *Passeri*), sont parmi peu de taxa vertébrés dont les vocalisations sont apprises parmi le développement (Jarvis, 2019). Contrairement à certains oiseaux chanteurs (par exemple, le Diamant mandarin ou le Bruant à couronne blanche), les **Moineaux du Japon** (*Lonchura striata domestica*) produisent des séquences d'éléments acoustiques ("syllabes") qui varient de chant en chant, ainsi qui parmi les oiseaux individus (Okanoya, 2004a, 2004b). Cependant, ces séquences ne sont pas aléatoires: les indices comportementaux et neuraux soutiennent le codage du contexte séquentiel pendant la génération des séquences syllabiques dans les Moineaux du Japon (Jin & Kozhevnikov, 2011; Warren et al, 2012). J'examine la variance en variabilité séquentielle parmi les oiseaux individus, et le rôle du de la variabilité séquentielle dans l'apprentissage. Je trouve que la variabilité séquentielle est contrainte par le contexte séquentiel parmi les oiseaux, tel que les oiseaux convergent plus étroitement quand le contexte séquentiel a été pris en compte, que quand il ne l'a pas. De plus, je trouve que la variabilité des séquences ou un syllable se trouve le plus souvent est reliée au succès d'apprentissage acoustique et séquentiel pour ce syllable.

Table of Contents

Abstract	
Table of Contents	3
i. Acknowledgements	4
ii. Contributions of the author	6
1. Introduction	7
1.1 Vocal sequencing in birdsong	7
1.2 Vocal sequence learning	15
1.3 Overview of research	16
2. Sequence Variability & History Dependence	19
2.1 Introduction	19
2.2 Data	21
2.3 Methods	21
2.4 Results	26
2.5 Summary	31
3. Sequence Variability & Vocal Learning	33
3.1 Introduction	33
3.2 Sequence learning	35
3.2.1 Methods	35
3.2.2 Results	37
3.3 Acoustic learning	39
3.3.1 Methods	39
3.3.2 Results	42
3.4 Summary	44
4. Discussion	45
5. Bibliography	49

i. Acknowledgements

First, I would like to thank my labmates, present and past; especially Logan James, Erin McKenzie Wall, and Anca Vochin. Logan and Erin showed me the ropes during my first year in the lab. Anca has been generous, and a constant source of humour and moral support.

I would like to thank the members of my committee: Michael Hendricks, Timothy O'Donnell, Jon Sakata, and Alanna Watt. They all provided much-needed feedback on my early progress (and lacks thereof), and thought-provoking questions and encouragement throughout. I have been lucky to attend a reading group organized by Tim on information theory, the content of which was of great use for this thesis. Most importantly, I am grateful to my supervisor, Jon Sakata, for taking me on as a grad student, as a linguist with little background in biology; for his tireless feedback on drafts of this thesis and numerous other documents, which has helped me think more clearly about scientific storytelling, and situate my interests in the field — both of which come less naturally to me.

I would like to thank my family, especially my sister Flannery Jefferson, who taught me to code for this project, and helped me think through my analytical framework for Chapter 3. I am also grateful to Omar Kassouar, without whose help my code for this thesis would have been swallowed up by GitHub, never to be seen again. I am grateful to my parents, Patrick Kennedy and Elizabeth Jefferson, for their inexplicable love and concern for me.

I am grateful to my previous mentors: my middle school science teacher Joe Hockin, who first made me love science; my grade 11 biology teacher, Ilke Van Hazel; my director of studies at Cambridge, Paula buttery, who encouraged me to pursue an honours project on animal communication, and shared in my enthusiasm for whale song; my honours thesis supervisor, Theresa Biberauer; my first research supervisor, Ellen Garland, who took me on for a summer internship in 2017, and taught me to annotate animal acoustic signals.

I am grateful to Glenn Gould, whose music fills me with joy and motivation.

ii. Contributions of the author

The song recordings analyzed in this thesis (described in §2.2) were collected and labelled by Logan James, a recent PhD graduate from the Sakata Lab, for a study on a set of related but complimentary questions (James et al, under revision). All analysis of the data beyond recording and labelling were my own work. Jon Sakata provided feedback on and revisions to all portions of the thesis.

1. Introduction

1.1 Vocal sequencing in birdsong

Many behaviours, both learned and innate, involve the serial ordering of a number of component subroutines (Lashley, 1951). Even for unlearned behaviours, such as the grooming patterns of *Drosophila melanogaster*, there is often variability in the ordering of subroutines, or sequence variability, both between and within individuals (Mueller et al, 2019).

Songbirds produce complex vocal displays (songs) that reuse a fixed inventory of discrete acoustic elements (syllables). Birdsong is not innate, but rather learned during development from a conspecific tutor, often (though not always) the bird's father (Williams, 2004; Brainard and Doupe, 2000, 2002). During this learning process, pupils learn to reproduce not only the acoustics of individual tutor syllables, but also the rules that govern the concatenation of syllables into sequences (Lipkind et al, 2013; Lipkind et al, 2017; James et al, under revision). Whereas syllable sequencing is stereotyped across renditions in many songbird species (e.g., zebra finch, white-crowned sparrow), other species produce songs that are similar to human speech and music in that they recombine fixed inventories of syllables to form sequences that are variable from rendition to rendition (e.g., **Bengalese finch (BF)**, *Lonchura striata domestica*; Okanoya, 2004a). These sequences are not entirely random, however, since which syllable appears next at a given point in the song is statistically influenced by what has come before it (Okanoya, 2004a, 2004b; Hanuschkin et al, 2011; James & Sakata, 2015).

Owing to its variability, BF song sequencing is usually modelled computationally by *stochastic processes*, which generate sequences incrementally by sampling from some probability

distribution over the syllables in the bird's repertoire. One class of simple stochastic models is a **first-order observable Markov model (FOMM)**, which encodes sequences using a two-part mechanism: 1) a set of *states* representing the syllables in a bird's repertoire, and 2) a transition matrix that defines the *transition probabilities* between various syllables. These can be visualized as weighted arcs connecting the states (Figure 1.1).



Figure 1.1. (A) A sample first-order Markov model with transition probabilities. **(B)** The transition matrix for (A), with transitions from a syllable in rows, and to a syllable in columns.

FOMMs have been widely applied to birdsong, and seem to be empirically adequate models for many species, including starlings, whose songs exhibit considerable sequence variability (Kershenbaum et al, 2014; Berwick et al 2011). The fit between a computational model and observed sequence data can be measured by the amount of sequence variability the model leaves unexplained. Information theory allows for the quantification of this unexplained sequence variability through metrics such as **entropy** (Shannon, 1948, 1951). Entropy estimates the average unpredictability of a sequential information source (such as a singing bird), by measuring the average spread of probability mass across possible transitions, in a stochastic model of that information source. Information theory has been applied to a range of animal vocal signals, including dolphin whistle sequences (Ferrer-I-Cancho 2012), humpback whale song (Suzuki 2006), human language (Shannon 1951, Jurgens 2016, Sainburg et al 2019), and birdsong (e.g., Yamashita et al 2008, Palmero et al 2014, Sainburg et al 2019; reviewed in Kershenbaum et al, 2014).

The application of information theory to Bengalese finch song has revealed significant variation in entropy among the different nodes (syllables) in a given bird's song (Sakata & Brainard, 2006; Warren et al, 2012; James & Sakata, 2015). Nodes with higher entropy are referred to as **branch points**, while nodes with lower entropy are referred to as **stereotyped** transitions. In addition to encapsulating the transition statistics of Bengalese finch song sequencing, FOMMs bear structural similarities to the avian brain circuitry responsible for sequence generation.

The brain areas involved in birdsong learning and production have been extensively mapped — much more so than in any other vocal taxa (Brenowitz et al, 1997; Mooney, 2009; Brainard & Doupe, 2013; Jarvis, 2019). The avian premotor nucleus **HVC (acronym used as a proper noun)** is thought to play a major role in syllable sequence generation in songbirds, and seems to be organized into populations of neurons that encode distinct syllable types in the bird's song (Nishikawa et al, 2008). These populations are organized into divergently and convergently connected feedforward circuits (Mooney & Prather, 2005), which are well-suited for the temporal precision necessary for syllable sequencing (Jin, 2009; Hanuschkin et al, 2011; Kornfeld, 2017). Neurons in each successive population are thought to project to neurons in downstream populations (Jin, 2009; Long et al, 2010; Yamashita et al, 2008; Hanuschkin et al, 2011). The synaptic weights between populations are thought to encode the transition probabilities encapsulated by FOMMs (Hanuschkin et al, 2011; Jin, 2014). Neurons also synapse with inhibitory neurons, ensuring that only one successor chain (and thus one syllable) is selected (Jin, 2009). Precisely how the interactions between populations of neurons regulate sequence variability in songbirds like the BF remains elusive.

Behavioural experiments have demonstrated that the sequence variability in BF song allows for short-term plasticity of transition probabilities, even in crystallized adult song. Transition probabilities to specific syllables can be experimentally manipulated by delivering aversive auditory feedback (**white noise, WN**) contingent on that syllable over a number of days (WN days) (Warren et al, 2012; Ali et al, 2013). For example, if two syllables ("d" and "e") can follow the sequence "abc", and a loud WN burst is played whenever the bird produced one of those two options (e.g., "d"), the transition probability to that target syllable will gradually decrease over several days, while the transition probability to "e" increases. This is thought to be achieved by manipulating inter-population synaptic weights (Warren et al, 2012; Jin 2014). Importantly, targeted WN playback experiments activate the same error learning circuitry that guides internally-guided song acquisition during development (e.g., Area X) (Hisey et al, 2018; Xiao et al, 2018); therefore, studying adult plasticity could provide insight into changes in syllable sequencing that naturally occur over development.

