

THE FUNCTIONAL SIGNIFICANCE OF MICROGEOGRAPHIC DIALECTS IN A HERMIT
HUMMINGBIRD

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Vinay A. Kapoor

February 2016

© 2016 Vinay A Kapoor

ALL RIGHTS RESERVED

THE FUNCTIONAL SIGNIFICANCE OF MICROGEOGRAPHIC DIALECTS IN A HERMIT HUMMINGBIRD

Vinay A. Kapoor, Ph.D.

Cornell University 2016

In this dissertation I examine the functional significance of microgeographic dialects in a lek-mating hummingbird, in four parts. First, I develop a novel song analysis tool, based upon techniques first developed for human speech recognition, to enable rapid, quantitative, unbiased, and repeatable measurement and description of vocal dialects. Then, using this tool, I analyze the songs of lekking little hermits (*Phaethornis longuemareus*) to quantify dialectal variation in this species. Second, I examine the potential role of male-male competition in structuring the dialects described in the previous chapter. Specifically, I use an experimental playback procedure to test predictions of the hypothesis that males recruiting to leks may use vocal mimicry to deceive territory holders about their residency status (i.e. the deceptive mimicry hypothesis). Males responding to playbacks 1) showed reduced aggression to neighbors' songs than to strangers (i.e. they showed a "dear enemy" response), 2) did not appear to recognize individuals based on their songs, and 3) responded less quickly, but not less aggressively to unfamiliar "mimic" songs. Together, these results support the deceptive mimicry hypothesis, and provide evidence that dialects on little hermit leks may form through the widespread use of song mimicry. Third, I test three hypotheses explaining the role of female-choice in the formation of song dialects on hermit leks: females use dialects 1) to facilitate comparisons between males, 2) to assess dialect-level phenotypes, and 3) to reveal male dominance patterns. I also test a fourth hypothesis that females use song to re-locate preferred males (i.e. "hotshots") and lower quality males deceptively mimic the song of these preferred males to deceive mate faithful females into mating with the mimic. I

found no support for the first three female choice hypotheses, but tentative support for the fourth. Females mated with different males within the same dialect at a rate greater than expected by chance, suggesting the possibility that females use song to re-locate males, and song sharing allows mimics the opportunity to intercept these females. Fourth, I describe a novel marking technique essential for the study of individual behavior in little hermits and other small hummingbirds.

BIOGRAPHICAL SKETCH

Vinay (Julian) Andrew Kapoor was a strontium-based failsafe device first devised by extremely small elves (2.8 ± 0.7 g, $n=56$) native to the northwestern isthmus of Protania on the poorly known island nation of Van Sveldsbørg in 1820. The inauguration of construction was scheduled for April 12th, 1820, however, due to extreme wartime shortages in the supply of construction materials, most notably calcium carbonate (known colloquially as “chalk”), in the winter of 1819-1820, the project was delayed significantly. The exact date upon which construction actually began is not known, but some historians have speculated that the date is likely to coincide with the holiday known as “Hhhgq,” held in late October, as vast quantities of chalk are consumed by the Denzier peoples of greater Snoeksborg, who, during this period, were known to trade heavily with the Protanian elves when trade routes along the tundra were sufficiently clear of snow and friendly relations permitted it. While nearing completion in the summer of 1823, several design flaws were discovered that led to structural instabilities across Kapoor’s seventh distal turnbuckle. Kapoor was subsequently condemned according to the ordinances of Protania at the time, due to safety concerns, until renovations could be completed. Kapoor languished, unfinished, for nearly two centuries. Although incomplete and badly rusted at the time, Kapoor was able to attend the University of California at Berkeley, where he completed his Bachelor of Arts degree in Integrative Biology and Psychology in 2004. Kapoor was subsequently shipped to Seewiesen, Germany, where he served alternately as a night-stand and writing desk for a period of 17 days. Subsequently, for a period of one year he was allowed to serve as a visiting researcher at the Max Planck Institute for Ornithology, where he studied acoustic communication in the Panamanian lance-tailed manakin, *Chiroxiphia lanceolata*. Tragically, the humidity of the Panamanian rainforest was the final stroke that caused Kapoor’s

seventh distal turnbuckle to give out finally, leading to the near destruction of several (previously very shapely) boxwood shrubberies. Were it not for the availability of Nigel Swartmann, a highly esteemed and skilled arborist taking cuttings of a Snowy Mespilus (*Amelanchier lamarkii*) in the town of Andechs nearby, the boxwoods would likely have perished or, at best, been permanently misshapen. The German government promptly replaced the damaged part, but, uncertain of the manufacturing quality of the remaining turnbuckles, decided to ship Kapoor to Cornell University in the early fall of 2006 in exchange for a bolt of unbleached muslin. There, with renewed vigor, Kapoor began to study a 2.7 X 1.1 m Formica table with plated brass legs as well as geographic variation in the vocalizations of the bearded bellbird, *Procnias averano*, in Trinidad and Venezuela. This work did not get very far, as this would be the time in Kapoor's life when he was in the late stages of telophase II of meiosis; the loss of paired sister chromatids was simply too traumatic for him to continue. Luckily, three of the resulting daughter cells fused to form a triploid mass of endosperm that would serve as nourishment for the fourth. By the time Kapoor's two cotyledons had absorbed most of the surrounding endosperm, he was well on his way to studying the functional significance of dialects in a tropical hummingbird native to Trinidad, the little hermit, *Phaethornis longuemareus*. Little did Kapoor know at the time, but the pages that follow in the thesis you hold now in your hands would detail yet another stage in Kapoor's already varied existence. Hermit crab.

ACKNOWLEDGEMENTS

Many thanks to Charlie Walcott, Paul Sherman, Bruce Land, Mike Webster, and Irby Lovette. Their advice and support through all stages of my dissertation made this work possible. Thanks to all of the behavior graduate students of NBB, especially Julie Miller, Jessie Barker, Caitlin Stern, Becky Cramer, Biz Turnell, Kevin “Lopez” Loope, Gil Menda, Josh LaPergola, Maria Modanu, Kristin Hook, Jay Falk, Geoff Broadhead, and all the participants of our research “jam sessions” for their friendship and invaluable support and advice. The many hours they spent with me brainstorming were a major source of the ideas that make up this body of work. Thanks also to the Fitzjames family: Carl, Kelly, Kahlil, and Kamala; they made (and continue to make) Trinidad a home for me, and have helped to keep me grounded in the most challenging of times in the field. Thanks to my ten field assistants: Valerie Morley, Stephanie Wheeler, Kelly Schaeffer, Lindsey Nietmann, Stephanie Coates, Rachel Neugarten, Emma Fraser, Phebe Quan, Katie Chenard, and April Williamson; without their dedication, tenacity, intelligence, companionship, and support none of this work would have been possible. My deepest gratitude to Julie Miller; my source of strength in all things. None of this work would have been possible without the cooperation of the little hermits; many thanks to them for letting me tell their story.

Many thanks also to my sources of funding: the SUNY Fellowship, ABS Grant, NBB Animal Behavior Grant, Sigma Xi Grant in Aid of Research, Cornell Chapter of Sigma Xi student research grant, Cornell Einaudi Center Travel Grant, Cornell Research Travel Grant, SICB Grant in Aid of Research, NSF GRFP and DDIG, Frank Chapman Memorial Grant, Paul Graduate Fellowship, Explorers’ Club Exploration Fund, NOC Travel Award, AOU Research Award, Cornell Lab of Ornithology Athena Fund, Walcott Lab funding, and Cornell Lab of Ornithology semester support.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH	iii
ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	x
CHAPTER	
1. GENERAL INTRODUCTION.....	1
1.1 The study of song dialects	1
1.2 Lek mating systems	2
1.3 <i>Phaethornis</i> hummingbirds.....	3
1.3.1 General biology.....	4
1.3.2 Natural history	4
1.4 Outline of the dissertation.....	5
2. MEASURING DIALECTS: AN ASSESSMENT OF NOVEL METHODS FOR THE AUTOMATED COMPARISON AND CLASSIFICATION OF CLUSTERED VOCAL VARIATION	6
2.1 Abstract.....	6
2.2 Introduction.....	6
2.3 Materials and methods	11
2.3.1 Field methods	11
2.3.2 Automated song characterization and comparison	13
2.3.3 Automated analysis of dialect variation.....	25
2.3.4 Validation of automated comparison and classification by human assessment	26
2.4 Results.....	28
2.4.1 Algorithm performance.....	28
2.4.2. Dialect structure	31
2.4.3 Comparison with human visual assessment.....	34
2.5 Discussion	36
2.5.1 Algorithm performance.....	36
2.5.2 Comparison to visual techniques	37
2.5.3 Comparison to visual approaches of other studies of hermit hummingbirds.....	43
2.5.4 Advantages of automated analysis.....	44
2.6 Conclusions.....	48

3. DECEPTIVE VOCAL MIMICRY IN A LEKKING HUMMINGBIRD	49
3.1 Abstract	49
3.2 Introduction	50
3.3 Materials and methods	55
3.3.1 General methods	55
3.3.2 Playback experiment	57
3.3.3 Statistical analysis	64
3.4 Results	69
3.4.1 Binary responses	69
3.4.2 Duration responses	76
3.4.3 Latency responses	76
3.4.4 Song similarity and dispersion	77
3.5 Discussion	79
3.5.1 Movement versus song responses	82
3.5.2 Effects of song dispersion / cohesion on detection of intruders	83
3.5.3 Deception by intruders or a password of group membership?	84
3.5.4 Could deceptive mimicry be common on leks?	85
3.6 Conclusions	89
4. COURTSHIP DIALECTS AND THEIR RELATIONSHIP WITH MALE REPRODUCTIVE SUCCESS IN A LEKKING HUMMINGBIRD	91
4.1 Abstract	91
4.2 Introduction	92
4.3 Materials and methods	98
4.3.1 Field methods	98
4.3.2 Male physiology	101
4.3.3 Spatial analysis	102
4.3.4 Song analysis	103
4.3.5 Reproductive success	108
4.3.6 Statistical analyses	113
4.4 Results	117
4.4.1 Paternity	117
4.4.2 Relationship between male quality and measures of song performance	117
4.4.3 Quality and song versus individual RS	119
4.4.4 Relative song performance and spatial centrality versus relative reproductive success (H1 and H2)	121
4.4.5 Dialect-level performance versus per capita reproductive success (H3)	121
4.4.6 Female mate fidelity, and male deceptive song mimicry (H4)	122
4.5 Discussion	123
4.5.1 The use of dialectal song variation in mate assessment	123
4.5.2 Dialects, mate fidelity, and the hotshot hypothesis	128

4.6 Conclusions.....	129
5. IMPROVED METHODS FOR COLOR-MARKING HUMMINGBIRDS.....	130
5.1 Abstract.....	130
5.2 Introduction.....	130
5.3 Methods	131
5.3.1 Back tag	132
5.3.2 Leg tag	134
5.4 Results.....	136
5.4.1 Back tag	137
5.4.2 Leg tag	138
5.5 Discussion.....	138
5.5.1 Back tag	138
5.5.2 Leg tag	139
BIBLIOGRAPHY	142

LIST OF FIGURES

Figure	Page
2.1. Spectrographic representation, measurement, and comparisons of hermit songs	17
2.2. Four TFR spectrograms of song elements used in the study	30
2.3. Hierarchical tree for the songs of all individuals on lek 5 in 2013, based on agglomerative clustering analysis of DTW results.....	33
2.4. Dialect map of lek 5 in 2013.....	42
3.1. An example schematic overview of an experimental playback setup, and the territorial locations of the five song stimuli presented to the focal bird.....	59
3.2. Spectrograms of a single song from an example focal bird and from males chosen to generate the five different stimuli	60
3.3. The relationship between stimulus type and the numeric score of pairwise similarity and dialect membership	63
3.4. Binary and continuous movement responses for both the categorical treatment models and the continuous song dissimilarity models	73
3.5. Kaplan-Meier cumulative failure plot for the latency of focal males to begin moving in response to the initiation of stimulus playback	77
3.6. Neighbor song dispersion negatively predicts the duration of movement in response to playback	79
4.1. Example spectrograms and waveforms of songs exemplifying high and low measures of individual song variability, and percent peak performance used to calculate amplitude performance.....	105
4.2. The calculation of amplitude performance for each song element	107
4.3. Relationship between male age and song variability.....	119
4.4. Relationship between tail extension and individual mating success	120
4.5. Expected and observed proportion of repeated matings with any male in the same dialect and with only different males in the same dialect	123
5.1. Design of the plastic back and leg tags	133
5.2. Correctly positioned back tag and leg tag on a male little hermit	136

