THE ROLE OF DAWN SONG IN TREE SWALLOWS AND ITS
PLACE IN THE DIVERSITY OF OSCINE SONG LEARNING

A Dissertation Presented
by
BENJAMIN N. TAFT

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 2011

Organismic and Evolutionary Biology
THE ROLE OF DAWN SONG IN TREE SWALLOWS AND ITS PLACE IN THE DIVERSITY OF OSCINE SONG LEARNING

A Dissertation Presented

by

BENJAMIN N. TAFT

Approved as to style and content by:

______________________________
Jeffrey Podos, Chair

______________________________
Ethan D. Clotfelter, Member

______________________________
Elizabeth M. Jakob, Member

______________________________
Paige S. Warren, Member

______________________________
Elizabeth R. Dumont, Director
Organismic and Evolutionary Biology
For my wife, Natalia, and my daughter, Lydia.
the news to my left over the dunes and
reeds and bayberry clumps was
fall: thousands of tree swallows
gathering for flight:
an order held
in constant change: a congregation
rich with entropy; nevertheless, separable, noticeable
as one event,
not chaos: preparations for
flight from winter,
cheet, cheet, cheet, cheet, wings rifling the green clumps
beaks
at the bayberries
a perception full of wind, flight, curve,
sound:
the possibility of rule as the sum of rulelessness:
the "field" of action
with moving, incalculable center:

from “Corson’s Inlet”, by A. R. Ammons
ACKNOWLEDGMENTS

Thanks to Jeff for many things, especially patience. Thanks also to Ethan, Beth, and Paige, for their tolerance of long silences broken suddenly by demands for instantaneous attention. Thanks to Dan Ardia and Ethan Clotfelter for access to and assistance at the Amherst College field site. Thanks to my parents and grandparents for unsurpassed help and assistance. Thanks to my wife, Natalia, for unswerving support.
ABSTRACT

THE ROLE OF DAWN SONG IN TREE SWALLOWS AND ITS PLACE IN THE DIVERSITY OF OSCINE SONG LEARNING

FEBRUARY 2011

BENJAMIN N. TAFT
B.A., CORNELL UNIVERSITY
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jeffrey Podos

Aspects of the behavioral ecology of bird song learning are examined in three parts. First, an approach from image analysis is extended to allow rapid, quantitative description of animal sounds. In this approach, sounds are summarized as sets of time-frequency-amplitude landmarks. Second, the role of dawn song in tree swallow (Tachycineta bicolor) breeding biology is examined. Song syllable sharing among tree swallows was found to be high among birds nesting at the same site, but sharing was lower between birds nesting at different sites. When birds nested at different sites, the distance between those sites was not related to the amount of difference between the birds’ syllable repertoire compositions. All tree swallow song repertoires did not remain constant during the breeding season; some individuals added new syllable types, others modified existing types. Singing performance was correlated with reproductive success in tree swallows: males that sang more precise repetitions of their syllable types attracted more extra-pair mates. Furthermore, pairwise comparisons between the social and genetic fathers of extra-pair young found that the genetic fathers averaged higher syllable consistency than the cuckolded males. Third, a comparative study of the phylogenetic distribution of vocal mimicry examined the evolutionary history of song learning in oscine passerines. Vocal mimicry, defined as the habitual incorporation of heterospecific sounds into song displays, was found in twenty-eight separate clades of oscines. These clades were found in every major oscine superfamily, but made up a higher proportion of daughter groups within the most ancient superfamilies of oscines. The most plesiomorphic lineages of oscines were found to contain many highly-skilled mimics. These observations support the hypothesis that the course of song learning in oscines has run repeatedly from permissive learning rules that permit mimicry to restrictive learning rules that limit mimicry.
# CONTENTS

<table>
<thead>
<tr>
<th>EPIGRAPH</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
</tbody>
</table>

## CHAPTER

1. **ACOUSTIC LANDMARKS: A QUANTITATIVE TECHNIQUE FOR ANALYZING SOUNDS AS SHAPES** ........................................ 1

1.1 Abstract .................................................. 1
1.2 Introduction ............................................... 2
1.3 Methods .................................................. 3
   1.3.1 Quantitative description using acoustic landmarks .......... 3
      1.3.1.1 Spectral peaks .................................. 4
      1.3.1.2 Events ......................................... 4
      1.3.1.3 Acoustic landmarks ............................. 5
      1.3.1.4 Analysis parameters ............................ 6
   1.3.2 Recording and digitization ................................ 6
   1.3.3 Vocal output .......................................... 7
   1.3.4 Vocal development .................................... 7
   1.3.5 Note type classification ................................ 9
1.4 Results .................................................. 10
1.5 Discussion ............................................... 12
   1.5.1 Potential advantages of the method ........................ 13
# 2. SPATIAL PATTERNS OF SONG SYLLABLE SHARING AND THEIR IMPLICATION FOR SONG LEARNING IN TREE SWALLOWS (*TACHYCINETA BICOLOR*)

- **2.1 Abstract** ............................................................... 24
- **2.2 Introduction** ......................................................... 25
- **2.3 Methods** ............................................................ 27
  - 2.3.1 Equipment ......................................................... 27
  - 2.3.2 Recording .......................................................... 28
  - 2.3.3 Sites ....................................................................... 28
  - 2.3.4 Acoustic analysis ................................................... 28
  - 2.3.5 Classification of syllable types ................................ 29
  - 2.3.6 Repertoire analysis ............................................... 30
  - 2.3.7 Geographic patterns ............................................ 31
- **2.4 Results** ................................................................. 32
  - 2.4.1 Syllable classification .......................................... 33
  - 2.4.2 Repertoire analysis ............................................... 33
  - 2.4.3 Geographic patterns ............................................ 34
- **2.5 Discussion** ............................................................ 34
  - 2.5.1 The function of post-dispersal learning .................... 37
  - 2.5.2 Implications about song learning in other swallows .... 40

# 3. COPYING AMONG MOVING TARGETS: SYLLABLE SHARING IN TREE SWALLOW DAWN SONG

- **3.1 Abstract** ............................................................... 49
- **3.2 Introduction** ......................................................... 50
- **3.3 Methods** ............................................................ 53
  - 3.3.1 Recording .......................................................... 53
  - 3.3.2 Landmarks .......................................................... 54
  - 3.3.3 Removing heterospecific sounds .............................. 55
  - 3.3.4 Classifying tree swallow syllables ......................... 56
  - 3.3.5 Defining repertoires .......................................... 58
  - 3.3.6 Syllable performance ........................................ 58
  - 3.3.7 Statistical tests ............................................... 59
- **3.4 Results** ................................................................. 60
- **3.5 Discussion** ............................................................ 62
4. SONG PERFORMANCE, PROVISIONING RATES, EXTRA-, AND WITHIN-PAIR REPRODUCTIVE SUCCESS ARE ALL CORRELATED IN MALE TREE SWALLOWS

4.1 Abstract .............................................................................................................. 68
4.2 Introduction ......................................................................................................... 68
4.3 Methods ............................................................................................................... 70
  4.3.1 Measures of singing performance ................................................................. 70
  4.3.2 Measures of reproductive performance ......................................................... 71
  4.3.3 Statistical analyses ....................................................................................... 71
4.4 Results .................................................................................................................. 72
4.5 Discussion ............................................................................................................ 74

5. A COMPARATIVE ANALYSIS OF THE DISTRIBUTION OF ALLOSPECIFIC VOCAL MIMICRY IN OSCINE PASSERINES ........ 85

5.1 Abstract ............................................................................................................... 85
5.2 Introduction ......................................................................................................... 86
  5.2.1 Mechanisms affecting the evolution of mimicry ............................................ 89
  5.2.2 Hypotheses ................................................................................................... 93
5.3 Methods ............................................................................................................... 93
  5.3.1 Literature survey ............................................................................................ 93
  5.3.2 Phylogenetic analyses .................................................................................. 94
5.4 Results .................................................................................................................. 96
  5.4.1 Two categories of mimicking clades ............................................................... 96
    5.4.1.1 Clades with many mimics ....................................................................... 97
    5.4.1.2 Clades with few mimics ......................................................................... 98
    5.4.1.3 Distinct evolutionary patterns for each category .................................... 98
  5.4.2 Distinct evolutionary patterns for each category ........................................... 98
5.5 Discussion ............................................................................................................ 99

BIBLIOGRAPHY ...................................................................................................... 110
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Measures of the vocal output of each swamp sparrow</td>
<td>22</td>
</tr>
<tr>
<td>1.2 Results from a principle component analysis of the parameters measured from swamp sparrow songs</td>
<td>23</td>
</tr>
<tr>
<td>2.1 Predicted relationships between repertoire similarity and spatial distance from four hypotheses about song learning</td>
<td>40</td>
</tr>
<tr>
<td>2.2 Sampling effort and syllable type occurrence at the four sites of the study</td>
<td>41</td>
</tr>
<tr>
<td>2.3 Mantel tests of divergence between syllable repertoires across different spatial scales</td>
<td>48</td>
</tr>
<tr>
<td>3.1 Expected changes in repertoire size and syllable performance according to three different hypotheses about post-dispersal song learning in tree swallows</td>
<td>67</td>
</tr>
<tr>
<td>3.2 Effects on syllable type performance consistency: syllable type, when it appeared in a bird’s repertoire, and how many week’s experience the bird has with the type</td>
<td>67</td>
</tr>
<tr>
<td>4.1 Non-exclusion probabilities for each of the nine separate microsatellite loci used for the paternity analysis (N = 75)</td>
<td>76</td>
</tr>
<tr>
<td>4.2 Results from linear models testing for correlations between measures of male singing traits and male reproductive traits in tree swallows</td>
<td>78</td>
</tr>
<tr>
<td>4.3 Results from mixed linear models testing for differences in performance traits between the actual sires of extra-pair chicks and their social fathers</td>
<td>80</td>
</tr>
<tr>
<td>5.1 Oscine families that contain many mimics</td>
<td>109</td>
</tr>
<tr>
<td>5.2 Taxa from families in which there are few or only one species of vocal mimic</td>
<td>109</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Two different representations of the same swamp sparrow trill.</td>
<td>16</td>
</tr>
<tr>
<td>1.2</td>
<td>The process of generating acoustic landmarks from a swamp sparrow trill</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>syllable.</td>
<td></td>
</tr>
<tr>
<td>1.3</td>
<td>Two swamp sparrow trills with different acoustic space densities, a measure of</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>singing consistency that is defined in the text.</td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>Hierarchical clustering with group number chosen by AIC.</td>
<td>19</td>
</tr>
<tr>
<td>1.5</td>
<td>Changes in an aggregate measure of trill performance in developing swamp</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>sparrows.</td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>Changes in an aggregate measure of note performance in developing swamp</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>sparrows.</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>The locations of the four sites where tree swallows were recorded for this study.</td>
<td>41</td>
</tr>
<tr>
<td>2.2</td>
<td>Spectrograms and aggregate landmark plots of each of the twenty-three syllable</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>types found in this study.</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Variation among birds that sang a specific syllable type (#2 from Figure 2.2).</td>
<td>43</td>
</tr>
<tr>
<td>2.4</td>
<td>The average saturation curve showing the cumulative number of syllable types in a</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>bout.</td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>Repertoire difference versus spatial distance (m) for pairs of tree swallows in which both birds nested within the same site during the same year.</td>
<td>45</td>
</tr>
<tr>
<td>2.6</td>
<td>Mean repertoire differences between pairs of birds are lowest within a site and equally high between sites separated by tens of kilometers as between sites separated by hundreds of kilometers.</td>
<td>46</td>
</tr>
<tr>
<td>2.7</td>
<td>The mosaic of syllable sharing groups at the Amherst College site.</td>
<td>47</td>
</tr>
<tr>
<td>3.1</td>
<td>The relationship between syllable amplitude and the distance of that syllable from the centroid of the type it belongs to.</td>
<td>65</td>
</tr>
</tbody>
</table>
3.2 Changes in the phonology of syllable type #9, as sung in different weeks by six male tree swallows. .......................................................... 66

3.3 Spectrograms showing examples documenting change over time of a syllable type within a single bird’s repertoire. ........................................... 66

4.1 Differences in amplitude consistency between days within two syllable types from a single bird’s repertoire. ...................................................... 77

4.2 The date on which a male tree swallow first begins to sing is a significant predictor of when his mate will lay her first egg. ................................. 78

4.3 The number of extra-pair mates taken by a female is negatively correlated with her social mate’s syllable consistency. .............................................. 79

4.4 The genetic sires of extra-pair chicks had greater syllable consistency when compared to the social fathers of those chicks. ........................................ 80

4.5 Feeding effort at a male’s own nest was greater for the genetic sires of extra-pair chicks when compared to the social fathers of those chicks. .............. 81

4.6 The genetic sires of extra-pair (EP) chicks had greater rank in terms of EP offspring when compared to the social fathers of those chicks. ............... 82

4.7 In tree swallows, there does not seem to be a tradeoff between gaining extra-pair paternity and maintaining within-pair paternity............................... 83

4.8 Tree swallow song output is affected by cold: males sing more syllables on warmer mornings. ................................................................. 84

5.1 The frequency of mimicry in different oscine clades, using a phylogeny based upon Jønsson and Fjeldså (2006). .................................................. 106

5.2 The observed amount of homoplasy in vocal mimicry among oscines compared to the distribution of homoplasy in 10000 simulations of neutral stepwise character evolution. ................................................. 107

5.3 Among oscine lineages that had diverged by 30 Mya, the younger groups have a lower proportion of daughter families with many mimics. ......................... 108
CHAPTER 1
ACOUSTIC LANDMARKS: A QUANTITATIVE TECHNIQUE FOR ANALYZING SOUNDS AS SHAPES

1.1 Abstract

Spectrograms present sounds as shapes, opening up the possibility of using image-processing techniques for analyzing sounds. One method for analyzing shapes that may appear in many different forms is to assign landmarks to analogous positions on each shape. The coordinates of each landmark then provide a fixed-dimensional set of quantitative data that can be used to compare all of the shapes of interest. This approach can be applied to sounds if each landmark is described in terms of time, frequency, and amplitude. A further advantage is that the process of assigning landmarks to sounds can be automated once a user inputs appropriate parameters for detecting sounds in a recording. To illustrate the effectiveness of this approach, developmental series of the songs of eleven juvenile swamp sparrows (Melospiza georgiana) are analyzed using five landmarks to describe each note. The automated approach made it possible to include more than 50 hours of recordings from each bird, containing an average of 3657 songs per bird, and a total of 1,106,905 notes. Measures of note and syntax consistency were calculated from the landmarks of each song’s notes. Note consistency was found to increase during April and level off in May, matching the findings of Clark et al. (1987) that note-types become crystallized early in the process of song learning. Syntax consistency improved steadily throughout the recording period, as previously observed by (Marler and Peters, 1982; Podos et al., 1999). Acoustic landmarks provide, with great speed and efficiency, biologically relevant acoustic information about animal sounds in a format that can be analyzed in a variety of powerful ways.
1.2 Introduction

The acoustic signals of animals are highly diverse because many different evolutionary processes interact to produce their forms. The study of acoustic communication has thus been a productive and integrative field of research (Podos et al., 2004a; Ryan et al., 2009). This success has come in the face of a methodological challenge. Animal sounds are so diverse that it can become a serious burden just to choose which acoustic features to extract and measure. Each different taxon or question suggests its own set of features, some of which are unique, and some of which are shared with other taxa and questions. There are two standard approaches to extracting features: measuring by hand (Esser and Schubert, 1998; Smith and Jones, 1997), or writing custom software (Chen and Maher, 2006; Deecke and Janik, 2006; Tanttu et al., 2006). Both of these options are time-intensive. This hinders quantitative approaches to questions that involve large sample sizes or the measurement of many features, including comparisons among taxa, populations, and developmental stages.

One solution to this problem is to describe sounds using techniques from the statistics of shape. Spectrograms render sounds as shapes (Koenig et al., 1946). Both geometric morphometrics (Bookstein, 1996) and motion capture (Brown et al., 2005) use landmarks to represent shapes quantitatively. In this chapter, I will introduce and test an algorithm that can create landmarks from the shapes that appear on spectrograms. Quantitative descriptions of the features of sounds can be thought of as measurements of those shapes. Some measures pertain to single shapes, such as the duration or mean frequency of a whistled note. Other measures pertain to collections of shapes, such as the trill rate or frequency bandwidth of a series of notes. Acoustic landmarks retain information about any feature that can be thought of as a measure of the position or shape of a sound in a spectrogram.

I will test the ability of the software to replicate a description of vocal development in swamp sparrows, *Melospiza georgiana*. To do this, I will try to replicate observations of four aspects of vocal development in this species. The first observation is that the stereotypy of note type reproduction increases during vocal development (Clark et al., 1987). The second observation is that the stereotypy of note type order within songs increases during vocal development (Podos et al., 1999). The third observation is that swamp sparrows increase the degree of difficulty of their songs
over the course of song development (Podos, 1996). The fourth observation is that the duration of swamp sparrow trills becomes less variable during vocal development (Marler and Peters, 1982). If I am successful in reproducing all of these observations, then I will have demonstrated the usefulness and flexibility of acoustic landmarks.

1.3 Methods

1.3.1 Quantitative description using acoustic landmarks

The process of generating acoustic landmarks is based on the idea of the spectrogram (Koenig et al., 1946). A digital spectrogram (Figure 1.2a) is created from a sound file by calculating a series of discrete Fourier transforms, or DFTs (Frigo and Johnson, 2005). The algorithm introduced here, named SoundPoints, performs several more steps after each DFT, as described in greater detail in the sections that follow. First, SoundPoints finds the spectral peaks, regions around the local maxima of a spectrum (Figure 1.2b). Next, the algorithm chains peaks together to form events (Figure 1.2c). In the present analyses of swamp sparrow songs, each event corresponds to a single trace on a spectrogram. If the user decides that a more complex unit is appropriate, he or she can choose settings so that events may include multiple traces. Finally, the algorithm creates acoustic landmarks by calculating a smoothed approximation of the chain of spectral peaks (Figure 1.2d). Each smoothed approximation contains the same number of points so each event in an analysis, regardless of its duration or complexity, is described by the same number of variables.

At each step in the analysis, the user can choose parameters that dictate the details of how the SoundPoints detects, measures, and groups sounds into events. A certain amount of trial and error is necessary in order to determine which parameters to use. A well-chosen set of parameters will produce a set of landmarks that matches the appearance of the spectrogram from which they were made (Figure 1.1). After verifying that the landmarks from several different files all match up with the appropriate spectrograms, the user can process a large number of files with those settings.
1.3.1.1 Spectral peaks

A spectral peak is a region in a frequency spectrum around and including a local amplitude maximum. Measuring the properties of spectral peaks is a common approach in studies of animals that make sounds with pure tones (Tanttu et al., 2006; Chen and Maher, 2006; Holy and Guo, 2005; Harma, 2003; Tchernichovski et al., 2000; Slabbekoorn and Ten Cate, 1998; Ito et al., 1996; Martindale, 1980). In such cases it is straightforward to identify local maxima in a spectrum, either by eye or with algorithms (Figure 1.2b). The SoundPoints algorithm calculates spectral peaks using a user-defined amplitude range. For example, in the current analysis of swamp sparrow notes, each peak is the region around the local maximum frequency where the amplitude is no less than 2 dB below the local maximum amplitude. SoundPoints calculates the mean amplitude, weighted mean frequency, and frequency bandwidth of each peak. The user can also specify the maximum number of peaks to extract from each spectrum. SoundPoints starts with the loudest peak, and continues to describe peaks until it reaches that number or until there are no more peaks left in the spectrum.

At each time step, SoundPoints filters the spectral peaks that it has calculated. The mean frequency of the peak is compared to a user-specified passband, and the mean amplitude is compared to a user-specified amplitude threshold. Any peaks that lie outside the passband, or that have amplitudes below the threshold, are removed from the analysis. Figure 1.2b shows the two peaks from a spectrum that meet the criteria for the swamp sparrow analysis. If a peak passes through the filter then the time that it occurs is added to the frequency and amplitude variables describing it.

1.3.1.2 Events

An event is a chain of consecutive spectral peaks that represent a single sound unit. The user specifies the criteria that SoundPoints uses to connect peaks together. For example, in the swamp sparrow analysis, an event corresponds to a single note (Figure 1.2b). The fact that each event is a chain of peaks simplifies the task of adding new peaks. The distance between a new spectral peak and an existing event is defined as the distance between the new peak and the final peak of the event.
SoundPoints calculates the distance between two peaks as the Euclidean distance between peaks in time-frequency-amplitude space. The user determines the relative importance of each of these axes by specifying weights for them. If the user wants events to consist of continuous notes, then distance can be scaled by the degree of overlap between the frequency bands of the peaks. In SoundPoints, the degree of overlap between peaks is calculated as the Jaccard similarity index of their frequency ranges.

Once the distances between a new spectral peak and each event have been calculated, two outcomes are possible. If distance between the new peak and the closest event to it is less one, then the new peak is added to the event. Otherwise, SoundPoints creates a new event that starts with the new peak. As SoundPoints processes a file, at some time the interval between the end of an event and the current time will be large enough that all new points will exceed the concatenation threshold. SoundPoints calculates and records the start time, duration, frequency band, and a set of acoustic landmarks for each event.

1.3.1.3 Acoustic landmarks

An acoustic landmark is defined here as a point in time-frequency-amplitude space. A small number of acoustic landmarks can describe a complex sound in the same way that a small number of spatial landmarks can represent a complex morphological feature (Bookstein, 1996). The user chooses the number of landmarks. SoundPoints calculates landmarks for an event by dividing its chain of spectral peaks into a series of groups. Each group contains the same number of spectral peaks. There is one group for each landmark. The landmark point from a group is the centroid of the group in time-frequency-amplitude space. The time value for each landmark is calculated relative to the beginning of the event, rather than the beginning of the file. The appropriate number of landmarks for an analysis depends upon the most complex sound in a data set. For example, the short, nearly linear notes of swamp sparrow trills require only five landmarks to provide a comprehensive description of the acoustic event (Figure 1.2d).
1.3.1.4 Analysis parameters

Eleven parameters govern the SoundPoints algorithm. Two parameters define the spectral resolution of the analysis. The *frequency resolution* defines the FFT size used to make spectra. The *time resolution* defines the time lag between when successive spectra begin. Two parameters define how peaks are extracted. The *dynamic range* defines the amplitude difference between the loudest and softest frequencies in a peak. Limiting the *number of peaks* extracted from each spectrum saves computation time and focuses the analysis on biologically relevant sounds. Three parameters define which peaks should be included in the analysis once they have been extracted: the *minimum frequency* and *maximum frequency* define a bandpass filter, and the *amplitude floor* defines the sound intensity level that separates relevant sounds from background noise. Three parameters define the way peaks are concatenated into events. The *time interval*, *frequency interval*, and *amplitude interval* all dictate the weights that SoundPoints gives to their respective variables when calculating the distance between peaks. This distance can be scaled by the *overlap function*. Peaks are never joined if their weighted distance is greater than one. The final user-specified parameter is the *number of landmarks* to be used per event.