The degree of feedback-dependent regulation and plasticity at a branch point during WN experiments in BFs has been shown to be predictable from the baseline sequence variability, or entropy, at the target branch point (Sakata and Brainard, 2006; Warren et al, 2012). Entropy is measured in *bits*, and is highest when transition probabilities are equally distributed among many possible transitions and lowest when transition probabilities are concentrated on one option — i.e. that syllable is certain to occur next. Sakata & Brainard (2006) observed larger changes in transition probabilities in response to auditory feedback perturbations at high entropy sequences than at low entropy sequences. Furthermore, Warren et al (2012) demonstrated that persistent WN playback led to more robust long-term sequence plasticity when transitions at branch points were targeted than when transitions within stereotyped sequences were targeted.

Despite the neural and behavioural relevance of first-order transition probabilities, FOMMs are unable to generate two important patterns attested in animal vocal sequences. The first is variable repetition, where the number of repetitions of a syllable follows a normal distribution (Kershenbaum et al, 2014; Jin & Kozhevnikov, 2011). The second is dependencies between non-adjacent elements (Suzuki et al, 2006; Fujimoto et al, 2011). While the first pattern of variable repetition cannot be remedied using standard Markovian approaches (Jin & Kozhevnikov 2011), the second pattern (statistical relationships that span non-adjacent syllables) can be accommodated by allowing the states of an observable Markov model to encode transition probabilities not from single syllables, but from sequences of *n* consecutive syllables. Such models are called *n*th-order Markov models. For example, Figure 1.2 shows a simple *second-order* Markov model, which tracks transitions between substrings of length two¹ ("bigrams").



Figure 1.2. (A) second-order Markov model, modified from Koumura & Okanoya (2016), with a probability matrix **(B)**, showing transitions from sequences of two syllables (bigrams) to single syllables. For each state in the diagram, the second syllable is the syllable currently being read, and the first syllable is the syllable read before it. Thus a single syllable, e.g. "A", is distributed across three states ("eA", "AA", and "BA") according to the syllable preceding that instance of "A". This allows us to store separate transition probabilities for "A" in each of those three contexts, as shown in (B).

¹ To be strictly second-order, imagine the state labelled "e" has an invisible *start of song* marker before it, as the authors likely do.

Alternatively, we can also allow states to bear a less transparent mapping to syllables or sequences: this is the case for **Partially-observable Markov models (POMMs)** and **Hidden Markov models (HMMs)**, both of which have been used to model BF song sequencing (Jin & Kozhevnikov 2011; Katahira et al, 2011). Whereas observable Markov models require uniformity of the sequence lengths encoded by all states, POMMs and HMMs allow their states to track various context lengths as needed². Therefore, POMMs and HMMs allow for more compact representations of data. They are also less susceptible to problems of data scarcity, since the number of conceivable substrings of *n* syllables that need to be modelled by an *n*th-order observable Markov model grows exponentially with *n*.

In addition to the finding that higher-order modelling of syllable BF song leads to more accurate models of syllable sequencing, the tracking of syntactic context beyond the current syllable has received support from both behavioural and neural experiments. In previous WN playback experiments, where one transition from a branch point is consistently targeted for WN playback, Warren et al (2012) showed that *history-dependent* transitions — where the next syllable can be predicted based on the syllable *before* the current syllable — are less plastic than transitions without this history dependence, even though both appear variable at the first order. In the example provided in Figure 1.3a, the branch point "a" is more likely to be followed by "b" if it was followed by "c" last time it occurred, and vice versa (in other words, the sequence transitions from "a" alternate between "ab" and "ac"). Across several such branch points, plasticity ("percent learning") correlated inversely with history dependence (Figure 1.3b). In other words, past transitions were found to influence not only the sequence statistics of future transitions, but also their plasticity.

² The distinction between POMMs and HMMs is that POMMs require a given state to encode only sequences ending in a single syllable, while HMM states need not meet this requirement and thus bear an even less transparent mapping to syllables.



Figure 1.3. Illustration of history dependence, from Warren et al (2012). **(A)** An example history-dependent branch point. The current syllable is "a", and transitions to "b" and "c" are influenced by what precedes it (i.e., "ab" or "ac"). **(B)** The relationship between history-dependence of branch points and their plasticity during WN playback experiments.

In addition to behavioural experiments, neural recordings at various sites in the song system have also suggested that BFs integrate over multiple syllables; however, there is some discrepancy as to the length of context encoded. Fujimoto et al (2011) studied the activity of HVC neurons projecting to the basal ganglia nucleus, Area X, a brain area involved in vocal plasticity. These neurons (HVC_x neurons) are thought to play a role in encoding learned syntactic rules of the sort tracked by the transitions in Markov models. They show burst spikes that are phase-locked with singing, with particular populations marking the onsets of particular syllables. In some populations, activity was impacted by upstream syllable arrangements in remote segments of the song, while in others, burst spikes encoded the identity of a specific syllable, independently of the sequential history before it. Still other neurons were found to fire selectively for specific inter-syllabic transitions. Interestingly, Fujimoto et al (2011) report that while some neurons responded differentially based on the previous one or two syllables, none responded to longer trajectories.

Sequential context also seems to impact activity in motor regions downstream of HVC, particularly in the robust nucleus of the arcopallium (RA). RA projects directly to motor neurons controlling the muscles of the syrinx (the vocal organ in birds). Bursts in RA neurons are driven by projections from HVC (and other song nuclei), and bursting in HVC RA-projecting (HVC_{RA}) neurons is time-locked with RA bursts and with song syllables (Fee et al, 2004). Wohlgemuth et al (2010) found that RA activity in BF song differs significantly for instances of the same syllable in different contexts. They observe larger effects from the convergent (preceding) than the divergent (following) context to a given syllable. Sequential context was also found to impact the acoustic rendition of a given syllable, with preceding context (convergence) conditioning acoustic variation more reliably than following context (divergence). In human speech, the acoustic realization of a phoneme can also be conditioned by the context preceding or following it — though the trend for directionality seems to lean the other way in humans: anticipatory assimilation (akin to divergence) is more common cross-linguistically than perseveratory assimilation (akin to convergence) (Javkin, 1977). Interestingly, Wohlgemuth et al (2010) found that phonetic variation could be conditioned by syllables as distant as three syllables back, a context that is somewhat longer than reported by Fujimoto et al (2011), but confirms the integration of sequential context over more than the preceding syllable.

The encoding of sequential context beyond two previous syllables is further supported by neurophysiological evidence from a perception study. Bouchard and Brainard (2013) recorded from HVC interneurons while BFs listened either to the **bird's own song (BOS)** or to synthetic BOS, in which BOS syllables had been spliced into novel sequences. A given syllable was found to elicit a stronger response profile if it was preceded by a context that often precedes it. Moreover, the impact of context for some syllables was significant up to 9 syllables (~1000 ms) before the current syllable, with the impact tending to decrease monotonically with each additional syllable.

Transition probabilities for a syllable are usually calculated using the probability distribution of transitions *from* that syllable (*divergence probabilities*). However, a complementary way of

analyzing vocal sequencing is to compute the variable of transitions *to* individual syllables. The latter are termed *convergence probabilities*, and are simply the transition probabilities after reversing sequencing within all the songs in a corpus ($abc \rightarrow cba$). Interestingly, Bouchard & Brainard (2013) found that convergence probabilities are better predictors of a contextualized syllable's HVC activation than its divergence probabilities. A syllable elicited the strongest response strength in HVC when divergence probability was highest — in other words, when it occurred in the context it occurs in most often.

1.2 Vocal sequence learning

While there has been considerable investigation of acoustic learning in songbirds, less is known about the learning of syllable sequencing. In particular, in species that exhibit sequence variability, the effects of this variability on learning and the degree to which it is replicated between tutors and pupils is poorly understood. A bird's sequence variability tends to decrease with age, such that juvenile birds' songs tend to become less entropic over time (Tchernichovski et al, 2004; James & Sakata, 2015).

This finding is congruous with proposals that sequence learning consists, in large part, in synaptic pruning. Hanuschkin et al (2011)'s proposal for the neural implementation BF song syntax starts with all-to-all connectivity between chains corresponding to individual syllables. Selective connectivity, and hence decreased sequence variability, emerge from subsequent strengthening of some excitatory pathways over others.

However, Lipkind et al (2013) argue based on behavioural evidence that the opposite seems to be true: that during learning in BFs, as well as zebra finches and human infants, transitions between segments are acquired in a stepwise manner, rather than all at once and subsequently being pruned. Thus attaining high sequence entropy at a syllable is the result of extensive development, rather than the default state. To implement transitions between syllables, Lipkind et al (2013) propose that the brain must first create chains corresponding to different phonetic segments, and

then a means by which neuronal activity can propagate from the tail of one chain to the head of another.