LIST OF TABLES

Table	Page
2.1. Descriptive statistics for dialects on individual leks and averages	31
2.2. Song analysis similarity validation with visually determined scores	34
2.3. Song analysis clustering validation with visually-determined scores	35
3.1. Component models used to generate final model-averaged estimates in Table 3.2.....	71
3.2. Standardized model averaged predictor coefficients for all six movement response models	72
3.3. Component models used to generate final model-averaged estimates in Table 3.4.....	74
3.4. Standardized model averaged predictor coefficients for all six song response models	75
4.1. Descriptive statistics for microsatellite loci used in the paternity analysis	111
4.2. Two song performance measures in relation to metrics of individual male physiology, spatial location, and activity levels	118
4.3. The relationship between male physiology, spatial location, and activity patterns on individual reproductive success	120
4.4. Male mating success relative to that of other dialect members and its relationship with three relative measures of song performance and one measure of dialect spatial centrality	121
4.5. Relationship between per capita reproductive success among dialects and dialect-level song measures, spatial features, and activity patterns	122
5.1. Perler beads used in my study of little hermits and whether bead color faded over a 2-year period	137

CHAPTER 1

GENERAL INTRODUCTION

1.1 The study of song dialects

The study of shared signals between groups of individuals (dialects) has helped ecologists to understand the social and genetic relationships that characterize those groups as well as the selective forces that maintain group boundaries (Baker 1982). Investigations of dialects in many oscine birds have revealed interesting commonalities across species: dialects are often implicated in territorial and dominance interactions (Baptista 1985). However, recent studies of less typical species have revealed that song dialects may also reflect the spread of particular songs that are especially effective in attracting mates (Trainer 1989; Kroodsma 2005; Fitzsimmons et al. 2008), and may even reflect non-sexual pressures (Balsby and Bradbury 2009). Only recently has interest in non-passerine vocal learners increased sufficiently to begin to shed light on dialectal variation in these poorly known groups.

Although the study of vocal dialects has had a large impact on our understanding of the selective forces at work within populations, almost all attention has focused on a few species of oscine passerines, most of which are temperate, and socially monogamous (Bretagnolle 1996). Relatively little work has been conducted on species with other mating systems, from non-temperate climes, and especially those from non-passerine families. Although part of this trend is undoubtedly related to the fact that fewer non-oscine birds have dialects, the few cases where dialects do occur provide us with a unique opportunity to test the generality of current models.

The study of dialects in *Phaethornis* hummingbirds will help advance our knowledge in this field in several ways. First, hummingbirds represent a third independent evolution of vocal learning in *Aves* (oscine passerines and parrots are the other two (Gahr 2000)), and can therefore provide us with a phylogenetically independent contrast (Felsenstein 1985) to test the generality of theories about the evolution of dialects. Second, hermit hummingbirds are tropical species and will thus fill a major gap in our understanding of the evolution of dialects in biomes other than temperate North America and Europe. Third, lek and resource defense polygyny mating systems of hummingbirds allow us to investigate how selective pressures affect sexual signals in non-monogamous birds.

1.2 Lek mating systems

In addition to expanding our understanding of the evolution and maintenance of song dialects, my dissertation work offers novel insights into the forces of selection acting within leks. Despite the fact that lekking behavior is rare among vertebrates (Höglund and Alatalo 1995), lek mating systems have been a popular focus of behavioral ecologists wishing to study the effects of intense sexual selection and reproductive skew on the evolution of male mating behavior and female choice. This interest has led to a great deal of theory about the key question: *why join a male aggregation when the chances of mating are so low?* Most studies which attempt to address this question do so through investigating the direct effects of individual male display behaviors on mating success (Bradbury et al. 1985; Wiley 1991). It has become increasingly apparent however, that a male's fitness on a lek also depends strongly on his interactions with his immediate environment (Uy and Endler 2004) and with his neighbors (Ryder et al. 2008; McDonald 2009).

The most recent work on sexual selection in lekking species has sought to incorporate this realization by expanding our knowledge of 1) how males interact with each other and the environment to increase their individual fitness and 2) which factors influence variation in these interactions. The most prominent examples of this include studies of association with kin on leks and lek-like aggregations (Petrie et al. 1999; Shorey et al. 2000; Krakauer 2005), cooperative male alliances (McDonald and Potts 1994; DuVal 2007), deceptive mimicry and parasitism (Trainer 1989; Cockburn et al. 2009), and frequency dependent mating strategies (Lank and Smith 1987; Sinervo and Lively 1996). Interestingly, few of these studies have sought to identify how these fitness-changing factors influence the organization of individuals on the lek itself. Addressing this gap in our knowledge will be vital if we wish to advance past current models of the adaptive significance of lekking behavior.

The *Phaethornis* genus represents an unparalleled opportunity to study the selective pressures that lead to individual organization on leks because, unlike many other lekking species, *Phaethornis* hummingbird leks show an overt pattern of sub-lek structure – microgeographic dialects. The discrete breaks between dialects within *Phaethornis* leks strongly imply that social or genetic groups have arisen through selection (Stiles and Wolf 1979; Mundinger 1982; González and Ornelas 2009), though little work has been conducted to investigate how selection on hermit leks could lead to such variation. These selective pressures may be common but unrecognized due to the lack of the vocal patterning in other species (most of which are non-learners) that acts as a marker of such pressures in hermits. The research proposed here will therefore provide a much needed investigation into selective pressures that may have implications for the evolution of lek mating in general.

1.3 *Phaethornis* hummingbirds

1.3.1 General biology

The hermits (Phaethornithinae) are a distinct group of neotropical hummingbirds that are characterized mainly by their relatively drab plumage (McGuire et al. 2009). Many hermits produce intricate vocalizations, in some cases rivaling that of songbirds in terms of vocal output and song complexity. Hermits, though poorly studied, have attracted a great deal of interest because of the large percentage of species which have obligate or facultative lek mating systems.

1.3.2 Natural history

Phaethornis hummingbirds breed during the tropical dry season in leks that range in size from fewer than ten males to well over fifty (Kapoor, unpublished data). Individuals are scattered at varying densities throughout the lek and display on low twigs from small territories. Males occupy these territories during the day for the duration of the breeding season and spend many hours singing to attract potential mates. Both males and females visit these display territories and are almost always greeted with a visual display consisting of the owner hovering in front of the perched bird. Once females have mated they leave the lek and raise their offspring without further assistance.

Perhaps the most intriguing aspect of the breeding behavior of hermits is their advertisement songs. The vocalizations vary considerably between species and range from a simple monosyllabic whine (*P. guy*) to a highly variable, multi-element and almost musical “twinkle” (*P. longuemareus*). Although only a few species have been studied in any detail the emerging pattern is that the songs of hermits vary markedly within leks. The best example of this is a pioneering study by Snow (1968) of the little hermit (*P. longuemareus*) in which he demonstrated that individuals on a single lek in Trinidad sang songs that fell into multiple discrete song dialects. Subsequent work in a handful of other species have identified that discrete

dialects are common in *Phaethornis* (Wiley 1971; Stiles and Wolf 1979; Ramjohn et al. 2003). Little theory has been developed that offers an explanation for how dialects can arise in such a small area as a lek and even less is known about possible selective forces on the lek that might maintain such variation.

1.4 Outline of the dissertation

My dissertation work focuses on the functional significance of vocal dialects in a *Phaethornis* hummingbird, the little hermit, and is comprised of four main sections. First, in Chapter 2, I develop and describe novel techniques for the analysis and description of dialectal variation, and test these techniques using songs recorded from a population of little hermits (three leks) in Trinidad. In Chapters 3 and 4, I test predictions of several hypotheses explaining the adaptive significance of the microgeographic dialects described in Chapter 2. Chapters 3 and 4 are separated into an investigation of the role of male-male competition and the role of female choice in the formation of dialects, respectively. Finally, in Chapter 5, I present a novel marking method that was used to mark individuals in my population.

CHAPTER 2

MEASURING DIALECTS: AN ASSESSMENT OF NOVEL METHODS FOR THE AUTOMATED COMPARISON AND CLASSIFICATION OF CLUSTERED VOCAL VARIATION

2.1 Abstract

The study of vocal dialects has been a popular focus for ecologists interested in the role of sexual and natural selection in both genetic and phenotypic divergence between groups of individuals. Unfortunately, there are few established analysis techniques for describing and studying dialects that are simultaneously precise, unbiased, accurate, and rapid. Here, I present novel song analysis methods based on dynamic time warping and agglomerative hierarchical clustering techniques that combine all of these features. I test these techniques on the vocalizations of *P. longuemareus* from three leks in Trinidad. The results of both element-wise and full-song comparisons, as well as dialect assignments generated by the algorithms described in this paper closely matched that of both naïve and experienced individuals asked to compare songs based on established visual scoring techniques. In addition, the results of the analysis of *P. longuemareus* vocal variation provide unbiased evidence of the existence of vocal dialects in this species.

2.2 Introduction

The study of geographic variation in signals among groups of animals has helped ecologists to understand the social and genetic relationships that characterize groups as well as

the selective forces that maintain group boundaries (Baker 1982; Baker and Cunningham 1985; Podos and Warren 2007; Wilkins et al. 2013). The phenomenon of discontinuous variation in vocal signals, i.e. vocal dialects, has received a great deal of attention because steep clines in sexually selected vocalizations may inhibit gene flow directly or reveal the presence of ecological selection inhibiting gene flow, and may lead to eventual reproductive isolation and speciation (Irwin et al. 2001; Slabbekoorn and Smith 2002a; Edwards et al. 2005; Price 2008). Although the majority of early work investigating the functional significance of dialects has concentrated on their role in the reinforcement of genetic boundaries (Baptista 1985; Macdougall-shackleton and Macdougall-shackleton 2001; reviewed in Podos and Warren 2007), other studies have revealed that song dialects may reflect the spread of songs that are especially effective in attracting mates (Payne 1985; Trainer 1989; Kroodsma 2005) or repelling rivals (Feekees 1977; Briefer et al. 2008a), and may even reflect non-sexual pressures such as social group cohesion (Balsby and Bradbury 2009).

Despite the decades of interest in the functional significance of vocal dialects, relatively little attention has been paid to the important tasks of defining and quantitatively measuring vocal variation to determine whether it varies continuously or forms discrete clusters (Lachlan et al. 2010). A recent review on the topic of dialects defines them as vocal geographic variation characterized by “sharp transitions in vocal parameters between localities, and consistency in vocal parameters within localities (Podos and Warren 2007).” What constitutes a “sharp transition” and in which vocal parameters with what weighting, remain open questions. In the absence of a framework for defining dialects and robust tools for measuring clusters in acoustic structure the classification of songs as “dialectal” will continue to be subjective and potentially misleading in animal behavior studies.

The existing methods employed by researchers to define and describe acoustic variation are varied, and differ in their advantages and pitfalls. Among the most common of these techniques involves the use of visual comparisons of spectrograms conducted by naïve and/or experienced researchers (e.g. Thorpe 1958; Trainer 1989; Nelson et al. 2004). This technique has the advantage that the human visual system is extremely good at pattern recognition and categorization, even when signals tend to be moderately noisy, but classification by human observers also suffers from a number of setbacks (Tchernichovski et al. 2000; Ranjard and Ross 2008; Lachlan et al. 2010). Inter-observer consistency can be difficult to maximize, especially for borderline cases when song variation is not entirely discrete (Janik 1999; Jones et al. 2001). This can arise because subjects may tend to use different criteria in making comparisons, which also has the effect of reducing objectivity. Also, features salient to the human visual system may not represent the relevant acoustic perceptual characteristics of the animal being studied. To minimize these effects, individuals are often primed with instructions designed to standardize criteria, however, which criteria are used, how they are weighted, and whether they are specifically selected to match animal perception are factors that are often unreported, are subject to inter-observer variability, and could potentially result in the loss of objectivity (Tchernichovski et al. 2000). Added to this is the large time investment in the process of visual scoring, which places constraints on the number of vocalizations that might be compared, imposing a potentially extreme handicap on large-scale studies.

Alternatives to visual scoring have sought to minimize these pitfalls, with varying success. One commonly employed technique is to parameterize vocal information by identifying features in time-frequency representations of sound that can be measured in all signals of interest, such as minimum/maximum frequency, duration, number of frequency inflections, etc.

(Payne and Budde 1979; Leader et al. 2000; Wright et al. 2008). Multivariate and clustering statistics can then be used to identify natural clusters in the dataset (e.g. Boughman 1997). One of the primary difficulties with this approach is in the selection of adequate numbers of acoustic landmarks to generate overall similarity scores with sufficient sensitivity (Cramer 2013a). The selection of these features becomes particularly difficult when signals are not homologous (i.e. they do not share similar shapes; Osiejuk et al. 2005). Furthermore, multivariate analysis of these features may not capture variation in their temporal patterning.