1.3.2 Recording and digitization

The application of SoundPoints software here focuses on song recordings of swamp sparrows, *Melospiza georgiana*. Swamp sparrow recordings were made as part of ongoing research on vocal development (Podos et al., 2004b). Subjects were eleven captive-reared male swamp sparrows that hatched in 2007. The recordings were made in from 8 April through 30 May of 2008. Each bird was recorded for 3.5 hours per morning on fifteen or sixteen separate days, for a total of 605.5 hours of recording with 52-56 recorded hours per bird (table 1.1). See Podos et al. (2009), for full details about the housing, recording, and animal care protocols for these birds. The recordings from these birds were stored on DVDs as uncompressed wav files. I processed these files by running the SoundPoints application on a Apple MacBook with a 2.1 GHz Intel Core 2 Duo processor and 1 GB of RAM.
1.3.3 Vocal output

The parameters that I chose for the swamp sparrow trill note analysis are based on the observation that this species’ notes are short in duration and can have rapid frequency modulation (Marler and Pickert, 1984). I chose a frequency resolution of 86.1 Hz, and a time resolution of 1.02 ms. I chose a dynamic range of 2 dB. I chose to extract the two loudest peaks from each spectrum because there can be some time overlap between the end of one note and the beginning of the next one (Figure 1.2a,b). I chose a frequency passband of 2-10 kHz, and an amplitude floor of -0.5 dB. I chose a time interval of 5 ms, a frequency interval of 1 kHz. I did not want to include amplitude when calculating the distance between events, so I set the amplitude weight to zero. I chose to use the overlap function. Finally, I chose to use five landmark points because the notes are short and relatively monotonic. An example of the landmarks produced by these settings can be seen in Figure 1.1. The close match between the note shapes in the spectrogram graphed in Figure 1.1a and the landmarks plotted in Figure 1.1b show that these settings are able to capture the acoustic features of the notes in a swamp sparrow trill.

1.3.4 Vocal development

Once the notes in a recording had been detected, I wanted to join the notes from each bout of song into analytical units. I used R for all remaining analyses in this paper (R Development Core Team, 2008). In the swamp sparrow analysis, I defined a bout as a series of notes that lasted more than 0.5 s with no intervals of silence between notes greater than 0.25 s (Marler and Peters, 1982). I used four different measures of the degree of vocal development shown in a bout.

First, I measured how densely clumped the notes of a bout were in acoustic space. This provides a measure of note stereotypy. This particular measure is independent of any note type classification. Towards this end, I divided the two-dimensional time-frequency acoustic space of a bout into a grid with the same resolution as the spectrogram that I used to generate the landmarks (Figure 1.3c,d). Each bout contains \( nk \) landmarks, where \( n \) is the number of notes and \( k \) is the number of landmarks per bout. These landmarks fall into \( b \) grid squares, where \( b \in [k, nk] \), since the landmarks for a single note will never occur in the same grid square. The measure that I used for the density of
notes in acoustic space was $\frac{nk-b}{b(n-1)}$. This measure ranges from zero when each landmark occupies its own bin to one when each note occupies the same $k$ bins.

Second, I measured the Euclidean distance between each note and the centroid of the note type to which it belonged. This is a second measure of note stereotypy, one that does depend on some way of classifying notes into types. The details of how I classified notes into types are given below. I described each note with five landmarks, so each note was described by ten variables: five time values and five frequency values. I log-transformed the frequency values to match them to the logarithmic way that vertebrates perceive sound frequencies (Deecke and Janik, 2006). I then z-transformed all of the variables so that variation in time was given the same weight as variation in frequency.

Third, I measured the predictability of a bout’s note type order. This is a measure of the stereotypy of note order within a bout of song. Each bout contains $n$ notes. These notes are divided into $k$ types. At each time $t \in [1, n]$, type $\tau_t \in [1, k]$. Thus, the frequency of transitions from $i$ to $j$, $p_{ij} = \frac{1}{n-1} \sum_{t=1}^{n-1} \left( \tau_t = i \right) \land \left( \tau_{t+1} = j \right)$. If the notes in a bout are not uttered in any particular order, then no transition type is more common than any other, and each $p_{ij} \approx p_i p_j$. The information entropy of the note transition matrix, $H = -\sum_i^k \sum_j^k p_{ij} \log_2 p_{ij}$, will be high. If the note order is not random, then some transitions will be common, and others will be rare or nonexistent. The mutual information of the note type at time $t$ and the note type at time $t + 1$, $I = -\sum_i^k \sum_j^k p_{ij} \log_2 \frac{p_{ij}}{p_i p_j}$, will be high. The ratio of these two values $\frac{I}{H}$ ranges from zero, meaning no note type order, to one, which means that note type order is completely predictable. However, it is not independent of the number of notes in the bout. The final step in creating the index is to scale the above ratio by the number of note types per note in the bout: $\frac{I}{H} - \frac{k}{n}$.

Fourth, I measured how closely a bout approached the performance limit that relates trill rate and frequency bandwidth. This is a measure of the degree of difficulty of a song in terms of a morphological constraint on trill performance (Podos, 1997). I calculated the frequency bandwidth of a bout as the frequency of the highest-pitched landmark in the bout minus the frequency of the lowest-pitched landmark in the bout. I calculated the trill rate of a bout as the number of notes in the bout divided by the bout’s duration. I calculated an upper-bound regression from these
measures, using trill rate bins that were 10 Hz wide, centered at trill rates ranging from 5 to 75 Hz (Blackburn et al., 1992). I calculated a bout’s distance from this line, $d$, using the following formula: 

$$d = \frac{\sqrt{(mx - y + b)^2}}{\sqrt{m^2 + 1}},$$

where $m$ and $b$ are the slope and intercept of the regression, $x$ is the trill rate of the bout, and $y$ is the frequency bandwidth of the bout.

Fifth, I measured the coefficient of variation of bout duration. This is a measure of consistency across a bird’s songs. Bout duration is the end time of the last note in a bout minus the start time of the first note in a bout. I calculated the coefficient of variation of bout duration among the songs from each day from each bird.

Finally, I performed a principal components analysis (PCA) on all five variables. Each of the variables is predicted to change in related ways during development, and a PCA can provide insight into the way those changes are linked. I used daily values from each bird for each variable for this analysis.

I used mixed linear models to test whether any of the features that I measured changed during development. I used the “lme” function from the R package “nlme” (Pinheiro et al., 2009). In the case of centroid distance, I used a model where Julian date was the independent variable and the random variables were bird and note type nested within bird. In all other cases I calculated the mean value of the measure for each bird on each day. I then used a model with Julian date as the independent variable and bird as the random variable.

### 1.3.5 Note type classification

The method that I chose to use in order to find the number of note types was agglomerative hierarchical clustering (Figure 1.4a). I used the same data for the classification analysis as I did when I calculated the Euclidean distance between notes. I used the “agnes” function from the R package “cluster” (Maechler et al., 2005). This method does not make any assumptions about the number of groups in the data (Duda et al., 2000). That makes it appropriate for this kind of analysis because I did not know the note type repertoire size of any of the subject birds. There are, however, two difficulties inherent in hierarchical clustering that I had to overcome in order to use the technique.
The first difficulty is that a hierarchical clustering analysis results in not one but many possible groupings, each nested within the next. In order to choose the optimal grouping from an analysis, I used an approach from model selection. For each analysis, I calculated an approximate AIC value for each grouping level using the formula $2gc + n \log \left( \frac{1}{n} \sum_{k=1}^{g} n_k \sum_{j=1}^{c} \left( \sum_{i=1}^{n_k} x_{ijk}^2 - \frac{1}{n_k} \left( \sum_{i=1}^{n_k} x_{ijk} \right)^2 \right) \right)$, where $n$ is the number of elements, $g$ is the number of groups, $n_k$ is the number of notes in group $k$, $c$ is then number of variables, and $x_{ijk}$ is the value of variable $j$ in the $i$th member of group $k$. I used the grouping with the smallest value of this score as the analysis’s number of groups (Figure 1.4b). This technique worked well because the examples of a note from a given type tended to group near each other, forming multivariate equivalents of roughly spherical clouds separated by regions of space that were much less dense (Figure 1.4c). Its usefulness is limited to this kind of easily-clustered space (unpublished simulations).

The second difficulty is that hierarchical clustering is an $O(n^2)$ algorithm because it is based on pairwise comparisons. I dealt with this difficulty by performing a series of cluster analyses on subsets of the data. I chose each subset so that it contained no more than 1200 notes. This number provided a good balance between an acceptable running time and a thorough sampling of the clusters in the represented acoustic space. After I chose the number of groups from the analysis on the subset, I chose the 10 examples from each group that were closest to the group centroid. I then combined these best examples from several subsets to form a new subset that represented a larger part of the data set. I repeated this process within each bird until I reached the point where the examples were drawn from entire output of that bird. The order of subset grouping was songs within days, days, weeks, and the whole bird. I did not pool the birds together because repertoire overlap was beyond the scope of this paper.

1.4 Results

The software identified a total of 1106905 swamp sparrow notes. The number of notes analyzed per bird ranged from 17874 to 208540 notes, with a median of 92063 notes and a standard deviation of 67270.05 notes (table 1.1). The birds sang 40237 bouts. The number of songs per bird ranged from 775 to 7290 songs, with a median of 2911 bouts and a standard deviation of 2363.817 songs.
Note type repertoire size ranged from 3 to 40 types, with a mean of 11.909 types / bird, and a standard deviation of 3.6318 types per bird (Table 1.1).

My analysis of the swamp sparrow songs matched each of the four published observations regarding how swamp sparrow songs change during development. I found that notes became more stereotyped as the swamp sparrows developed. The mean number of landmarks per grid square per note increased at a rate of $1.80 \times 10^{-4} \pm 4.70 \times 10^{-5}$ units per day ($F_{1,162} = 14.6, p = 1.92 \times 10^{-4}$). The mean distance of a landmark from its type centroid decreased at a rate of $-788 \pm 131$ units per day ($F_{1,162} = 434, p = 1.04 \times 10^{-47}$). I also found that the order of note types within trills became more stereotyped as the swamp sparrows developed. The mean predictability of note type order increased at a rate of $0.00105 \pm 1.90 \times 10^{-4}$ bits per day ($F_{1,162} = 9977, p = 1.92 \times 10^{-147}$). In terms of the performance limit that relates trill rate to frequency bandwidth, the swamp sparrows sang increasingly challenging songs as they developed. The mean distance to the performance limit decreased at a rate of $-0.0145 \pm 0.00349$ units per day ($F_{1,162} = 17.2, p = 5.48 \times 10^{-5}$). Finally, the duration of swamp sparrow songs became less variable during vocal development. The coefficient of variation of swamp sparrow song duration decreased over the course of the recording period at a rate of $-0.373 \pm 0.0585$ notes per day ($F_{1,162} = 40.536, p = 1.9283 \times 10^{-9}$).

The principal component analysis (PCA) found a high degree of covariation among the variables (Table 1.2). The first two principle components (PCs) explained 27.3 % and 25.4 % of the variation in the date, respectively. The first PC loaded strongly on three variables: negatively on variability in song duration, positively on predictability of note order, and negatively on distance from the performance limit. PC1 increased at a rate of $0.0488 \pm 0.00511$ units per day ($F_{1,161} = 91.3, p = 2.05 \times 10^{-17}$, Figure 1.5). The second PC also loaded strongly on three variables: positively on density of acoustic space, negatively on distance from note type centroid, and positively on distance from the performance limit. PC2 increased at a rate of $0.0294 \pm 0.00561$ units per day ($F_{1,161} = 27.4, p = 5.10 \times 10^{-7}$, Figure 1.6).
1.5 Discussion

In this study, acoustic landmarks made it possible to describe song development with quantitative, biologically meaningful measures of over one million swamp sparrow notes. By replicating published results about swamp sparrow development, I provided evidence that acoustic landmarks are a biologically informative method for analysis. In addition, the scale of the analysis, in terms of sample size and features measured, demonstrate the practical usefulness and flexibility of acoustic landmarks. These capabilities suggest statistical approaches and biological questions that are either new or newly tractable. In essence, acoustic landmarks store the same amount of information as a spectrogram, but do so in a much more useable form, freeing researchers from the time-consuming task of measuring spectral features by hand.

The five different statistical measures of song development support three conclusions. First, swamp sparrow note performance increases during vocal development (Figure 1.6). Both measures of note stereotypy changed during development in the manner that indicates increased stereotypy. This is corroborated by the second component of the PCA, the loadings of which make it a good index of note performance. These results support the findings of Clark et al. (1987). Second, swamp sparrow trill performance increases during vocal development (Figure 1.5). Stereotypy of note order and stereotypy of song duration, and the degree of difficulty of songs all increased during vocal development. This is corroborated by the first component of the PCA, the loadings of which make it a good index of trill performance. These results support the findings of Podos et al. (1999) and Podos (1996). Third, these two patterns proceed along different time courses. The trill performance PC steadily increases during development (Figure 1.5), while the note performance PC increases during the early part of development, and then remains steady or even decreases slightly (Figure 1.6). This matches well with the seven stages of song development proposed by Marler and Peters (1982). The period in which note performance and trill performance improve simultaneously corresponds to subsong (stage VII of Marler and Peters) and subplastic song (stages VI and V of Marler and Peters). The period in which note stereotypy stops improving while trill performance continues to improve corresponds to plastic song (stages IV-II of Marler and Peters). This makes sense because one needs recognizable note types before one can make hypotheses about
syntactic rules. These results replicate, in a quantitative way, the qualitative observation that swamp sparrow development occurs in discrete stages.

The observation that note performance stabilizes before trill performance is consistent with the putative role of these two aspects of performance in adult behavior. Notes seem to be perceived categorically in swamp sparrows (Nelson and Marler, 1989), which suggests that swamp sparrows may not have the perceptual or cognitive ability to make fine judgements about note performance. Trill performance, on the other hand, has been shown to be involved in mate choice (Ballentine et al., 2004), probably in part because it signals male age and size (Ballentine, 2009). Trill performance may be a more honest signal of quality than note performance, either because it is intrinsically harder or because the longer amount of time that it takes to achieve high trill performance to develop means that it represents a better sample of a bird’s history.

1.5.1 Potential advantages of the method

These results demonstrate the usefulness of acoustic landmarks in analyzing swamp sparrow song. In general, acoustic landmarks have three main advantages: speed, reuse, and information capture.

Acoustic landmark analysis is automated, making its speed much greater than manual techniques for measuring spectrograms. In total, it took the software, running on a typical laptop from 2008, 25 hours to convert the 600 hours of recordings into a million database entries. This represents an average measurement rate of 257 parameter measurements per second, far in excess of possible human performance. The measurements represented as acoustic landmarks are also contain a large amount of useful information. As the results from the analysis of swamp sparrow development show, acoustic landmarks retain information about collections of notes, note order, and the shape of notes. Acoustic landmarks store information in a way that can capture different kinds of variation in shape. They are useful when note shape varies either continuously, as in the early stages of swamp sparrow vocal development. They are also useful when note shapes can be assigned to discrete types, as in the late stages of swamp sparrow vocal development. They also provide a means for achieving the biologically important but mathematically vexing task of comparing sounds whose variation in
shape lies between continuity and discrete types. They can even be used to determine how many types appear in a data set.

The reusability of acoustic landmarks lends itself to classification analyses. One of the most common, and effective, techniques currently used to classify sounds is to have humans visually classify spectrograms and then verify those categories by post-hoc statistical analyses (Nowicki and Nelson, 1990; Jones et al., 2001; Baker and Logue, 2003). Clustering techniques that use acoustic landmarks can provide similar results to subjective human classification, but with several advantages. First, the algorithm for assigning shapes to different groups is mathematically explicit, rather than psychologically implicit. This makes it possible to test how robust a classification is to variations in classification method or statistics used. For example, I could have used a different criterion for cutting the clusters into groups, such as choosing the number of groups that gave the most predictable note order. Second, acoustic landmarks makes it easy to explicitly change which features are used to classify notes. For example, I used log-transformed frequency values when I classified the swamp sparrow notes in this study. If I had wanted to focus on note shape, to the exclusion of pitch, I could have subtracted the mean frequency of each note from its landmarks before I ran the cluster analysis. Finally, while subjective classifications can successfully group sounds, statistical classification provides a way to test biological hypotheses. Each different clustering technique or set of feature measurements represents a hypothesis about how the study organism perceives and remembers sounds. Acoustic landmarks thus transform sound classification from a black-box bottleneck into a field for generating and testing hypotheses.

The acoustic landmark approach is in direct contrast to efforts based on speech-recognition. In human speech, it is always possible to ask someone for the meaning of a sound. Thus, an efficient way to design a speech recognition system is to create a large library of known sounds, and then devise very rapid ways to summarize new sounds and compare them to the library (Placer et al., 2006; Somervuo et al., 2006; McIlraith and Card, 1995). This approach has been used in a number of automated surveying applications (Farnsworth, 2007; Mellinger and Clark, 2000). This approach will fail to identify sounds that fall outside of the range of variation within the library. Acoustic landmarks describe sounds without reference to anything but the way they appear on a spectrogram,
and as such represent an excellent way to analyze novel sounds. The techniques described in this paper could be used to provide a quantitative summary of the acoustic signals of a newly described species. Library-based approaches depend on knowing the range of variation in advance. Acoustic landmarks provide a way of quantitatively describing sounds that may contain features that no one has thought of before.

Finally, acoustic landmarks capture more information than a typical set of hand-measured parameters because it is possible to interpolate between landmarks. Scaling landmarks by the mean time and frequency of a note yields a unitless description of note shape (Rohlf, 1999). The geometric morphometrics literature represents a rich resource for ideas about how to describe variation in note shape (Bookstein, 1996), as well as ways to use shape statistics in both comparative and developmental frameworks (Rohlf, 1998). Acoustic landmarks can facilitate this research by removing the otherwise onerous effort of measuring spectrograms by hand (Price and Lanyon, 2002).

Acoustic landmarks that measure time, frequency, and amplitude capture the lion’s share of features that appear in the bioacoustics literature. However, there are several ways to extend the usefulness of acoustic landmarks. In its most general form, the process consists of three steps: regular sampling, joining samples into events, and smoothing events into landmarks. While keeping track of the temporal features of events seems universally important, there is nothing inherent in the process that limits it to measuring frequency and amplitude. It would be straightforward, for example, to apply the technique to a quefrency cepstrum (Tchernichovski et al., 2000), or any other signal processing alternative to the spectrogram.

Acoustic landmarks provide speed, flexibility, and biological relevance. They can be a powerful tool for bioacoustics research, whether in furtherance of existing research, or in opening up previously intractable questions.
Figure 1.1. Two different representations of the same swamp sparrow trill. (a) A spectrogram of a swamp sparrow trill. FFT size 512, overlap 91.2%. (b) Acoustic landmarks from the notes of the trill. Each landmark is represented by a circle. The landmarks from each note are connected by line segments.
Figure 1.2. The process of generating acoustic landmarks from a swamp sparrow trill syllable. (a) A spectrogram of the trill, with time resolution of 1 ms and frequency resolution of 86 Hz. The vertical lines enclose the part of the recording used to create (b), a frequency spectrum for one time window of the spectrogram. Two spectral peaks are illustrated (dashed boxes), with asterisks at the amplitude-frequency centroid of each peak. (c) The syllable as a series of peaks, with each peak drawn as a circle at its central frequency and a vertical line showing its frequency bandwidth. There are seven groups of contiguous peaks, three of which correspond to notes from the syllable. (d) A set of five acoustic landmarks for each note. Each landmark (circle) is a time, frequency, amplitude point at the centroid of one fifth of the peaks that make up its note.
Figure 1.3. Two swamp sparrow trills with different acoustic space densities, a measure of singing consistency that is defined in the text. (a) A swamp sparrow trill with high acoustic space density (0.083072). The trill is composed of notes, and each note is represented by landmarks (open circles) connected with line segments. (b) A swamp sparrow trill with low acoustic space density (0.00070373). (c) The acoustic space occupied by the landmarks from the first trill. Each rectangle represents a grid square with duration $\frac{45}{44100}$ s and bandwidth $\frac{44100}{512}$ Hz. Darker grid squares contain more landmarks. (d) The acoustic space occupied by the second trill.
Figure 1.4. Hierarchical clustering with group number chosen by AIC. In all cases, symbols on the plot denote the group to which a note belongs. (a) A dendrogram showing the agglomerative hierarchical cluster made using Ward’s method and the landmarks from the notes found in figure 1.1. (b) The approximate AIC values for each level of the cluster. (c) The acoustic space of the trill as represented by the first two principal components from a PCA of the landmarks from the trill’s notes. (d) An aggregate landmark plot showing the landmarks from each note type plotted on top of one another. The notes are spaced according to the average inter-note interval between the respective types. This kind of plot is useful for verifying that quantitatively determined types make up biologically reasonable categories.
Figure 1.5. Changes in an aggregate measure of trill performance in developing swamp sparrows. The principle component shown here loads strongly on stereotypy of song length, stereotypy of note order, and distance from the trill rate / frequency bandwidth performance limit (see table 1.2). This makes this PC an analog of trill performance. On average, trill performance increases steadily during the period of vocal development observed in this study. The trend line is a locally weighted regression and smooth (LOWESS) (Cleveland, 1979)
Figure 1.6. Changes in an aggregate measure of note performance in developing swamp sparrows. The principle component shown here loads strongly on two measures of note stereotypy: density in acoustic space and distance from note type centroid (see table 1.2). This makes this PC an analog of note performance. On average, note performance increases early, then remains constant or even declines later on in vocal development. The trend line is a locally weighted regression and smooth (LOWESS) (Cleveland, 1979)
Table 1.1. Measures of the vocal output of each swamp sparrow

<table>
<thead>
<tr>
<th>Bird</th>
<th>Songs</th>
<th>Total Song Duration (s)</th>
<th>% Time Singing</th>
<th>Note Types</th>
<th>Notes</th>
<th>Total Note Duration</th>
<th>Note % Time Singing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7290</td>
<td>11659.0</td>
<td>5.78340</td>
<td>11</td>
<td>200913</td>
<td>4968.7</td>
<td>2.46470</td>
</tr>
<tr>
<td>2</td>
<td>7083</td>
<td>11759.0</td>
<td>5.83320</td>
<td>6</td>
<td>208540</td>
<td>6494.7</td>
<td>3.22160</td>
</tr>
<tr>
<td>3*</td>
<td>775</td>
<td>1144.2</td>
<td>0.60542</td>
<td>4</td>
<td>17874</td>
<td>619.94</td>
<td>0.32801</td>
</tr>
<tr>
<td>4</td>
<td>2911</td>
<td>5008.9</td>
<td>2.48460</td>
<td>8</td>
<td>147199</td>
<td>3933.2</td>
<td>1.95100</td>
</tr>
<tr>
<td>5</td>
<td>3023</td>
<td>4157.6</td>
<td>2.06230</td>
<td>4</td>
<td>66056</td>
<td>2316.2</td>
<td>1.14900</td>
</tr>
<tr>
<td>6</td>
<td>1824</td>
<td>2256.9</td>
<td>1.11950</td>
<td>4</td>
<td>35899</td>
<td>1488.0</td>
<td>0.73808</td>
</tr>
<tr>
<td>7</td>
<td>2379</td>
<td>3238.3</td>
<td>1.60630</td>
<td>29</td>
<td>113814</td>
<td>2153.0</td>
<td>1.06800</td>
</tr>
<tr>
<td>8*</td>
<td>2705</td>
<td>4171.2</td>
<td>2.20700</td>
<td>16</td>
<td>64792</td>
<td>1877.1</td>
<td>0.99318</td>
</tr>
<tr>
<td>9</td>
<td>5156</td>
<td>4666.0</td>
<td>2.31450</td>
<td>40</td>
<td>92063</td>
<td>1780.1</td>
<td>0.88300</td>
</tr>
<tr>
<td>10</td>
<td>947</td>
<td>1134.0</td>
<td>0.56251</td>
<td>3</td>
<td>21758</td>
<td>742.21</td>
<td>0.36816</td>
</tr>
<tr>
<td>11</td>
<td>6144</td>
<td>8668.8</td>
<td>4.30000</td>
<td>6</td>
<td>137997</td>
<td>3819.7</td>
<td>1.89470</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Note % Time Singing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3657.9</td>
<td>2363.8</td>
<td>1.3690</td>
</tr>
</tbody>
</table>

Nine of the birds were recorded for a total of 56 hours over 16 days. The two birds marked with * were recorded for 52.5 hours over 15 days.
Table 1.2. Results from a principle component analysis of the parameters measured from swamp sparrow songs. PC1 loads strongly on duration stereotypy, note order stereotypy, and trill performance, making it an axis of trill performance. PC2 loads strongly on note stereotypy, making it an axis of note performance.