Lipkind's findings sit uneasily with the finding that transition entropy tends to decrease, rather than increase, with age (Tchernichovski et al, 2004; James & Sakata, 2015). If, as Lipkind et al claim, the number of bigrams beginning with a given syllable increases during maturation, transition entropy should increase, rather than decrease with age. Further work is required to reconcile these two findings.

Pupils' sequence statistics bear a significant resemblance to those of the song they were tutored with (Yamashita et al, 2008; James et al, under revision), even if the tutor is not the pupil's biological father (James et al, under revision). Sequence statistics of the tutor seem to impact some aspects of sequence learning in pupils. Sequence learning is not perfect, and variation in the fidelity of sequence learning is observed across multiple sequences within a bird, as well as across birds. However, it is not known to what extent sequence variability in the tutor impacts sequence learning success in the pupil, or the ability of the pupil to learn and reproduce the acoustic features of individual syllable types.

1.3 Overview of research

Taken together, both behavioural and neural evidence supports the encoding of sequential context over more than the preceding syllable during sequence generation. In this thesis, I investigate the extent of higher-order sequencing in BF song, and assess how sequence variability could impact learning. This capitalizes on the fact that songbirds are one of a small number of vertebrate taxa that learn their vocalizations over development (see Jarvis, 2019 for a recent review), as well as the fact that BF songs exhibit considerable sequence variability.

In Chapter 2, I use various nth-order observable Markov models to track long-range statistics in BF song. In systems where long-range dependencies exist, the *entropy rate* (average entropy per syllable) tends to decrease as the order of the Markov model is increased (Shannon, 1951), since

increasing the model's "memory"³ leaves less sequence variability unexplained. I analyzed how entropy rate in BFs changes with Markov model order. I hypothesized that entropy rate would decrease with order, assuming that some degree of high-order sequence processing is universal in BF song (Okanoya, 2004a; Okanoya, 2004b; Jin & Kozhevnikov, 2011). I also anticipated diminishing returns for increases in order, given the asymptotic relationship between context length and HVC response strength in Bouchard & Brainard (2013).

It is known that Bengalese finches learn sequence statistics from their tutors, but with fidelity that varies within and between individuals (James et al, under revision). However, little is known about the role of the tutor's sequence variability in this learning process. Thus a second line of inquiry in my thesis (Chapter 3) concerns the role played by the sequence entropy of a tutor syllable in the learnability of that syllable by the pupil. In order to encompass two major dimensions of learning in my analysis, I quantified both acoustic learning — the similarity of the sequence learning — the similarity of the probability distributions for that syllable in the tutor and pupil's songs.

Bouchard & Brainard (2013) observed more robust neural activations in response to syllables when they occurred in more common contexts. Although this was found in adults listening to their own song, I conjectured that a similar effect of convergence statistics might influence learning in juveniles. In particular, I hypothesized a negative relationship between the entropy of a syllable in the tutor's song (i.e., the variability of the sequence in which that syllable is embedded), and the success with which his pupil learns the acoustic and sequential properties of that syllable. I specifically predicted that pupils would learn the acoustic and sequence structure of a syllable better if it is embedded within a stereotyped (low-entropy) sequence. Additionally, given that convergence (backwards) sequence statistics predict neural activation in HVC more strongly than divergence statistics (Bouchard and Brainard, 2013), I predicted that *convergence entropy* (an estimate of the degree of sequence variation *to* a focal syllable; Alliende et al, 2017) would predict learning better than divergence entropy.

³ I use this term loosely. Markov models do not have memory (e.g. a stack or a register), but they are able to distinguish between one element embedded in various contexts, by representing them in different states.

In Chapter 4, I review my findings and discuss the degree to which they support existing models of vocal production, development, and plasticity. In addition, I propose new experimental lines of inquiry to address questions left unanswered by my analysis.

2. Sequence Variability & History Dependence

2.1 Introduction

In human language, the occurrence of an element at one place in a sequence is often dependent on, and statistically constrained by, what has come before it (Chomsky, 1957; Saffran, 2002). Such dependencies are thought to be due to several distinct, hierarchically nested processes: for example, phonological (between-sound) dependencies are usually local and can be captured using simple Markovian processes (Heinz & Idsardi, 2013), while syntactic (between-word) dependencies often span longer distances and cannot be efficiently captured by Markov models (Chomsky, 1957; Jurafsky & Martin, 2000). Statistical dependencies have also been found to be widespread in learned animal vocal sequences, including birdsong (Suzuki et al, 2006; Noad et al, 2015; Katahira et al, 2011; Kershenbaum et al, 2014). In birdsong, statistical dependencies spanning non-adjacent syllables are referred to as **history dependence** (Jin & Kozhevnikov, 2011; Warren et al, 2012; Wohlgemuth et al, 2010; Fujimoto et al, 2011).

Information theory allows for quantification of the amount of sequence variability that is unexplained by a given model, through metrics such as entropy. Information theory has been used to support for the existence of both local and long-range dependencies between vocal elements: for example, in humpback whale song (Suzuki et al, 2006), in dolphin whistles (Ferrer-i-Cancho & McCowan, 2012) in four songbird species (Sainburg et al, 2019), and in humans (Shannon, 1951; Sainburg et al, 2019)⁴.

While it is known that Bengalese finches consistently produce songs with variable but history-dependent sequencing (Jin & Kozhevnikov, 2011; Warren et al, 2012; Bouchard & Brainard, 2013), little is known about inter-individual variation in sequence variability and

⁴ The first and third studies mentioned employ mutual information (I), another metric from information theory. Specifically, the mutual information of two random variables X and Y is the reduction of uncertainty in one from knowing the other (Cover & Thomas, 2016); while the conditional entropy of X given Y is the uncertainty that remains after that reduction has been made.

history dependence, or the transmission of sequence variability from tutor to pupil during the vocal learning process. For this reason, I calculated entropies for a dataset of 29 birds (9 tutors and 20 pupils) using Markov models of orders 1 through 5, to capture the relationship between the span of sequential context a model has access to, and the amount of sequence variability it is able to explain. In addition, I examined the degree to which the entropy of tutors was transmitted to their pupils during vocal learning.

In line with findings that Bengalese finch song sequencing is history-dependent (Jin & Kozhevnikov, 2011; Warren et al, 2012; Wohlgemuth et al, 2010; Fujimoto et al, 2011), I predicted that as Markov models were given access to longer sequences on which to condition their predictions, the amount of unexplained sequence variation (entropy) would decrease. Thus I predicted a negative relationship between entropy and Markov order, for all birds. It should be noted that entropy can never *increase* between subsequent Markov orders (Cover & Thomas, 2006; Smith, 2016) — however, in a system with no history dependence (e.g. randomly scrambled letters), it could remain constant across Markov orders, or decrease only slightly.

More specifically, I predicted that Markov order would decrease entropy with diminishing returns: that is, the largest decrease in entropy would be between the first and the second order, followed by that between the second and third, and so on. This is due to findings by Bouchard & Brainard (2013) that BFs integrate over at least 7 syllables during perception of the bird's own song, but with diminishing returns.

A final prediction was that entropy would not decrease to 0 bits in any bird — that is, that there would be some residual sequence variability in every bird that could not be explained even by the highest-order Markov model analyzed. This would support the assumption that Bengalese finch syllable sequencing is inherently variable, possibly due to unstable bifurcation points in BF neural circuitry, from which two or more trajectories are possible (Yamashita et al, 2008; Jin, 2009; Hanuschkin et al, 2011).

2.2 Data

For my analysis of entropy and learning biases, I used a dataset of songs from 29 male BFs from 9 different nests (9 adult tutors, 20 juvenile pupils). All tutors were the biological fathers of their pupils, and tutors and pupils were housed together in a semi-natural acoustic environment, where neighbouring birds could be heard but not seen. Song recordings were made in sound-attenuating boxes, and songs were segmented into syllables using a constant amplitude threshold. Individual syllables were then classified manually, in isolation from the contexts in which they occurred. Categories were agreed upon by independent raters, as described in James et al (under revision).

There were an average of 26 song recordings per bird (range: 16-42 songs), with a mean corpus length of 1,745 syllables per bird, and a total corpus length of 52,372 syllables.

2.3 Methods

In order to estimate entropy values for a particular Markov order n, I first calculated the conditional probability for each state in the Markov model, x, which corresponds to a single syllable in a first-order model and a string of n syllables or **n-gram** in an nth-order model. For each **transition** y from x, the probability of y given x is calculated from a corpus using the following formula:

$$p(y \mid x) = \frac{p(x,y)}{p(x)}$$

, where p(x,y) is the joint probability distribution, or the probability of the substring *xy*. The conditional probability distribution for a bird's song is encoded by a **transition matrix**, which lists the conditional probabilities of various transitions, which sum to 1 for each n-gram. We can calculate the uncertainty or **entropy** *H* at each n-gram *x* by taking the expected value of the *surprisal* of that n-gram, where the surprisal of each transition *y* is the negative \log_2 of the conditional probability of that transition (Cover & Thomas, 2006). The expected surprisal or entropy of an n-gram is taken by weighting each transition *y*'s surprisal by its likelihood of occurring:

$$H(x) = -\sum_{y \in Y} P(y|x) * log_2 P(y|x)$$

, where *Y* is the set of individual transitions *y* from *x*. This formulation of entropy or *transition entropy* has been widely used as a metric for sequence variability for specific syllables or n-grams in Bengalese finch song (James & Sakata, 2015; Sakata & Brainard, 2006; Hanuschkin et al, 2011).