The technique of spectral cross correlation (SPCC), rather than focusing on a reduced set of parametric measurements, uses the entire spectrographic representation of a vocal signal, allowing all of the spectral structure to contribute to a holistic pair-wise similarity score that can then be used to cluster songs into groups (Clark et al. 1987). Unfortunately, this technique is typically useful only when comparing signals that vary little in time; small changes in temporal patterning, and shifts in frequency or its modulation rate that reduce the pixel-by-pixel overlap between signals that are otherwise nearly identical can result in extremely low similarity scores (Tchernichovski et al. 2000; Meliza et al. 2013). SPCC is also often plagued by high sensitivity to noise, since shared noise profiles can overwhelm similarity measures between signals. Both of these issues are problematic given the ability for animals to perceive similarities between sounds that vary markedly in their temporal structure, and to separate relevant sounds from noise (Dooling 1982; Dooling et al. 2000; Lohr et al. 2003).

Progress in dealing with these difficulties has been made in the development of two bioacoustic techniques. The first is in the vectorization of signals into separate time-varying acoustic features so that the effects of noise can be reduced or eliminated, and the weightings and choice of vector representations can be specifically tailored to match an hypothesis about how

those features may be biologically important (Tchernichovski et al. 2000; Ranjard and Ross 2008). For instance, several studies have revealed that tonal sounds are perceived by a pitch that corresponds to the fundamental frequency (F0; Cynx and Shapiro 1986; Shofner 2005), and that the shape of a tonal sound (i.e. relative pitch; Hurly et al. 1990; Osiejuk et al. 2005) and its temporal structure and patterning (Briefer et al. 2013) are important in identification of sounds as similar to another. These factors (and many others) can be separately measured, weighted, and included in the calculation of similarity in a way that can be decomposed to identify salient features. The second advance is in the use of sequence alignment algorithms such as dynamic time warping (DTW), which allow temporal distortions of time series data so that they can be aligned and compared even if their temporal structure is markedly different (Itakura 1975; Sankoff and Kruskal 1983). This capability clearly has analogs to animal auditory perception, enabling more biologically relevant measures of similarity between signals. The combination of these techniques has potential to provide a consistent, rapidly calculated, repeatable measure of similarity, with explicitly stated parameters, that can be used with classic clustering methods both to evaluate the presence of and to describe dialects, yet remains a seldom utilized method for their analysis.

Here, I present novel bioacoustic methods based on vectorization, DTW, and hierarchical clustering to compare and classify tonal acoustic signals, and use these methods to analyze the songs of little hermits (*Phaethornis longuemareus*), a species with putative dialects. Using the results of this analysis I present a procedure for statistically evaluating the existence of song dialects. I then compare the results of the automated techniques to traditional visually based classification methods, and to previous work on *P. longuemareus* dialects. Finally, I discuss the advantages of the methods I present in this paper over existing techniques for analyzing and

describing vocal dialects, and provide suggestions for the adoption of an analytical framework for the study of dialects.

2.3 Materials and methods

2.3.1 Field methods

2.3.1.1 Study site and study duration

The study was conducted on three leks of *P. longuemareus* in a 400 ha area dominated by mixed secondary and primary growth lower montane and montane rainforest surrounding the village of Brasso Seco in the Northern Range of Trinidad (10°45' N, 61°16' W), between 130 and 400 m above sea level. The primary peak in breeding in *P. longuemareus* in Trinidad coincides with the dry season (i.e. January to May; Snow 1968), and all field work was conducted during this period between 2008 and 2014. Due to the non-independence between years in male tenure on each lek (and the resulting non-independence of inter-year spatial and vocal variation), I chose to examine data from a single year (2013) for which the most complete dataset was available.

2.3.1.2 Study species

P. longuemareus is a small (mean \pm SD: 2.93 \pm 0.19 g, $n = 399$) trap-lining hummingbird, which occurs from Trinidad and north-east Venezuela to French Guiana. In Trinidad this species occupies a range of habitats from semi-open scrub to dense cloud forest and occurs at elevations from near sea level to the upper slopes of the Northern Range (ca. 940 m, personal observation). This species has a lek mating system throughout its range, and in mountainous regions the leks are invariably located along ridges ($n = 16$). Males aggregate on leks during the breeding season (December to June) where they defend 1 to 5 horizontal twigs

(1.5 – 4 mm in diameter), between 22 and 58 cm (mean = 33.3 cm) from the forest floor, that are used regularly as singing perches (Wiley 1971). Inter-male spacing on leks varies greatly, ranging from 2.5 to 27.4 m (mean = 8.13 ± 4.1 m). Males often use the same twigs for several years (personal observation), and may be present on their territories for as much as 70% of daylight hours during the breeding season, producing songs at an average rate of 0.5 songs per second (Snow 1968).

2.3.1.3 Capture and marking

The 78 lekking males included in the analysis were captured in one of two ways. During the start of capture effort on a lek, ultra-fine mesh mistnets (Ecotone, Sopot, Poland) were set in a linear array along the contour of the slope below each lek. When individuals left their territories to feed they would occasionally be captured passively. This technique lost effectiveness after one or two days of netting as hummingbirds that hit the net but escaped or saw the net upon approach were far less likely to be captured in this manner, and capture rates dropped precipitously. I therefore developed a second technique for targeting individuals using a modification of a Russell net trap (Russell and Russell 2001, p. 34), in which a mistnet enclosure (open on one side) was placed around the primary song perch of a targeted male. When a male alighted on the primary perch, he was chased farther into the enclosure and retrieved. All captured males were marked with a metal band and plastic colored leg tag and/or back tag (details in Kapoor 2012).

2.3.1.4 Sound recordings

Once marked, I recorded the display vocalizations of each lekking male on each lek by placing an omnidirectional Sennheiser K6/ME67 microphone (Sennheiser Electronics, Old Lyme, CT), connected to a Shure FP11 microphone-to-line amplifier (Shure Brothers, Inc.) and

Marantz PMD671 solid-state recorder (Marantz, Mahwah, NJ), under the primary song perch of a focal individual. Individuals were allowed to return to their perches and recorded for as long as they continued to sing, up to approximately ten minutes. I recorded individuals periodically throughout the season, aiming to obtain recordings for all lekking males in each month of the breeding season to capture variation in singing behavior within a year. Because the unstructured songs of recruiting juveniles are not representative of their eventual crystallized song structure only adult males were included in the analysis. Digital recordings were made with 16 bit resolution, a 48 kHz sampling rate, and were stored as uncompressed PCM WAVE files. The location of each individual recorded was measured using standard surveyor techniques (details will be provided in a subsequent manuscript).

2.3.2 Automated song characterization and comparison

2.3.2.1 Element segmentation and terminology

Fifteen contiguous songs were chosen for each bird for the song analysis. The block of songs chosen was the section of a full recording with the highest signal-to-noise ratio (SNR; resulting in the clearest spectrographic representation). Here, a song is defined as a series of elements that form the shortest regular repeating pattern of elements possible (e.g. sequence ‘abac’ from the full sequence ‘abacabacabacab’). Elements are single vocalizations of at least 23 ms duration isolated on both sides by a gap of > 6.5 ms. Elements were then segmented automatically using the ‘band limited energy detector’ of the sound analysis software program RAVEN PRO version 1.5 (Cornell Laboratory of Ornithology, Ithaca, NY) using the following settings: minimum frequency 1500 Hz, maximum frequency 12000 Hz, minimum duration 0.023 s, maximum duration 0.35 s, minimum separation between elements 0.0065 s, minimum occupancy 60%, SNR threshold 15.0 dB, noise block size 3 s, hop size 0.5 s, and noise percentile

15.0. The automated segmentation was conducted on spectrograms with a Hann window of 312 samples, and 3 dB filter bandwidth of 221 Hz, a frequency grid with a DFT size of 512 samples and 93.8 Hz spacing, time grid with 156 sample hop size and 50% overlap, and no spectral averaging.

2.3.2.2 Spectrographic representation

All signal comparisons were based on time and frequency parameters (and their derivative features). A first step in the spectral analysis of natural sounds is in balancing the resolution of time and frequency measures, features whose resolutions inherently trade off with each other due to the Heisenberg-Gabor inequality. An improvement of time and frequency resolution over short-term Fourier analysis (STFA) can be achieved with the technique of time-frequency reassignment (TFR; Auger and Flandrin 1995). TFR uses only the energy spectrum in calculations, whereas STFA uses this and the signal's phase spectrum to sharpen estimates of the power of each time-frequency point. High temporal and spectral analysis resolution is especially important when signals are strongly frequency modulated (Meliza et al. 2013) as they are for some elements in the little hermit (up to ca. 230 Hz). I therefore modified code distributed by Xiao and Flandrin (2007) written in MATLAB version R2009b (MathWorks, Natick, MA) to generate multitaper reassigned spectrograms with 6 Hermitian tapers (each with a window length of 199 samples), a time step of 20 samples, and 44.9 Hz spacing to improve temporal and spectral estimates of signals. All signals were high-pass filtered at 1500 Hz before the spectrograms were calculated to remove low frequency noise.

Three spectral-temporal contours (i.e. vectors containing the value of an acoustic parameter for each time step in a signal) were chosen to represent song elements: pitch, pitch change, and time. In species where vocalizations are tonal, such as those of little hermits, the

perception of pitch corresponds to F_0 of the vocalization, even when there is little or no acoustic energy at F_0 when harmonics are present (Cynx and Shapiro 1986; Shofner 2005; Meliza et al. 2013). Several aspects of the pitch of acoustic signals - including the production of vocalizations with loud low frequencies, singing with consistent pitch between song renditions, and the ability to vary pitch rapidly, among others - have been implicated in mate attraction and territorial defense in birds (Ballentine et al. 2004; Byers 2007; Hall et al. 2013). Multiple lines of evidence suggest that relative pitch (i.e. how a signal's pitch changes in time) is also an important feature in the perception and information content of tonal acoustic signals (Hulse and Cynx 1985; Hurly et al. 1990; Christie et al. 2004). I also incorporated temporal characteristics of signals into estimates of similarity since element length is a major component of variability that is likely to carry information (Cardoso et al. 2007; Cardoso et al. 2009).

Finally, the scale on which acoustic parameters are measured (e.g. linear versus logarithmic) can be an important consideration when comparing vocal signals. Current models of vertebrate acoustic perception and production suggest that the perception of sound frequency, the production of resonant frequencies, and the differences in fundamental frequencies between vocalizations of individuals of different age or size classes should be characterized on a logarithmic scale, but that production of fundamental frequencies within an individual are no better characterized by a log-scale than a linear scale (reviewed in Cardoso 2013). Therefore when comparing elements within individuals the pitch estimates and their derivative measures were not log transformed, but all comparisons between individuals used log transformed spectral values. To ensure that temporal features followed Weber's law, I adopted the method of Lachlan (2013, p. S4), which normalizes temporal measures by dividing time values by the total duration of the longer signal in each pairwise comparison.

2.3.2.3 Element measurement

I measured song elements using a custom MATLAB pitch-tracking algorithm. This algorithm is a modified version of one used by Wang and Seneff (2000) and Meliza et al. (2013). Briefly, the algorithm uses cross correlation with a single-pulse template to identify all candidate spectral peaks (if they exist) that may be a lower harmonic or the F_0 of the main peak in each spectral slice of a signal. Once these peaks have been identified each is evaluated for the possibility that it is either 1) a potential lower harmonic of the main peak, or 2) a lower harmonic or F_0 of a previously identified lower harmonic candidate peak, based on the property that harmonic frequencies are integer multiples of F_0 . The process is repeated until the lowest candidate peak that is harmonically related to the main peak is found and assigned as F_0 . Once F_0 estimates have been assigned to each spectral slice, the algorithm smoothes estimates by first locating the longest block of F_0 estimates that are not broken by large jumps in frequency (here, 15 frequency bins) that could be the result of pitch halving / doubling errors, or noise. Starting from this block, all F_0 estimates are forced to the frequency bin with the maximum power within 15 bins of the previously estimated adjacent slice's F_0 estimate. The smoothed estimates of F_0 result in a contiguous pitch contour that can be used in subsequent comparisons. The other two measures used to compare elements (pitch change and time) were derived from the F_0 vectors. The derivative of acoustic parameters with respect to time has been used effectively as a method for estimating the similarity in the acoustic “shape” between signals (Keogh and Pazzani 2001), and was therefore incorporated as the pitch change feature in the overall analysis (see Figure 2.1(a) for a graphical representation of the two spectral measures).