<table>
<thead>
<tr>
<th></th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landmarks per grid square per note</td>
<td>0.14704</td>
<td>0.59176</td>
<td>0.49718</td>
<td>0.61726</td>
<td>-0.0018371</td>
</tr>
<tr>
<td>Distance from note type centroid</td>
<td>-0.26250</td>
<td>-0.60614</td>
<td>0.49418</td>
<td>0.24408</td>
<td>-0.5098100</td>
</tr>
<tr>
<td>Coefficient of variation of song duration</td>
<td>-0.50515</td>
<td>0.083312</td>
<td>-0.63262</td>
<td>0.54946</td>
<td>-0.1891200</td>
</tr>
<tr>
<td>Note order predictability</td>
<td>0.66949</td>
<td>0.049534</td>
<td>-0.29829</td>
<td>0.03127</td>
<td>-0.6777700</td>
</tr>
<tr>
<td>Deviation from performance limit</td>
<td>-0.45396</td>
<td>0.522510</td>
<td>0.13934</td>
<td>-0.50650</td>
<td>-0.4949200</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.31380</td>
<td>1.21710</td>
<td>0.95544</td>
<td>0.75689</td>
<td>0.55389</td>
</tr>
<tr>
<td>Cumulative proportion of variance</td>
<td>0.27387</td>
<td>0.52759</td>
<td>0.72676</td>
<td>0.88454</td>
<td>1.00000</td>
</tr>
</tbody>
</table>

23
2.1 Abstract

The relationship between the distance between a pair of songbirds and the amount of repertoire overlap between them is affected by life history traits such as dispersal, as well as song learning traits such as the duration of the sensitive period. In birds with well-studied life histories, such as the tree swallow (*Tachycineta bicolor*), song learning traits can therefore be inferred through studying patterns of song type sharing in space. Syllable repertoires were defined for 57 male tree swallows nesting among four sites in the northeastern United States. The sites were located in Dryden, NY (42.47° N, 76.44° W, 305m), Turners Falls, MA (42.59° N, 72.58° W, 45m), Deerfield, MA (42.49° N, 72.58° W, 49m), and Amherst, MA (42.36° N, 72.51° W, 52m). The sample of 24025 syllables contained 23 different syllable types. Repertoire size ranged from 1 to 10 syllable types, and averaged 3.78 types. Mantel tests found significant correlations between physical distance and repertoire similarity for comparisons between birds within sites ($r = -0.140, p = 0.01$), and significantly higher repertoire sharing between birds that nested at the same site than between pairs of birds nesting at different sites ($r = 0.1846, p = 0.000484$). There was, however, no difference in sharing between birds nesting at separate sites within 50 km of each other than between birds nesting at sites separated by more than 500 km (Mantel $r = -0.000887, p = 0.988$). Ten of the eleven syllable types found in the New York site were also found in sites in Massachusetts. These results suggest that tree swallows share songs with neighbors that nest within hearing range of each other. This spatial pattern could be produced by post-dispersal repertoire changes or by associative nest-site selection based on pre-existing repertoire similarities.
2.2 Introduction

Patterns of song learning in oscine passerines has been an important source of insight in studies of speciation (Huber et al., 2007), behavioral ecology (Beecher and Brenowitz, 2005), neurobiology (Williams, 2004), and endocrinology (Buchanan et al., 2003). Much of this research has focused on a few species that can be kept in captivity, particularly the zebra finch, *Taeniopyga guttata*, the European starling *Sturnus vulgaris*, finches in the genus *Serinus* and New World sparrows in the genera *Zonotrichia* and *Melospiza*. Song learning is an extraordinary feat, and these studies have been particularly effective at elucidating the mechanisms that make it possible (reviewed in Podos et al., 2004a). They have also shown that the details of song learning vary among species, particularly in terms of the timing of learning and the scope of possible models for learning (Podos et al., 2004b; Putland et al., 2006). One current goal in the study of oscine song is to understand the evolutionary forces that shape this diversity of song learning strategies. This requires a comparative approach (Price and Lanyon, 2002), which depends on knowledge about song learning in far more species than have been studied to date. Relatively few songbirds are well-suited to life in captivity, and those that do make up a narrow and unrepresentative sample of songbird diversity. In order to extend our knowledge of song learning throughout the oscine clade, we must have ways to investigate song learning that do not depend on keeping birds in captivity.

One way to infer details about song learning is to look at the geographic distribution of repertoire elements. In songbirds that learn by imitation, there are two mechanisms through which repertoire elements can move around in space. A repertoire element can be transmitted from bird to bird via imitation (Marler and Tamura, 1964), or it can be carried from site to site when a bird disperses (Payne, 1983). In other words, the geographic distribution of song elements is the result of interactions between song learning rules and life history traits (Kroodsma, 1996). These interactions are complex, and involve many different factors (Podos and Warren, 2007). The available information about a species’ life history traits sets limits on the deductions about song learning rules that can be made by studying the geographic distribution of song elements.

The tree swallow, *Tachycineta bicolor*, is a species in which it is likely that we can deduce song learning rules from geographic patterns of song element sharing. Tree swallows readily nest in
artificial nest boxes, facilitating research on all aspects of their life history, from lifetime reproductive success to dispersal (De Steven, 1980; Robertson et al., 1992; Jones, 2003). While tree swallows migrate long distances, they are most likely to breed within 10 km of their natal site (Winkler et al., 2005) and surviving adults generally return to breed at the same site each year (Winkler et al., 2004). These movements are sufficient to create a population that is genetically homogenous across all of North America east of the Rocky Mountains (Stenzler et al., 2009). At the same time, the movements are restricted enough that almost all of a breeding tree swallow’s neighbors will have been born within 15 km of the breeding site (Hosner and Winkler, 2007).

While many workers study the life history traits of swallows, the singing behavior of tree swallows, and of swallows in general, has not been the subject of as much research. Among swallow species, there is evidence of adult song learning in both purple martins, *Progne subis* (Morton, 1985), and barn swallows, *Hirundo rustica* (Galeotti et al., 2001). These are the only two species of swallow for which anything is known about song learning. Horn (1996) made a thorough description of the dawn song behavior of tree swallows, but he reached no conclusions about their form of song learning. He showed that the repertoire elements in this species are 10 millisecond syllables that are uttered repetitively while the bird either flies above or perches upon the nest. He found that there was sharing of syllable types among males, but he did not identify any geographic pattern of syllable sharing. All of his study sites were within 15 km of one another, and it is possible that geographic patterns of syllable sharing occur on larger scales.

I will attempt to distinguish between four different hypotheses about the way tree swallows learn their songs. Each of these hypotheses makes a distinct set of predictions about the relationship between geographic distance and repertoire similarity among tree swallows (Table ??). To test these predictions, I will examine syllable sharing among tree swallows at three different geographic scales. I will compare syllable sharing among neighbors within a site, between birds nesting at sites within 45km from one another, and between birds nesting at sites more than 300 km from each other. The first hypothesis is that tree swallows learn their song repertoires through improvisation. This form of song learning is found in nomadic sedge wrens (Kroodsma et al., 1999). If tree swallows improvise their song repertoires in a similar manner, then there should be very little sharing of syllables
between individuals. In addition, there should be no relationship between distance and repertoire similarity at any geographic scale. The second hypothesis is that tree swallows learn their songs during a critical period that ends before natal dispersal. This form of song learning has been studied in detail in species including Melospiza sparrows (Peters et al., 2000). The pre-dispersal hypothesis predicts high levels of syllable sharing among birds within the same site, moderate sharing among birds nesting at nearby sites, and low sharing among birds nesting hundreds of kilometers from each other. The third hypothesis is that tree swallows learn their songs before dispersal to their breeding site, but bias their vocal output toward shared syllables. This is the selective attrition hypothesis of Nelson and Marler (1994). It has similar predictions to the pre-dispersal hypothesis, but with the added prediction that syllable sharing will be a function of distance between nest locations within sites. The fourth hypothesis is that tree swallows change their repertoires after dispersal by imitating some of the syllables they hear at the new site. This may result from a lifelong process of learning (Hasselquist, 1998), or just through a critical period that is longer than one year (Bitterbaum and Baptista, 1979). This hypothesis predicts that birds will share the most syllables with neighbors whose songs they can hear. Syllable sharing among birds that cannot hear one another’s songs should be low, whether they nest at opposite ends of a large site or hundreds of kilometers apart from one another.

2.3 Methods

2.3.1 Equipment

I used three different combinations of equipment to make my recordings. Recording setup “A” consisted of a Sony TCM-5000 cassette recorder, a Sennheiser ME40 microphone, and a Telinga Parabola. Recording setup “B” consisted of a Sony TCM-5000 cassette recorder and two Radio shack 33-3022 boundary microphones. Each of the microphones was placed directly on top of a nest box. Recording setup “C” consisted of an Olympus DM-20 voice recorder and an Olympus ME-51S microphone. Both pieces of equipment were placed in a quart-sized zip-top plastic bag, and the bag was placed directly on top of a nest box.
2.3.2 Recording

Tree swallows sing before dawn while flying in orbits above their nest sites. Tree swallows sing in flight while it is completely dark, but they usually perch on their nest box and continue to sing at the onset of civil twilight (the period of time when the sun is below the horizon but it is not dark, e.g. the stars are no longer visible). All singing activity stops with the appearance of the sun, and is replaced by mating and agonistic behaviors. For this reason, I made all my recordings of tree swallow song during civil twilight, when each bird could be positively identified as belonging to the nest on which he perched. In Massachusetts and New York, civil twilight occurs from 5:00-6:00 am in May and from 4:00-5:00 am in June.

2.3.3 Sites

I recorded tree swallows at four sites over three different years (Figure 2.1). The first site is in Dryden, NY (42.47° N, 76.44° W, 305m). There are 22 boxes at the Dryden site. I used recording setup “A” at the Dryden site in 1999. The second site is in Turners Falls, MA (42.59° N, 72.58° W, 45m). There are 20 boxes at the Turners Falls site. I used recording sets “A” and “B” at the Turners Falls site in 2004 and recording sets “A”, “B”, and “C” in 2005. The third site is in Deerfield, MA (42.49° N, 72.58° W, 49m). There are 20 boxes at the Deerfield site. I used recording sets “A” and “B” at the Deerfield site in 2004 and recording sets “A”, “B”, and “C” in 2005. The fourth site is in Amherst, MA (42.36° N, 72.51° W, 52m). There are 115 boxes at the Amherst site. I used recording sets “A”, “B”, and “C” at Amherst in 2005.

At each Massachusetts site, I used a handheld Garmin GPS device to record the location of each nest. I used these data to calculate the distance, in meters, between each pair of nest boxes, including distances between nest boxes at different sites. I used aerial photos in Google Maps to estimate distances (to the nearest 20m) between nest boxes in Dryden, NY.

2.3.4 Acoustic analysis

I digitized my analog recordings using Audacity (audacity.sourceforge.net) on a Macintosh G4 computer. I used a sample rate of 441000 kHz, and saved all the resulting files in 16-bit PCM wav or aiff files. I filtered all of these files using a 1500 Hz high-pass filter. I created dynamically
homogenous cuts from the filtered files. Each cut is a continuous stretch of sound from a field recording that has both relatively constant background noise levels and vocalizations from the focal bird with relatively constant amplitude. Cuts varied from 5-30 seconds in length.

I analyzed the digitized recordings using the method of acoustic landmarks (see Chapter 1). The resolution of the analysis was 0.385 ms by 86.1 Hz. In order to detect events, I used an amplitude floor of 85% of the file’s amplitude range and a minimum interval between events of 18 milliseconds. I used 20 landmarks to describe each syllable. After calculating the landmarks with SoundPoints, I performed all subsequent statistical analyses in the R statistical environment (R Development Core Team, 2008).

In order to remove heterospecific and abiotic sounds from the data set, I calculated the mean time and frequency for each detected sound. I retained all sounds with durations between 0 and 0.2 seconds and with mean frequencies between 3 and 5 kiloHertz. Next, I removed all sounds that were below the 5th percentile or above the 95th percentile in either duration or mean frequency. Finally, I removed all syllables from any bird represented by forty or fewer syllables. At the end of this culling process, my data set consisted of 24025 syllables from 57 birds.

### 2.3.5 Classification of syllable types

I used the iterative hierarchical classification procedure described in the previous chapter. In the first iteration, I did separate cluster analyses for each bird’s syllables. In the next iteration, I did separate cluster analyses for example syllables from each site. I included all syllables from Dryden for its analysis (n = 1679). I included the best twenty examples of each type from each bird for the Amherst analysis (n = 2270), and the best forty syllables within each bird-type combination for both the Deerfield (n = 1240) and Turners Falls (n = 1113) analyses.

At this stage, I visually inspected examples of each within-site type and found that one potential syllable type from Amherst was actually a bell call from a blue jay *Cyanocitta cristata*. I excluded the syllables of this type from the final clustering analysis.

For the final iteration, I chose the seventy syllables closest to each site-type centroid as the whole-population training set (n = 2118). I set the nearest-neighbor portion of the iteration to poll
the ten nearest neighbors of each syllable and to designate a syllable as unclassifiable unless at least two neighbors of those neighbors were of the same type.

2.3.6 Repertoire analysis

I defined a bird’s repertoire size in a way that excludes syllable types that made up a very small part of the bird’s repertoire. Some syllables from each bird’s vocal output were assigned types that were rare for that bird. These assignments may have occurred because a syllable has unusual phonology. For example, a syllable could be cut short as it is uttered, masked by another sound in the environment, or garbled through an error in production. The result is that the syllable is a poor fit in any of the well-defined types and is essentially assigned to a type at random. These random assignments represent noise that can lead to overestimation of repertoire size, and therefore must be removed. In order to remove them, I calculated a preliminary repertoire size for each bird that included the rare types. The inverse of the preliminary repertoire size is the expected frequency of each syllable type if all of that bird’s syllables were assigned a type at random. I excluded the syllable types that appeared with a lower frequency than expected by chance. Finally, I calculated the proportion of the corrected repertoire that consisted of each syllable type.

I used each bird’s corrected repertoire to determine the rate at which new syllable types appear as the number of syllables sampled increases. To do this, I divided the recordings of each bird into bouts. I defined the start of a new bout as any pauses between syllables of at least two seconds duration. I included rare types when determining pauses between bouts, but I did not include rare syllable types when I counted number of syllables sampled or number of types present in a bout.

I calculated a modified version of the Bray-Curtis repertoire difference between the repertoires of each pair of birds, including comparisons between birds found at different sites. I used the *ecodist* package to calculate differences between repertoires (Goslee and Urban, 2007). If each repertoire is a vector of syllable type frequencies with sum equal to one, then $\Delta(a, b)$, the modified Bray-Curtis difference between repertoire $a$ and repertoire $b$, is:
\[ \Delta(a, b) = 1 - \sum_{i=1}^{n_{\text{types}}} \begin{cases} a_i + b_i & \text{if } a_i > 0 \text{ and } b_i > 0 \\ 0 & \text{otherwise} \end{cases} \]

This measure combines features of Euclidean distance and Jaccard dissimilarity because, when two birds share a syllable type in common, the difference between the two birds depends on how often each bird sings that type. In contrast, the birds are considered maximally different when only one of them uses a syllable type in his repertoire (Quinn and Keough, 2002). For example, in a system with three syllable types, repertoires \( A = (0.5, 0.3, 0.2) \) and \( B = (0, 0.3, 0.7) \) have a Euclidean distance of 0.71, a Bray-Curtis difference of 0.5, and a Jaccard dissimilarity of 0.33.

### 2.3.7 Geographic patterns

In order to look for geographic patterns of repertoire similarity, I compared the matrix of differences between birds’ repertoires to matrices that describe the spatial distance between birds’ boxes. Comparisons between pairwise distance matrices must account for the fact that the values within each matrix are not statistically independent. Mantel’s test achieves this by using resampling to compare the observed correlation of a set of matrices to a range of possible correlations calculated by randomly rearranging the elements of the matrices (Mantel, 1967). A Mantel test with categorical distance matrices is analogous to a \( t \)-test, and a Mantel test with continuous distance matrices is analogous to a linear regression (Sokal and Rohlf, 1995). In each of the Mantel tests described below, I used one million permutations to calculate the distribution of possible correlation values. I used the ecodist package to calculate distances between boxes and to perform the Mantel tests (Goslee and Urban, 2007). I looked for a relationship between repertoire similarity and spatial distance at three different spatial scales.

At the most local geographic scale, I tested the relationship between spatial separation and repertoire difference for pairs of birds that nested within the same site during the same year. Spatial distance, for this comparison, was calculated as the number of meters between boxes. Both the selective attrition and the post-dispersal learning hypotheses predict that birds nesting farther from one another within a site will have greater repertoire differences than birds nesting closer...
to one another. The improvisation and the critical-period imitation hypotheses predict no spatial pattern of repertoire differences within sites (see Table ??).

At the intermediate geographic scale, I compared repertoire differences among pairs of birds that nested at the same site to repertoire differences among pairs of birds that nested at separate sites. Spatial distance in this comparison was zero if both birds nested at the same site and one if they nested at separate sites. All three imitation hypotheses predict greater mean repertoire differences between sites than within sites. Only the improvisation hypothesis predicts an equal amount of repertoire difference both within and between sites.

At the largest geographic scale, I compared repertoire differences among pairs of birds that nested at separate sites within 50 km of one another to repertoire differences among pairs of birds that nested more than 500 km from one another. I used a partial Mantel test for this comparison, in which the spatial distance between birds was set to one if they nested more than 500 km from each other, zero if both birds nested less than 500 km for each other, while removing any potential effect of within-site versus between-site comparisons. The critical-period and the selective attrition hypotheses predict higher mean repertoire differences between birds in separate states than in birds in different sites within the same state. Both the improvisation and the post-dispersal imitation hypotheses predict that there will be no difference in repertoire dissimilarities at these two spatial scales.

2.4 Results

I recorded 24025 syllables from 57 birds across all of my study sites. I recorded 1354 syllables from eight birds at the Dryden, NY site. I recorded 2789 syllables from four birds at the Turners Falls site in 2004 and 710 syllables from four birds in 2005. In 2004, I recorded 2546 syllables from three birds at the Deerfield site, and I recorded 1153 syllables from four birds in 2005. I recorded 10529 syllables from 34 birds at the Amherst site in 2005. On average, each bird was represented by 398 syllables, with a range of 41 to 2127 syllables. My sampling effort was lower at Dryden than at the other sites (Table 2.2), but still sufficient to capture each bird’s full syllable repertoire (see below).
2.4.1 Syllable classification

I identified twenty-three different types in the set of tree swallow syllables. These types varied in both overall time-frequency envelopes and in internal structure (Figure 2.2). The average syllable duration ranged from 40 to 171 milliseconds, with a mean of 110 ms. The average syllable frequency ranged from 3.35 to 4.42 kHz, with a mean of 3.92 kHz. The average frequency bandwidth of the syllables ranged from 1.06 to 2.94 kHz, with a mean of 1.8 kHz. These measures all agree with those of Horn (1996). The syllable types also varied in complexity, containing from 1 to 11 notes, with a mean of 4.5 notes. While there is variation among note types, as well as variation among individual birds’ renditions of a syllable type (Figure 2.3), each bird seems to produce the types in its repertoire in a highly stereotyped way. This can be seen through the similarity between aggregate landmark plots and syllable spectrograms (Figure 2.2, 2.3). My analysis yielded substantially larger number of syllable types than Horn found. The most likely explanation for this discrepancy is that Horn used a small number of qualitative features to determine his among-birds syllable type categories, while I used a large number of quantitative features. The combination of individual stereotypy and complex variation among types suggest that tree swallow syllables may encode a large amount of information.

2.4.2 Repertoire analysis

The tree swallows that I recorded sang their songs with immediate variety. In a typical bout of song, a male tree swallow alternates between his two most common syllable types, occasionally interrupting this regular pattern with one of his less-common types. These interruptions occur often enough that, on average, a bird will have presented his entire repertoire by the thirtieth syllable of a bout (Figure 2.4). These observations are in general agreement with those of Horn (1996), although this study found slightly larger individual repertoires, and therefore required recording a correspondingly larger number of syllables in order to capture a bird’s entire repertoire.

Each bird sang a unique combination of syllable types. Repertoire size ranged from 1 to 10 syllable types, with a mean of 3.78 types per bird. On average, birds shared at least one syllable type with thirty-three other birds. This repertoire overlap is not just a result of the classification
process: one bird from Turners Falls shared no syllable types with any other bird. This bird had a repertoire of two distinct types, each of which were unique to him. Of the twenty-three syllable types, three were found at all four sites, six were found at three sites, five were found at two sites, and nine were found at only one site, including the two types unique to a single individual (Table 2.2).

2.4.3 Geographic patterns

Among pairs of birds nesting at the same site, there was a significant positive correlation between repertoire difference and distance between nest boxes (Mantel $r = 0.140025$, $p = 0.009999$). While some pairs of birds had high repertoire differences across the full range of within-site distances, the lowest repertoire differences (i.e., greatest repertoire similarities) were only found among birds nesting within 300 m of one another (Figure 2.5). Pairs of birds nesting at the same site had significantly lower repertoire differences than pairs of birds nesting at separate sites (Mantel $r = 0.184640$, $p = 0.000484$). Repertoire differences among pairs of birds nesting at separate sites within 50 km of one another (both birds in Massachusetts) were just as high as repertoire differences among pairs of birds nesting at sites 500 km apart from one another (one bird in MA, the other in NY) (Mantel $r = -0.000887$, $p = 0.988097$). Large repertoire differences were found between birds separated by all three degrees of spatial distance, while the lowest repertoire distances were found only among pairs of birds nesting at the same site (Figure 2.6).

2.5 Discussion

Physical proximity between birds is correlated with repertoire similarity in tree swallows, but only at small spatial scales. Within a site, syllable repertoires are most similar to those of other birds that nest nearby (Figure 2.5). Syllable repertoires are also more similar within sites than they are between sites. However, when comparing repertoires from different sites, the distance between sites is not related to repertoire similarity (Figure 2.6). The average degree of repertoire difference among pairs of birds increases as the distance between birds increases at small spatial scales, but is constant at large distances. In addition, the range of repertoire differences is similar at all scales.
Many birds that nest at the same site do not share any syllable types in common. The spatial pattern that emerges from these observations is that tree swallows that cannot hear each other have a random level of repertoire similarity, and tree swallows that nest within potential hearing range have an above-average degree of repertoire similarity.