The **entropy rate** of a bird's song is calculated by taking a weighted average of the entropies for each n-gram, weighted by the unconditioned probability of that n-gram (how often it occurs relative to other n-grams of the same length). Entropy rate allows us to measure and compare the entropies of individual birds' songs as wholes, rather than the heterogenous n-grams that make them up.

To investigate the effect of context length on entropy, I calculated each bird's entropy rate with respect to Markov models of order 1 through 5, using Python software I developed for this project (available on GitHub). The fifth order was chosen somewhat arbitrarily as the maximum order for analysis, due to data scarcity problems that worsen as order increases. For human language, it is estimated that the number of possible meaningful English sentences of n words is around 10^n (Pinker, 2004). In Bengalese finches, this growth is less steep, but still significant: the mean number of unique observed n-grams for different values of n, across all birds, is shown in Figure 1.



Figure 2.1. The average number of unique observed n-gram types as a function of Markov order, across all birds in the dataset.

More to the point, as n increases, the token frequency of the least frequent n-grams decreases (Ha et al, 2003). For example, n-gram types that are only observed once, or *hapax legomena*, increase in proportion to all n-grams. This generally exerts a downward bias on entropy estimates (Miller, 1955; Panzeri & Treves, 1996; Smith, 2014), because one is likely to underestimate the diversity of contexts in which an n-gram occurs, given fewer observations. For example, the entropy estimate of a hapax legomenon (an n-gram observed once) is necessarily 0 bits, since only one transition follows it with a probability of 1, and entropy is zero for certainty. In order to estimate the extent of this bias, I calculated the average proportion of n-gram tokens across all birds in the dataset with less than 5 or 10 tokens, as a function of Markov order (shown in Table 2.1).

Table 2.1. Proportion of n-gram tokens across all birds belonging to n-gram types with more than 5 and more than 10 tokens with which to estimate entropy values. As order increases, the proportion of n-gram tokens belonging to rare n-gram types increases.

	>5 tokens	>10 tokens
1st order	99.9%	99.8%

2nd order	99.1%	98.1%
3rd order	97.5%	95.1%
4th order	95.3%	91.6%
5th order	92.5%	87.3%

In order to reduce the downward bias exerted by rare n-grams, especially at higher orders, I built a minimum count threshold into my entropy estimator, such that n-grams with fewer than that threshold of tokens could be excluded from the final entropy rate estimate for that bird. However, the exclusion of rare n-grams could introduce a bias of its own, if rare n-grams have systematically different entropies than more frequent ones. Thus I calculated entropy rates for all birds, for orders 1 through 5, both with and without a minimum count threshold of 10, and observed the impact on estimates (Table 2.2).

Table 2.2. Decreases in entropy rate observed between threshold-free estimates and estimates with a minimum count threshold of 10, across Markov orders 1-5. Averaged across all 29 birds.

	Entropy decrease
1st order	0.02%
2nd order	0.60%
3rd order	2.29%
4th order	4.89%
5th order	7.86%

Since (following Table 2.1) proportionately more n-grams are observed a small number of times as order increases, the discrepancy between estimates was highest at the fifth order, and near 0 for the first order (nearly every syllable occurred more than 10 times).

To directly assess the effect of corpus size on my entropy rate estimates, I calculated entropy rate using subsets of the entire corpus, incrementally increasing the corpus size by 10 songs after each estimate and observing the effect on entropy estimates. Again, low-order estimates were fairly robust to increases in sample size, while higher-order estimates showed an upward trend. Figure 2.2 shows entropy rate profiles after successive incrementations of the sample size by ten song files for one bird, starting with a corpus of a single song (in red), and incrementing up to a total of 50 songs (3,975 syllables).



Figure 2.2. Entropy rate estimates for one bird after successive incrementations of corpus size by 10 songs.

While successive additions of songs continued to raise entropy values throughout, they did so with diminishing returns, and even the fifth-order values appear to be converging toward a limit (points become increasingly close together). However, all high-order entropy estimates used in this thesis come with the caveat that they are not entirely robust to changes in sample size, and so are intended primarily for comparisons between birds, rather than as absolute metrics of

sequence variability. It bears repeating that, on average, 1,745±97 syllables were analyzed per bird (range:845-2,828).

2.4 Results

I calculated entropy rates for all 30 birds in the dataset, for Markov orders 1 through 5. Entropy rates decreased monotonically with model order for all birds(Figure 2.3). The mean decrease in entropy rate between the first and fifth orders was $46\% \pm 2.6\%$ (range: 18%-69%).



Figure 2.3. Entropy rate as a function of Markov model order across the 29 birds studied (solid lines, colour-coded by nest). The black dashed line shows the mean entropy rate across birds.

The relationship between the mean entropy rate across birds (H) (see dashed line in figure 2.3) and Markov order (n) is best fit by a transformed multiplicative inverse (y = 1/x) function,

$$H = 0.32 + \frac{0.41}{n}$$

with a correlation coefficient of 0.98. This suggests that entropy rates asymptote at an average of \sim 0.3 bits (the y-intercept of that equation), but further work is required to control for the bias that sample size might exert on estimates of entropy rate (i.e., this estimate might be lower than the true entropy rate). As predicted, the entropy rate did not reach 0 bits for any of the 29 birds, even at the 5th order: the lowest 5th-order entropy rate was 0.20 bits. This means that for every bird, there was residual sequence variability that could not be perfectly explained by taking the previous five syllables into account.

Also of note is that the standard deviation in entropy across birds is twice as large at the first order than it is at the fifth (Figure 2.4). At the first order, the standard deviation of entropy estimates is 0.22, while at the fifth order, the standard deviation was 0.10. This could be due to a lower bound on high-order entropy, due to processing constraints (see Chapter 4 — Discussion).



Markov Order

Figure 2.4. Entropy rate as a function of Markov Order. Each dot represents the value for a bird at a specific Markov Order. Standard deviations across orders are: first: 0.22; second: 0.14; third: 0.12; fourth: 0.11; fifth: 0.10.

Since songbirds like BFs learn sequence statistics from their tutors, I predicted that some of the between-bird variation in entropy rate could be explained by the nest a bird belonged to. At the first order, the entropy rate of pupils (sons) did not significantly correlate with that of their tutor's song. However, when I compared the entropies of individual syllables in the tutor's song with entropies of corresponding syllables in their pupil's songs, I did find a significant positive correlation (F_{148} =45.0, p<.0001).



Figure 2.5. The entropy for syllables in the pupils' songs as a function of that syllable's entropy in their tutor's song.

As a general observation, the entropy of a pupil's song tended to be higher than that of his tutor's song. This is likely because pupils were recorded as young adults (4-5 months), while their tutors were substantially older (>2 yrs) at the time of recording, and because sequence entropy at particular branch points tends to decreases with the age of a given bird (James and Sakata, 2014, 2015).

There was considerable variability in tutor-pupil entropy differences across the nine nests examined. Figure 2.6 plots the entropy rate for the tutor (dotted line) and pupils (solid lines) from each nest, for Markov orders 1 through 5. While nests A through D support the trend for higher

entropy in pupils' songs, nests E through H show a less straightforward relationship: pupils' entropy values straddle those of the tutor, with some falling above and some below, depending on the order. Nests J and I show the opposite trend, with pupils producing more stereotyped syllable sequences than their tutors.



Figure 2.6. Entropy learning within and across nests (n=9 nests). Entropy rate for each bird is measured on the y-axis, and Markov model order is on they x-axis. Dashed lines depict data for tutors, solid lines depict data for pupils.

I hypothesized that this variability in entropy differences across nests might be due to a reversion toward the mean, whereby higher-than-average entropy rates are lowered and lower-than-average rates raised during the learning process. To test the extent to which this kind of reversion could explain between-nest variability in entropy change, I correlated the tutor's entropy relative to the mean of all tutors with the difference between the tutor's entropy and the pupil's (for each pupil), and found a significant negative correlation at each of the 5 Markov orders (Figure 2.7).



Figure 2.7. The relationship between tutor entropy rate (relative to the mean of all tutors), and the change in entropy rate from tutor to pupil, across 5 Markov orders for entropy estimation. The negative relationship observed suggests that pupil entropies tend to revert toward the mean during the vocal learning process.

In other words, the songs of pupils whose tutors had higher-than-average sequence entropy and those with lower-than-average entropies were reverting toward the population mean during learning, similar to the inheritance of normally distributed phenotypic traits such as human height (Galton, 1886)⁵. Interestingly, this negative correlation was even more robust when calculated on second-order (and other higher order) entropy values (r=-0.68, p=0.001; Figure 2.7b). This suggests that higher-order statistics may play a more central role in this reversion process than first-order statistics, in line with previous findings that BFs track high-order statistics of songs in production and perception (Jin & Kozhevnikov, 2011; Warren et al, 2012; Bouchard & Brainard, 2013).