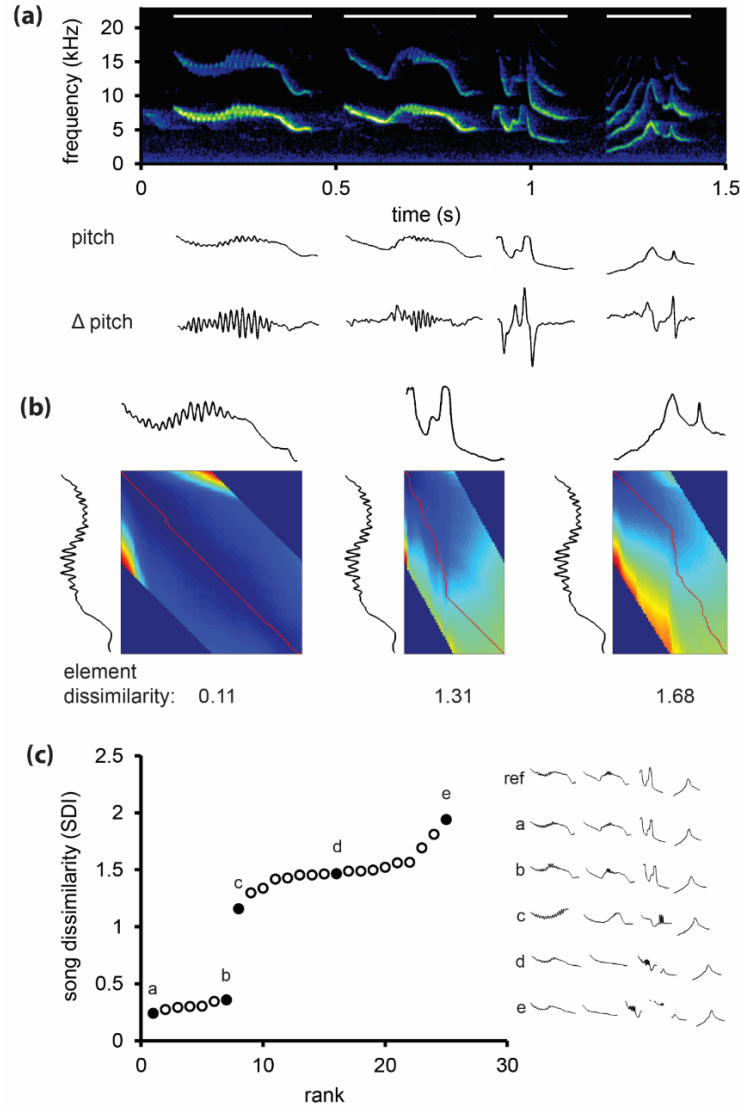


Figure 2.1. Spectrographic representation, measurement, and comparisons of little hermit songs. (a) A spectrogram of the song of a male *P. longuemareus*. Solid white lines indicate elements, and traces of two spectral parameters chosen as factors in the DTW analysis (pitch and pitch change) are provided below each element. (b) Three examples of DTW comparisons between a reference element (vertical axis) and a comparison element (horizontal axes). The blue triangles in the lower left and upper right corners of the plots are Sakoe-Chiba bands (see Methods) designed to prevent extreme warping. The hue of the diagonal areas represent the dissimilarity between the signals at the time points being compared, with warmer hues (i.e. closer to red) representing greater dissimilarity. The red diagonal trace in each plot represents the optimal warping path along the matrix. The overall element-wise dissimilarity scores (unitless) are provided below each plot. (c) A plot of the SDI scores between the song of a randomly chosen reference individual and all other members of that individual's lek. Pitch trace exemplars of the reference male (ref) and five other lekking males are provided, and refer to their respective pairwise SDI scores (filled circles with matching letters). Note the large jump in dissimilarity between individuals b and c, which is coincident with a dialect boundary (see Results).

2.3.2.4 Element comparisons

I used a DTW approach to compare elements using modifications of algorithms by Lachlan et al. (2013) and Meliza et al. (2013). DTW is a two-step process in which 1) the pairwise differences between each time point in two signals are used to generate a matrix of distances for a particular acoustic feature, and then 2) an optimal warping path through the matrix is chosen that minimizes the cumulative distance between the two signals (Sankoff and Kruskal 1983). In the simplest form of DTW, the similarity score between the two signals is then the average of the distance scores along this path. It is important, however, that the degree of warping (i.e. how much temporal stretching or compression was required to align the signals) be incorporated into the overall estimates of this distance. For instance, two signals that share spectral characteristics but that differ in length by a factor of two should certainly be considered less similar than two similar signals of the same length. The two main DTW approaches that have been employed to solve this problem are 1) the application of edit distances which add a cost factor for certain types of warping (in some cases compression, deletion, expansion, and insertion all possess different costs (Ranjard and Ross 2008), whereas others have employed exponentially increasing costs for each warping step (Meliza et al. 2013)), and 2) the use of normalized time values themselves as components of the distances between sampled points (Lachlan et al. 2013). The former approaches possess the desirable quality that they penalize moves that involve the warping of signals but do not penalize sections of the signal that are not warped in time. The latter approach enables time to be normalized so that the effects of differences in time and the effects of differences in other normalized acoustic parameters can be scaled to similar magnitudes. I therefore used a combination of these two methods to penalize time warping and compare the overall similarity between two signals as described below.

First, to compare two given elements A and B (of vector lengths n and m), the two spectral vectors (i.e. F_0 and F_0 change) were normalized and weighted as in Lachlan et. al (2013, p. S4; See equation 1). Then, to reduce the error in distance calculations associated with rapidly modulated signals I used symmetrical time interpolation to compare sampling points for both of these vectors separately (See Sankoff and Kruskal 1983, pp. 156–157). This approach essentially identifies the shortest distance between segments of the two signals where each segment is the line drawn between a pair of adjacent sampling points. These two vectors were then combined to form a Euclidean Distance (ED) representing the spectral measurements as:

$$\partial_{i,j} = \sum_h^2 (X'_{A,i,h} - X'_{B,j,h})^2 \quad (1)$$

where $\partial_{i,j}$ is the squared ED between the normalized spectral features at time points i and j , and $X'_{A,i,h}$ is the normalized value of element A at time point i in spectral feature h .

As the measure of temporal distance between segments of the two signals depended on the warping path, time was not normalized or interpolated in the same manner as for the spectral measures. Instead, I first defined the edit distances of particular warping operations as:

$$t(i,j) = \begin{cases} -1, & \text{if } t(i-1,j) = 0, \\ t(i-1,j) - 1, & \text{if } t(i-1,j) > 0, \end{cases} \quad \text{"deletion of } A_i \text{"}$$

$$t(i,j) = \begin{cases} 0, & \text{if } (i-1,j-1) \text{ is defined}, \\ 1, & \text{if } t(i,j-1) = 0, \\ t(i-1,j) + 1, & \text{if } t(i,j-1) > 0, \end{cases} \quad \begin{matrix} \text{"substitution of } A_i \text{ by } B_j \text{"} \\ \text{"insertion of } B_j \text{"} \end{matrix} \quad (2)$$

where t is the time warping edit distance, and i and j represent the respective time points being compared in elements A and B . These edit distances apply linearly increasing costs for warping (i.e. deletions or insertions), but no cost when no warping occurs (i.e. moving one time step forward in both signals [substitutions]). Note that when a move occurs along the diagonal of the

distance matrix comparing the two signals, the time penalty resets to zero, preventing time warping penalties early in the comparison from being weighted more heavily than penalties associated with equivalent warping later in the comparison (compare to Lachlan et al. 2010; and Meliza et al. 2013). These edit distances were then converted to normalized time penalties as:

$$\tau_{i,j} = \left(t(i,j) * \left(\frac{w_t}{s_t} \right) \right)^2 \quad (3)$$

where τ is the squared normalized time penalty of time point i of element A and time point j of element B , and, following Lachlan et al. (2013), w_t and s_t are the time weighting (here, 20) and the duration of the longer of the two elements being compared, respectively. The value of w_t was chosen through simulations with a test dataset to approximate the results of psychophysical tests demonstrating that birds are approximately ten times more sensitive to changes in frequency than to changes in the duration of sounds (Dooling and Haskell 1978; Dooling 1982). This method allows the time penalties to be normalized and scaled appropriately so that they can later be incorporated with the other ED measures into an overall distance score.

The two squared distance components $\partial_{i,j}$ and $\tau_{i,j}$ (spectral and temporal distances, respectively) were then combined into a single ED measure ($d_{i,j}$) to score the overall dissimilarity between time points i and j in the two elements. Since the temporal distance depends upon which of the three legal edit operations was chosen (i.e. moving one time point in B and zero time points in A [deletion of A_i], moving one time point in both A and B [substitution of A_i by B_j], or moving one time point in A and zero time points in B [insertion of B_j]), the three possible values of d are:

$$\begin{aligned}
d_{i-1,j} &= \sqrt{\partial_{i,j} + \tau_{i-1,j}}, & \text{"deletion of } A_i \text{"} \\
d_{i-1,j-1} &= \sqrt{\partial_{i,j} + \tau_{i-1,j-1}}, & \text{"substitution of } A_i \text{ by } B_j \text{"} \\
d_{i,j-1} &= \sqrt{\partial_{i,j} + \tau_{i,j-1}}, & \text{"insertion of } B_j \text{"}
\end{aligned} \tag{4}$$

These operation-specific distances were then used to calculate the recursion equation, which calculates the cumulative distance of the optimal alignment of the two elements using the following dynamic programming algorithm:

$$\gamma_{i,j} = \min \begin{cases} \gamma_{i-1,j} + d_{i-1,j}, \\ \gamma_{i-1,j-1} + d_{i-1,j-1}, \\ \gamma_{i,j-1} + d_{i,j-1}, \end{cases} \tag{5}$$

The final distance, or dissimilarity score, between the two elements A and B was then given by:

$$D_{AB} = \gamma_{n,m}/l \tag{6}$$

which is simply the final cumulative distance $\gamma_{n,m}$ divided by the length of the path followed from $\gamma_{1,1}$ to $\gamma_{n,m}$ by the dynamic programming algorithm. Note that the measure of the distance between elements is symmetrical, i.e. $D_{AB} = D_{BA}$.

An important consideration in the use of DTW to compare signals is the use of constraints on the movements through the distance matrix. The first constraint was a continuity constraint that forced the beginnings and ends of the signals being compared to align. This constraint effectively forces each comparison to consider the similarity of the entirety of both elements with one another (i.e. there is no partial matching between elements). The second constraint was a global limit on the degree of overall time warping that was permitted, prohibiting the stretching or compression of one signal by more than half the length of the other signal (Sakoe and Chiba 1978; see Figure 2.1(b)). This constraint prevents unrealistic alignments in which a very short segment of one signal is matched to a long segment of another signal that is

coincidentally similar. Examples of pitch trace comparisons using the DTW algorithm and resulting dissimilarity scores are provided in Figure 2.1(b).

The computation time required to compare signals using DTW increases exponentially with the size of the elements being compared. To limit computation time, and to smooth noise in acoustic measurements, all elements were compressed by a factor of 7, though signals were not allowed to compress to vector lengths below 10 (after Lachlan et al. 2013).

2.3.2.5 Element classification

Each individual bird's song elements were classified using the Ward method agglomerative hierarchical clustering algorithm (Ward 1963) on the dissimilarity scores calculated by the DTW algorithm. The hierarchical tree was then pruned at different cutoff heights (i.e. different "inconsistency" values; Zahn 1971) to generate cluster assignments for every possible number of clusters from one to the number of elements being compared. The best clustering solution was chosen by identifying the number of clusters that maximized the Global Silhouette Index (GSI; a metric of the similarity of elements within clusters and dissimilarity between clusters ranging from 0-1; Rousseeuw 1987). The standard GSI becomes uninformative in the presence of singletons which incorrectly score a perfect 1. In the presence of many singletons the GSI will always be high, regardless of the quality of other clusters (Almeida et al. 2011). The GSI was therefore modified to include only non-singleton clusters in the estimation of the global score. Because tree pruning occurred at the same height (i.e. inconsistency value) for all clusters, the presence of elements that were connected to an otherwise tight cluster above the cutoff point occasionally resulted in a sub-optimal partitioning of elements. In these cases, I manually merged these sub-optimally split clusters, recalculated the GSI, and accepted the result if the new score was higher than the original GSI. The result of this clustering analysis was a set

of unique categories into which each element in the song bout of an individual fit (i.e. an individual-level repertoire library).