The improvisation hypothesis is not consistent with the observation that tree swallows share syllable types and by the observation that physical proximity is correlated with the extent of syllable sharing. If each bird had a totally improvised repertoire then there would be no sharing of types between birds (Kroodsma et al., 1997). It is possible that the syllable types I found in my analysis are not imitated, but instead the result of coincidental improvisations of similar syllables. That is unlikely to be the case, for three reasons. First, the range of variation among syllable types is so large that it seems improbable that tree swallows would independently choose identical syllables by chance (Figure 2.2). Second, the classification method did identify one individual as having a distinct repertoire with two syllable types that he did not share with any other individuals. This is evidence that the classification method is not simply lumping loosely similar individual types together. The third, and most important contradiction is the increased repertoire similarity among neighbors. If birds are improvising random types and my analysis is lumping those types together, there is no reason that birds nesting close to one another should have a higher chance of sharing syllable types. A likely explanation for higher sharing among neighbors is that tree swallows are changing what they sing based on what they hear from their neighbors.

The critical-period hypothesis is not consistent with the combination of repertoire similarity among neighbors within sites and the lack of association between repertoire differences and distances between sites. The observation of no large-scale spatial pattern in repertoire similarity is also inconsistent with the predictions of the selective attrition hypothesis. Within-site patterns of sharing with neighbors are observed in species with short critical periods and low dispersal (Wilson et al., 2000). Adult tree swallows prefer to breed at or near the nest box that they used in previous years (Winkler et al., 2004; Shutler and Clark, 2003), but in a critical-period learner, adult philopatry can only lead to similarity between neighbors if the adults shared syllable types when they first chose their nest sites. There is no evidence that first-year tree swallows choose nest sites adjacent to
their parents’ nest (Hosner and Winkler, 2007; Winkler et al., 2005). Furthermore, if tree swallows crystallize their repertoires before dispersal then the distance between sites should be correlated with the amount of repertoire similarity among birds from those sites. Dispersal across large distances is rare in tree swallows (Hosner and Winkler, 2007; Winkler et al., 2005), so syllable types that have been carried by such events should also be rare (Cicero and Benowitz-Fredericks, 2000; Tracy and Baker, 1999). Instead, I found that birds from one of my three sites in western Massachusetts were just as likely to share a syllable type with a bird in Dryden, NY as they were to share a type with a tree swallow from another site in the same state.

The observed spatial patterns of repertoire similarity supports the hypothesis that tree swallows can change their repertoires by imitating syllable types that they hear after dispersal. According to this hypothesis, repertoire similarity within sites is a direct result of copying by neighbors. Repertoire differences between sites are a result of the combination of the random pool of available syllable types and idiosyncratic copying choices at each site (Kroodsma et al., 2002). The uniqueness of each site means that adjacent sites are just as different from each other as they are from distant sites. This combination of a strong local relationship and no large-scale relationship between physical distance and repertoire difference has been used to infer postdispersal learning in a diverse collection of bird species. These taxa include parrots such as yellow-naped amazons, *Amazona auropalliata* (Wright and Wilkinson, 2001) and budgerigars, *Melopsittacus undulatus* (Hile et al., 2000); and passerines such as chowchillas, *Orthonyx spaldingii* (Koetz et al., 2007), blackeyed bulbuls, *Pycnonotus barbatus* (Lloyd et al., 1999), and house finches in California, *Carpodacus mexicanus* (Bitterbaum and Baptista, 1979). These species span a wide range of breeding phenologies, social systems, and migratory behaviors, but in each case there seems to be selective pressure for singers to sound like their neighbors.

The variation that I observed among birds with respect to repertoire size and organization can also be explained by post-dispersal learning. A few birds sang a large number of syllable types, while others sang only one syllable type. If the birds that I recorded were at different stages of post-dispersal song learning then that would explain why they differed from one another in repertoire organization. The sampling process that I used for this study makes it is impossible to say which,
if any, of these birds were in the process of modifying their repertoires. However, if the different stages of song learning in tree swallows have detectable differences in repertoire organization, then those differences may provide useful information about a bird’s arrival time, cognitive ability, or some other aspect of male quality related to song production.

### 2.5.1 The function of post-dispersal learning

Several hypotheses exist about the adaptive value of post-dispersal learning. Two common hypotheses are particularly unlikely in tree swallows. First, in many species, lifelong learning produces a correlation between repertoire size and age (van Dongen, 2006; Gil et al., 2001; Hasselquist, 1998; Eens et al., 1991a). Tree swallows have small repertoires and the majority of each bird’s vocal output involves rotating one at a time through 1-3 syllable types. This type of organization of song output is poorly suited to showing off repertoire size. It is therefore unlikely that tree swallows are learning after dispersal as part of a lifelong process of increasing repertoire size. Second, in many species, shared song elements are used as directed signals towards a particular neighbor during territorial interactions (Searcy and Beecher, 2009). As mentioned above, tree swallow dawn song consists of a small number of syllable types repeated in a regular syntax. This organization is not well-suited as a directional signal because shared syllable types are constantly recurring. Additionally, the physical location of tree swallows as they sing dawn song does not lend itself to directional interactions. Tree swallows sing at dawn while either flying in orbits around or perched directly adjacent to their nest sites (Horn, 1996). Since they are either in constant motion or at the center of their small territories, they are not in a good position to focus their singing at a particular neighbor. Finally, tree swallows have a well-documented vocabulary of agonistic calls that are distinct from their dawn song syllables (Robertson et al., 1992). I have observed that a tree swallow will interrupt its dawn song and switch to this category if it flies near a neighbor during an orbit, or if a conspecific approaches the nest cavity (Taft, personal observation). One possible function of post-dispersal learning is to signal group membership. Call convergence is an important part of group cohesion in many species of birds (Farabaugh et al., 1994; Nowicki, 1989). In some colonial species of swallow, parents and offspring use vocalizations to maintain contact in crowded colonies.
and flocks (Medvin et al., 1992; Medvin and Beecher, 1986; Beecher et al., 1985). Tree swallows form huge flocks during migration. The contact calls that they use in these flocks are similar to dawn song syllables (Taft, personal observation).

It is possible that smaller social groups exist within the huge migratory flocks. If they do exist, then group cohesion may be facilitated by call convergence through vocal imitation. If these groups could maintain their association from fall migration until nest site selection in the spring, then groups of birds with similar contact calls could nest together, creating the spatial patterns of repertoire similarity described above. Any repertoire similarities between neighbors would not have any signal value because they would be epiphenomena of group membership (Warren, 2003).

This kind of non-breeding association could be beneficial to younger birds if it allows them to follow older birds to good nest sites. Older or more dominant birds might in turn benefit from increased opportunities for extrapair paternity. All of the birds would benefit from having established positions in a dominance hierarchy, making the process of nest site competition smoother and less costly. Some significant costs might also be associated with this kind of association. More birds means more competition for scarce nest sites. Younger or subordinate birds could lose paternity, or even the opportunity to claim a nest cavity. Finally, a bird in a large migrating flock does not need to belong to a stable subgroup in order to follow older birds to a potential nest site.

A more plausible function of post-dispersal learning is as a signal of how long a bird has been at the breeding site. Song learning is not an instantaneous process. The time required to learn a new syllable type is a constraint that enforces honesty. If a pair of birds share a syllable type, one of them must have been a resident long enough to learn and copy one of the elements of his neighbor’s repertoire, and the second bird has been a resident for at least as long as the first one. The ability to signal arrival time could benefit tree swallows in both intra- and intersexual selection.

A male tree swallow could also benefit by signaling territory tenure during competition for nest sites with other males. Tree swallows are obligate cavity nesters, but lack the ability to create their own cavities. As a result, competition for nest sites is intense, and a population of floaters exists at many sites. A bird that has occupied a nest cavity long enough to imitate his neighbors’ repertoires might appear less of a threat to those neighbors than a nestless floater. This is the classic "dear
enemy” scenario (Temeles, 1994). One test of this role for syllable sharing could be made at sites where tree swallows and violet-green swallows, *Tachycineta thalassina*, occur in sympatry. Since the two species compete for nest sites, heterospecific syllable sharing would have the same benefits as conspecific syllable sharing. This phenomenon may occur in Scandinavian *Ficedula* flycatchers (Gelter, 1987).

A male tree swallow could benefit from signaling territory tenure while trying to attract extrapair copulations. There is a high degree of extrapair paternity in tree swallows (Kempenaers et al., 2001). When nest sites are scarce, females tree swallows may choose a resident male simply because he controls a place where she can lay eggs. She may look to neighboring males for the genes to fertilize those eggs. One criterion she might be interested in is arrival time. The ability to forage early in the season, when food is rare, is a key factor in female fitness (Bowlin and Winkler, 2004). Females arrive on the breeding site after males do, so females cannot directly assess male arrival time. Females could therefore benefit from using repertoire sharing with neighbors as an index of male arrival time. A female choosing on the basis of repertoire sharing would be able to identify recent male arrivals because their repertoires would be the least similar to their neighbors. The disadvantage of standing out from the repertoire neighborhood could also be the reason why the earliest males do not change their repertoires to remain distinct.

If post-dispersal imitation does serve the purpose of signaling arrival time, then several possible strategies are available to a male when he arrives at the breeding site. After the first male arrives, the next male will probably choose a nest that is beyond hearing range from the first male’s nest box. Each of these males could then maintain their own distinctive repertoires and appear equally early. Successive male arrivals would each choose an isolated box, ultimately serving as a nuclei of separate neighborhoods of repertoire similarity. These neighborhoods would only begin to fill in after all of the isolated nest cavities had been taken. A bird’s physical proximity to the center of a repertoire neighborhood would also signal arrival time. The latest males would find themselves at the edges of neighborhoods, matching some of their neighbors closely and not sharing any syllable types with others. The result of this process would be a complex mosaic of repertoire neighborhoods
and birds with intermediate repertoires. This is precisely the spatial pattern that I observed at the Amherst site (Figure 2.7).

2.5.2 Implications about song learning in other swallows

The fact that tree swallows migrate and breed synchronously has influenced all of the adaptive explanations that I have advanced for the role of post-dispersal song learning in tree swallow reproductive biology. Many other species in the genus *Tachycineta*, including two of the tree swallows’ three closest relatives (Whittingham et al., 2002), are tropical birds that do not migrate and have less synchronous breeding seasons. These species would therefore encode very different information into their repertoires by learning elements after dispersal. For example, open-ended learning allows some long-lived, sedentary, tropical oscines signal age via repertoire size because birds accumulate new syllable types throughout their lives (Coleman et al., 2007). It is not clear from this study whether tree swallows are open-ended learners, or if they have a delayed critical period that ends after dispersal. It is also possible that song learning parameters vary widely between different swallow species, as is the case in their cousins, the wrens (Kroodsma et al., 1999).

Table 2.1. Predicted relationships between repertoire similarity and spatial distance from four hypotheses about song learning. These hypotheses stem from the observations that tree swallows are migratory birds with low philopatry and long-range natal dispersal distances.

<table>
<thead>
<tr>
<th>scale</th>
<th>correlation between repertoire difference and spatial distance</th>
<th>continuous distance between boxes</th>
<th>categorical within vs. between sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>same site</td>
<td>different sites</td>
</tr>
<tr>
<td>pre-dispersal</td>
<td>0</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>selective attrition</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>post-dispersal</td>
<td>+</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>process</td>
<td>hearing</td>
<td>dispersal</td>
<td>both</td>
</tr>
</tbody>
</table>
Figure 2.1. The locations of the four sites where tree swallows were recorded for this study.

Table 2.2. Sampling effort and syllable type occurrence at the four sites of the study. If a syllable type is found at only one site, then it is uniquely present at that site. If a syllable type is found at three sites, then it is uniquely absent from the site where it is not found.

<table>
<thead>
<tr>
<th>site</th>
<th>birds</th>
<th>syllables</th>
<th>syllables per bird</th>
<th>syllable types</th>
<th>unique types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amherst, MA</td>
<td>34</td>
<td>10529</td>
<td>214.0</td>
<td>18</td>
<td>4  0</td>
</tr>
<tr>
<td>Deerfield, MA</td>
<td>7</td>
<td>3699</td>
<td>297.5</td>
<td>7</td>
<td>0  2</td>
</tr>
<tr>
<td>Turners Falls, MA</td>
<td>8</td>
<td>3499</td>
<td>390.0</td>
<td>13</td>
<td>4  2</td>
</tr>
<tr>
<td>Dryden, NY</td>
<td>8</td>
<td>1354</td>
<td>105.0</td>
<td>11</td>
<td>1  2</td>
</tr>
</tbody>
</table>
Figure 2.2. Spectrograms and aggregate landmark plots of each of the twenty-three syllable types found in this study. The landmarks of every syllable from that type are plotted on top of one another to create an aggregate syllable shape on the left. A spectrogram of an example syllable is shown on the right. The example syllables were the closest to the type centroid among the loudest 50% of the syllables of that type. Syllable types are listed in order of frequency. Syllable type numbers are arbitrarily assigned by the clustering algorithm and have no biological significance.
Figure 2.3. Variation among birds that sang a specific syllable type (#2 from Figure 2.2). Each syllable is represented in the same way as in Figure 2.2. Syllables from eight birds are shown, with each site represented by two birds.
Figure 2.4. The average saturation curve showing the cumulative number of syllable types in a bout. Tree swallows sing a small repertoire of syllable types with immediate variety, so that nearly all of the repertoire has been sampled by the tenth syllable in a bout.
Figure 2.5. Repertoire difference versus spatial distance (m) for pairs of tree swallows in which both birds nested within the same site during the same year.
Figure 2.6. Mean repertoire differences between pairs of birds are lowest within a site and equally high between sites separated by tens of kilometers as between sites separated by hundreds of kilometers. The boxes show the medians and quartiles for each range. Whiskers enclose all points less than 1.5 times the interquartile range from the median. Small circles represent outliers beyond that distance from the median. Different letters indicate whether there was a significant difference in repertoire dissimilarity between different categories of spatial distance. Pairs of tree swallows nesting within the same site have lower mean repertoire differences between each other (higher syllable sharing) than pairs of birds nesting at different sites (Mantel $r = 0.184, p < 0.0003$). Pairs of tree swallows nesting in different states have the same mean repertoire differences as pairs of birds nesting at different sites within the same state (Mantel $r = -0.004, p = 0.9432$).
Figure 2.7. The mosaic of syllable sharing groups at the Amherst College site. The repertoire of each bird is represented by a barplot in which the height of each color is proportional to the frequency of a syllable type. The location of each bar corresponds to a bird’s nest site. Note the variation in repertoire composition among birds. Spatially homogenous groups of birds with similar repertoires are connected with straight dark lines. Groups of birds with similar repertoires with disjunct spatial distributions are connected with curved light lines. Repertoire groups were chosen from a Ward’s hierarchical cluster of repertoire differences.
Table 2.3. Mantel tests of divergence between syllable repertoires across different spatial scales. The within-site comparison tested for a correlation between repertoire differences and distances between boxes. The other two comparisons tested for differences in average repertoire differences among pairs of birds separated by different spatial scales. In each test, the sample size was 57 birds. All tests used one million permutations.

<table>
<thead>
<tr>
<th></th>
<th>Mantel $r$</th>
<th>95% conf. int.</th>
<th>two-tailed $p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within sites (0 – 1.6km)</td>
<td>0.140</td>
<td>0.091 – 0.190</td>
<td>0.00100</td>
</tr>
<tr>
<td>Same (0 – 1.6 km) vs. separate (17.6 – 516.4 km) sites</td>
<td>0.185</td>
<td>0.142 – 0.238</td>
<td>0.00048</td>
</tr>
<tr>
<td>Nearby (17.6 – 42.0 km) vs. distant (506.6 – 516.4 km) sites</td>
<td>-0.001</td>
<td>-0.044 – 0.036</td>
<td>0.98810</td>
</tr>
</tbody>
</table>
CHAPTER 3
COPYING AMONG MOVING TARGETS: SYLLABLE SHARING IN TREE SWALLOW DAWN SONG

3.1 Abstract

Tree swallows (Tachycineta bicolor) share song syllables with neighbors that nest within hearing range, but are no more likely to share syllables with birds nesting a kilometer away than with one nesting 500 kilometers away. Three possible explanations for this spatial pattern of repertoire similarity are nest-site selection based on pre-dispersal repertoire overlap, post-dispersal imitation of neighbors’ syllables, or post-dispersal convergence among neighbors via syllable-type modification. In order to distinguish among these hypotheses, I recorded and classified 11782 dawn song syllables from 41 tree swallows during May 2006. Each bird was recorded multiple times, with a median of 49.5 syllables per day on 4 separate days spread across 3 different weeks. Individuals’ syllable type repertoires changed from day to day and week to week, demonstrating post-dispersal repertoire change. This change in repertoire composition did not result in any increase in average repertoire size, and approximately equal proportions of birds showed increases (0.338), no changes (0.323), or decreases (0.338) in repertoire size. Syllable performance did change, showing significant improvement during the month ($F_{1,236} = 24.9, p < 0.0001$), regardless of whether the syllable in question was a recent addition or had been in a bird’s repertoire on the first day it sang. These results show that the syllables of tree swallow dawn song are not static song types but plastic repertoire elements that can be modified through imitative learning even after the first year of a bird’s life, and thus support the post-dispersal convergence hypothesis.
3.2 Introduction

The standard model of a bird song repertoire is a set of crystallized element types. The set itself may or may not be mutable, able to accept new elements or eliminate unwanted ones, but the elements themselves do not change after they have been mastered during a process of sensorimotor learning. However, this model is not the simplest one that can explain some observed cases of vocal learning. Call convergence in parrots (Farabaugh et al., 1994; Scarl and Bradbury, 2009), chickadees (Nowicki, 1989), and crossbills (Sewall, 2009) can be explained as modification of plastic, rather than crystallized, element types. There is also evidence for plasticity in the occurrence and sequence of elements in song like those of canaries (Leitner et al., 2001), and starlings (Chaiken and Bohner, 2007), which are series of many elements of different types. The element types themselves do not seem to be plastic in these species. To date, there is no strong evidence in any species for plasticity of elements used in full adult song.

A documented case of a species in which there are changes to existing adult song elements, rather than the addition of new elements, would have important implications for how we think about song learning. First, it would widen the scope of information that we currently believe songs can carry. Crystallized song elements contain information about the past, such as where a bird lived when it learned its song (Podos and Warren, 2007), and what its health was at that time (Nowicki et al., 1998). Plastic song elements, in contrast, would encode more up-to-date information about the bird, such as the social group it belongs to (Ford, 1991), and, potentially, its current ability to produce a challenging song (Podos et al., 2004b). Second, the existence of plastic elements in adult repertoires would expand the theoretical possibilities of what the song template is. A long-standing view describes the crystallized song template as an static neural representation that is no longer able to be influenced by outside stimuli (Adret, 2004). A plastic template would still accept such stimuli. Since there are clearly species, such as chickadees and crossbills, in which such stimuli can still effect template change, as well as species such as song sparrows in which crystallization is truly rigid (Nordby et al., 2002), there is most likely a continuum of song template mutability, if the call templates of chickadees and crossbills are stored in the same way as are the song templates of sparrows. In some species, it may be under constant influence from a wide variety of sources,
while in others only a few things can change the template, and in still others it cannot change at all. Third, it raises questions of unitary storage of song templates. In the case of crystallized types, the idea of a single unitary location for storage is tidy, and there is some evidence for it in species where neurons within RA fire in association with specific song elements (Mooney, 2009). However, a composite representation that integrates information from several different brain regions makes particular sense for plastic song template, since changes within different parts of the composition provides a mechanism for partial changes to the acoustic properties of the template. This idea is supported by the observation that portions of both HVC and RA have activity related to specific acoustic properties at a lower level than that of the song element during both song perception and song performance (Mooney, 2009).

Tree swallows (*Tachycineta bicolor*) are a species in which song elements may be plastic. The elements, called syllables, of tree swallow dawn song are short call-like bursts of sound that are repeated three times a second in flight or while perched (Horn, 1996). In Chapter 2, I showed that tree swallows share syllable types with their nearby neighbors, and are much less likely to share syllable types with individuals that nest beyond hearing range, whether those birds are found a few hundred meters away at the same site, or at a different site hundreds of kilometers away. Syllable plasticity should have effects on large-scale geographic patterns of similarity within repertoire elements. The presence of large-scale clines of acoustic properties within notes, a pattern that is seen in the African warbler *Cisticola erythrosparsus* (Benedict and Bowie, 2009), suggests a stepping-stone model where each bird is only slightly different from its neighbor. In the complex mosaic of habitats that characterizes any species’ range, syllable plasticity should break up any such smooth transitions into transient local variations, which is the pattern found in tree swallows. Many parrots show micro-geographic patterns of repertoire similarity such as this, and the cause of this is ascribed to call convergence among flock mates (Wright, 1996). This combination of appropriate acoustic properties and promising geographic patterns of sharing makes it worthwhile to investigate whether or not tree swallows syllable repertoires change over time, and if they do, whether or not existing syllable types are plastic or crystallized within individual bird’s repertoires.
Other learning processes besides syllable plasticity could explain the observation that tree swallows nesting near one another share more syllable types. Each learning process will in turn have different effects on three different aspects of how tree swallow singing might change over the course of the breeding season (Table 3.1). First, changes in repertoire size are only expected if some tree swallows imitate their neighbors, adding copied syllable types to their neighbors. Second, imitation through sensorimotor learning also means that syllable performance should improve over time (Tchernichovski et al., 2001). Third, if imitated notes are added to a repertoire of already crystallized notes, then there should be a difference in the rate of performance improvement between original syllable types and added ones. Examining these three aspects of singing performance in tree swallows will allow me to determine which hypothesized learning process can best account for how tree swallows match their neighbor’s repertoires.

The first hypothesis is that there is no song learning after dispersal in tree swallows. In this case, repertoire similarity would be a result of assortative nest site selection. Song sparrows (*Melospiza melodia*) are an example of this song learning process because they have a pre-dispersal critical period (Nordby et al., 2002), but share song types with neighbors because song type sharing is involved in mediating disputes during territory acquisition (Wilson et al., 2000). The repertoire stasis hypothesis predicts that there will be no change in repertoire size or syllable performance.

The second hypothesis is that tree swallows match their neighbors by expanding their repertoire to include some of their neighbors’ syllable types. Repertoire expansion is the mechanism through which age and repertoire size are correlated in lifelong learners (Mountjoy and Lemon, 1995). Birds with this learning mechanism have two categories of repertoire elements – existing elements that have reached a more or less crystallized form, and new elements that must still be practiced before they can be formed with complete consistency (Chaiken and Bohner, 2007). Consequently, the repertoire expansion hypothesis predicts that repertoire size should increase over the course of the season. Furthermore, it suggests that syllable performance will improve with time, but only in any syllable types that a bird adds to its repertoire, not in any syllable types found in its repertoire on the first day that it sings.
The third hypothesis is that tree swallows match their neighbors through a process of modifying existing syllable types to achieve convergent types that they share with their neighbors. This kind of partial matching can be seen at the within-song level in species with complex songs with multiple elements, such as canaries (*Serinus canaria*) (Lehongre et al., 2009). It is more frequently associated with contact call convergence in social species including black-capped chickadees (*Poecile atricapillus*) and budgerigars (*Melopsittacus undulatus*) (Nowicki, 1989; Hile et al., 2000). The syllable convergence hypothesis does not make any specific hypotheses about repertoire size because it is reasonable to think that birds which are able to alter their stored syllable templates would also be able to add or subtract syllable types from their repertoires. However, it does make the strong prediction that all syllable types, both those found in a bird’s original repertoire as well as any that are subsequently added, could potentially show improved performance over time as birds gain more experience singing their modified syllables.