2.5 Summary

By comparing the entropy rates of 29 birds across Markov model orders 1 through 5, I found that all birds showed decreases, suggesting some degree of history-dependence beyond a single previous syllables in their songs. I found that while pupil's songs tended to be more entropic than their pupils', confirming previous findings (Tchernikovski et al, 2004; James & Sakata, 2015), not all nests displayed this trend. I found that some of this variance was explained by whether the tutor's entropy fell above or below the population mean, suggesting that pupils' entropies tend to revert toward the mean during vocal learning. This effect was particularly strong for higher-order entropy values. Additionally, birds clustered more closely around the mean as Markov order increased, suggesting that an upper and/or lower bound might be at play for those values. Potential mechanisms for this effect are discussed in Chapter 4.

⁵ I make this comparison between the statistical phenomena, not the mechanisms — there may or may not be a genetic component to sequence variability.

While the relationships between the entropies of tutors and pupils is noteworthy, it does not unambiguously point to the reproduction of the tutor's sequence statistics in the pupil's song, since a single entropy value can be achieved by many different probability distributions. For example, an n-gram "a" with transitions to "b" (P=1/4) and "c" (P=3/4) has the same entropy (0.81 bits) as an n-gram "d" for which the two probabilities of "b" and "c" have been switched. In other words, entropy estimates capture variability, but are blind to the particular statistics of one transition or another. For this reason, in Chapter 3 I directly compare sequence statistics (i.e., transition probabilities between syllables) in tutors and pupils, in order to observe the role of tutor sequence variability on sequence learning.

3. Sequence Variability & Vocal Learning

3.1 Introduction

During development, songbirds learn to reproduce both the acoustic features of individual syllables from their tutor's song and the rules that govern the sequencing of those syllables. However, neither learning process is perfect. To start with, while pupils' songs are significantly similar to their tutors', both acoustically and syntactically, only a subset of a tutor's syllable repertoire is usually replicated in a recognizable form in the pupil's song — thus we can subdivide the tutor's syllable inventory into *retained* and *dropped* syllables with respect to one of his pupils (e.g., James et al, under revision, Hough & Volman, 2002; Zevin et al, 2004). In addition, pupils may introduce *novel* syllables that were not present in the tutor's song. This three-way distinction is illustrated in Figure 3.1.



Figure 3.1. (from James et al, under revision). Three categories of syllables. Only retained syllables — those that occur both in the tutor and pupil song — were analyzed here.

In order to study acoustic and sequence learning on a syllable-by-syllable basis, I limited my analysis of learning success to retained syllables, since there are no clear grounds for tutor-pupil comparison in either dropped or novel syllables. James et al (under revision) studied the similarities between tutors (fathers) and their pupils (offspring) in semi-natural breeding conditions and report that on average, 59% of Bengalese finch tutor syllables are retained by their pupils. Conversely, 74% of syllables in the pupil's repertoire consisted of syllables that were retained from the tutor's song. Even within retained syllables, learning outcomes (e.g., the degree of acoustic similarity, or the similarity of the sequences in which these syllables are embedded) vary between pupils, as well as between specific syllables and sequences from the tutor's song. I sought to explain some portion of this variability by investigating the sequence variability (entropy) in which various syllables were embedded.

For each retained syllable type in each bird, I quantified the sequence complexity (entropy), using the method described in §2.3. Given previous findings that neurons in the Bengalese finch brain are sensitive to both divergence statistics (what follows that syllable) and convergence statistics (what precedes it) (Bouchard & Brainard, 2013), I investigated both aspects of sequence statistics in the tutor's song, yielding estimates of **divergence entropy** and **convergence entropy**. Divergence entropy measures the diversity of syllables that follow a given syllable, whereas convergence entropy captures the diversity of syllables that precede it. Divergence entropies were calculated by first reversing the order of all song sequences (*abc* \rightarrow *cba*), then calculating entropy using the same method described in §2.3.

Only first-order entropies (for individual syllables) are computed and analyzed here, for the sake of retaining a viable sample size: the proportion of polysyllabic sequences (n-grams) retained in the pupil's song from the tutor's is far smaller than the proportion of single syllables retained, since if an n-gram contains even one dropped syllable (in the tutor) or one novel syllable (in the pupil), it cannot be considered retained in its entirety, making grounds for comparison less clear.

Bouchard and Brainard (2013) found that the transition probability of an embedded syllable positively predicted response strength during perception of a bird's own song. Although this relationship used *probabilities* of embedded syllable tokens, rather than convergence entropy, the two are related metrics. Because the entropy of a syllable is the expected value of the log probability to that syllable, the two will tend to have a negative relationship. Thus syllables with high convergence entropy tend to have lower convergence probabilities, and vice versa. For this reason, I predicted that learning would be better for syllables with lower entropy — that is, syllables whose sequential context is stereotyped across renditions.

One caveat to generalizing Bouchard & Brainard's findings to learning is that the relationship they observed was in adult birds listening to their own song, whereas learning involves juveniles listening to a song for whose sequencing, at least at first, they have not yet developed statistical representations. However, it seemed plausible that a similar mechanism might be at play.

A secondary prediction originates from the finding that the convergence probability of an embedded syllable more closely predicted response activations than its divergence probability (Bouchard & Brainard, 2013). Thus I predicted that convergence entropy would more significantly impact learning success than divergence entropy.

3.2 Sequence learning

3.2.1 Methods

In order to investigate the effect of tutor sequence variability on sequence learning, I quantified sequence learning success between corresponding tutor and pupil syllables using three complementary measures of *sequence dissimilarity*, all roughly measuring the difference between the first-order transition matrix for each syllable in the tutor's song, *P*, and the transition matrix of the corresponding retained syllable in the pupil's song, **Q**. These are described below:

1. **Euclidean distance.** This measures the amount of probability mass that has to be redistributed among transitions to transform the tutor's probability distribution for that syllable

into the pupil's (or vice versa, symmetrically). It is the sum, over all possible transitions from that syllable, of the absolute difference between the tutor's probability for that transition and the pupil's. If a transition from a given syllable is present in the tutor's song but absent in the pupil's, or vice versa, a value of 0 is substituted for the absent transition's probability, so the addition of a novel transition and the dropping of an existing one are equally weighted. The Euclidean distance is given by:

$$D_E(P,Q) = \sum_{x \in X} |P(x) - Q(x)|$$

, where X is the union of the transitions x in the tutor's transition matrix (P) and the pupil's (Q).

2. **Log distance**. Since entropy scales logarithmically with changes in outcome probabilities, and the correlation techniques used assumed a linear relationship, I also computed the Euclidean distance described above the log probabilities for each transition:

$$D_L(P,Q) = \sum_{x \in X} \left| log(P(x)) - log(Q(x)) \right|$$

3. **Kullback-Leibler Divergence (D**_{KL}) (Kullback & Leibler, 1951). D_{KL} is drawn from information theory, and measures the expected loss in encoding efficiency from using a code based on the parent distribution P when the true distribution is Q. The formula used is as follows:

$$D_{KL}(P || Q) = \sum_{x \in X} P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

Unlike the other two metrics used, D_{KL} is not symmetric: that is, $D_{KL}(P \parallel Q)$ does not necessarily equal $D_{KL}(Q \parallel P)$. The tutor's probability distribution was used as P, and the pupil's as Q.

All three metrics were strongly correlated. In particular, D_{KL} correlated with Euclidean distance with a coefficient of 0.91, suggesting that the two were comparable metrics of sequence

similarity. Log distance correlated with Euclidean distance with a coefficient of 0.69, and with D_{KL} with a coefficient of 0.64.

In choosing a statistical test for the relationship between entropy and sequence learning, I deemed that the syllable repertoires of different birds are not necessarily independent data points with regards to sequence entropy and acoustic features. This is because many of the birds' songs shared a common vocal ancestor, one generation or more in the past. Thus a small patterning of sequence entropy and learnability due to chance could have propagated through the vocal learning process. To reduce pseudoreplication, I analyzed the relationship between sequence entropy and sequence learning using a Linear Mixed-effects Model (LMM), with the pupil's bird ID as a random effect.

3.2.2 Results

Regression analysis using a Linear Mixed-effects Model revealed that none of the three metrics for sequence distance bore a significant relationship with convergence entropy ($F_{1,146}=2.0$, p=0.158). Divergence entropy, on the other hand, was found to be significantly and positively related to both Euclidean Distance ($F_{1,143}=9.8$, p=0.0022; Figure 3.2) and Log Distance ($F_{1,143}=63.0$, p<0.0001), and bore a near-significant positive relationship with D_{KL} ($F_{1,142}=2.7$, p=0.10). Results for Euclidean Distance are plotted in Figure 3.2.



Figure 3.2. Relationships between tutor first-order entropy and Euclidean distance, a measure of sequence learning, for both divergence entropy (**A**) and convergence entropy (**B**).

Because the Euclidean distance inversely reflects sequence similarity, these results suggested that the sequence statistics of the tutor's song were learned more accurately for low-entropy (stereotyped) transitions than for more entropic transitions ("branch points"). Indeed, this binary delineation (branch points vs stereotyped sequences) aptly described the data with regards to entropy. The distributions of both convergence and divergence entropy values among tutors appeared to be bimodally distributed. Histograms of both distributions (Figure 3.3) revealed a primary peak just above 0 bits for both divergence and convergence entropy, and a secondary peak between 1.1 and 1.2 bits for divergence entropy and between 0.9 and 1.0 bits for convergence entropy. Both distributions are sparsely populated between 0.4 and 0.8 bits, where the population means (~0.5 bits) lay for both distributions.



Figure 3.3. Histograms for divergence entropy **(A)** and convergence entropy **(B)** values across all syllables in all tutor songs.