A single exemplar of each element type in an individual's repertoire was then chosen for downstream comparisons between the full songs of individuals. Because a randomly chosen exemplar may be relatively far from the “acoustic centroid” of a given cluster of elements, I selected the “median element” from each cluster as the exemplar. To do this I calculated the average of the elements in the cluster using the DTW algorithm described above and the method of Sankoff and Kruskal (1983, pp. 157–159). I then identified the element in the cluster that most closely matched this average sequence as determined by DTW. The use of the median (instead of the average) element allowed for the retention of the fine-scale structure of an original element that is smoothed during averaging. This median element was used in subsequent between-individual analyses.

2.3.2.6 Song / bout comparisons

In order to compare songs between individuals within leks, accounting for differences in syntax, I adapted the method by Lachlan et al. (2010, p. 108; see section on chaffinch song). Briefly, the dissimilarity scores between all elements in a given song A (bird one) and song B (bird two) were recorded as described above. Then, I scored the dissimilarity between element transitions for song A and song B by calculating the pairwise element dissimilarity scores for sets of adjacent elements, weighted by the relative durations of the elements:

$$T_{A_i B_j} = w_{A_i} * D_{A_i B_j} + w_{A_{i+1}} * D_{A_{i+1} B_{j+1}} \quad (7a)$$

$$w_{A_i} = \frac{t_{A_i}}{t_{A_i} + t_{A_{i+1}}} \quad (7b)$$

where $T_{A_i B_j}$ is the transition distance between element i and $i + 1$ in song A and j and $j + 1$ in song B, and t_{A_i} is the duration of element i of song A. Transition scores are not symmetrical because $w_{A_i} \neq w_{B_j}$. I therefore calculated both $T_{A_i B_j}$ and $T_{B_j A_i}$ for all transitions in the two songs being compared. All adjacent elements, including the last and first element in each song, were considered valid transitions. Although moving from the final element of one song to the first element of the following song cannot be considered a true element transition in the sense of within-song transitions, scoring these cases as transitions allows for the similarity between two songs in their first and last elements to be incorporated into the overall song dissimilarity scores. This may be especially important for songs delivered in quick succession, such as those of *P. longuemareus*, where the beginnings and ends of songs are not well defined temporally. To measure the overall dissimilarity between a pair of songs, I first found the most similar element transition (i.e. the one with the smallest distance, $T_{A_i B_j}$) in song B for each element transition in song A, and vice versa. Then, to prevent transitions containing relatively short elements from receiving equal weight in the overall song distance score I weighted each transition score by its duration relative to the total duration of all elements in each transition for that song:

$$X_{AB} = w_{A_i} * T_{A_i B_j} + w_{A_{i+1}} * T_{A_{i+1} B_{j+1}} + \dots + w_{A_n} * T_{A_n B_n} \quad (8a)$$

$$w_{A_i} = \frac{t_{A_i}}{t_{A_i} + t_{A_{i+1}} + \dots + t_{A_n}} \quad (8b)$$

where n is the total number of elements in A. To combine the two song distance scores X_{AB} and X_{BA} into a single metric I calculated the mean distance, \bar{X}_{AB} .

Finally, to compare all 15 songs in the selected bout from bird one with the 15 songs of bird two I identified the most closely matched pair of songs in the two sets (based on the mean distance score above), then the next most closely matched pair, and so on, until each song in one bout was paired with a song in the other. I then took the mean of the dissimilarity scores of these 15 pairs of songs as the overall Song Dissimilarity Index (SDI) between these two birds. This method provides a way to penalize differences between songs in which the element structure is similar but element syntax varies; a condition that can result in consistent song differences between males. Figure 2.1(c) provides an example of the pairwise SDI scores and pitch traces of a reference individual and all other males on his lek.

2.3.3 Automated analysis of dialect variation

2.3.3.1 Dialect assignment and significance testing

To assign each lekking male to a single dialect on his lek I used the Ward method of agglomerative hierarchical clustering of the SDI scores, and identified the number of clusters that maximized the GSI (as in the song element classification step in the song analysis). Each of the three leks was analyzed separately.

Since clustering algorithms will assign groups to random data that do not possess actual structure it is essential to perform a significance test on the dialect cluster assignment results to confirm that they are non-random. To do this, I followed the general approach recommended by McShane et al. (2002) and Lachlan et al. (2010). Briefly, I subjected the song dissimilarity matrix for each lek to a nonmetric multidimensional scaling (NMDS) procedure to decompose the dissimilarity scores into an n-dimensional set of components. To avoid issues associated with sparse data in high dimensions (see Handl et al. 2005) I limited the number of components to retain by calculating the minimum number that led to a stress value of less than 0.1. I then

generated simulated component values by sampling from Gaussian distributions with means and standard deviations in each component estimated from the original data. The simulated components were then used to calculate a dissimilarity matrix using Euclidean distances. For each simulated dissimilarity matrix I used the clustering algorithm (described above) to partition the data into the same estimated number of clusters as the original dataset. The original GSI score was then compared to the distribution of simulated GSI scores calculated from 9999 simulated datasets.

To measure the extent to which the dialect clusters corresponded to the SDI values of the original dissimilarity matrices, I calculated the adjusted R^2 between a model matrix representing dialect assignments (a value of '0' represents individuals with the same dialect, a '1' represents individuals in different dialects) and the original SDI matrix for each lek. This test is roughly analogous to the cophenetic correlation of Sokal & Rohlf (1962), and should be considered a descriptive statistic in that it cannot be used for significance testing (Legendre and Legendre 2012).

2.3.4 Validation of automated comparison and classification by human assessment

Human validation of automated song analysis results was targeted at three hierarchical levels of song analysis: element dissimilarities, full song dissimilarities, and song clustering (i.e. dialect assignment), respectively. Each level contained its own protocol for human subjects, all subjects (other than the one subject experienced in song analysis with hermit hummingbird vocalizations: J. Kapoor) were different between the three analyses and were naïve to the identity of the birds and to the dialect assignments generated by the computational algorithms, and each

of the three analyses comprised songs recorded from individuals on three different leks, respectively.

2.3.4.1 Element dissimilarity comparison

To compare the similarity between the automated analysis and visual-based comparisons, I first randomly selected a set of 30 pairs of elements from a dataset of songs sung by all members of a single lek in 2013 (lek 7), and generated pairwise dissimilarity scores for each pair using the procedure outlined in the element comparisons section above. Next, four individuals naïve to the purpose of the study, and one experienced researcher (J. Kapoor), scored the pairwise similarity between the same 30 pairs of song elements based on visual comparisons of their spectrograms. All subjects were instructed first to inspect the spectrograms of all elements included in the comparison set to become familiar with the diversity of structure among the elements. Then, each subject was asked to score similarity for each pair on a scale ranging from 1 (totally dissimilar) to 10 (identical). To make these similarity scores compatible with the element dissimilarity scores generated by the DTW algorithm I transformed the scores to range from 0-1 (identical to totally dissimilar). Finally, I calculated the average correlation coefficient between the dissimilarity scores from all five human scorers, between the four naïve scorers, between each naïve scorer and the automated analysis scores, and the correlation coefficient between the experienced researcher's and the automated analysis scores.

2.3.4.2 Song similarity comparison

To compare the similarity between the automated analysis of full songs (rather than individual elements) and visual-based comparisons, I randomly selected a set of 20 pairs of songs from among all the members of lek 6 in 2013, and generated pairwise dissimilarity scores for each using the song comparisons procedure outlined above (in this case, however, only a

single song was chosen rather than 15). Three naïve individuals, and J. Kapoor scored the pairwise similarity between pairs of full songs using the same visual representation, scoring system, and score transformation procedures as in the element similarity analysis. Inter-observer and observer-algorithm statistics were generated as above.

2.3.4.3 Dialect comparison

For the dialect-level validation analysis I used the methods outlined above for generating dialect assignments, again using a single song per individual rather than 15 for all individuals recorded on lek 5 in 2013. Three naïve individuals and J. Kapoor then classified the same set of 26 songs (representing all individuals on the lek) by inspecting spectrographic representations of each song, and sorting the spectrograms into groups within which all individuals shared similar song structure. Groups of any size and number were allowed, from 1 (all 26 individuals belong to a single dialect) to 26 (each individual possesses a unique song). Scorers were instructed to classify the groups by whatever criteria seemed most appropriate to them to avoid influencing the clustering decisions. Agreement between scorers and between a consensus measure of the scorers and the computational algorithm were measured using the % agreement and % occurrence agreement metrics outlined by Jones et al. (2001).

2.4 Results

2.4.1 Algorithm performance

The TFR spectrograms used by the MATLAB script to sharpen time and frequency estimates generally sharpened spectrograms markedly, however, the computational complexity required a large amount of time to process relatively short sound clips (49.9 ± 14.2 s per s of sound data, $n=86$). The TFR algorithm, which generates data for all sound files to be compared

before any subsequent part of the analysis is resumed, allowed this data be generated in one period.

The pitch-tracking algorithm reliably tracked F_0 of the elements being measured, requiring 3.0 ± 1.1 s per s of sound data (n=86) of computation time. Most recordings were of very high quality [Figure 2.2(a) and (c)], as birds were typically recorded singing between 0.25 and 1 m from the microphone. Occasionally recordings were of lower quality, and although the pitch tracking algorithm typically succeeded in tracing the approximate course of F_0 with time down to a SNR as low as 0 dB, spectral smearing caused by reverberations and ambient noise tended to reduce the temporal resolution of notes with rapid frequency modulation (FM) when the SNR was below ca. 6 dB [Figure 2.2(b) and (d)]. Importantly, the pitch tracking algorithm reliably traced F_0 of notes with prominent harmonics that were of greater amplitude (either partially or wholly) than the fundamental. In cases when such errors occurred a manual pitch tracing feature allowed F_0 to be measured. Pitch trace smoothing generally functioned well to avoid large jumps between harmonics. Of a sample set of 86 elements, 0% had errors associated with misidentification of a harmonic for the fundamental frequency, and 13.0% required minor manual pitch smoothing due to a low SNR (10.6%) or due to the presence of reverberations (2.4%; see Figure 2.2(d) for an example).

Despite the large number of computations involved in the element- and song-wise comparisons each comparison was completed rapidly (0.29 ± 0.03 s per comparison, n=3655). However, since subsequent song classification steps require a full dissimilarity matrix to be calculated between all individuals, the computation time required increases by the square of the number of individual songs to be compared. For datasets comprising more than 30-40 individuals, the computation time required may reach several hours. For the test set of 86

elements, the total time required to generate the full element dissimilarity matrix (3655 comparisons) was 17.67 min. The vector smoothing and segment comparison features of the DTW algorithm reduced the effect of low fidelity pitch traces (due to lower quality recordings) in comparison scores. The greatest element dissimilarities were between short elements with minimal FM and long elements that were highly modulated. This tended to result in large element transition scores when a song possessing a short-duration element was compared to a song that was otherwise nearly identical but did not possess a short-duration element. The mean amount of time required for the completion of the element transition comparisons and clustering algorithm combined was 28.4 ± 2.8 s per lek ($n=26$ individuals in each lek).