The information content in a tree swallow’s dawn song depends to a great extent on the process it used to learn that song. Innate syllable types carry very different information than do learned ones, especially because the location (Podos and Warren, 2007) and conditions (Nowicki et al., 1998) in which a bird learns encode information into those songs. An important step towards identifying the information content by which a bird’s songs affect the behavior of those that hear it is to identify the mechanisms through which a bird learns its songs.

### 3.3 Methods

#### 3.3.1 Recording

I recorded tree swallows around sunrise at fields belonging to Amherst College in Amherst, MA (42.36° N, 72.51° W, 52m) throughout May 2006. There are 115 boxes at the Amherst site, spread across wet meadows that slope down from woods to a small river. I made all my recordings in 2006 with a Sennheiser ME40 microphone, a Telinga Parabola, and a Marantz PMD750 digital audio recorder. I recorded directly to 16-bit PCM WAV-format files. Using this handheld setup, I would quietly approach a bird while it perched on or near its nest box and sang during civil twilight. If there were many birds singing, I would count syllables as I recorded a male, and after recording
at least one hundred of them I would move on to the next bird. The site, as well as the bird population, were too large for me to record every bird during the 15 to 90 minutes before dawn when birds sang while perched, so I focused my recording efforts on opposite ends of the site on alternating days. After recording each day, I transcribed any spoken notes that I had made while recording, then split the continuous recordings into cuts labeled according to time, date, and nest box and archived the files on an Apple Macintosh laptop for future analysis.

### 3.3.2 Landmarks

I used SoundPoints software to generate a set of acoustic landmarks for each tree swallow syllable. I provided SoundPoints’ event detector with time, frequency, and amplitude parameters that it used to separate tree swallow syllables from background noise. The time resolution of the analysis was 0.998 ms. The frequency resolution was 86.1 Hz. At each time step of the analysis, SoundPoints calculated the frequency and amplitude of the loudest spectral peak. A peak was only included in the analysis if its amplitude was more than 0.25 dB above the average peak amplitude over the previous 75 ms and its frequency was between 1500 and 9000 Hz. Two frequency traces were treated as separate syllables if their adjacent ends were separated by more than 75 ms, 2500 Hz, or 0.5 dB. The set of landmarks describing each syllable contained thirty time-frequency-amplitude points. I also calculated the start time, duration, frequency bandwidth, and dynamic range for each syllable.

I loaded the resulting landmarks into R to transform the data and conduct further statistical analyses (R Development Core Team, 2008). I transformed the landmark data so that the time, frequency, and amplitude values, which have very different scales, could be used in the same analysis. First, I log$_{10}$-transformed the frequency data so that they conformed to vertebrate perceptions of pitch (Deecke and Janik, 2006). Next, I calculated the mean amplitude of each syllable’s landmarks, and then subtracted its mean from those values. This transformation converted the amplitude variables from measures of absolute loudness, which is only related to the distance of the bird from the microphone, to measures of the relative change in loudness over the course of the syllable, which is a useful diagnostic feature. I then removed the thirtieth amplitude measure from the analysis.
because its value was totally determined by the mean amplitude and the twenty-nine other relative amplitudes. Finally I z-transformed all of the variables, removing the potential for spurious effects caused by the different scales of milliseconds, Hertz, and decibels.

### 3.3.3 Removing heterospecific sounds

The settings that I used detected more than just tree swallow syllables. Consequently, I inspected the syllables using two statistical approaches for detecting false positives. First, I compared the descriptions of the syllables in my data set to the known range of variation in tree swallow song syllables (Robertson et al., 1992; Sharman et al., 1994; Horn, 1996). I removed all sounds with durations greater than 250 ms, and all sounds with a mean frequency outside of the range between 3000 and 5000 Hz. I did not remove any syllables based upon amplitude measures because amplitude, especially in field recordings, is heavily influenced by the distance between the subject and the microphone (Wiley, 1991). Second, I performed a series of classification analyses in which I plotted aggregates of the landmarks of each sound type and removed sound types that clearly consisted of sounds other than tree swallow syllables. I used the same classification procedure as described below, except that, instead of performing cluster analyses within individual bird’s song output, I compared sounds within each day. I chose this approach for two reasons: first, I wanted to avoid any bias that might have occurred by removing sounds on an individual-by-individual basis, and second, by combining sounds from the same day I hoped to increase the chances of the cluster analysis detecting any background sounds that were particularly loud on that day. Aggregate landmark plots are very similar in appearance to spectrograms when they include a large enough number of samples. Visual inspection of these plots allowed me to quickly remove sounds that were clearly not made by tree swallows. Since the presence of a moderate number of heterospecific sounds radically affects the shape of the multivariate space used by the clustering algorithm, I repeated the analysis on each day’s sounds until there were no more types that were unambiguously made by some other species.
3.3.4 Classifying tree swallow syllables

Once my data set consisted only of tree swallow syllables, I used a series of hierarchical clusters to find local population’s set of syllable types. It was necessary to use a series of cluster analyses, rather than combining all of the population’s syllables into one analysis, because cluster analyses require time and computer memory proportional to the square of the number of samples in the analysis (Duda et al., 2000). There is also a biological advantage to starting within birds because each syllable type in a single bird’s repertoire must, by definition, occupy a unique location in multivariate acoustic space. If they occupied overlapping locations then ‘they’ would be a single, variable, syllable type. Syllable types from many different birds have no such limitation, and any overlap among several types from different birds will make it less likely for a hierarchical cluster to correctly place syllables into different types. The disadvantage of separately classifying each bird’s syllables is that there is no way to compare types among several different analyses. To overcome these problems, I took a subset of each bird’s classified syllables, and performed a hierarchical cluster on that subset. I then classified all of the syllables using the subset as a training set for a nearest-neighbor clustering algorithm.

I used the twenty-nine scaled variables to create hierarchical clusters which described the relative similarity among syllables in each bird’s repertoire. I only calculated a syllable cluster for a bird’s repertoire if I had recorded at least 30 of his syllables. I used Ward’s method of hierarchical clustering, which attempts to minimize the variance within groups relative to the variance among groups at each level of the hierarchy (Ward, 1963), implemented in the ‘agnes’ function from the R package ‘cluster’ (Maechler et al., 2005). Choosing the best number of groups from a hierarchical cluster analysis is not straightforward because the approach creates a dendrogram that links all its members together in a tree that could be cut into as many different groups as there are samples (Duda et al., 2000). In this analysis, I chose to use an approach based on the elbow method of choosing the number of groups that matches up with the angle of the L-shaped plot relating the number of groups to the amount of additional variance explained by the new group (Thorndike, 1953). To do this, I calculated the L-shaped plot, and then for each number of groups, split the L-shaped plot into the set of points above and below that number, and fit a pair of lines through
the groups. The curvature of the plot means that a pair of lines is a poor fit for the graph unless the number of groups is close to the elbow. Therefore, I chose the number of groups that created the split that produced the pair of lines that most closely approximated the L-shaped graph. This approach strikes a good balance between explaining the variance in the data without over-fitting by using too many groups.

A classification scheme describing the syllable types within each bird’s repertoire provides a foundation for classifying the entire population of syllables. The best syllables for the whole-population subset are the syllables closest to the acoustic centroid of each type within each bird’s repertoire. These syllables are packed closely together in acoustic space, reducing the chances that they will overlap with syllables from a type in another bird’s repertoire that is similar to, but not exactly the same as, the original bird’s syllable type. Ideally, groups of syllables from different birds will only overlap each other when they are truly from a shared syllable type. To this end, I chose the 9 syllables from each syllable type within each bird’s repertoire that were closest to the time-frequency-amplitude centroid of that type. I then made a hierarchical cluster with this subset as the data, and determined the number of groups as described above.

At this stage, only some of the syllables were classified according to population-wide criteria. In order to extend the classification from the subset to the entire population, I used a nearest-neighbor classification approach. Nearest-neighbor classifications are not as computationally intensive as hierarchical clustering, but they do require a training set of examples of known type (Duda et al., 2000). Creating this training set was the purpose of all of the previous hierarchical clustering. I classified the types using the ‘knn’ (k-nearest neighbors) routine from the built-in R package ‘class’ (Venables and Ripley, 2002). Each syllable from the entire data set was assigned a syllable type by comparing it to the twenty syllables from the training set that were closest to it in acoustic space. If at least ten of the twenty nearest neighbors in the training set were of the same type, then the syllable was assigned to that type.
3.3.5 Defining repertoires

While the most straightforward definition of tree swallow’s syllable repertoire would be the proportion of its songs made up by each syllable type, that approach is overly simplistic because the classification procedure described above produces a small but significant number of misclassifications. For example, if there is a misclassification rate of 5%, then my classification procedure would frequently find eight syllable types in a sample of 100 syllables from a bird has a true repertoire size of three syllable types. Closer inspection of that hypothetical data set would show that three of the types each contained about 30 syllables, and five of the types contained only one or two syllables. The five infrequent types are clear examples of misclassified syllables. A straightforward way to remove misclassified syllables is to calculate the expected number of syllables per type that each type would have if they were all sung at random. This can be estimated by dividing the number of syllables sampled from the bird by the original repertoire size that includes potentially spurious extra types. Types represented by fewer syllables than would be expected by chance are almost certainly misclassifications and can be removed. I used this approach to define the number of syllable types in a birds’ repertoire, as well as the frequency of each of those types in his repertoire. I calculated a bird’s repertoire over each week in the month of May, 2006. Specifically, recordings made between 29 April and 7 May were used for week one, week two corresponded to recordings made between 9 and 12 May, week 3 to 17 through 21 May, and week 4 to 24 through 31 May. A period of bad weather in which there was no singing separates each range of dates.

3.3.6 Syllable performance

The measure that I used to describe how well a bird was able to sing its syllables was the distance of each syllable from its type centroid. This distance is equivalent to the square root of the sum of squared residuals from a MANOVA relating the transformed landmark coordinates to syllable type. The greater the distance between a syllable’s landmarks and its type centroid, the more different is its shape from the type standard (Rohlf, 1999). The distance values can potentially range from zero to infinity, with a great deal of rightward skew. Log-transforming the distance values converts this
distribution to a close approximation of normality, so I log-transformed all of the syllable centroid distance values.

In order to compare differences in syllable type performance among individuals, I needed to remove the effects of variation in recording amplitude as well as differences in performance driven by differences in repertoire composition. Syllables recorded at higher amplitude have a greater signal-to-noise ratio, which means the acoustic landmarks approximating their shapes are less affected by random noise. As a result, there is a straightforward linear relationship between a syllable’s mean absolute amplitude and its distance from its type centroid (Figure 3.1). I also chose to remove any effects of different syllable types on syllable performance. Given their diversity in acoustic structure, it is to be expected that there will be differences in average centroid distance among different syllable types, as well as in how much that centroid distance changes with amplitude. This interaction between syllable type and amplitude is a potential source of serious noise, and so it must be removed along with the effect of amplitude alone. However, I did not wish to remove differences in mean centroid distance that existed among syllable types, since that is a potentially useful piece of information. My final measure of performance for each syllable was its residual from the linear model ‘distance = amplitude + type × amplitude - type + error’, multiplied by -1 so that greater syllable performance was denoted by positive residuals.

3.3.7 Statistical tests

Testing the change in time of either repertoire size or syllable performance involves a mixed-models approach analogous to repeated-measures ANOVA, in which bird is a random effect. I used the R package ‘nlme’ to perform these tests (Pinheiro et al., 2009). There were up to four fixed effects in each test: syllable type, experience, newness, and the interaction between newness and experience. Syllable type is used to remove differences in average syllable performance among the syllable types found in the classification described above. Experience is the number of weeks a bird has been singing that type, up to and including the week in which its syllable performance is being assessed. For example, in week 3, bird 85 sang syllable types 9 and 12. He first sang type 9 during week 2, and so he had two weeks of experience with it on week 3. It was the first week that he sang
type 12, so on week 3 he had only one week’s experience with that type. Finally, the bird itself was first recorded during week 1, so on week 3 he had three week’s experience singing at the site. Newness is a categorical variable that distinguishes between original syllable types (those that are found in a bird’s repertoire during the first week in which he is recorded, and added syllable types) which do not appear until later.

The test for change in repertoire size used a bird’s experience as the fixed effect and bird identity as the random effect. The test for change in syllable performance used syllable type, experience, newness, and the interaction between experience and newness as fixed effects, and bird as a random effect. In order to perform the second test without pseudoreplication, I found the mean performance for each syllable type within each possible combination of bird, syllable type, and week. I used this set of average syllable performance values in the test above.

3.4 Results

In total, I recorded 11782 syllables from 41 tree swallows. The number of syllables per bird per week ranged from 12 to 1167, with a median of 229. I recorded 736 syllables from 9 birds on week 1, 2326 syllables from 10 birds on week 2, 3927 syllables from 12 birds on week 3, and 4793 syllables from ten birds on week 4.

Twenty-two syllable types emerged from the clustering analyses. The number of syllables per type ranged from 2 to 2167, with a median of 308.5. Collectively, the tree swallows sang eleven different syllable types during week one, including one unique type that only appeared that week. They sang sixteen types during week two, including five that did not appear in week one, a unique type that appeared only during week two, and omitting three that had appeared in week one. They sang nineteen types during week three, including two novel types, two types that had only appeared during week one, but no unique types. In week four, the population repertoire size was also nineteen, with one new type and one type omitted. Over the course of the season, three syllable types were found in only one week, three in only two weeks, six types were found in three out of four weeks, and ten of the twenty-two syllable types were found in all four weeks.
Overall, the correspondence between number of birds, number of syllables, and number of syllable types recorded per week suggests that the data are a representative, but not perfect, sample of the vocal output of the population of birds at this site. Thus, it is not possible to chart the precise course of every bird’s repertoire evolution, but it is possible to infer what sorts of changes took place overall.

Many repertoires did change in size over the course of the season, but there was no trend to the direction of repertoire change. Out of 64 total comparisons in my sample between successive weekly repertoire sizes, an individual’s repertoire size increased 22 times, remained constant 21 times, and decreased 22 times. The test for changes in repertoire size found no overall effect of experience ($F_{1,63} = 0.408, \ p = 0.525$).

In contrast, syllable performance did improve during the season (Table 3.2). As expected, there were significant differences in syllable performance among different syllable types ($F_{21,236} = 2.94, \ p < 0.0001$). Tree swallows that had been singing their syllables for more weeks had higher syllable performance ($F_{1,236} = 24.9, \ p < 0.0001$). There was no difference in average syllable performance between syllable types that were originally in a bird’s repertoire and those that were added in later days ($F_{1,236} = 2.50, \ p = 0.115$), nor was there any detectable difference between original and added types in the rate at which performance improved ($F_{1,236} = 1.54, \ p = 0.216$). In other words, equivalent amounts of ongoing improvement took place both in syllables types that were added to a bird’s repertoire and in those that had been in a bird’s repertoire when he started singing at the site.

In many cases, changes in syllable phonology were subtle, but consistent, as illustrated in figure 3.2. This figure shows the complete output of syllable type nine, grouped by week, of six tree swallows. It consists of aggregate landmark plots, in which the landmarks from each syllable are plotted on top of one another. This illustrates the acoustic shape of a syllable in much the same way as an averaged spectrogram does (Clark et al., 1987). In addition, the sharpness of the image reflects how consistently a bird sang the syllable type because the landmarks of very similar syllables will fall on top of one another. It is clear from this figure that any modifications to syllable templates are subtle. Nevertheless, the consistency with which changes are reproduced is evidence that, despite
their subtlety, they are real. The case of bird 62 is particularly striking because he produces a very consistent syllable type on week 2, is more variable during week 3, and then on week 4 is once again consistent, but with different syllable phonology (Figure 3.3). This is precisely the kind of pattern one would expect if the male transitioned gradually between the two subtly different phonologies over the course of week 3.

3.5 Discussion

Adult tree swallows are able to modify their syllable repertoires through at least two different kinds of imitative vocal learning. First, they are able to add new syllable types to their repertoires by copying neighbors. Second they can modify the phonology of existing syllable types. Both of these forms of vocal learning can serve to increase the similarity between a bird’s own dawn song and that of his neighbors. However, the neighborhood of singers around a tree swallow nest box is dynamic and complex. Additional birds arrive throughout the breeding season, occupying boxes and bringing with them unshared syllable types. Nest failures can also displace birds, leading to lacunae in sharing at the site of the failed nest and a sudden decrease in sharing at the site of any attempts to renest. This has consequences both at the population level, where it was difficult to quantify any overall increase in sharing in 2006, and at the individual level, where a bird’s syllable repertoire and phonology may undergo continuous change. Despite the dynamic nature of their social environments, tree swallows using these imitative learning processes should converge towards a common, site-specific repertoire of syllable types.

The syllable plasticity shown by tree swallows seems like it might be the result of changes in the bird’s mental representation of what the syllable should sound like. A form of plasticity in syllable performance has been observed in urban house finches (Carpodacus mexicanus) that truncate the lowest-frequency parts of some of their syllables in noisy parts of Mexico City (Bermudez-Cuamatzin et al., 2009). It is less convincing as a demonstration of template plasticity, however, because the truncation was observed to correlate to immediate increases in environmental noise while individual males were singing. When the noise disappeared, the birds returned to singing the full frequency range of the syllable type. This suggests that it is not the song template that is modified, but
that parts of the motor pattern for producing the template are left off in noisy environments. The changes in tree swallow syllable involve more than just truncation – pitches, relative amplitudes, and patterns of frequency modulation can all change between days (Figure 3.2). This modification of the target of song learning, rather than in how the bird approximates that target with its vocal tract, is similar to the instances of call convergence described in the introduction. In contrast, song repertoire elements are generally thought of as immutable after a sufficient amount of sensorimotor practice (Marler, 1997).

It is possible that tree swallow dawn song syllables are plastic because contact call have been co-opted for a song-like purpose in this species. The calls of migrating tree swallows are quite similar to dawn song syllables. I have observed fledgling swallows, which were presumably siblings because they were banded and staying close to a recently successful nest, giving calls that resembled dawn song in both syllable structure and the cadence with which syllables were uttered. Furthermore, I have seen both tree swallows and white-rumped swallows (Tachycineta leucorrhoa) perform this behavior. It would make sense for both tree and white-rumped swallows to have a means for keeping together with a group of familiar birds while they migrated in their enormous wintering flocks. These flocks can number in the hundreds of thousands, at least in tree swallows, and a bird could substantially reduce the costs from potentially endless dominance interactions by sticking close to a group of known individuals with a stable dominance hierarchy. Call convergence would be an excellent way for tree swallows to keep together in such an association, much the same way as penguins do in their huge breeding colonies (Aubin and Jouventin, 2002). In a huge migrating flock, two birds with very different repertoires could simply ignore each other, which would help to maintain call similarity among associates. During the breeding season, in contrast, a male cannot ignore a rival, and so would be constantly engaged in dominance interactions that could provoke imitative vocal learning. Syllable type sharing in breeding tree swallows could thus be an hold-over from vocal learning rules for flocking calls that are driven by the demands of migration.

It seems equally likely that, in tree swallows, there is no clear distinction between what, acoustically, makes call and song elements. There are clear contextual clues, of course, which is the basis for treating dawn song as a phenomenon distinct from the vocalizations made during the constant
agonistic encounters that take place each day of the breeding season (Horn, 1996). If there is no distinction between call and song elements then that could be the result of selection in the two different contexts for similar learning processes. Selection for vocal convergence during flocking is easy to imagine, as I described above. Selection for vocal convergence during the breeding season is well-known in many species of songbirds, where it can serve many roles, including reducing the costs of fighting with known neighbors (Temeles, 1994), targeting specific individuals with shared repertoire elements (Molles and Vehrencamp, 1999), and signaling territory tenure (Beecher et al., 2000). Using the same learning rules for the different contexts could be advantageous to tree swallows because if it can somehow eliminate extra brain space that would be needed for additional vocal learning processes. It also suggests that there may be no fundamental distinction between calls and songs in oscines, at least in terms of learning. Rather, oscine vocalizations may lie within a range of degrees of plasticity, regardless of their social function. Some call types crystallize early in life and some song elements are subject to change throughout life. The fact that many calls seem innate, while songs are learned, could be the result of selection for different degrees of phenotypic plasticity, rather than differences in genetic versus environmental contributions, because of the effects learning has on the information encoded in a vocalization.

In tree swallows, it is still not clear what information in encoded in male dawn song. It may be that a bird’s precise repertoire is not significant, merely the fact that he is able to sing early on cold mornings in the spring, before many insects are available. The substantial variation in which day birds first began to sing makes this hypothesis a plausible one. Repertoire overlap seems less appealing, as can be seen in 2006 when the high density and rapid turnover of territories prevented the formation of any strong patterns of local repertoire sharing within the site. However, syllable consistency still seems like a plausible source of information for mate choice. It is known to be a factor in mate choice in many other species (Holbeck et al., 2008; Botero et al., 2009b; de Kort et al., 2009), presumably because of the challenge of minimizing random differences between utterances (Nowicki et al., 1998). Tree swallows seem able to rapidly learn new, or modify existing, syllable types, and so birds that spend the greatest proportion of their time singing with high consistency
could be the best learners and performers, even in the cacophony of constant repertoire change that seems to take place during a tree swallow breeding season.

Figure 3.1. The relationship between syllable amplitude and the distance of that syllable from the centroid of the type it belongs to. The greater signal-to-noise ratio of louder, closer syllables is reflected in their significantly lower centroid distances ($m = -0.0709, p < 2.2 \times 10^{-16}$).
Figure 3.2. Changes in the phonology of syllable type #9, as sung in different weeks by six male tree swallows. Each plot shows the landmarks of every syllable belonging to that type sung by a bird in a given week. A blurred appearance indicates variability, while sharper-appearing aggregates are the result of high syllable performance.

Figure 3.3. Spectrograms showing examples documenting change over time of a syllable type within a single bird’s repertoire. Day twelve is in week two, days seventeen and nineteen are in week three, and day twenty-seven is in week four.
Table 3.1. Expected changes in repertoire size and syllable performance according to three different hypotheses about post-dispersal song learning in tree swallows

<table>
<thead>
<tr>
<th>Repertoire size</th>
<th>Repertoire expansion</th>
<th>Syllable convergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>early → late</td>
<td>constant</td>
<td>increase</td>
</tr>
</tbody>
</table>

Syllable performance

| old vs new | constant | old > new | increase |
| early → late | constant |          |          |
| old vs new × early → late | constant | old : constant new : increase | no difference |

Table 3.2. Effects on syllable type performance consistency: syllable type, when it appeared in a bird’s repertoire, and how many week’s experience the bird has with the type.