For this reason, I performed a second analysis where I binned entropy into "High" and "Low", with 0.6 bits as the dividing line. This analysis yielded comparable results to when scalar entropy values were considered in the analysis: Euclidean distance ($F_{1,139}$ =6.8, p=0.01) and Log distance ($F_{1,139}$ =52.9, p<0.0001) remained significantly related to divergence entropy, and there was a trend for D_{KL} to be different between High and Low entropy transitions in the tutor's song.

3.3 Acoustic learning

3.3.1 Methods

Within a pupil's retained syllable repertoire, there is variation in how well the acoustic features of the corresponding tutor syllable are reproduced. As with sequence learning, I hypothesized that sequence entropy might explain some of this variation, with better learning outcomes for more stereotyped (low-entropy) syllables. In order to quantify acoustic learning of syllable types between tutors and pupils, I measured the **acoustic distance** between the tutor's and pupil's median renditions of a given syllable, for all retained syllables. To calculate acoustic distance, I

first extracted five acoustic features for each rendition of each syllable, across all birds in all nests, using <u>autolabel</u>, an open-source features-based syllable classifier. These features are:

- 1. Mean frequency
- 2. Spectral density entropy
- 3. Duration (in milliseconds)
- 4. Loudness entropy
- 5. Spectral-temporal entropy.

Median values were then calculated for each syllable type in each bird (whether retained or not). A principal component analysis (PCA) was then performed on the median values for all birds (tutors and pupils), in order to create a standardized multidimensional space in which acoustic comparisons between retained syllable types could be made across all tutor-pupil pairs. Principal component analysis is often used to visualize how syllable tokens within a dataset cluster into types, by reducing acoustic features into a low-dimensional acoustic space (e.g. Dunlop et al, 2007; Wohlgemuth et al, 2010). Figure 3.4 shows the distribution of syllable tokens for retained syllable types in a sample tutor-pupil pair from my dataset.



Figure 3.4. Exemplar distributions of syllable renditions plotted in a 2D PCA space, for 6 syllable types in a tutor's song (**A**), and the 6 corresponding retained syllables in the pupil's song (**B**). White circles are the centroids of the clouds. The PCA for this figure was conducted on a dataset that include acoustic features for every rendition of retained syllables between a tutor and one of his pupils.

Note that some corresponding clusters (such as syllable "b", in green) occupy roughly the same portion of PCA space in the tutor and the pupil, while others (such as syllable k, in magenta) have been displaced in the pupil relative to the tutor.

In order to analyze these differences ("acoustic distances") between corresponding syllable types in tutors and pupils, I performed a PCA not on syllable tokens, but on the median feature values for all syllable types in all birds. Before computing principal components, I first transformed all feature medians for all syllables into *z*-scores, so that the PCA would not be biased toward higher-variance features. I then calculated all 5 principal components for the data using the decomposition.PCA module of the Python machine learning library scikit-learn. For each pupil in each nest, I then found the set of retained syllables — i.e. those tutor syllables that also occur in the pupil's song (see Figure 3.1). In order to quantify acoustic learning for each retained syllable, I calculated the Euclidean distance (in PCA space) between the tutor's and pupil's centroid for that syllable, as illustrated in Figure 3.5.



Figure 3.5. The centroids for retained syllable types in the tutor (black) and the corresponding centroid for the pupil (white) plotted in PCA space. The distance between corresponding centroids illustrates the measure of acoustic distance used here. N.B. The distances here are shown in a 2-dimensional PCA space, for ease of visualization. Actual distances were calculated across all five PCA dimensions.

The Euclidean distance between syllable centroids in PCA space served as my measure of the acoustic distance between the tutor and pupil's version of the syllable. I then fit two linear mixed-effects models to the data with Tutor entropy (first divergence entropy, then convergence entropy) as independent variables, and Bird ID as a random effect.

3.3.2 Results

Unlike sequence learning, acoustic learning was not significantly related to divergence entropy ($F_{1,141}$ =0.6, p=0.4354). However, it was significantly and positively predicted by convergence entropy ($F_{1,140}$ =5.2, p=0.0243; Figure 3.6).



Figure 3.6. Relationships between divergence **(A)** and convergence **(B)** entropy of transitions in the tutor's song and acoustic distance, a measure of acoustic learning in the pupil.

In other words, the more entropic the context preceding a syllable in the tutor's song, the more accurate the acoustic imitation of the syllable by the pupil. This ran contrary to my hypothesis that learning would be better for stereotyped syllables.

Several possible confounds could mediate this apparent link between convergence entropy and acoustic learning success. One is the proportion of the number of times that a syllable is produced in tutor's song, relative to the total number of syllables produced, or *tutor prevalence*. It could be that syllables occurring in more variable contexts are learned better because they are also heard more often. Indeed, I found that a syllable's prevalence in the tutor's song negatively predicted acoustic distance ($F_{1,145}$ =8.2, p=0.0049). In other words, syllables that were more frequently produced in the tutor's song were more accurately imitated by the pupil. I also found a significant positive correlation between tutor prevalence and tutor *divergence* entropy ($F_{1,153}$ =51.1, p<0.0001). However, tutor prevalence also predicted tutor *divergence* entropy ($F_{1,152}$ =56.0, p<.0001). Because divergence entropy did not significantly predict acoustic learning need not be mediated by tutor prevalence.

To directly assess the relationship between convergence entropy and acoustic learning, independently of syllable prevalence, I computed the residuals (unexplained variance) from the correlation between prevalence and acoustic distance, and asked whether convergence entropy significantly explained these residuals. I found that there was no significant relationship $(F_{1,142}=0.63, p=0.4298)$. Thus, it is possible that the relationship between convergence entropy and acoustic learning was mediated by the prevalence of syllables in the tutor's song.

Bouchard and Brainard (2013) observed that convergence *probabilities* related to the degree of auditory responses of HVC neurons. While convergence entropy is a fairly proxy for the expected convergence probability at a syllable, it is not identical (Figure 3.7).



Figure 3.7. The negative relationship between entropy and expected probability, across all retained syllables in all birds.

Therefore, I computed expected convergence probabilities for retained syllables, and found that convergence entropy was significantly and negative related with expected convergence probability ($F_{1,146}$ =4925, p<0.0001). In addition, I observed that expected convergence probability is also a significant predictor of acoustic distance ($F_{1,139}$ =5.6, p=0.0195). However,

like convergence probability, it failed to significantly explain the residuals for from the correlation between prevalence and acoustic distance.

3.4 Summary

I hypothesized that variation in the sequence complexity of a tutor's song contributes to variation in learning success in the pupil. I investigated the extent to which the variation in both acoustic and sequence learning of individual syllables can be predicted by the complexity of the sequence contexts in which syllables are embedded in the tutor's song. I calculated two measures of sequence variability for each syllable in the tutor's song: divergence entropy (the diversity of contexts following that syllable) and convergence entropy (the diversity of contexts preceding it). I found that divergence entropy negatively predicted sequence learning success, while convergence entropy related positively to acoustic learning success, but this relationship could be mediated by a confound such as the prevalence of a syllable in the tutor's song.

In Chapter 4, I discuss the implications of the findings presented here and in Chapter 2, as well as future experimental work that might help answer unresolved questions posed by the work in this project.

4. Discussion

Songbirds are among a small number of vertebrate taxa that learn their vocalizations over development (Jarvis, 2019). Unlike some songbird species (e.g. zebra finch, white-crowned sparrow), Bengalese finches produce song sequences that are variable from redition to rendition, as well as across individuals (Okanoya, 2004a, 2004b). However, these sequences have been found to be history-dependent (Jin & Kozhevnikov, 2011; Warren et al, 2012), such that transition probabilities from a given syllable depend on preceding sequential context. The extent of inter-individual variability in sequence variability, as well as its role in song learning, is poorly understood. For this reason, I investigated sequence variability and history-dependence across a sample of 29 birds (Chapter 2). I also investigated the effect of sequence variability on learning success (Chapter 3).

In Chapter 2, I used nth-order observable Markov models to investigate the extent of history dependence in BF song. I found that entropy rate decreased for all birds across Markov orders, with decreases for individual birds clustered closely around the mean (mean: 46%, SEM: 2.6%). The implications of this are twofold: first, all birds showed some degree of history dependence — otherwise decreases near 0% would have been observed (the minimum decrease was 18%). Second, no birds showed history dependence across the first through fifth orders that entirely explained the sequence variability observed at the first order — otherwise entropy would have declined to 0 for some birds.

Furthermore, I found that between-bird variance decreased as a function of Markov order. In other words, birds converge more tightly around the population mean at higher orders than at the first-order. The reversion of entropy values toward the mean between tutor and pupil, which was found to be more pronounced for higher-order entropy, could explain the decrease in between-bird entropy variance as a function of Markov order.

Consequently, these findings point to an attractor state for high-order entropy, which is a balance between history dependence and high-order variability. If an attractor for history dependence exists, the causal mechanism for it is not clear.