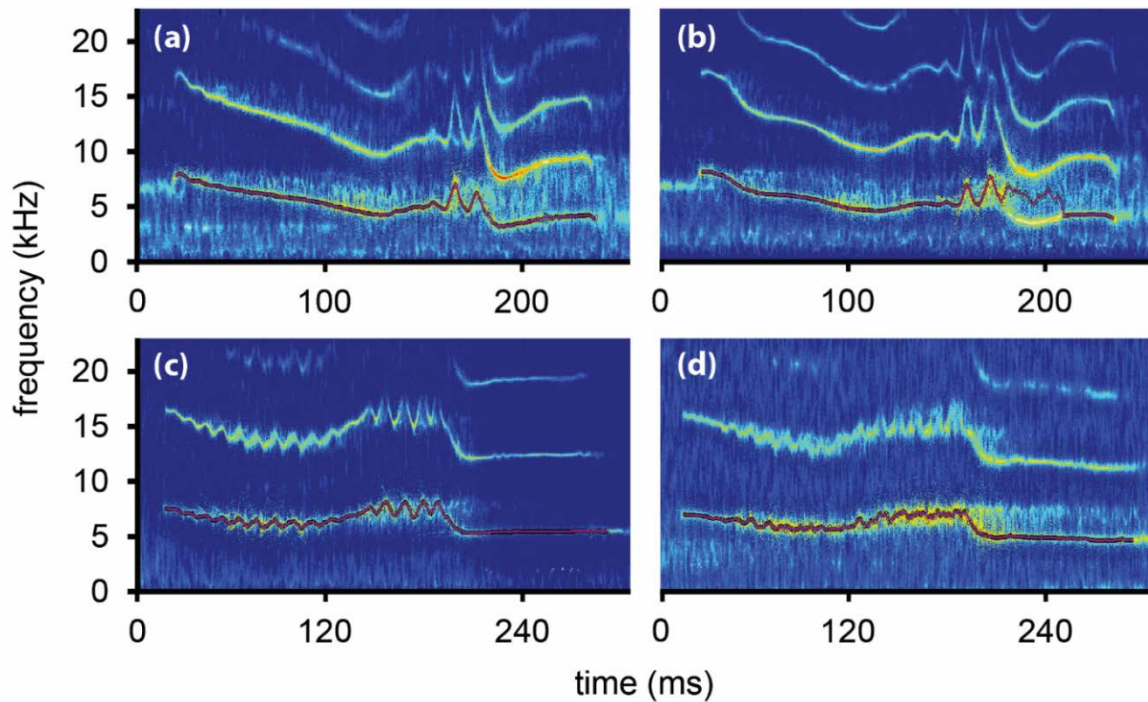


Figure 2.2. Four TFR spectrograms of song elements used in the study with original (i.e. not manually modified) fundamental frequency estimates (purple traces). In spectrogram (a) note that the trace smoothly tracks F_0 , even at time points when the peak power lies in a harmonic (black arrow). Spectral smearing (reverberation; black arrow) in (b) has caused the F_0 estimate to deviate from the true F_0 . Manual tracing of F_0 is necessary in these cases. Note that in spectrogram (c) the rapid FM of F_0 (near the black arrow) is faithfully reproduced by the pitch trace estimate. A similar element is represented in (d), however the low SNR has degraded the F_0 estimate so that the fine-scale FM structure of the element (black arrow) is lost. Manual tracing was generally moderately successful in recovering some of this structure.

2.4.2. Dialect structure

The hierarchical clustering algorithm used to generate groups of individuals that share similar songs (i.e. individuals with the same dialect) created significant clustering for all three leks tested [Table 2.1; 9999 permutations; Fisher's combined $p < 0.001$]. The GSI scores for the clustering results on all three leks were greater than 0.5 [Table 2.1], indicating that dialects were well-defined (Grimsley et al. 2012). All three of the mean GSI scores for the simulated data, representing the null expected value for each lek, were below 0.32.

Table 2.1. Descriptive statistics for dialects on individual leks and averages. The GSI score is provided for the chosen clustering solution on each lek, along with the null GSI expected by chance. The adjusted R^2 measures the proportion of variance in the SDI score by dialect assignments.

Lek	No. dialects	Dialect size	Range of dialect size	GSI (Null GSI)	Adjusted R^2
5	6	4.3 ± 3.1	1 – 8	0.63 (0.31)*	0.79
6	5	5.2 ± 4.1	2 – 12	0.64 (0.31)*	0.81
7	7	3.7 ± 2.2	1 – 7	0.56 (0.32)*	0.69
average \pm SD	5 ± 1	4.3 ± 3.0	1 – 12	0.61 ± 0.044	0.76 ± 0.064

* $p = 0.0001$

The coefficients of determination (adjusted R^2) between each lek's SDI score matrix and the corresponding dialect matrix averaged 0.76 [Table 2.1], indicating high correspondence between the matrices, and that approximately 76% of the variation between individuals' songs is explained by between-dialect differences.

Basic descriptive statistics for the 18 dialects identified (number of dialects per lek, number of individuals per dialect) are provided in Table 2.1. Eleven of the 15 multi-individual dialects comprised individuals who all shared the same numbers of closely matched elements with the same syntax. The remaining four non-singleton dialects contained one individual that divided one element into two ($n=1$) or contained individuals with an additional (i.e. unshared) and/or missing element and with recombined syntax ($n=3$). There were three singleton (i.e. single

individual) dialects between the three leks included in the analysis. In one of these cases the singleton dialect was the result of the fusion of two elements and the repetition of a sub-sequence of elements shared with the most closely-matched dialect. The two other singletons all possessed novel elements, two with novel short-duration elements, and one with a truncated element reminiscent of the introductory note of the sister dialect and a missing element (pitch traces of the latter two dialects are adjacent to the brown and magenta circles in column f of Figure 2.3). All four factors included in the pairwise distance calculations (pitch, shape, duration, and syntax) contributed towards the clustering solution, as evidenced by systematic exclusion or modification of one of these factors to explore the resulting effect on clustering results.

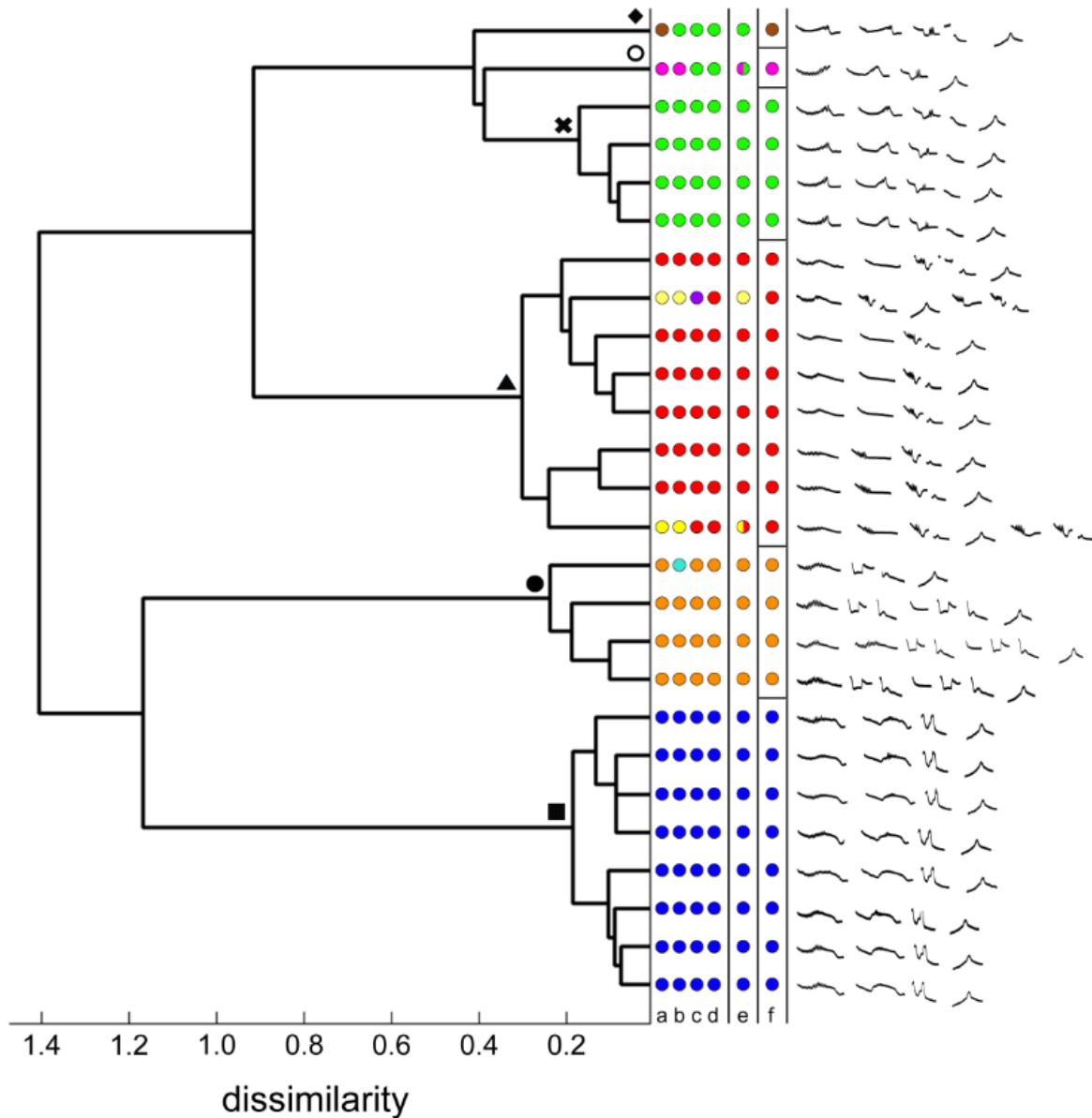


Figure 2.3. Hierarchical tree for the songs of all individuals on lek 5 in 2013, based on agglomerative clustering analysis of DTW results. The length of each branch is proportional to the dissimilarity (SDI score) between the song pairs being connected. Each leaf of the tree is flanked on the right by a row of four colored circles representing the dialect assignments of the three naïve individuals (a-c) and one experienced researcher (d) who classified songs visually, followed by a circle representing the consensus (see Table 2.3 for details) of all four scorers (e; split assignments are represented by circles containing two colors). The final circle in each row (f) is the assignment given by the DTW and clustering algorithms (horizontal bars were added between dialects for clarity). Pitch traces of exemplar songs are included on the far right. Note that within-individual variability in syntax is not represented by the pitch traces.

2.4.3 Comparison with human visual assessment

The inter-observer and observer-algorithm correlation results for both the element-level and song-level similarity scoring trials are presented in Table 2.2. In general, the agreement between observers for both tests was high, and roughly similar to values reported by other studies (Tchernichovski et al. 2000; Jones et al. 2001; Lachlan et al. 2010), and the correlation between these scores and the DTW algorithm scores were similarly high. Four of the five highest element dissimilarity scores generated by the DTW algorithm consisted of a comparison between a short upswing element and a long element containing sections with a high degree of FM. Four of these pairs were also among the top five most dissimilar element pairs identified by visual scorers (on average). The algorithm typically generated relatively larger dissimilarity scores when short elements were compared with long ones than the analogous visual scores. The five lowest dissimilarity scores generated by the algorithm comprised medium or long elements with very similar durations, similarly complex shapes, and FM structure. The average visually-based dissimilarity score placed three of these five element pairs as the most similar. The song-level analysis revealed slightly less agreement of song differences between visual assessment and the automated computer algorithms. Only two of the five song-pairs scored as most dissimilar by the computer program were also scored as the most dissimilar by visual assessment. Assessment of the most similar songs was more convergent between the two methods; four of five of the most similar song pairs were shared between the two methodologies.

Table 2.2. Song analysis similarity validation. Values are inter-observer and observer-DTW song distance score correlation coefficients.

Analysis	Correlation coefficient (r)			
	All scorers	All naïve scorers	Naïve scorers vs. DTW	Experienced scorer vs. DTW
Element similarity	0.88	0.85	0.80	0.85
Song similarity	0.81	0.80	0.78	0.84

In the visually-based clustering task, the visual scorers all identified different numbers of dialects (4, 5, 7, and 8), whereas the automated algorithm identified six dialects (note that this is the average number of dialects identified by the visual scorers). I measured the agreement between the DTW/clustering algorithms and the visual clustering analysis by using metrics recommended by Jones et al. (2001; Table 2.3). Overall % agreement between observers and with the algorithm was generally high, however, this metric over-emphasizes agreements in scoring unshared songs (i.e. those that are in different dialects). I therefore also measured the % occurrence agreement, which only counts scores in which one or both scores being compared indicated that the song pair should be grouped together. The % occurrence agreement was lower than % agreement across all groups compared, but highest between the experienced scorer and DTW algorithm. Scorers tended to differ in the decision of where in the hierarchical tree to prune, rather than how the branches were connected together, as evidenced by strong skew towards either disagreements where the visual scorer assigned two songs to be clustered, but the automated assignments separated them (1-0 mismatches), or vice versa (0-1 mismatches). On average, 83% of each individual's scores fell into either the 0-1 or 1-0 categories. Overall, the average proportion of 1-0 to 0-1 mismatches with the algorithm across all scorers was 0.5:1, indicating a greater tendency for the individual scorers to separate songs than the algorithm.

Table 2.3. Song analysis clustering validation. Values are inter-observer and observer-algorithm clustering agreement coefficients.

Measure	All scorers	All naïve scorers	Consensus score [†] vs. DTW	Experienced scorer vs. DTW
% agreement	94.2	94.3	95.2	97.2
% occurrence agreement	73.8	72.3	78.9	88.3

[†]Consensus scores for all naïve scorers were calculated by assigning a score of a pairwise match (1) when a majority of scorers indicated the song pair was the same and assigning a score of a mismatch (0) when a majority indicated the two songs being compared were different. Ties (i.e. half of the scorers scoring two songs as the same, half scoring them as different) were randomly assigned a score of 0 or 1 (five permutations of random assignment of scores for ties were averaged for the final agreement scores).