<table>
<thead>
<tr>
<th>Value</th>
<th>Estimate + s.e.</th>
<th>Statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable type</td>
<td></td>
<td>$F_{21,236} = 2.94$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>Present first day</td>
<td></td>
<td>$F_{1,236} = 2.50$</td>
<td>0.115</td>
</tr>
<tr>
<td>yes</td>
<td>$-0.0291 \pm 0.0210$</td>
<td>$t_{236} = -1.90$</td>
<td>0.0585</td>
</tr>
<tr>
<td>no</td>
<td>$-0.0152 \pm 0.00967$</td>
<td>$t_{236} = -1.45$</td>
<td>0.149</td>
</tr>
<tr>
<td>Experience</td>
<td></td>
<td>$F_{1,236} = 24.9$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td>$F_{1,236} = 1.54$</td>
<td>0.216</td>
</tr>
<tr>
<td>slope, 1st-day types</td>
<td>$0.0185 \pm 0.00466$</td>
<td>$t_{236} = 3.98$</td>
<td>$&lt; 0.0001$</td>
</tr>
<tr>
<td>slope, added types</td>
<td>$0.0317 \pm 0.0106$</td>
<td>$t_{236} = 1.24$</td>
<td>0.216</td>
</tr>
</tbody>
</table>
CHAPTER 4
SONG PERFORMANCE, PROVISIONING RATES, EXTRA-, AND WITHIN-PAIR REPRODUCTIVE SUCCESS ARE ALL CORRELATED IN MALE TREE SWALLOWS

4.1 Abstract

Tree swallows (Tachycineta bicolor) experience intense competition for nest sites and high rates of extra-pair paternity. Sexual selection should therefore be intense in this species, with high-quality males under strong pressure to honestly advertise their desirability as mates. Song is a likely channel for that honest signals in monomorphic species such as tree swallows. If there is a tradeoff between mate-guarding and soliciting extra-pair copulations then tree swallow song should contain different messages depending on the strategy chosen by the singer. If both within- and extra-pair paternity are mediated through an ecological performance character, such as foraging ability, then males with high-quality song should have high rates of paternity in their own nest and among the nests of their neighbors. Comparisons between the social and genetic fathers of 29 extra-pair chicks from 17 nests found that the genetic fathers averaged significantly higher syllable consistency scores ($F_{1,28} = 5.35, p = 0.028$), food-delivery rates ($F_{1,16} = 12.4, p = 0.0029$), and extra-pair fertilizations ($F_{1,28} = 12.3, p = 0.0016$). There is no evidence for tradeoffs between reproductive strategies in tree swallows. Instead, the best male tree swallows excel at all aspects of reproductive performance: singing, provisioning, and extra-pair copulation.

4.2 Introduction

Mate choice in birds is a complex process fraught with risk for both sexes. When rearing offspring is too difficult for a single parent to do it successfully, both sexes are constrained to choose a single partner. This can be observed in many species of pelagic birds, including penguins (Moreno
et al., 2000). In contrast, when one sex can provide all of the care that offspring need, multiple mating can evolve (Goth, 2007), which can benefit individuals of the promiscuous sex in terms of both fecundity and offspring fitness (Jennions and Petrie, 2000). Complexity and risk lie in the ecological space that falls between these two extremes. The benefits of cooperative offspring rearing – as compared to individualistic mate seeking – vary across species, sites and even years. Choosing an unfaithful mate could be particularly detrimental in a year when parental care is costly, while settling for a suboptimal mate could be just as big a problem in other years. Display traits such as bird song can help ameliorate this risk, for both males and females, by giving the listening sex enough information to make an appropriate choice.

The songs of passerines are known to encode multiple messages, some for members of the opposite sex, and some for competitors of the same sex (Catchpole and Slater, 1995). Males may signal the amount of parental care they are capable of giving by performing energetically expensive songs (Buchanan and Catchpole, 2000). They may also produce songs that showcase the good genes that allowed them to develop the neuromuscular apparatus required for consistent singing (Nowicki et al., 1998). Ecological conditions will determine the relative importance of each of these messages.

Tree swallows (Tachycineta biclor) are secondary cavity nesters, lacking the ability to create their own nest holes. Consequently, they face a chronic shortage of nest sites and compete intensely among themselves for possession of the cavities that they do find. Males spend more time in territorial behavior, and are more vigorous when they are doing so, so a female generally cannot defend a nest site on her own (Rendell and Robertson, 1994). As a result, extra-pair copulation is high in tree swallows, and results in more than 50% of young being sired by extra-pair fathers (Morrill and Robertson, 1990; Conrad et al., 2001; Kempenaers et al., 2001). In addition to the high rate of cuckoldry, there is also a great deal of variation among males in the percentage of chicks sired within their own nest, which can range from 0-100%. This variance in male mating success could impose strong sexual selection on males.

The strategies that male tree swallows use to entice females to mate with them will be driven by what females need in order to maximize their own reproductive success. By recording male dawn song, and comparing it to measures of male parental effort and paternity, it should be possible to...
infer whether song is a trait that might inform female choice (Zann and Dunstan, 2008; Botero et al., 2009b). Females that are capable of raising their chicks on their own, without male help, may choose both their social mates, as well as the actual sires of their chicks, at random. If that is the case then there would be no relationship between song characteristics and any measures of male parental effort or mating success. Alternatively, females could choose mates in order to maximize the genetic quality of their offspring. In that case, there should be a correlation between song performance and male reproductive traits, but no correlation between male song and parental effort. Finally, females could depend on male efforts in order to successfully fledge a brood (Whittingham et al., 1994), in which case they might choose mates who will provide the most care. Singing performance would, in that case, be correlated with both male parental effort and reproductive success.

4.3 Methods

4.3.1 Measures of singing performance

The field work for this study was conducted in fields at Amherst College from April to July of 2006. See the Chapter 3 for information about the location of the site, as well as the methodology used to record, classify, and measure the consistency of dawn song. For the purposes of this study, it is necessary to emphasize that I began visiting the field in the last week of April 2006, before there were any tree swallows singing. As I visited the site on each day that there was no rain, and covered the entire range of boxes over pairs of successive days, the date on which I first recorded a bird singing from his box should be a close estimate of the day when he first began singing.

One elaboration on the previous chapter’s approach to measuring syllable consistency was to separately analyze temporal, frequency, and amplitude consistency by using only the corresponding subsets of the variables making up the acoustic landmarks. There is reason to think that each of these separate sub-measures could correspond to different performance constraints on syllable consistency. In addition, visual data exploration indicated that syllable bandwidth, in addition to syllable amplitude, is correlated with centroid distance, and so the consistency measures used in this study have had the effects of both syllable amplitude and bandwidth removed by linear regression.
4.3.2 Measures of reproductive performance

Nests were monitored according to the methods described in Ardia et al. (2009) for information about timing of breeding, adult physiology, fecundity, and nestling growth. In brief, nest boxes were checked on alternate days until the end of laying, daily after two weeks of incubation, and then every three days post hatching until the chicks were twelve days old. Foraging activity was monitored for 60 minutes periods on the eighth and tenth days post hatching, and foraging effort was calculated as the mean residuals from a linear model that included the effects of differences among observations in weather, insect availability, and time of day (Perez et al., 2008). Adult females were captured on the fourth day after clutch completion, while males were captured in wig-wag traps (http://golondrinas.cornell.edu/Data_and_Protocol/WigwagAdultTrap.html) on the tenth day post hatching. Blood was collected from males on that occasion, and from chicks and females on the eleventh day post-hatching, by piercing the brachial vein with a sterile needle and blotting onto unbleached coffee filter paper (Big Y Foods, Inc., Springfield, MA).

I genotyped adults and chicks using nine microsatellite markers from the set described by Makarewich et al. (2009). Using these genotypes and the software package CERVUS (Marshall et al., 1998; Kalinowski et al., 2007), I estimated the most likely pair of parents for each chick. Based on these assignments of parentage, I then calculated the number of within- and extra-pair young for each nest, the number of chicks outside of his nest known to be sired by each male, and the number of different males each female mated with.

4.3.3 Statistical analyses

I compared singing performance to reproductive success using two different perspectives: among males and among extra-pair chicks. The comparisons among males using a series of linear models in each of which a measure of male reproductive success was the response variable and a measure of male singing performance was the predictor variable. Measures of singing performance that I used included first singing date, overall syllable performance, and syllable performance looking at only time, frequency, or amplitude. Measures of male reproductive success that I used included clutch initiation date, feeding effort, total fecundity, within-pair paternity rate, the number of mates his
social mate took, and each male’s rank, among those sampled, in number of extra-pair young sired. The comparisons among extra-pair chicks used a series of mixed linear models which tested for a significant difference between performance traits of genetic sires and performance traits of social fathers. In these tests, the fixed effect was the type of father and the random effects were female and chick, nested within female. Traits that show significant differences between genetic and social fathers could either be cues that females use directly when choosing extra-pair mates, or correlated with such cues. I did not use the full suite of male performance traits described above as response variables in these analyses. Instead, I chose to use only those song traits that showed significant correlations with ecological traits in the among-male analyses. I also tested all of the ecological traits. All statistical tests were conducted in R (R Development Core Team, 2008), and the mixed linear models used the ‘lme’ function from the package ‘nlme’ (Pinheiro et al., 2009).

4.4 Results

Two different measures of male singing performance were significantly correlated with measures of reproductive success (Table 4.2). Male arrival time was correlated with the eventual clutch initiation date of his nest ($F_{1,39} = 7.87, p = 0.00779$, Figure 4.2), but it was not correlated with any of the other measures of reproductive success. In contrast, clutch initiation date is strongly correlated with reproductive success by female tree swallows (Winkler and Allen, 1996). The absence of a significant results in these data may be explained by a nonexistent correlation between brood size and clutch initiation ($F_{1,29} = 0.0371, p = 0.849$).

Extrapair paternity was common, with a great deal of variation among nests in the degree of cuckoldry. Allele frequencies and non-exclusion probabilities from the CERVUS analysis are presented for each locus in Table 4.1. The allele frequencies at the nine loci combined to give non-exclusion probabilities of $2.27 \times 10^{-3}$, $6.03 \times 10^{-5}$, and $5.00 \times 10^{-8}$ for maternal, paternal, and bi-parental identity assignments, respectively. The average paternity rate was 38.5%. More than a third $\frac{15}{38}$ of males sired none of the chicks in their nest, while five males sired all of their social offspring. Extra-pair success was also highly variable, as twenty-four males sired no extra-pair
offspring, while the most successful male sired 10 extra-pair young among the nests that I sampled. The average number of extra-pair young sired per male was 1.15.

Neither overall syllable performance, nor performance with respect to time or frequency measures alone were correlated with any measure of reproductive success. A male’s syllable performance in terms of amplitude was significantly correlated with the number of times his mate cuckolded him ($F_{1,26} = 6.59, p = 0.0164$, Figure 4.3). Males with greater syllable-amplitude consistency thus suffered less infidelity by their mates.

Three different comparisons between the performance of chicks’ social and genetic fathers showed significant differences (Table 4.3). On average, an extra-pair chick’s genetic father had greater syllable-amplitude consistency than its cuckolded social father did ($F_{1,28} = 5.35, p = 0.028$, Figure 4.4). The amount of difference, 0.0145 units, was within one standard error of the standard deviation of syllable-amplitude consistency found among males, 0.0160. By itself, this seems like a small effect, but the combination of this result with the correlation with cuckoldry suggests that the ability of a male to produce syllables with a consistent amplitude course is a factor in female mate choice in tree swallows.

Females also seemed able to identify males who, later in the season, would be good feeders at their own nests. The genetic fathers of extra-pair chicks had higher average feeding rates than did the cuckolded social fathers ($F_{1,16} = 12.4, p = 0.0029$, Figure 4.5). The degree of difference between the two categories of father, 2.83 units, was within one standard error to the amount of variation among fathers, 2.02 units. This result cannot be explained by lower feeding effort by males with low within-pair paternity because feeding effort and paternity rate were not correlated with one another ($p > .8$), which matches results from previous studies of feeding effort and paternity rate in tree swallows (Kempenaers et al., 1998).

Finally, male success in gaining extra-pair copulations did not lead to reduced within-pair paternity. The genetic fathers of extra-pair chicks had higher average ranks in terms of extra-pair young sired than did the cuckolded social fathers ($F_{1,28} = 12.3, p = 0.0016$, Figure 4.6). In addition, there was a significant positive correlation between a male’s within-pair rate of paternity and his extra-pair mating rank ($F_{1,24} = 5.16, p = 0.0324$, Figure 4.7).
Female tree swallows seem to prefer males who are the best at everything: singing, provisioning young, and wooing extra-pair mates.

4.5 Discussion

There seems to be no trade-off between direct or indirect benefits in tree swallows because, in this species, there are positive associations between male singing performance, feeding efforts, within-pair and extra-pair reproductive success. The tree swallows who are best at one set of behaviors are the best at all of them. Female tree swallows seem to choose the genetic fathers of their offspring accordingly, preferring as the fathers of their chicks males that have higher singing consistency and provisioning ability than that of their social mates. In addition, male tree swallows that have many extra-pair offspring also have a higher rate of within-pair paternity. This difference is probably driven by female preferences for high-quality males, rather than superior mate-guarding ability of high-quality males, because repeated copulation, rather than mate-guarding, seems to be the preferred strategy for paternity insurance in tree swallows (Beasley, 1996).

In contrast to its relationship with extra-pair reproductive success, male quality seems to play little role in how tree swallows form of social pairs. There is variation among males in how early in the season they can begin singing at dawn, which is most likely driven by the physiological challenge of singing in the dark, cold mornings of early spring. To illustrate, the number of syllables that I recorded from each male increased by an average of 2.9 syllables for every additional degree Celsius in the overnight low ($F_{1,17} = 10.1, p = 0.00546$, Figure 4.8). Despite this potential physiological link between timing of arrival and male quality, I found no relationship between male arrival time and any measure of female preference. I did find that the females who began their nests early in the season were more likely to nest with males who had begun singing early (Figure 4.2), but I found no evidence to suggest that females tree swallows do anything besides choose the first single male that is guarding a box when she arrives. Since female tree swallows need a nest cavity in order to reproduce (Whittingham and Dunn, 2001), a female who spends too much time looking for a nest box defended by a high-quality male could potentially miss an entire reproductive year if she is pre-empted by a series of less-picky females. In this nest-site limited species (Robertson et al., 2001).
1992), it makes sense for females to indiscriminately grab a place to put their eggs, and then worry later about who will fertilize them.

There could be physiological constraints on amplitude modulation that maintain amplitude consistency as an informative signal that females can use when choosing the fathers of their chicks. Tree swallow syllables are composed of several short notes that have rapid modulations in both frequency (see previous two chapters) and amplitude (Figure 4.1). With wide, short bills that can rapidly snap closed to catch insects, tree swallows’ performance limit on frequency bandwidth during trill production may actually lie too close to their perception limits for trill rate to be an informative character in this species (Ballentine et al., 2004). The five cycles of amplitude modulation in 50 ms seen in the exemplar tree swallow syllable, on the other hand, gives an AM rate of 100 Hz, close to the physiological limit of 125 Hz of AM that is driven by syringeal muscles, rather than acoustic interference (Suthers et al., 1999). In addition to being potentially difficult to perform, rapid changes in sound intensity make it easy to identify the location of a sound source (Marler, 1955). Giving a highly localizable call is clearly useful in the dense breeding grounds, or even denser migrating flocks, in which tree swallows live and vocalize. Despite, or perhaps originally driven by this clear function for amplitude modulation in non-breeding contexts, it is possible that there is a predation cost to giving highly localizable calls while flying around in the dark, in which case the humble tree swallow chirp could function in mate attraction like a peacock’s tail.

It is perhaps surprising that there are not more elaborate ornaments among tree swallow males, given the degree of disparity in this species between high-quality individuals, which excel at all aspects of life, and low-quality individuals, which are in the bottom percentiles in every measure of quality. Variation in the brightness male plumage does correlate with age in tree swallows (Bitton and Dawson, 2008), but this is a subtle amount of variation given the range of plumage elaboration found in swallows (Bitton and Dawson, 2008). One explanation for the relatively low phenotypic difference in male display traits could be the large impact of stochastic weather on the health and success of tree swallows. As aerial insectivores, tree swallows are highly susceptible to changes in their food supply that are driven by changes in the weather (Bowlin and Winkler, 2004). Precipitation, low air temperatures, and high wind speeds all dramatically reduce supplies of flying
insects while at the same time increasing metabolic costs (Slagsvold and Rohwer, 2000). Conversely, when the weather is good, food supplies are abundant and maintenance costs are low (Thomas and Shutler, 2001). Each cycle of weather, from good to bad and back to good again, could act like a cycle in an economy governed by laissez-faire fiscal policies: as the downturn progresses, those who were worst off to begin with suffer the most harm, and those who started out in the best shape will recover the most quickly as conditions improve. Thus, in tree swallows, each round of bad weather throughout the year should amplify any initial differences in quality among birds, until any costs incurred through tradeoffs between behaviors are negligible compared to the variation in condition among individuals.

**Table 4.1.** Non-exclusion probabilities for each of the nine separate microsatellite loci used for the paternity analysis. (N = 75)

<table>
<thead>
<tr>
<th>Locus</th>
<th>Alleles</th>
<th>Heterozygosity</th>
<th>PIC</th>
<th>Non-exclusion Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>obs.</td>
<td>exp.</td>
<td>mother</td>
</tr>
<tr>
<td>TaBi 1</td>
<td>11</td>
<td>0.693</td>
<td>0.735</td>
<td>0.699</td>
</tr>
<tr>
<td>TaBi 25</td>
<td>14</td>
<td>0.68</td>
<td>0.69</td>
<td>0.668</td>
</tr>
<tr>
<td>TaBi 34</td>
<td>21</td>
<td>0.893</td>
<td>0.883</td>
<td>0.867</td>
</tr>
<tr>
<td>TaBi 4</td>
<td>14</td>
<td>0.547</td>
<td>0.837</td>
<td>0.811</td>
</tr>
<tr>
<td>TaBi 8</td>
<td>10</td>
<td>0.76</td>
<td>0.827</td>
<td>0.798</td>
</tr>
<tr>
<td>Tal 6</td>
<td>11</td>
<td>0.72</td>
<td>0.742</td>
<td>0.709</td>
</tr>
<tr>
<td>Tal 8</td>
<td>23</td>
<td>0.867</td>
<td>0.917</td>
<td>0.904</td>
</tr>
<tr>
<td>Tle 16</td>
<td>12</td>
<td>0.787</td>
<td>0.78</td>
<td>0.748</td>
</tr>
<tr>
<td>Tle 19</td>
<td>16</td>
<td>0.773</td>
<td>0.867</td>
<td>0.847</td>
</tr>
</tbody>
</table>

all 14.67 0.747 0.809 0.784 2.27e-3 6.03e-5 5.00e-8 3.03e-12 1.06e-4

PIC = polymorphic information content
Figure 4.1. Differences in amplitude consistency between days within two syllable types from a single bird’s repertoire. In these aggregate landmark plots, time (relative to the beginning of the syllable) is on the abscissa, and amplitude (relative to the mean syllable amplitude) is on the ordinate. Less blur means more consistency because individual landmarks from very similar syllables are more likely to fall on top of one another. The consistency values from 20 May (more blurred) are 0.0416 and -0.00871 for 9 syllable types 9 and 12, respectively, and the values for 28 May (less blurred) are 0.0682 and 0.0689.
Figure 4.2. The date on which a male tree swallow first begins to sing is a significant predictor of when his mate will lay her first egg. Date is given as the number of days after 30 April, 2006.

Table 4.2. Results from linear models testing for correlations between measures of male singing traits and male reproductive traits in tree swallows.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Slope</th>
<th>st. err.</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clutch initiation date</td>
<td>0.255</td>
<td>0.0910</td>
<td>0.147</td>
<td>$F_{1,39} = 7.87$ 0.00779</td>
</tr>
<tr>
<td>First singing date</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Times cuckolded by mate</td>
<td>-17.1</td>
<td>6.66</td>
<td>0.172</td>
<td>$F_{1,26} = 6.59$ 0.0164</td>
</tr>
<tr>
<td>Syllable consistency, amplitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extra-pair fertilizations, paternity rank</td>
<td>0.0201</td>
<td>0.00886</td>
<td>0.1425</td>
<td>$F_{1,24} = 5.16$ 0.0324</td>
</tr>
</tbody>
</table>
Figure 4.3. The number of extra-pair mates taken by a female is negatively correlated with her social mate’s syllable consistency. Syllable consistency is the average, across all of a male’s syllable types, of how much individual syllables deviated in amplitude modulation from the type centroid.
Figure 4.4. The genetic sires of extra-pair chicks had greater syllable consistency when compared to the social fathers of those chicks. Syllable consistency is defined as in Figure 4.3.

Table 4.3. Results from mixed linear models testing for differences in performance traits between the actual sires of extra-pair chicks and their social fathers. Mother and chick nested within mother were random effects.

<table>
<thead>
<tr>
<th>Male trait</th>
<th>Social father - Genetic father</th>
<th>Within-nest</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable consistency, amplitude</td>
<td>-0.0145</td>
<td>0.00629</td>
<td>0.0160</td>
</tr>
<tr>
<td>Feeding effort</td>
<td>-2.83</td>
<td>0.805</td>
<td>2.02</td>
</tr>
<tr>
<td>Extra-pair fertilizations, rank</td>
<td>-6.81</td>
<td>1.95</td>
<td>2.90</td>
</tr>
</tbody>
</table>
**Figure 4.5.** Feeding effort at a male’s own nest was greater for the genetic sires of extra-pair chicks when compared to the social fathers of those chicks. Feeding effort is the calculated from the number of visits made during one-hour observations of nests when the chicks were ages 8 and 11 days old, and takes into account effects of variation in weather, time of day, and female effort.
Figure 4.6. The genetic sires of extra-pair (EP) chicks had greater rank in terms of EP offspring when compared to the social fathers of those chicks. Males with no EP offspring had a rank of one, and the male with the most EP offspring (10 chicks), had a rank of 41.
Figure 4.7. In tree swallows, there does not seem to be a tradeoff between gaining extra-pair paternity and maintaining within-pair paternity. Instead, males who succeed at one task also succeed at the other.
Figure 4.8. Tree swallow song output is affected by cold: males sing more syllables on warmer mornings.
CHAPTER 5
A COMPARATIVE ANALYSIS OF THE DISTRIBUTION OF ALLOSPECIFIC VOCAL MIMICRY IN OSCINE PASSERINES

5.1 Abstract

Song learning requires adaptations for the perception, selection, and production of species-appropriate sounds. Species that are vocal mimics have songs that incorporate sounds uttered by many other species. Vocal mimics thus have an unusually permissive definition of what a species-appropriate sound is. The breadth and accuracy of vocal imitations performed by individuals belonging to species such as the northern mockingbird (Mimus polyglottos) and superb lyrebird (Menura novahollandiae) has provoked hypotheses that vocal mimicry is an extreme form of display elaboration. If vocal mimicry is an extreme display then broader and more accurate vocal mimesis evolves via accumulation of vocal production mechanisms. An alternative hypothesis is that vocal mimicry is the byproduct of pre-existing permissiveness in which sounds are selected for learning. Examining the distribution of vocal mimicry across the oscine phylogeny can help distinguish between these two alternatives. In a survey of singing behavior in eighty-eight clades of songbirds, representing all oscines (except for laughing-thrushes and song-babblers 50 spp.), mimicry was found in twenty-eight clades scattered throughout the tree. Two out of the three clades within Menuroidea, the most ancient oscine superfamily, are excellent mimics. These observations suggest that mimicry is not the result of selection for elaborate vocal production, but is in fact a sign of pre-existing permissiveness in selectivity for song templates. Furthermore, the ancient occurrence of this permissiveness suggests that overproduction of a wide variety of sounds, rather than highly constrained species-typical sounds, may be the ancestral condition of oscine song learning.
5.2 Introduction

Mimicry of North American birdsong by introduced European Starlings (Sturnus vulgaris) provided crucial proof that wild passerines learn their songs (Townsend, 1924). Once the phenomenon of song learning was an accepted fact, speculation emerged about why mimics learn songs from other species. The ability to learn songs does not necessarily mean that a bird will learn to reproduce every sound it hears (Marler and Tamura, 1964). The norm among songbirds is cultural transmission of vocalizations within a species (Jenkins, 1978). For more than a century, ornithologists have proposed ways that sounding like another species might benefit a vocal mimic. These hypotheses have been produced through analogy with unlearned forms of mimicry among birds. Macronesian orioles (Oriolus sp.) with plumage and calls that resemble figbirds (Sphecotheres sp.) suffer fewer attacks from the larger, more aggressive species when the two frugivores forage together (Diamond, 1982). European cuckoo (Cuculus canorus) chicks give begging calls that sound like an entire brood of begging host chicks, and that resemblance spurs the host parents to greater feeding efforts (Davies et al., 1998). In these cases evolution, rather than learning, has produced mimicry that is functionally, if not consciously, referential.