One possible explanation is that there are neuroanatomically-enforced upper bounds on entropy decreases (relative to the first order entropy) for a given bird. Bengalese finches diverged from white-rumped munias (the wild variety of their species) ~200 years ago, and their songs involve greater sequence variability than those of the munia (Okanoya, 2004b). The neural circuitry for BF sequencing evolved from those controlling first-order stereotyped songs, where one syllable unambiguously predicted the following syllable. However, the BF nervous system is now tasked with producing a variable transitions, where that bijection no longer holds due to the existence of branch points. Limits on history-dependence could be due to inherently unstable bifurcation points in the neural circuitry responsible for sequencing, from which two or more trajectories are possible (Yamashita et al, 2008; Jin, 2009; Hanuschkin et al, 2011).

One potential experimental test of this hypothesis would involve tutoring juvenile BFs with synthetic song stimuli that are entropic at the first order, but stereotyped at higher orders, and analyze whether BFs insert high-order variability into the songs, or are able to reproduce the difference between first-order and high-order statistics present in the input.

Another complimentary explanation for an attractor state for high-order entropy is that a narrow range of high-order entropy values have been selected for by female preferences. One previous study (Kato et al, 2010) tested female preferences on synthetic song stimuli varying in first-order entropy, and did not find consistent preferences for more variable songs. However, this does not rule out a preference for specific higher-order entropy values. A future probe into female preferences might present synthetic song sequences that are matched for first-order entropy, but differ in higher-order entropy.

In Chapter 3, I investigated the extent to which syllable sequencing (especially the variability of syllable sequencing) in the tutor's song influences the fidelity of song learning in pupils. I investigated the relationship between two measures of (first-order) entropy in the tutor's song — divergence and convergence entropies — and two measures of learning — the learning of the sequencing and acoustic structure of syllables. I quantified sequence learning success by comparing the transition probability matrices of the pupil with that of the tutor for each syllable. I found that the first-order divergence entropy of a syllable significantly predicted the success with which the sequencing of that syllable was learned.

The relationship between the sequence statistics in the tutor, and sequence learning in the pupil, has been explored in a previous study (James et al, under revision). One type of sequential change between a tutor and pupil's song analyzed there is the insertion of a syllable into a novel context, relative to the tutor song. James et al (under revision) found that insertions are significantly more likely to be at the beginning of a stereotyped sequence (*motif*) in the tutor's song than in the middle of one, and a trend for insertions to take place more often at the end of motifs. Since motifs are defined as sequences with low internal transition entropy (as opposed to branch points, which are higher entropy transitions), the beginnings and ends of motifs, where syllable insertions were observed most often, coincide with high-entropy transitions. Because syllable insertions by pupils negatively impact all three measures of sequence learning used, these findings are consistent with my observations.

In addition, I also found that convergence entropy *positively* predicted acoustic learning success. This was counterintuitive given the finding in Bouchard & Brainard (2013) that a syllable elicits a stronger neural response when embedded in a sequence it occurs in often. Stemming from those findings, I had predicted that syllables with low convergence entropy should elicit stronger responses in HVC and, thus, stronger encoding of syllables. This prediction rests on the assumption that HVC activity in juveniles resembles HVC activity in adults and that the strength of HVC responses to song covaries with the strength of learning, but these assumptions might not be true.

However, it remains unclear what could mediate the relationship between convergence entropy and syllable learning. For example, convergence entropy is positively related to the prevalence of a syllable in the tutor's repertoire, which is also positively related to the fidelity of acoustic learning. The latter finding is consistent with a previous finding that the prevalence of a syllable in the tutor's song predicts whether the syllable will be retained or dropped in the pupil's song (James et al, under revision). The causal mechanism for this effect could be purely sensory or sensorimotor. A purely sensory account would be that increased tutor prevalence leads to a better representation of the acoustic features of a syllable, and a better template for prediction error learning. A sensorimotor explanation would be that since increased prevalence in the tutor leads to increased prevalence in the pupil (James et al, under revision), it leads to more iterations of prediction error learning. To disentangle these two possibilities, future work could investigate whether tutor prevalence or pupil prevalence better predicts acoustic learning success.

Despite my analysis of higher-order entropy values in Chapter 2, I limited my analysis of entropy as a factor in learning to first-order entropies for individual syllables in the tutor and pupil's songs. This was because the sample size for retained n-grams between tutors and pupils decreases rapidly, relative to that of retained single syllables. However, future studies might work around this problem by averaging the *n*th-order entropies of all n-grams ending in a given syllable, and using those average higher-order entropy values to investigate learning success.

The selective pressures, if any, that have created sequence variability and history-dependence in BFs are unclear. Across songbird species, Robinson et al (2019) argue that females select for syllable repertoire size, and in so doing, indirectly drive the evolution of adult plasticity. To investigate whether sequence variability and history-dependence have undergone similar selective pressures, similar comparative work could be undertaken across vocal learning taxa with variable sequencing, including other songbirds (Sainburg et al, 2019) and cetaceans (Ferrer-I-Cancho 2012; Suzuki 2006).

5. Bibliography

- Ali, F., Otchy, T. M., Pehlevan, C., Fantana, A. L., Burak, Y., & Ölveczky, B. P. (2013). The Basal Ganglia Is Necessary for Learning Spectral, but Not Temporal, Features of Birdsong. Neuron, 80(2), 494–506. https://doi.org/10.1016/j.neuron.2013.07.049
- Alliende, J., Giret, N., Pidoux, L., Del Negro, C., & Leblois, A. (2017). Seasonal plasticity of song behavior relies on motor and syntactic variability induced by a basal ganglia–forebrain circuit. Neuroscience, 359, 49–68. https://doi.org/10.1016/j.neuroscience.2017.07.007
- Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. Trends in Cognitive Sciences, 15(3), 113–121. https://doi.org/10.1016/j.tics.2011.01.002
- Bouchard, K. E., & Brainard, M. S. (2013). Neural Encoding and Integration of Learned Probabilistic Sequences in Avian Sensory-Motor Circuitry. Journal of Neuroscience, 33(45), 17710–17723. https://doi.org/10.1523/JNEUROSCI.2181-13.2013
- Brainard M. S., Doupe A.J. (2000) Alteration of auditory feedback causes both acute and lasting changes to Bengalese finch song. Soc Neurosci Abstr 26:269.266
- Brainard, M. S., Doupe, A. J. (2002). What songbirds teach us about learning. Nature 417, 351–358 doi:10.1038/417351a
- Brainard, M. S., Doupe, A. J. (2013). Translating Birdsong: Songbirds as a Model for Basic and Applied Medical Research. Annual Review of Neuroscience, 36(1), 489–517. https://doi.org/10.1146/annurev-neuro-060909-152826

Brenowitz, E. A., Margoliash, D., & Nordeen, K. W. (1997). An introduction to birdsong and the avian song system. Journal of Neurobiology, 33(5), 495–500. https://doi.org/10.1002/(SICI)1097-4695(19971105)33:5<495::AID-NEU1>3.0.CO;2-#

Chomsky, N. (1957). Syntactic Structures. The Hague: Mouton De Gruyter.

- Cover, T. M., Thomas, J. A. (2006). Elements of Information Theory, 2nd Edition. Wiley.
- Dunlop, R. A., Noad, M. J., Cato, D. H., & Stokes, D. (2007). The social vocalization repertoire of east Australian migrating humpback whales (Megaptera novaeangliae). The Journal of the Acoustical Society of America, 122(5), 2893. https://doi.org/10.1121/1.2783115
- Fee, M. S., Kozhevnikov, A. A., & Hahnloser, R. H. R. (2004). Neural mechanisms of vocal sequence: Generation in the songbird. Annals of the New York Academy of Sciences, 1016, 153–170. https://doi.org/10.1196/annals.1298.022
- Ferrer-I-Cancho, R., & McCowan, B. (2012). The span of correlations in dolphin whistle sequences. Journal of Statistical Mechanics: Theory and Experiment, 2012(6), 1–12. https://doi.org/10.1088/1742-5468/2012/06/P06002
- Fujimoto, H., Hasegawa, T., & Watanabe, D. (2011). Neural Coding of Syntactic Structure in Learned Vocalizations in the Songbird. The Journal of Neuroscience, 31(27), 10023–10033. https://doi.org/10.1523/JNEUROSCI.1606-11.2011
- Ha, L. Q., Sicilia-Garcia, E. I., Ming, J., & Smith, F. J. (2003). Extension of Zipf's Law to Word and Character N-grams for English and Chinese. Computational Linguistics and Chinese Language Processing, 8(1), 77–102. Retrieved from http://www.aclweb.org/anthology/O03-4004
- Hanuschkin, A., Diesmann, M., & Morrison, A. (2011). A reafferent and feed-forward model of song syntax generation in the Bengalese finch. Journal of Computational Neuroscience, 509–532. https://doi.org/10.1007/s10827-011-0318-z