Among the disagreements between the visual scorers and the algorithm were a case where the algorithm placed songs in separate groups while a majority of the scorers had placed songs in the same group [Figure 2.3, column e, green circles] and another where scorers were split over the decision to place the songs in the same group [Figure 2.3, column e, magenta/green circle]. Similarly, there was one case in which a majority of scorers placed a song outside a group in which the algorithm had placed them together [Figure 2.3, column e, yellow circle], and another where scorers were again split over whether to include the songs in the same group [Figure 2.3, column e, yellow/red circle]. In all these cases of disagreement between the human scores and the algorithm's scores there was at least one scorer who agreed with the clustering decision made by the algorithm. Two of the four songs in these disagreements involved cases in which syntax differed but the majority of elements were shared with the other members of the contested group, and in two cases the element syntax was essentially the same, but a single short-duration element was either inserted or omitted.

2.5 Discussion

2.5.1 Algorithm performance

In general, the algorithms used in this study performed well, despite the existence of several steps involving substantial computation time. The step requiring the most computation time (TFR spectrogram generation) did not require human input, so could be allowed to run unattended. The pitch tracking algorithm typically performed well when the signal had at least twice as much power as noise (i.e. 6 dB), and relatively good approximations could be produced through manual tracing in even noisier conditions (down to 0 dB). Recordings used here were of high quality (ca. 30 dB), so an exhaustive analysis of the susceptibility of the algorithms to high

noise conditions is beyond the scope of this study. The pitch tracking algorithm developed by Meliza et al. (2013) reported comparable levels of noise tolerance. Currently, the pitch tracking algorithm uses a continuity constraint to smooth pitch traces by selecting the longest series of contiguous F_0 estimates (i.e. those without large jumps in frequency between time slices) as the starting point for estimating the full trace. One possibility for improving noise tolerance could be to allow multiple contiguous segments to serve as starting points. There were relatively few song elements in the test set where the pitch trace estimate had to be manually corrected (13.0 %), and the majority of these manual corrections involved tracing very brief segments of the element, providing further support that the algorithm performed adequately. Nearly all subsequent steps in the analysis – element comparison, element classification, full song and song bout comparisons, and dialect classification – were completed rapidly and without the need for user input, excepting several cases in which manual merging of element clusters was required to maximize the GSI (see element classification in Methods).

2.5.2 Comparison to visual techniques

Although high inter-observer agreement and high correspondence between visual scoring results and those of automated analyses do not supplant the need for external validation that these techniques provide biologically meaningful measures of vocal similarity or clustering, this correspondence does lend support that these techniques are reliable measures of acoustic structure (Jones et al. 2001). In the absence of direct measurement of the perceptual discrimination of the organism being studied, human judgements of similarity are the most widely accepted validation technique (reviewed in Jones et al. 2001). Despite this, consistent differences between the results of automated and visual techniques need not suggest that the

results of the automated technique are incorrect, especially when automated techniques are designed specifically to follow perceptually relevant rules.

At the element level the visual and automated song dissimilarity scores produced very similar results, with correlations that are well within the range typical for inter-observer correlations. The primary difference between visual and algorithm comparisons at the element level was that automated comparisons produced dissimilarity values that were not bounded between 0 and 1, resulting in a long right-tailed distribution. More studies of animal perception are required to determine whether there is an upper limit to the degree of dissimilarity between sounds, which would provide an argument for setting a threshold value beyond which all comparisons would share equal dissimilarity values. The fact that the correlation was high between the unbounded automated scores and the bounded visual scores suggests that this difference is unlikely to have a practical impact on clustering results.

At the song level, the correlations between visually-based and algorithm dissimilarity scores were slightly, though consistently, lower. This suggests that human scoring of element transition distances, and more generally, the penalties applied for syntax differences and differences between songs with missing or additional unshared elements, introduced some additional variance not explained by differences in element measurement between the two methods. Again, the overall correlations between the two song-level methods were high, suggesting little practical significance of discrepancies.

Correspondence was also very high between visual and computer algorithm clustering results. The automated analysis estimated a number of dialects that was intermediate to the individual visual scorers' estimates, providing support that the algorithm performs similarly to the average visual scorer. This conclusion is further supported by the fact that the % agreement

and % occurrence agreement scores between the automated analysis and the consensus assignments were higher than the same agreement metrics among the naïve observers. Regarding the composition of song clusters, five of 26 of the individuals' songs analyzed produced some level of clustering disagreement between scorers and the automated analysis results, with three of these regarding the stringency of criteria for assigning a match, and the remaining two regarding the branching structure of the hierarchical tree. In the latter two disagreements at least half of human scorers placed an individual that was more peripheral in the hierarchical tree, inside a group in which a more central individual had been excluded. Below, I discuss each case separately:

In the first case, the automated algorithm placed two individuals in their own singleton dialects ("diamond" and "O" dialects, respectively) and four others in a third dialect ("X" dialect) [Figure 2.3, column f, brown, magenta, and green circles, respectively]. All of the visual scorers placed the song that was most peripheral in this group (i.e. the "diamond" dialect individual) in the same cluster (i.e. Figure 2.3, column e, green circles), but half placed a more central individual (i.e. the "O" dialect individual) in a separate dialect (i.e. Figure 2.3, column e, magenta/green circle). It is logical to ask which of these two incompatible classifications appears to have the most support. The first thing to note is that the two individuals in question have branches that connect to the "X" dialect cluster at very similar tree depths, suggesting that the automated algorithm treats these two songs as nearly equal in their distance from it. Had the "O" dialect scored as only slightly more dissimilar from the "X" dialect the branching order would no longer be in conflict with the visual classification. Second, the visual scorers appear to have emphasized the number of differences in their placement of the "O" dialect individual closer to the "X" dialect. The "diamond" dialect individual possesses an extra short-duration element

compared to the “X” dialect (i.e. 1 difference), whereas the “O” dialect has a truncated first element and is missing a down-sweep element (i.e. 2 differences). The automated analysis appears to take a more nuanced approach in comparing element structure as well. Despite the truncation of the first element in the “O” dialect individual’s song, the shape of this element is a closer match to the same part of the first element of the “X” individuals, than the “diamond” dialect individual’s first element, which is spectrally flatter. The second element follows this same pattern. Considering these factors, the determination of which of these songs should be considered more similar to the “X” dialect becomes more complicated. Finally, a comparison of the spatial arrangements of the individuals in question reveals that the “O” dialect individual is spatially nested within the “X” dialect, while the “diamond” individual resides more distally.

Although spatial distance only partially correlates with song structure (see below), this provides further support for the automated analysis placement of the “diamond” dialect individual as more peripheral acoustically. Importantly, the algorithm placed both of these individuals as singletons, given the tight cohesion of the rest of the group, which circumnavigates the problems associated with including one of the singleton individuals with the rest of the “X” dialect.

The second case of disagreement between human scorers and the automated analysis involved a song that visual scorers agreed should be clustered with another song that did not share a direct branch in the automated hierarchical tree [Figure 2.3, column e, yellow and yellow/red circles]. In this case, visual scorers appear to have given precedence to the presence of shared elements and syntax, and gave less weighting to similarity in the fine-scale structure of shared elements. The two songs that visual scorers clustered together clearly shared a certain degree of their syntactical structure in that they introduced three shared elements after the

elements making up the conclusion of all other members of the “triangle” dialect. The automated algorithm, placing these two songs in separate sub-clusters, clearly did not emphasize this similarity, instead emphasizing similarities in the fine-scale structure of the long introductory elements. Specifically, all individuals of the sub-cluster containing the “yellow” individual (i.e. the individual represented by the yellow circle in Figure 2.3, column e) shared a first introductory element with more FM than the second sub-cluster’s first element. Similarly, the second element of the “yellow-red” individual’s song (i.e. the individual represented by the yellow/red circle in Figure 2.3, column e) shared marked FM with his own sub-cluster that was absent in the first sub-cluster. The element that is clearly unique to the two separated individuals in question makes a definitive conclusion difficult. Again, an analysis of the spatial arrangement of these individuals, relative to others in the “triangle” dialect may help to provide an explanation; these two individuals are represented by the triangles on the map of Figure 2.4 with a dashed line between them. Despite the visual classification, which suggests these two individuals might be associated with one another, they are not nearest neighbors. Furthermore, the automated analysis places one of these birds in the acoustic hierarchical tree near his spatial neighbors.

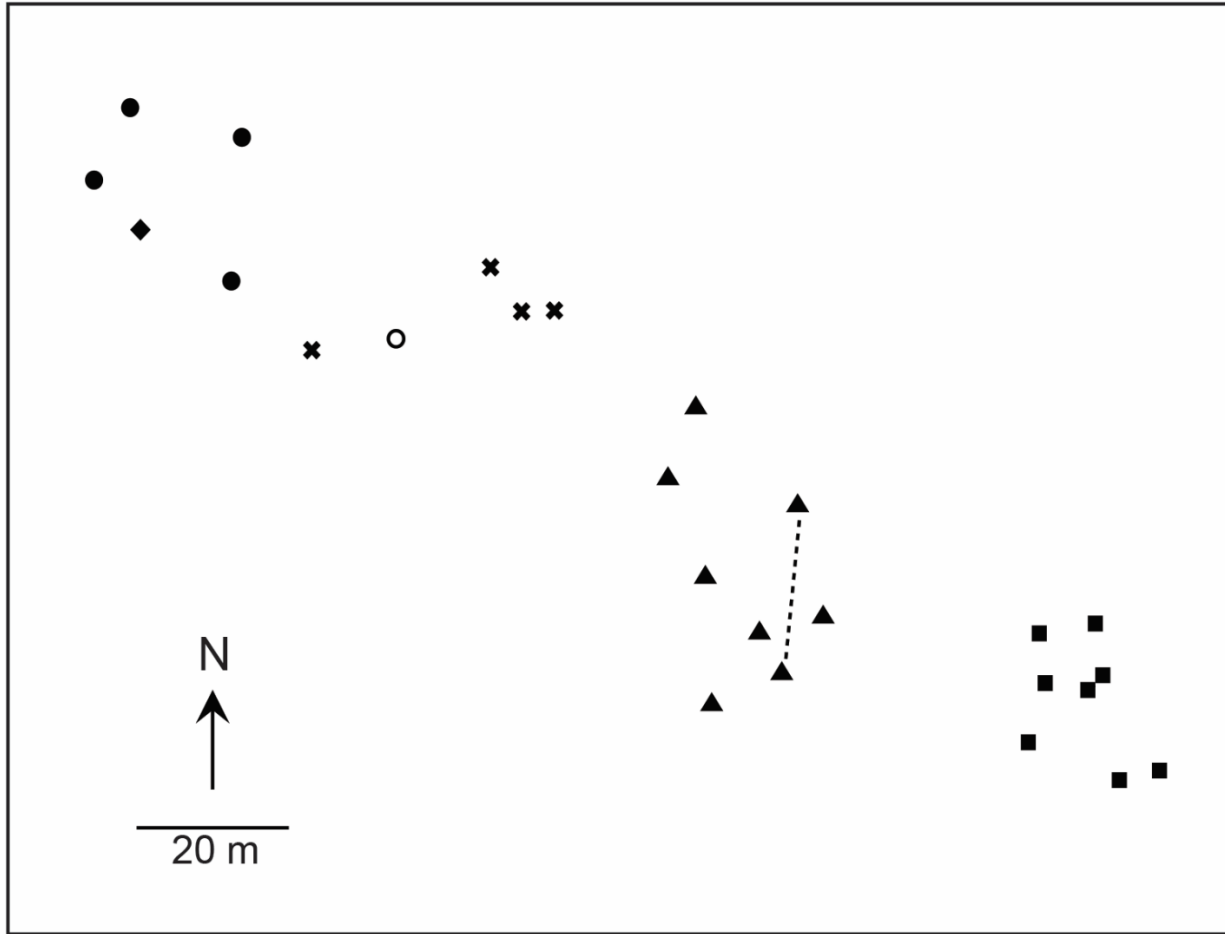


Figure 2.4. Dialect map of lek 5 in 2013. Each marker represents the primary song perch of a lekking male, with different symbols representing different dialects as determined by the DTW and clustering algorithms. Symbols match those in the hierarchical tree of Figure 2.3. The dashed line connects the two individuals that were placed in their own group by half of the visual scorers (represented by yellow and yellow/red circles in Figure 2.3, column e).

These two examples of disagreements between visually-based clusters and the automated algorithm suggests the possibility that the song acquisition process in *P. longuemareus* is not limited to learning from a single tutor or tutors from the same dialect group. Songs could instead be formed by piecing together elements and syntax learned from multiple individuals with different songs. The two alternative clustering results may reflect the possibility that not all songs are hierarchically related to one another, rather than reflect that one technique's results are invalid.