The idea that vocal mimicry might be functionally referential is exciting because vocal mimicry is learned. Using vocal mimicry to refer to another species must therefore involve some kind of decision-making process on the part of the mimic. There are anecdotes of northern mockingbirds (Mimus polyglottos) suddenly starting to sing a particular species’ song when an individual of that species appears, and even countersinging between mockingbirds and other species (Townsend, 1924; Baylis, 1982). Dobkin (1979) took this idea and created an elaborate scheme of categories of mimicry based on different one-to-one ecological and social relationships between model and copier. Rather than appearing in Dobkin’s two-species contexts, the two known examples in which experiments have shown a definite effect of mimicry on other species both involve multi-species interactions. Greater racket-tailed drongos (Dicrurus paradiseus) mimic the contact calls of other members of mixed-species flocks (Goodale and Kotagama, 2006a), and playback of drongo songs that include mimicry attract more potential flock-mates than songs that lack mimicry (Goodale and Kotagama, 2006b). Phainopeplas (Phainopepla nitens) in distress will mimic both other species’
alarm calls and the calls of predators (Chu, 2001b). Playback of Phainopepla distress calls that included mimicry provoked more mobbing of a predator mount than distress calls without mimicry (Chu, 2001a). In both of these studies, the greater response to mimicry by other species was not driven by any particular species responding to its own calls, and the majority of responders in both cases were species whose vocalizations were not mimicked at all. Thus, even when it has an interspecies function, the effect of vocal mimicry does not seem to come from referencing particular species, but rather from producing a medley of sounds from many different species. The archetypal mimic reels off just such a medley from its song perch (Chisholm, 1946), apparently using mimicry in the same way that other birds use their non-mimetic song.

A medley of many different species songs seems to serve as a form of vocal elaboration, but no more so than in other species where large repertoires of consistently performed elements are favored. Songs that are complex in this way are physiologically difficult to perform (Gil and Gahr, 2002; Podos et al., 2004a). The handicap principle predicts that such signals will retain their information value over evolutionary time (Zahavi and Zahavi, 1997). In non-mimic species, repertoire size is larger in birds that are older (Cosens and Sealy, 1986) or more efficient at learning songs (Ward et al., 1998), a result that is also found in mimics (Eens et al., 1991a; Hasselquist et al., 1996; Coleman et al., 2007; Zann and Dunstan, 2008). The ability to repeat repertoire elements in a consistent way is also a signal of both age (de Kort et al., 2009; Botero et al., 2009b), and learning ability (Hughes et al., 2002; Holveck et al., 2008; Takahasi and Okanoya, 2010) in birds that learn their songs without mimicking, as well as in mimics (Coleman et al., 2007; Zann and Dunstan, 2008). Incorporating mimicked elements does not seem to add anything extra to the complexity of songs, whether one compares songs among species (Botero et al., 2009a) or among individuals (Hindmarsh, 1986). It is still not clear why some birds copy other species’ vocalizations when so many birds are able to broadcast the same messages while learning only their own species’ song.

One compelling hypothesis suggests that mimicry is the selectively neutral result of copying errors (Hindmarsh, 1986). These errors are particularly likely to occur in a species with large repertoires of disparate sounds (Kroodsma and Pickert, 1984). To the human ear, copies of other species’ vocalizations stand out, but to a mimic, they may be unremarkably within the acoustic zone.
of species-typical song elements. Mimicry, in this view, is just an epiphenomenon of acoustically catholic taste in song models. Two recent reviews came to exactly this conclusion when they failed to find any correlations between ecological conditions and the occurrence of mimicry (Garamszegi et al., 2007; Kelley et al., 2008). This hypothesis would be satisfactory if all mimics had repertoires that included just a few mimicked sounds taken from a limited acoustic subset of the sounds in their environments. This is certainly the case in European Starlings (Hindmarsh, 1984) and Lawrence’s Goldfinches (Carduelis lawrencii) (Remsen et al., 1982). However, extraordinary mimics such as the Superb Lyrebird (Menura novaehollandiae) have repertoires made up of more than seventy percent mimicry (Robinson, 1975), and the mimicked sounds come from a phylogenetically and acoustically broad range of other species’ vocalizations (Zann and Dunstan, 2008). Mimicry is probably not simply a result of copying errors in these species, but rather driven by selection for copying other species’ sounds.

One of the difficulties in trying to find a functional explanation for the evolution of vocal mimicry is that it begs the question of how difficult vocal mimicry is to evolve in the first place. Typically, when a theorist is speculating about the origins of vocal mimicry, he or she begins with the assumption that mimicry is really difficult to do. When one considers the singing of lyrebirds and mockingbirds, this assumption is understandable. Words like ‘virtuoso’ and ‘extraordinary’ are typical, and justified, adjectives used to describe the range and precision of vocal imitation by mimics such as these. Nevertheless, vocal mimicry in many other species seems less like virtuoso elaboration and more like the accumulation of inadvertant copies. The European starling is an excellent example of this. It can make very precise copies of vocalizations that are within its species-appropriate acoustic range, but its varied vocal repertoire is mostly composed of starling-typical notes, with only a relatively small proportion of its repertoire being made up of elements copied from other species. Vocal mimicry through copy errors seems like a very reasonable explanation when the putative mimic already has a large and diverse repertoire of song elements. In contrast to the very common and highly precise mimicry in lyrebirds, the relatively rare copying by starlings does not seem particularly difficult. Obviously, terms like “particularly difficult” are not helpful in
formulating testable hypotheses. A more useful way to frame the question of difficulty is in terms of mechanisms of vocal production.

Any explanation of mimicry must start with an explanation for why some birds are more permissive than others in their acoustic tastes. Even in the cases where mimicked sounds do have a function, those sounds still had to get into a mimic’s repertoire through a process of learning that allows copying with less regard than is usual for the characteristic properties of both the sound and the tutor. Understanding the evolution of permissivity in learning is not a matter of finding the fitness payoff of specific instances of copying. Rather, it requires an investigation of the learning mechanisms involved when a bird, confronted with the cacophony of sounds in its environment, chooses a subset of those sounds to include in its repertoire. In other words, instead of thinking about how generalized sound learning might somehow be adaptive, we can think about specific mechanisms of vocal learning that serve to restrict which sounds a bird can learn.

### 5.2.1 Mechanisms affecting the evolution of mimicry

The process of vocal learning is complex because it involves feedback at many stages, including memory, proprioreception, and, of course, perceiving one’s own song. Nevertheless, for the purposes of this paper it is sufficient to break vocal learning into three basic stages – perception, cognition, and production. Sounds can be eliminated from the set of potential models at each of these stages. If a bird cannot hear a sound then that sound will be eliminated at the perception stage. If a sound is not stored as a template for song learning than it is eliminated at the cognition stage. If a sound is stored as a template, but is physically impossible to perform, then it is eliminated at the production stage. In contrast to our lack of understanding about the ecological benefits of copying other species’ sounds, there is a large body of research about how sounds pass through each of these stages of the vocal learning process. By examining each of these stages in detail, we can better understand what mechanisms underly mimicry, the most permissive form of vocal learning.

Perception is the stage of vocal learning during which sounds are transduced from patterns of pressure to patterns of nerve impulses. As sound waves pass through the inner ear of a bird, they cause patterns of pressure that stimulate the hair cells that line the cochlea. The length of the
cochlea, number and density of hair cells, and the thickness of the membrane on which the cells rest all affect which sound frequencies a bird can detect (Bradbury and Vehrencamp, 1998). There seems to be little variation among songbirds in the ranges of frequencies that they can detect, although there is variation among species in which frequencies they are most sensitive to (Henry and Lucas, 2008). This makes sense given the general usefulness of sound for monitoring one’s whole environment – too much specialization for perceiving a narrow frequency range or set of acoustic features will leave a hole in one’s defenses that predators can exploit. The primary difference among species in perception is which sound features the hindbrain is best able to extract from sounds. Any set of neurons that serves as a feature detector must choose between high sensitivity to temporal or frequency features. A feature detector that is tuned to detect rapid changes in amplitude or pitch must work with short inputs so that the brief instances are homogenous with respect to the feature of interest. A feature detector capable of detecting very small differences in pitch, in contrast, requires longer samples of sound to get an accurate estimate of frequency. While there is some room for specialization in the cochlea for one or both of these, as is seen in bats with different echolocation styles (Carew, 2000), the auditory brainstem can take inputs from anatomically similar cochleas and create different representations of similar sounds by combining different kinds of feature detector. This is strikingly illustrated by seasonal plasticity in feature detection among a range of small woodland birds in North America (Lucas et al., 2007). The capacity for annual change in auditory sensitivity within a single bird, combined with homogeneity in pitch sensitivity across orders (Lucas et al., 2002; Pytte et al., 2004), suggests that it is unremarkable that vocal mimics can accurately perceive the vocalizations of other species. Constraints on vocal mimicry therefore are more likely to be found among the cognition or production stages of vocal learning.

I am using ‘cognition’ as a catch-all term for any neuronal process that takes place between when a bird perceives a song and when it starts to practice singing that song. The cognition stage is when a bird uses other sources of information to decide whether or not to use a sound as a template for sensorimotor learning. The sources of information can be genetically encoded preferences for sounds with specific acoustic properties, or that appear in species-appropriate sequences, or even biases toward copying certain tutors. If there is no filtering at the cognitive stage, then
the vocal output of a bird should be a random sample of the bird’s acoustic environment, perhaps weighted by the frequency with which model sounds occur (Hartshorne, 1973). Genetically encoded preferences for songs with specific acoustic features are actually poorly documented in the literature. Swamp sparrows (*Melospiza georgiana*) have historically been put forward as a model of this type of learning, both because of a species-universal set of note types (Marler and Pickert, 1984), and because swamp sparrows preferentially learn conspecific syllables from artificially created tutor tapes that combine swamp and song sparrow (*M. melodia*) syllables and syntax (Marler and Peters, 1977). It is true that swamp sparrows typically do not learn notes that are long, constant-frequency whistles, but more recent studies show that the range of variation among swamp sparrow notes is continuous, rather than discrete (Clark et al., 1987). Furthermore, swamp sparrows’ ability to pick species-appropriate syllables out of artificial songs with song sparrow syntax is evidence for endogenous preferences for note sequences, rather than notes with species-typical structure.

Preference for note sequences is a much more common and well-documented phenomenon among songbirds, having been found in zebra finches, *Taenopygia guttata* (Tchernichovski and Mitra, 2002); white-crowned sparrows, *Zonotrichia albicollis* (Soha and Marler, 2000); common nightingales, *Luscinia megarhynchos* (Hughes et al., 2002); and brown-headed cowbirds, *Molothrus ater* (Hauber et al., 2001). The case of the cowbirds is particularly interesting because, as a brood parasite, it faces an unusual challenge in finding appropriate song tutors. Indigobirds (*Vidua spp.*), which are also brood parasites, first copy their host father, who is from another species, and then copy the songs of other indigobirds that sing the same host song (Payne et al., 1998). Indigobirds are clearly integrating both visual and acoustic information when they choose song tutors, which is an important mechanism for maintaining species-specificity in song learning (Hultsch et al., 1999). Songbirds also use social information when choosing songs to learn and tutors to learn from (Williams, 2004). In song sparrows, this information can include the number of tutors that sang a particular song type, as well as the number of other juveniles that copied that song type (Nordby et al., 2000). While cognition may serve to emphasize a particular song or a particular tutor, for the most part the cognitive processes serve as filters, eliminating sounds as potential templates for vocal learning unless they appear in appropriate contexts. Even a preference for a certain conspecific
song or tutor can be explained as a preference for the closest match to a restrictive filter. What this means for mimicry is that highly specific learners must have more cognitive filters than vocal mimics do. From a cognition point of view, then, mimicry can be thought of as the simplest case, and preferences for specific sounds as extra filters added to a default behavior in which all perceived sounds are passed on to the production apparatus as potential templates.

In order to produce accurate copies of a template sound, a bird must learn to control and coordinate muscles throughout its body (Suthers and Zollinger, 2004). Muscles associated with respiration control air pressure in the thoracic air sac, which in turn affects the frequency at which the syrinx vibrates (Amador et al., 2008). Oscine passerines are also able to affect the vibration frequency of the syrinx with a set of associated muscles (Goller and Suthers, 1996). The trachea and esophagus above the syrinx can serve as a bandpass filter, removing unwanted elements from the harmonic series of tones produced by the vibrating syrinx (Beckers et al., 2003). Opening and closing the beak can affect the tension and length of the trachea (Daley and Goller, 2004), which is one way that a songbird can affect the filtering properties of its vocal tract (Riede et al., 2006). A bird must be able to change the state of its vocal tract rapidly and repeatably in order to produce even a moderately complex song (Westneat et al., 1993). Stereotyped motor patterns are associated with each different vocalization in a bird’s repertoire after a period of overproduction of and selection for motor patterns that produce sounds that match the template for that vocalization (Marler, 1991). Strikingly, the motor patterns associated with several different mimicked elements of northern mockingbird song match the patterns used by the model species to produce the same sound (Zollinger and Suthers, 2004). There is also a within-bird correspondence between motor pattern and acoustic similarity in both brown thrashers (Toxostoma rufum) and gray catbirds (Dumetella carolinensis), such that motor patterns that lead to similar sounds are more alike than motor patterns that produce acoustically different songs (Suthers et al., 1996). Mimics that produce a wider acoustic range of songs than non-mimics must therefore possess a wider range of vocal motor patterns than non-mimics do. In contrast to the cognition stage, elaborate vocal learning mechanisms at the production stage may be needed for vocal mimicry to evolve.
5.2.2 Hypotheses

There are two main contrasting hypotheses about the function of vocal mimicry. The first hypothesis is that mimicry evolves as an incidental byproduct of permissive song learning. From a learning rules perspective, permissive song learning should be easy to achieve, but from a motor point of view, producing diverse sounds is difficult. Thus, the byproduct hypothesis is more likely if song learning evolution is driven by changes in cognitive filters that affect which sounds a bird learns. The byproduct hypothesis predicts that mimicry should be common and widespread because it evolves through the loss of complex filters. The second hypothesis is that mimicry evolves under selection for an elaborate display of vocal complexity. For elaboration to be honest, and therefore evolutionarily stable, it must be costly. One possible link between elaboration and physiological cost is the mapping between vocal motor patterns and specific sounds, such that more elaborate songs require greater numbers of endogenous motor patterns (Suthers et al., 1996). The elaboration hypothesis therefore predicts that mimicry should be rare, and found mostly in a few derived clades that have accumulated many vocal motor patterns. The frequency and distribution of mimicry in oscines can thus provide insight into how song learning has evolved in birds.

5.3 Methods

5.3.1 Literature survey

In order to document the phylogenetic distribution of mimicry in oscines, I used narrow definitions of allospecific sounds and mimicry. I defined an allospecific sound as a sound that is clearly attributable to some acoustic source that is not a member of the putative mimic’s species. I only considered a species to exhibit mimicry if its members’ vocal output consistently includes allospecific sounds from multiple different sources. It is important to note that this definition excludes three specific cases of imitation that previous studies have labelled mimicry. First, while brood-parasitic indigobirds in the genus *Vidua* imitate the songs of their host fathers (Payne, 1973), I am not including that behavior in this study because they do not mimic the songs of any other species. This seems like a special, and intriguingly beneficial, case of misimprinting, which is another form of imitation that I am excluding. There are many reports in the literature of unusual individuals
from one species incorporating another species’ songs or calls into its unique repertoire (many exam-
amples as early as Townsend (1924)). If anything, these reports of misimprinting provide evidence
against the possibility of vocal mimicry in their respective species because these individuals stand
out from their species’ norm. Finally, I did not include species which have only been recorded to
mimic in non-song contexts, such as pine grosbeak subsong (Taylor, 1979) and phainopepla distress
calls (Chu, 2001b). While these cases provide useful information about the presence of some of the
classes of learning rules described above, it is impossible to infer syntactical rules from mimicry
given outside of the context of song.

Using these restrictions, I searched for reports of mimicry in three different kinds of publications.
I used descriptions of vocalizations in field guides of regional avifauna as a preliminary screen for
mimics. When they report mimicry, most field guides do not provide clear attribution of who
judged that the species was a mimic. For that reason, I did not use a field guide as the sole source
describing a species as a mimic unless it explicitly cited an observer, e.g. Kennedy et al. (2000). I
also used reviews, which included reviews of mimicry and regional compilations of species accounts.
These sources documented their descriptions of vocal behavior with multiple references. Finally, I
used primary literature sources to confirm mimicry as often as possible, although I was unable to
find some of the older and more regionally focused ornithological journals. I only counted a species
as a mimic if at least one source gave a list of allospecific models whose sounds appeared in an
individual bird’s songs.

5.3.2 Phylogenetic analyses

I constructed a supertree of oscines to look at the phylogenetic distribution of the species that
I had identified as mimics. I used the tree from Jønsson and Fjeldså (2006) as the backbone of my
tree, and then used more recent phylogenetic analyses to determine relationships within superfam-
gy groups. I used the topology in Gardner et al. (2010) for relationships within Meliphagoids. For
relationships within the core Corvoidea (Barker et al., 2004), I used trees from Norman et al.
(2009) and Dumbacher et al. (2008). I used several complementary studies of relationships within
the Muscicapoidea, which focused on bombycillids (Spellman et al., 2008), muscicapids (Zuccon and
Ericson, 2010a; Outlaw et al., 2010; Zuccon and Ericson, 2010b), and turdids (Pan et al., 2007). Finally, I combined several recent studies to form a hypothesis about the interrelationships among members of the Sylvioidea (Fregin et al., 2009; Nguembock et al., 2007; Jérôme Fuchs et al., 2006; Cibois, 2003). Passeroidea was comparatively well-sampled in Jønsson and Fjeldså (2006), so I used their topology without further modification.

Obviously, it is currently impossible to create a fully-resolved species-level phylogeny of oscines, even using the supertree approach. For this reason, I assigned each of the terminal clades in my supertree into one of three categories, based upon the number of mimics that occurred in that group. The first category contained all terminal clades in which I could find no evidence for mimicry. The second category contained all of the terminal clades in which there was only one species of mimic, or a single clade of closely related mimics. The final category contained all the clades in which there were many different mimicking species. I treated these three categories as strictly ordered character states (no mimicry < some mimicry < much mimicry), and estimated the history of the character in the oscine tree using parsimony reconstruction of ancestral states in Mesquite (Maddison and Maddison, 2009).

In addition to surveying the number of vocal mimics among oscines, I also wanted to quantify whether mimicry has repeatedly and independently evolved, or if it is a character of just a few closely-related groups. The byproduct hypothesis predicts that mimicry could have evolved many times, and the mimicry traits would therefore show high levels of homoplasy. In contrast, the elaboration hypothesis predicts that mimicry should have evolved few times, in closely related taxa, so that there will be low or no homoplasy in mimicry. Quantifying the degree of homology or homoplasy of a trait in a tree involves counting how many times the trait changes in excess of the minimum possible number of changes (Prum, 1997). This count can be expressed as the Consistency Index (CI), which is the ratio of the minimum number of state changes to the actual number of changes (Kluge and Farris, 1969). The CI is equal to one when there is no homoplasy in a character, and approaches zero as the number of homoplasies increases. I used repeated simulations of random trait evolution to compare mimicry’s observed CI in oscines to the distribution of CI’s expected from neutral evolution of a trait. Using the same tree topology in each run, I simulated
the evolution of the three mimicry states using a rate of 0.48 changes per million years, which was the rate of change in the data (38 state changes over 80 million years). The character states evolved according to a strictly ordered model, which meant that any evolution between the ‘no mimicry’ and ‘much mimicry’ states had to pass through the ‘some mimicry’ state. I used ten thousand iterations in each simulation. I performed all of the simulations in Mesquite 2.72 (Maddison and Maddison, 2009). If mimicry’s observed CI falls above the distribution of simulated values then it is more phylogenetically concentrated than would expected by chance. This result would support the elaboration hypothesis. An observed CI within or below the distribution of simulated values would support the prediction of the byproduct hypothesis that mimicry evolves in oscines essentially at random.

5.4 Results

Mimicry occurs throughout the oscine radiation. Fifteen of the eighty-eight terminal clades in this study have many mimics in them (Table 5.1), while thirteen of the terminal clades had a small number of mimics (Table 5.2). The clades with many mimics are not nested within larger groups that contain few mimics. On the contrary, three of the four most ancient clades of oscines – Menuridae, Atrichornithidae, and Ptilonorhynchidae – contain many mimics. The position of these groups on the oscine tree causes mimicry (either many mimics or few mimics) to be recovered as the most parsimonious estimate of the ancestral state of all oscines (Figure 5.1).

The consistency index of mimicry was 0.0645. This was well within the distribution of CI values from the simulation: 85.68 % of the runs had CI values lower than that of oscines (Figure 5.2). Thus, there is less parallelism and fewer character reversions in oscine mimicry than would be expected from completely random evolution.

5.4.1 Two categories of mimicking clades

Mimicry is found in each of the major subdivisions of oscines (Figure 5.1). Some of the most famously adroit mimics are found in the ancient lyrebird lineage (Robinson and Curtis, 1996). The spiny wrens, which are probably sister to lyrebirds (Jønsson and Fjeldså, 2006), are also accom-
plished mimics (Chisholm, 1932). The next most ancient songbird lineage contains the bowerbirds, which are mimics (Coleman et al., 2007; Frith and McGuire, 1996), and the logrunners, which are not mimics. The phylogenetic evidence supports the idea that the ancestor of oscines was likely to learn its songs according to permissive rules. This permissive learning, even if it was not itself mimicry, would have been conducive to the evolution of mimicry in its descendants. The remaining lineages of mimicking oscines are all nested within clades in which the common ancestor was most likely not a mimic.

5.4.1.1 Clades with many mimics

Several oscine clades contain many mimics, despite their probable descent from non-mimicking ancestors (Table 5.1). These families are characterized by both large numbers of mimicking species and by having members whose mimicry is striking and obvious (see references in (Table 5.1). The Australian warblers (Acanthizidae) are the lone group of mimics from the Meliphagoidea. Drongos (Dicruridae), crows, nutcrackers, and jays (Corvidae), and true shrikes (Laniidae) are all mimic-rich families from the core Corvoidea. Four families from the Muscicapoidea contain many mimics: the starlings (Sturnidae), mockingbirds (Mimidae), Old-World flycatchers plus chats (Muscicapidae), and thrushes (Turdidae). The muscicapoid mimics form a clade in which only the only family lacking mimics are the dippers (Cinclidae). Two families from the Sylvioidea, the larks (Alaudidae) and the warblers of Old-World wetlands (Acrocephalidae), have many mimics. Finally, one small clade in the Passeroidea, the leafbirds (Chloropseidae) are predominantly mimics. The wide range of ecological and social systems encompassed by the mimics in these unrelated groups is further support for the idea that there is no common environmental factor favoring mimicry in oscines (Hindmarsh, 1986).