- Heinz, J., & Idsardi, W. (2013). What Complexity Differences Reveal About Domains in Language. Topics in Cognitive Science, 5(1), 111–131. https://doi.org/10.1111/tops.12000
- Hisey, E., Kearney, M. G., & Mooney, R. (2018). A common neural circuit mechanism for internally guided and externally reinforced forms of motor learning. Nature Neuroscience, 21(4), 589–597. https://doi.org/10.1038/s41593-018-0092-6
- Hough, G. E., & Volman, S. F. (2002). Short-term and long-term effects of vocal distortion on song maintenance in zebra finches. Journal of Neuroscience, 22(3), 1177–1186. https://doi.org/10.1523/jneurosci.22-03-01177.2002
- James L. S, Sun H, Wada K, & Sakata J. T. (under revision). Statistical learning for vocal sequence acquisition in a songbird. Unpublished manuscript.
- James, L. S., Sun, H., Wada, K., & Sakata, J. T. (under revision). Statistical learning for vocal sequence acquisition in a songbird. Unpublished manuscript.
- Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. Science, 366(6461), 50–54. https://doi.org/10.1126/science.aax0287
- Javkin, H. R. (1977). Phonetic Universals and Phonological Change. UC, Berkeley. Retrieved from https://escholarship.org/uc/item/08f2z7np
- Jin DZ. Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. Phys Rev E Stat Nonlin Soft Matter Phys 80: 051902, 2009. doi:10.1103/PhysRevE.80.051902.
- Jin, D. Z. (2014). Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC Generating variable birdsong syllable sequences with branching chain networks in avian premotor nucleus HVC. Physical Review., (March). https://doi.org/10.1103/PhysRevE.80.051902

- Jin, D. Z., & Kozhevnikov, A. A. (2011). A Compact Statistical Model of the Song Syntax in Bengalese Finch. PLoS Computational Biology, 7(3). https://doi.org/10.1371/journal.pcbi.1001108
- Jurafsky, D., & Martin, J. H. (2000). Speech and Language Processing. Prentice Hall.
- Jurgens, A. (2016). Entropy in Written English, 1–22.
- Katahira, K., Suzuki, K., Okanoya, K., & Okada, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. PLoS ONE, 6(9). https://doi.org/10.1371/journal.pone.0024516
- Kato, Y., Hasegawa, T., & Okanoya, K. (2010). Song preference of female Bengalese finches as measured by operant conditioning. Journal of Ethology, 28(3), 447–453. https://doi.org/10.1007/s10164-010-0203-7
- Katahira, K., Suzuki, K., Okanoya, K., & Okada, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. PLoS ONE, 6(9). https://doi.org/10.1371/journal.pone.0024516
- Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: not the Markov chains we thought they were. Proceedings of the Royal Society B: Biological Sciences, 281(1792), 20141370. https://doi.org/10.1098/rspb.2014.1370
- Kornfeld, J., Benezra, S. E., Narayanan, R. T., Svara, F., Egger, R., Oberlaender, M., ... Long, M. A. (2017). EM connectomics reveals axonal target variation in a sequence-generating network. ELife, 6. https://doi.org/10.7554/eLife.24364
- Koumura T, Okanoya K. (2016). Automatic Recognition of Element Classes and Boundaries in the Birdsong with Variable Sequences. PLoS ONE, 11(7): e0159188. https://doi.org/10.1371/journal.pone.0159188

- Koumura T., Okanoya K. (2016). Automatic Recognition of Element Classes and Boundaries in the Birdsong with Variable Sequences. PLoS ONE 11(7): e0159188. https://doi.org/10.1371/journal.pone.0159188
- Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. The Annals of Mathematical Statistics, 22(1), 79–86.
- Lashley, K.S (1951). The problem of serial order in behavior. In L.A. Jeffress (ed.) Cerebral Mechanisms in Behavior. Wiley, New York.
- Lipkind, D., Marcus, G. F., Bemis, D. K., Sasahara, K., Jacoby, N., Takahasi, M., ... Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. Nature, 498(7452), 104–108. https://doi.org/10.1038/nature12173
- Lipkind, D., Zai, A. T., Hanuschkin, A., Marcus, G. F., Tchernichovski, O., & Hahnloser, R. H. R. (2017). Songbirds work around computational complexity by learning song vocabulary independently of sequence. Nature Communications, 8(1). https://doi.org/10.1038/s41467-017-01436-0
- Long, M. A., Jin, D. Z., & Fee, M. S. (2010). Support for a synaptic chain model of neuronal sequence generation. Nature, 468(7322), 394–399. https://doi.org/10.1038/nature09514
- Miller, G.A. Note on the Bias of Information Estimates. In Information Theory in Psychology: Problems and Methods. Quastler, H., Ed.; Free Press: Glencoe, IL, USA, 1955; pp. 95–100
- Mooney, R. (2009). Neural mechanisms for learned birdsong. Learning & memory (Cold Spring Harbor, N.Y.). 16. 655-69. 10.1101/lm.1065209.
- Mooney, R., & Prather, J. F. (2005). The HVC microcircuit: The synaptic basis for interactions between song motor and vocal plasticity pathways. Journal of Neuroscience, 25(8), 1952– 1964.

- Mueller, J. M., Ravbar, P., Simpson, J. H., & Carlson, J. M. (2019). Drosophila melanogaster grooming possesses syntax with distinct rules at different temporal scales. PLoS Computational Biology, 15(6), 1–25. https://doi.org/10.1371/journal.pcbi.1007105
- Nishikawa J, Okada M, Okanoya K (2008). Population coding of song element sequence in the Bengalese finch HVC.Eur J Neurosci 27:3273–3283.
- Noad, M. J., Rekdahl, M. L., Dunlop, R. A., & Garland, E. C. (2015). Non-song social call bouts of migrating humpback whales, 3042. https://doi.org/10.1121/1.4921280
- Okanoya, K. (2004a). Song Syntax in Bengalese Finches: Proximate and Ultimate Analyses. Advances in the Study of Behavior, 34, 297–346. https://doi.org/10.1016/S0065-3454(04)34008-8
- Okanoya, K. (2004b). The Bengalese finch: A window on the behavioral neurobiology of birdsong syntax. Annals of the New York Academy of Sciences, 1016, 724–735. https://doi.org/10.1196/annals.1298.026
- Palmero, A. M., Espelosín, J., Laiolo, P., & Illera, J. C. (2014). Information theory reveals that individual birds do not alter song complexity when varying song length. Animal Behaviour, 87(C), 153–163. https://doi.org/10.1016/j.anbehav.2013.10.026
- Panzeri, S., & Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. Network: Computation in Neural Systems, 7(1), 87–107. https://doi.org/10.1088/0954-898X/7/1/006

Pinker, Steven. (1994). The Language Instinct. New York: William and Morrow. Chapter 4.

Robinson, C. M., Snyder, K. T., & Creanza, N. (2019). Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning. ELife, 8, 1–26. https://doi.org/10.7554/eLife.44454

- Saffran, J. R. (2002). Constraints on statistical language learning. Journal of Memory and Language, 47(1), 172–196. https://doi.org/10.1006/jmla.2001.2839
- Sainburg, T., Theilman, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. Nature Communications, 10(1), 3636. https://doi.org/10.1038/s41467-019-11605-y
- Sakata, J. T., & Brainard, M. S. (2006). Real-Time Contributions of Auditory Feedback to Avian Vocal Motor Control. The Journal of Neuroscience, 26(38), 9619–9628. https://doi.org/10.1523/JNEUROSCI.2027-06.2006
- Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27, 379-423.
- Shannon, C. E. (1951). Prediction and entropy of printed English. Bell System Technical Journal, 30, 50–64.
- Smith, R. (2014). Complexity in animal communication: Estimating the size of N-gram structures. Entropy, 16(1), 526–542. https://doi.org/10.3390/e16010526
- Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs. The Journal of the Acoustical Society of America, 119(3), 1849–1866. https://doi.org/10.1121/1.2161827
- Tchernichovski, O., Lints, T. J., Derégnaucourt, S., Cimenser, A., & Mitra, P. P. (2004). Studying the song development process: Rationale and methods. Annals of the New York Academy of Sciences, 1016, 348–363. https://doi.org/10.1196/annals.1298.031
- Tumer, E. C., & Brainard, M. S. (2007). Performance variability enables adaptive plasticity of 'crystallized' adult birdsong. Nature, 450(December). https://doi.org/10.1038/nature06390

- Warren, T. L., Charlesworth, J. D., Tumer, E. C., & Brainard, M. S. (2012). Variable Sequencing Is Actively Maintained in a Well Learned Motor Skill. Journal of Neuroscience, 32(44), 15414–15425. https://doi.org/10.1523/JNEUROSCI.1254-12.2012
- Williams, H. (2004). Birdsong and singing behavior. Annals of the New York Academy of Sciences, 1016(2004), 1–30. https://doi.org/10.1196/annals.1298.029
- Wohlgemuth, M. J., Sober, S. J., & Brainard, M. S. (2010). Linked Control of Syllable Sequence and Phonology in Birdsong. Journal of Neuroscience, 30(39), 12936–12949. https://doi.org/10.1523/JNEUROSCI.2690-10.2010
- Xiao L, Chattree G, Oscos FG, Cao M, Wanat MJ, & Roberts TF. (2018). A Basal Ganglia Circuit Sufficient to Guide Birdsong Learning. Neuron. Apr 4;98(1):208-221.e5. doi: 10.1016/j.neuron.2018.02.020.
- Yamashita, Y., Takahasi, M., Okumura, T., Ikebuchi, M., & Yamada, H. (2008). Developmental learning of complex syntactical song in the Bengalese finch : A neural network model. Neural Networks, 21(9), 1224–1231. https://doi.org/10.1016/j.neunet.2008.03.003
- Zevin, J. D., Seidenberg, M. S., & Bottjer, S. W. (2004). Limits on reacquisition of song in adult zebra finches exposed to white noise. Journal of Neuroscience, 24(26), 5849–5862. https://doi.org/10.1523/JNEUROSCI.1891-04.2004