These families are also disparate in their collective vocal behavior. Some of the groups, such as thrushes, mockingbirds, and larks, are beloved for the melodiousness of their songs, while others have many species that vocalize primarily through harsh calls, such as drongos, crows, and starlings. While many of these families have characteristic vocal flavors that aid in identification, each family has members whose vocalizations would be a better fit in a different group. For example, in some
species of *Mimus*, the vocal repertoire is limited to harsh contact calls, while others are excellent mimics with huge repertoires Botero et al. (2009a). The wide range of both within and among-species variation found throughout these clades suggests that the common ancestors of each of these families had permissive rules about what kinds of song elements they could learn.

5.4.1.2 Clades with few mimics

Another set of oscine clades each contain a small number of mimicking species among a much greater number of non-mimics. These species are listed individually in Table 5.2. Few of these mimics are famous for their imitative ability, but each of the referenced accounts contains forceful and detailed enumeration of imitated species (see references in Table 5.2). Nearly all of them are described in field guides as having unusual variability or variety among the elements of their songs. For example, the two *Carduelis* are reported to string allospecific calls together into long and complex songs that resemble the songs of their congener, the American goldfinch (*Carduelis tristis*). European starlings songs are similar in that they can feature some allospecific elements in a starling-typical syntax (Bohner and Todt, 1996). Like starlings, gray catbirds (*Dumetella carolinensis*) are members of a family with many mimics, and they are known to include mimicry in their songs (Suthers et al., 1996). However, catbirds can improvise functionally species-typical songs in the absence of conspecific tutors (Kroodsma et al., 1997). At least some of the species in this set of putative mimics may be, for the most part, improvisers, in which case the mimicked elements in their songs would be neutral incorporations (Hindmarsh, 1986).

Improvisation is particularly plausible in the case of the mistletoebird *Dicaeum hirundinaceum* and euphonias (*Euphonia sp.*). These species are specialists on mistletoe (Santalales: Misodendraceae, Lorranthaceae, and Santalaceae). They are seminomadic because of the patchy and unpredictable distribution of mistletoes (Ward and Paton, 2007; Carlo and Aukema, 2005). Nomadism is thought to promote improvisation in wrens (Kroodsma et al., 1999), and may do so in these species.

5.4.1.3 Distinct evolutionary patterns for each category

The two sets of mimicking clades seem to have different phylogenetic distributions. The taxa in Table 5.1 are listed in approximate order of age, and the table shows that superfamilies that
have diverged from each other more recently have fewer families with many mimics. The taxa in Table 5.2 are also ordered by approximate age, and there is no relationship between taxon age and number of daughter clades with mimics. To test if this observation was quantitatively true, I counted the number of descendants from each branch on the oscine tree present thirty million years ago, just after the three largest superfamilies diverged from one another. I calculated the proportion of lineages within each of these groups that fell into each type of mimicking clade. I used generalized least squares models with Brownian motion (Martins and Hansen, 1997) to test if there was a significant relationship between divergence time and the frequency of either much or some mimicry. I performed the analysis using the ‘APE’ package (Paradis et al., 2004) in R (R Development Core Team, 2008). Younger superfamilies do have a lower proportion of families with many mimics ($F_{1,11} = 10.67, p = 0.0075$), but they do not differ from older superfamilies in the numbers of daughter families with just a few mimics ($F_{1,11} = 2.65, p = 0.1321$).

The byproduct hypothesis can account for these two separate patterns. If reversions to simple cognitive filters occur at some low, random rate, then a species with the cognitive potential for mimicry should appear whenever it has lost all of its cognitive filters. This can account for the observed proportion of families with just a few mimics. Families in which many species are mimics therefore might include many species in which repertoire complexity is favored by sexual selection, which would maintain permissive learning rules with regards to song models.

5.5 Discussion

Vocal mimicry has arisen dozens of times in oscines (Tables 5.1, 5.2), which is evidence that it takes relatively little evolutionary change for a non-mimic to become a mimic. Mimicry is found throughout the oscine tree, rather than being limited to a few clades with particularly elaborate vocalizations (Figure 5.1). The apparent ease with which mimicry can evolve suggests that many, if not most, oscines have the ability to produce a wide acoustic range of sounds. In this view, mimicry is not limited by production constraints, but is rather enabled by lax rules for choosing which sounds to learn, and has the potential to appear in any lineage that has lost such rules. There are many possible ways to identify a sound as species-appropriate, and the usefulness of any
such mechanism depends upon the ecological and social contexts the song learner lives in (Beecher and Brenowitz, 2005). Consequently, there are many possible evolutionary paths to losing cognitive filters, copying other species’ vocalizations, and becoming a vocal mimic. This can explain why, in contrast to other complex traits such as seasonal migration or cooperative breeding, there are no clear ecological or social environments associated with the occurrence of vocal mimicry (Garamszegi et al., 2007; Kelley et al., 2008). Nevertheless, there are still cases where, once copying other species has evolved, a mimic can benefit from uttering those sounds (Atkinson, 1997; Chu, 2001b; Goodale and Kotagama, 2006b). It also suggests that the occurrence of mimicry in a taxon may be a signal that substantial evolution of song learning rules has taken place in its lineage.

As a songbird is processing potential targets for vocal learning, many different kinds of information can influence which sounds it will try to add to its repertoire. As described above, this information can include the acoustic properties of the sound itself, characteristics of the sound source, and the sequence of sounds that a particular vocalization occurs in. The reliability of these sources is determined by the environment a bird lives in. For example, the number of song types a male song sparrow shares with his neighbors indicates his fighting ability because a tough competitor can establish territories closer to his natal territory, where he learned his songs (Wilson et al., 2000). This pattern only holds true in sedentary subspecies, however. Migratory song sparrows have much higher dispersal distances and as a consequence almost never share any song types with their neighbors (Beecher et al., 2000). In the sedentary subspecies, juvenile song sparrows exposed to multiple song tutors preferred to learn song types that were shared among many tutors (Nordby et al., 2000). Given enough time and reproductive isolation, the migratory population should lose this preference because selection cannot maintain a bias toward learning shared song types when song type sharing is nonexistent. In general, for any kind of information channel used to rule out inappropriate tutors, there will always be some environment where that trait, channel does not distinguish between appropriate and inappropriate tutors. It is therefore straightforward to imagine a lineage producing a vocal mimic if its evolutionary history took place in a series of environments that successively removed all rules for choosing conspecific tutors, but the contingent nature of such an evolutionary course means that we can’t use this insight to predict whether a species will mimic.
The relationship between the ecological environment and the cognitive filters that eliminate potential templates is not as straightforward as in the case of tutor selection. Although some acoustic features of bird song can evolve to match the sound-transmission properties of a species’ habitat (Ey and Fischer, 2009), there is no intrinsic meaning to either the exact acoustic properties of song elements, or the sequence in which those elements appear (Bradbury and Vehrencamp, 1998). To use an analogy from human language, knowing that a word contains the letter “n” does not give any information about its meaning, and a sentence’s meaning is unrelated to the number of times the letter “n” occurs in it. Song elements have information value only in the context in which they appear, where they can serve as audible manifestations of a birds’ ability to develop and practice the complex neuromotor task of singing. A single, down-swept note doesn’t signify anything by itself, but when it is repeated with consistency as part of a trill (de Kort et al., 2009), or in a diverse repertoire (Botero et al., 2009b), it helps carry useful information. The two contexts that are most likely to shape the information value of a sung element are the immediate element sequence and the total repertoire that an element appears in.

The sequence of notes in a complex song carries so much information in itself that it limits the information value of individual element types, making mimicry less likely to appear in species with elaborate syntax. Any kind of syntax depends on element type categories because syntactical rules describe which element types can appear at any location within a sequence of elements (Leger, 2005). Both the singer and the listener must use the same rules to classify elements before they can agree on the syntax of a song, which can impose selection for a simple set of rules that can rapidly and unambiguously classify each song element into its appropriate type. Swamp sparrows are a perfect example of this. Given a series of notes that vary continuously in duration, a swamp sparrow will classify them into two note types according to a threshold duration value (Nelson and Marler, 1989). All swamp sparrow songs are composed with a set of just six of note types that are found across the entire species (Marler and Pickert, 1984). Possessing this simple, limited set of element types is probably what enables swamp sparrows to pick out species-appropriate syllables from artificially synthesized tutor songs that are mash-ups of syllables from different Melospiza songs (Marler and Peters, 1977). In principle, willingness to copy repertoire elements from other
species should not hinder a vocal mimic’s ability to produce such songs, because a small number of acoustic rules can just as easily be used to classify heterospecific as conspecific sounds. This appears to be the case in Lawrence’s goldfinch (Carduelis lawrencii) and European starlings, which both have songs with complex syntax that incorporate mimicked elements (Remsen et al., 1982; Eens et al., 1991b). However, even a small number of acoustic rules for organizing note types can limit what sounds a bird learns. For example, cross-fostered zebra finches fail to learn species-typical syntax and also subsequently show reduced ability to distinguish between songs of different species (Campbell and Hauber, 2009) The song elements of their Bengalese finch foster-fathers must have lacked the appropriate combination of features that would have allowed the zebra finches to organize those elements into a species-appropriate syntax. The ability, which is the same thing as the need, to learn species-typical syntax is widespread in oscines (Catchpole and Slater, 1995). In vocal mimics that must learn syntax, such as the starling (Chaiken and Bohner, 2007), mimicked elements will only appear during a note sequence in positions that are appropriate given that element’s acoustic properties (Hindmarsh, 1986). These limits on the acoustic range of elements that can occur in a song should become more strict as the complexity of the rules governing song syntax increases. Consequently, there should thus be a negative correlation between the syntax complexity of a species’ song and the amount of mimicry incorporated into that song. This can help explain why mimicry is comparatively rare among both wrens and nine-primaried oscines, groups in which there is a great deal of acoustic variability in repertoire elements, but relatively little mimicry.

Song element repertoire size should be positively correlated with mimicry because it is correlated with acoustic variability. The main factor influencing the evolution of a species’ typical repertoire size is the average individual’s time budget for song learning and evaluation. All other things being equal, the individual that invests the most time in song learning will have the largest repertoire size. Since nothing is ever equal in nature, however, repertoire size positively correlates with a bird’s age (Mountjoy and Lemon, 1995; Gil et al., 2001) and degree of brain development (Nowicki et al., 1998; Doutrelant et al., 2000). These correlations hold because age and brain development affect a bird’s time spent practicing and songs learned per unit time, respectively. In addition to their importance
during repertoire assembly, time constraints also affect a receiver’s ability to evaluate repertoire size because it requires more listening time to accurately estimate the size of a large repertoire than a small one (Fessl and Hoi, 2000). Consequently, repertoire size should be strongly influenced by the costs of evaluating a repertoire size, which will be influenced by factors including the length of the breeding season, the audible range of the song, and the population density of a species. For example, at intermediate densities, a receiver can simultaneously evaluate the repertoire sizes of several singers, without being overwhelmed by destructive interference among them.

Senders should also be under selection to make it as easy as possible for a receiver to accurately estimate the sender’s repertoire size. A receiver that hears repertoire elements from acoustically distinct types will need fewer repetitions of each type in order to assess two important aspects of song performance, repertoire size (Cosens and Sealy, 1986; Eens et al., 1991a; Hasselquist et al., 1996; Ward et al., 1998; Coleman et al., 2007; Zann and Dunstan, 2008), and element type consistency (Hughes et al., 2002; Coleman et al., 2007; Holveck et al., 2008; Zann and Dunstan, 2008; Botero et al., 2009b; de Kort et al., 2009; Takahasi and Okanoya, 2010). Both of these features are used as cues for mate choice in species with large repertoires, presumably because of the physiologically challenging nature of these tasks (Gil and Gahr, 2002; Podos et al., 2004a). Greater acoustic separation between element types will thus benefit both the receiver and the sender, by making mate choice more efficient for both parties. Thus, in species with large repertoires, there should be directional selection for larger, more multidimensional acoustic spaces that can accommodate more distinct element types. According to this analysis, strong selection to produce many distinct repertoire elements causes an expansion of the cognitive criteria that individuals use to identify species-appropriate song element models. The species with the most extreme examples of this expansion of criteria are then more likely to be vocal mimics. This is a synthesis of the ‘copying error’ hypothesis, which states that there is no distinct role for mimicked elements (Hindmarsh, 1984; Garamszegi et al., 2007; Kelley et al., 2008), and the elaboration hypothesis, which states that including mimicked elements makes repertoires seem larger and more complex (Witchell, 1896; Townsend, 1924; Chisholm, 1937; Robinson, 1975; Baylis, 1982). It seems like an adequate account for mimicry in species like European starlings, in which mimicked elements form a small proportion
of the repertoire, but it still leaves unexplained the high proportion of stunningly accurate mimicry found in species like superb lyrebirds and northern mockingbirds.

Once an individual begins to incorporate mimicked elements in its repertoire, the singer may be at an advantage because his mimicked notes can be assessed by the receiver without the sender having to repeat them. Female satin bowerbirds prefer both large repertoires and consistent element performance, and they seem to lend extra weight to the mimicked part of a males song when evaluating those features (Coleman et al., 2007). A bird that has a large repertoire of song elements is faced with a tradeoff between displaying the size of his repertoire or the consistency of his elements. Demonstrating consistency requires repetition, and each time a bird repeats an element he sacrifices an opportunity to display a new element from his large repertoire. If a bird were instead to sing without any repetition at all, it would show the largest possible sample of its repertoire but make it impossible to tell whether he was repeating crystallized elements or producing sounds at random. Mimids, even non-mimicking mimids, ameliorate this problem by repeating notes a small number of times in rapid succession, which facilitates assessing consistency (Botero et al., 2009b). Repertoire size is also clearly important for mockingbirds, since it is during courtship that they maximize the number of element types they sing per unit time (Derrickson, 1988). However, a mockingbird that repeats each note three times can, at best, demonstrate only one third of the repertoire size as a bird that sings without repetition. Mimicry makes it possible to sing without repetition and still allow the listener to assess consistency. In effect, the mimic is using the repetition provided by the species he has copied when the models are singing themselves. A listener that has already learned an element that the mimic sings can instantly estimate the singer’s consistency by comparing its own memory of the sound to the mimic’s element. This is a potentially huge advantage, but it only works when the listener can correctly identify the common template sound. A listener from a different acoustic environment might not be able to assess a mimic’s singing consistency because it wouldn’t have memorized the mimicked sounds. Even worse, from the perspective of both the mimic and the receiver, the mimic and listener could each have memorized a different variation on the mimicked sound. The mismatch between the two versions would make the mimic appear unfairly inconsistent to the listener. Therefore, a prediction of the broken trade-off hypothesis
is that mimics will prefer to mimic sounds that are easy to hear and unlikely to vary. Another prediction is that species with higher dispersal will have higher repetition rates because they are more likely to encounter conspecifics from different acoustic environments.

Mimicry should therefore be most common in clades that tend to have large repertoires sung with simple syntax, and least common in clades that tend to have complex syntax with a small repertoires. The results of my survey of mimicry in oscines supports this hypothesis, although a rigorous estimate of syntax complexity across all oscines is beyond the scope of this paper. Furthermore, three of the most ancient lineages of oscines, lyrebirds, scrub-birds, and bowerbirds, are accomplished mimics that sing their repertoire elements with simple syntax. The most parsimonious interpretation of this observation is that the common ancestor of oscines was a vocal mimic. This breaks down into three hypotheses about the ancestral state of song learning mechanisms in oscines. First, it is likely that this ancient bird was capable of perceiving, memorizing, and producing a broad acoustic range of sounds. Second, it is likely that this bird sang its elements with a simple sequence, with little or no association between successive element types. Third, it is likely that the bird had few cognitive filters on which tutors it copied, instead learning sounds from a variety of sources. This combination of features resembles the subsong of a wide variety of taxa, from Australian warblers (Gilbert, 1937) to pine grosbeaks (Taylor, 1979), and so makes a pleasingly ontogeny-recapitulating hypothesis. It also inverts the traditionally assumed ancestral states of two aspects of song learning. First, a restricted set of simple song elements may be a derived condition resulting from selection for complex repertoire organization. Second, severe restrictions on appropriate tutors for song learning may also be a derived condition resulting from selection for copying tutors who are particularly likely to sing informative songs. In both cases, according to the view granted by thinking about mimicry, the ancestral condition was permissive learning, and the derived conditions are complex combinations of rules that impose restrictions on what a bird can learn.
Figure 5.1. The frequency of mimicry in different oscine clades, using a phylogeny based upon Jønsson and Fjeldså (2006). Branch lengths are in millions of years. Clades with more mimics are indicated with darker colors. Ancestral states were estimated using parsimony on characters ordered by increasing amounts of mimicry. There are many clades with only a few mimics, and the clades with many mimics are not nested within clades with few mimics. This suggests that mimicry is a plesiomorphic character, rather than a derived one.
Figure 5.2. The observed amount of homoplasy in vocal mimicry among oscines compared to the distribution of homoplasy in 10000 simulations of neutral stepwise character evolution. A low consistency index (CI) signifies more homoplasy (Kluge and Farris, 1969). Homoplasy of mimicry in oscines was within the range of CI values generated by simulations of stepwise evolution in which an intermediate state of low levels of mimicry came between the no-mimicry state and the high-mimicry state. This result suggests that mimicry is not simply a randomly evolving epiphenomenon of copy errors (Hindmarsh, 1986).
Figure 5.3. Among oscine lineages that had diverged by 30 Mya, the younger groups have a lower proportion of daughter families with many mimics. The proportion of families with only a few mimics has not changed over oscine evolution. The horizontal axis shows the dates of interior nodes from Figure 5.1 that are older than 30 Mya. The vertical axis shows the phylogenetically independent contrasts (Felsenstein, 1985) of the proportion of each category of mimicking family. The trend lines are locally weighted smooths with kernel lengths of 0.75. The proportions of families with many mimics are shown with circles and solid lines, while the proportion of families with few mimics are shown with x’s and dashed lines.
### Table 5.1. Oscine families that contain many mimics

<table>
<thead>
<tr>
<th>Superfamily</th>
<th>Family</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Atrichornithidae</td>
<td>(Higgins et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>Menuridae</td>
<td>(Robinson and Curtis, 1996)</td>
</tr>
<tr>
<td></td>
<td>Ptilonorhynchidae</td>
<td>(Frith and McGuire, 1996)</td>
</tr>
<tr>
<td>Meliphagoidea</td>
<td>Acanthizidae</td>
<td>(Higgins and Peter, 2002)</td>
</tr>
<tr>
<td>Corvoidea</td>
<td>Artamidae + Cracticida</td>
<td>(Higgins et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>Corvidae</td>
<td>(Cramp, 1980)</td>
</tr>
<tr>
<td></td>
<td>Dicuridae</td>
<td>(Rasmussen and Anderton, 2005; Roberts et al., 1985)</td>
</tr>
<tr>
<td></td>
<td>Laniidae</td>
<td>(Cramp, 1980; Roberts et al., 1985)</td>
</tr>
<tr>
<td>Muscicapoidea</td>
<td>Mimidae</td>
<td>(Allard, 1939; Ridgely and Tudor, 1989)</td>
</tr>
<tr>
<td></td>
<td>Muscicapidae</td>
<td>(Cramp, 1980)</td>
</tr>
<tr>
<td></td>
<td>Sturnidae</td>
<td>(Allard, 1939; Hindmarsh, 1984)</td>
</tr>
<tr>
<td></td>
<td>Turdidae</td>
<td>(Rasmussen and Anderton, 2005; Ridgely and Tudor, 1989)</td>
</tr>
<tr>
<td>Sylvoidea</td>
<td>Acrocephalida</td>
<td>(Cramp, 1980)</td>
</tr>
<tr>
<td></td>
<td>Alaudida</td>
<td>(Cramp, 1980)</td>
</tr>
<tr>
<td></td>
<td>Chloropseida</td>
<td>(MacKinnon et al., 2000)</td>
</tr>
</tbody>
</table>

### Table 5.2. Taxa from families in which there are few or only one species of vocal mimic

<table>
<thead>
<tr>
<th>Superfamily</th>
<th>Family</th>
<th>Genus</th>
<th>Species</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corvoidea</td>
<td>Oriolidae</td>
<td><em>Sphecotheres</em></td>
<td><em>viridis</em></td>
<td>(Higgins et al., 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Oriolus</em></td>
<td><em>sagittatus</em></td>
<td>(Higgins et al., 2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>xanthornus</em></td>
<td>(MacKinnon et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>Pachycephalidae</td>
<td><em>Falcunculus</em></td>
<td><em>frontatus</em></td>
<td>(Higgins and Peter, 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Colluricincla</em></td>
<td><em>megarhyncha</em></td>
<td>(Higgins and Peter, 2002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Pachycephala</em></td>
<td><em>nudigula</em></td>
<td>(Coates et al., 1997)</td>
</tr>
<tr>
<td>Vireonida</td>
<td></td>
<td><em>Vireo</em></td>
<td><em>griseus</em></td>
<td>(Adkisson and Conner, 1978)</td>
</tr>
<tr>
<td>Sylvoidea</td>
<td>Sylviiinae</td>
<td><em>Parisoma</em></td>
<td><em>subaeaculeum</em></td>
<td>(Roberts et al., 1985)</td>
</tr>
<tr>
<td></td>
<td>Zosteropinae</td>
<td><em>Zosterops</em></td>
<td><em>lateralis</em></td>
<td>(Higgins et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>Timaliinae</td>
<td><em>Ptiloichla</em></td>
<td><em>mindanensis</em></td>
<td>(Kennedy et al., 2000)</td>
</tr>
<tr>
<td></td>
<td>Polioptilidae</td>
<td><em>Polioptila</em></td>
<td><em>carae</em></td>
<td>(McNair, 1985)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>damica</em></td>
<td>(Narosky and Yzurieta, 2003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>lembeyei</em></td>
<td>(Garrido et al., 2000)</td>
</tr>
<tr>
<td>Passeroidea</td>
<td>Dicaeidae</td>
<td><em>Dicaeum</em></td>
<td><em>hirundinaceum</em></td>
<td>(Higgins et al., 2006)</td>
</tr>
<tr>
<td></td>
<td>Fringillidae</td>
<td><em>Euphonia</em></td>
<td><em>violacea</em></td>
<td>(Snow, 1974)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>lanirostris</em></td>
<td>(Remsen, 1976; Morton, 1976)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>psaltria</em></td>
<td>(Remsen et al., 1982)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td><em>lawrenci</em></td>
<td>(Remsen et al., 1982)</td>
</tr>
<tr>
<td></td>
<td>Emberizidae</td>
<td><em>Serinus</em></td>
<td><em>leucopterus</em></td>
<td>(Roberts et al., 1985)</td>
</tr>
<tr>
<td></td>
<td>Icteridae</td>
<td><em>Pezopetes</em></td>
<td><em>capitalis</em></td>
<td>(Skutch, 1967)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Icterus</em></td>
<td><em>cayancensis</em></td>
<td>(Fraga, 1987)</td>
</tr>
<tr>
<td></td>
<td>Parulidae</td>
<td><em>Icteria</em></td>
<td><em>virens</em></td>
<td>(Ritchison, 1988)</td>
</tr>
</tbody>
</table>


Cicero, C., Benowitz-Fredericks, Z. M., 2000. Song types and variation in insular populations of Lincoln’s sparrow (Melospiza lincolnii), and comparisons with other Melospiza. Auk 117 (1), 52–64.