2.5.3 Comparison to visual approaches of other studies of hermit hummingbirds

Several authors have previously investigated song dialects (often called “song groups”) in hermit hummingbirds, and a comparison between the quantitative techniques described here and the mostly visually based techniques of these authors may be illuminating. The first extensive investigation of song variation in *P. longuemareus* was conducted by Snow (1968), who documented song dialects in this species. Snow used a visual gestalt, separating individual singers into different groups based on the overall complement of elements that were “recognizably different” between groups. No indication about how divergent groups of individuals had to be to obtain “song group” status was given. Wiley (1971) analyzed the song dialects in a more quantitative manner, using a visual approach to identify commonalities in fine-scale features of elements shared between individuals. As in the former study, Wiley does not provide criteria for the choice of the level of clustering that was used to identify dialects, though verbal descriptions provide some indication that songs within groups bore “strong resemblance” to one another, or were “virtually indistinguishable,” and that songs in different groups were “radically different.” Munding (1982) takes a different approach, borrowed from the field of linguistics, to reanalyze Wiley’s dataset and determine whether the term “dialects” is reasonable. In his study he identifies shared elements (again based on visual examination of spectrograms), then, using a map of individuals on the lek, draws lines around individuals sharing a song element (known as an “isogloss”), and repeats this procedure for each unique element. The result is a series overlapping (and occasionally intersecting) lines that can be used to identify groupings (i.e. dialects) in which multiple parallel lines form a bundle. How variation between elements within an isogloss is considered, and how many overlapping isoglosses are necessary to define a

dialect are unclear. All three of these studies propose a hierarchical structure of similarity among individuals on the leks, and then seek to identify the possible song learning program based on the observed patterns. All three authors conjecture that individuals from one or more founding groups may have moved to new locations and introduced variants of existing songs, thus starting new dialects.

The results of the automated song analysis provide a general picture that is highly compatible with the results of these other studies. First, although the measure of song dissimilarity in this study included features not explicitly defined in these other studies (e.g. the absolute pitch of elements and similarity in element durations), and even gave importance to a feature omitted by Mundinger (i.e. element and song syntax), similar numbers of dialects were identified in all studies. Also, in the two leks included in the visual analysis-based studies, spatial cohesiveness of individuals was high within song groups (i.e. individuals within dialects were each other's closest neighbors), as was the case for the individuals in 16 of the 18 dialects described for the three leks in the current study. The hierarchical tree generated by the automated algorithm is directly analogous to the verbal argument of the three authors in their descriptions of which dialects are most closely related and how this structure relates to the spatial locations of individuals with different dialects. On all three leks (only data for lek 5 shown) the most closely related song groups were as likely to be separated by more acoustically dissimilar song groups as they were to be nearest neighbors. The results of this analysis provide further evidence for the hypothesis that dialects are generated through founder effects rather than exclusively through budding of variants on the periphery of existing dialects.

2.5.4 Advantages of automated analysis

Although high correspondence between visual and automated song analysis techniques suggests that automated techniques are compatible, this correspondence does not, by itself, provide an argument for the adoption of automated methods. My goal in this study was to present and evaluate a measurement and sound classification system for tonal vocalizations, validated by visual techniques, but that is superior to them in its repeatability, objectivity, the explicit definition of its parameters, and in the usefulness of its quantitative measures. The automated song analysis techniques described here have very few steps requiring human input, and use explicitly defined parameters and weights to calculate distance metrics, making solutions more repeatable than techniques requiring extensive human input. Visually based (and to some degree spectrographic measurement based) analysis techniques use human judgement to estimate resemblance and to make clustering decisions, which accounts for significant drops in repeatability. Using the automated algorithm, only a small subset of song elements required manual tracing of a short section of F_0 . During the course of this study, I manually re-measured song elements several times, with no changes in the resulting clustering solution (data not shown).

The reduction of steps requiring human input also has the effect of increasing the speed of comparisons, enabling the analysis of large datasets. In the study of dialects, it is often important to compare many individuals in a pairwise manner to fully describe dialectal structure. Computationally this can be problematic because every additional individual included in an analysis increases the number of comparisons exponentially, which severely limits the number of individuals that can be compared using human visual scoring (Kogan and Margoliash 1998). It should be noted that although the practical sample size limit for a computer-based procedure is much larger, it suffers from the same limitation. For very large datasets the use of evolving tree

clustering methods (Pakkanen et al. 2006; Ranjard and Ross 2008) has proven to be an invaluable alternative to classical clustering methods.

The explicit definition of vectorized parameters for both the comparison and classification steps (e.g. the inclusion of pitch, and spectral shape with equal weighting in calculating pair-wise distances) allows the maximization of the use of objective criteria that are treated consistently across all signals included in an analysis. Of course, a fair number of somewhat arbitrary decisions (e.g. the choice of parameter weightings, the element transition scores, the clustering algorithm, and cluster validation statistics) have to be made to fill in the gaps in our current understanding of how animals process and classify sound, however, at worst, this results in subjectivity on par with that of visually based analyses (Tchernichovski et al. 2000; Lachlan et al. 2010).

Another important advantage of the use of vectorized spectral-temporal features and explicitly defined parameters is that this enables the song analysis procedure to be tailored to match perceptual classifications obtained from the animals themselves. For instance, the exclusion or addition of certain features or the modification of weights in the analysis of vocal structure can be used to investigate the importance of each of those factors in generating the structure and composition of perceptual classes (for an example see Ito and Mori 1999). In the absence of external validation of clustering results from psychophysical assays, as in this study, feature and parameter selection can be based upon perceptually relevant information such as Weber's Law (Lachlan et al. 2010), the perception of F_0 in tonal signals (Meliza et al. 2013), and other articulatory variables involved in sound production (Ho et al. 1998; Tchernichovski et al. 2000). These features are especially important since the selection of arbitrary analysis parameters can often generate technically good classification results that are not likely to be relevant to the

organism being studied. This capability represents an improvement over visual approaches, and SPCC approaches not using stated parameters.

Finally, one of the most important advantages of the automated analysis used here over visually-based analyses is the quantification of acoustic metrics beyond simple cluster assignments. First, the vectorization of each signal allows a number of different metrics to be calculated automatically (such as wiener entropy, harmonicity, vibrato amplitude and frequency, etc.) that can then be used in downstream analyses. Next, the DTW algorithm generates a full dissimilarity matrix, allowing both pair-wise and group-wise comparisons, which can be used to provide measures of the acoustic cohesion of a group of individuals, the degree of separation between an individual and its own group's centroid (i.e. the average song within a dialect) or to other groups. Pairwise comparisons can enable the estimation of the accuracy of song learning in particular parameters such as syntax or element shape, and the identification of potential tutors. This analysis technique can be applied towards temporal analyses as well, enabling the tracking of individual song development and population-level changes over time. The DTW algorithm can also be used to compare songs within an individual to estimate repertoire sizes and to measure song consistency for many explicitly defined parameters. This allows the combination of the sensitivity of SPCC for whole element consistency measures with the specificity of parameterization of time-frequency measures (see Cramer 2013a). The DTW algorithm is also capable of generating new synthetic signals that represent the average between two existing signals for use in playback studies. Hierarchical clustering techniques create not only group assignments (i.e. dialects), but estimate the order of relationships between individuals within and between dialects, making it possible to test hypotheses about whether a hierarchy possesses significant clustering.

2.6 Conclusions

Analysis tools inspired by recent advances in the field of human speech processing, such as the feature extraction, DTW, and clustering algorithms presented in this paper, have made it possible to improve bioacoustic methods to perform comparably with more traditional visually-based methods. Moreover, this approach has distinct advantages in its speed, ability to incorporate explicit models of animal acoustic perception, and the level of quantitative detail produced. The empirical test of these algorithms on the highly structured vocalizations of *P. longuemareus* demonstrates the utility of this method to describe vocal variation and help provide an analytical framework to statistically test for the existence of dialects in species with tonal vocalizations. Despite the fact that many of the steps involved in this and similar methods involve some level of subjectivity, and remain far from perfect, it is my hope that the use of these methods, combined with psychophysical confirmation of song analysis results, will allow for a more rigorous and quantitative approach to the study of animal vocal dialects.

CHAPTER 3

DECEPTIVE VOCAL MIMICRY IN A LEKKING HUMMINGBIRD

3.1 Abstract

Among birds that learn their songs, microgeographic dialects, or shared songs between neighboring birds, are common. Two dominant hypotheses have been proposed to explain this pattern: First, dialects could result as an epiphenomenon of the use of shared songs in modulating aggressive interactions between territorial males. Second, dialects may result from the use of song mimicry by intruding individuals in an attempt to deceive existing territory holders into treating those intruders like “dear enemies” (i.e. the deceptive mimicry hypothesis). Although general support has been found for the former hypothesis in many oscine taxa, no strong support has been found for the latter. Here, I test whether the deceptive mimicry hypothesis may be a plausible mechanism for the generation of dialects in lek mating systems, where the intense selection for obtaining a display territory may make deceptive mimicry especially likely to evolve. To test predictions of the deceptive mimicry hypothesis I used an experimental playback design in which I compared aggressive responses by lekking male little hermits (*Phaethornis longuemareus*) to playbacks of shared (“mimic”) and unshared (“non-mimic”) songs. I found strong empirical support for the conclusion that deceptive song mimicry by intruders can produce a “dear enemy” response in territorial males. More generally, the results of this study provide evidence that deceptive mimicry may be a potentially common mechanism for the formation of dialects in lekking species.

3.2 Introduction

Discrete geographic variation in acoustic signals (i.e. dialects) has been a major focus of evolutionary biology as researchers attempt to uncover the processes leading to such variation, and its possible adaptive significance (Slabbekoorn and Smith 2002b; Podos and Warren 2007). Whereas patches of shared vocal signatures across broad geographic scales may be a reflection of the effects of ecological selection (Podos and Warren 2007; Badyaev et al. 2008; Derryberry 2009), acoustic adaptation in different habitats (Slabbekoorn and Smith 2002a; Seddon 2005), or founder effects and drift in isolated populations (Grant and Grant 1996; Lachlan et al. 2013), patterning in vocal signals at the level of social groups has been studied extensively in the context of social and sexual selection (reviewed in Podos and Warren 2007). This microgeographic variation in signals has sparked considerable debate: are shared signals epiphenomena of the learning process and settlement patterns or socially adaptive (Baker and Cunningham 1985; Slater 1989)?

One of several hypotheses that have been put forward to explain how social selection for song sharing might generate microgeographic dialects is the deceptive mimicry hypothesis: a male recruiting to a population copies local song(s) in order to give the deceptive impression of being an established male, and either enjoys reduced aggression from other males through ‘dear enemy’ effects or deceives site-faithful females into mating (Payne 1982; Payne 1985; reviewed in Rainey and Grether 2007; Cockburn et al. 2009). For intrasexual deceptive mimicry through song sharing to be favored by selection, several conditions must be met. First, territorial males must show reduced aggression rates or intensity towards neighbors than non-neighbors, a phenomenon known as the ‘dear enemy’ effect (Fisher and Fisher 1954). Second, individual recognition must be masked through accurate copying of songs or through constraints on

recognition abilities, otherwise mimics would be immediately identified as such (Wilson and Vehrencamp 2001). Third, species showing this form of mimicry should have natural history traits in which successful deception about residency status would contribute significantly to fitness (McGregor and Krebs 1984).

General support for the presence of a ‘dear enemy’ effect has been found in a wide range of taxa and mating systems including territorial oscines (Stoddard et al. 1990; Brindley 1991; Briefer et al. 2008a) and suboscines (Lovell and Lein 2004), cooperative breeders (Botero et al. 2007), polygynous birds (Falls and McNicholl 1979), and many non-avian taxa (Leiser and Itzkowitz 1999; Kaib et al. 2002; Lesbarrères and Lodé 2002; Husak and Fox 2003). In many territorial songbirds, the ‘dear enemy’ effect manifests as a reduction in aggression rates and intensity between established territorial neighbors, as these individuals typically represent a lower level of threat of territorial take-overs with respect to unknown floaters who are prospecting for opportunities (reviewed in Stoddard 1996). More generally, species in which there exists a difference in relative risk of territory usurpation or competition between two or more classes of individuals are expected to exhibit a ‘dear enemy’ effect (reviewed in Temeles 1994) and may therefore be candidates for the evolution of deceptive mimicry.

Support for the second requirement of the deceptive mimicry hypothesis (i.e. that vocal deception of identity must be possible) is more mixed. First, in a captive psychophysical experiment Stoddard et al. (Campbell et al. 1992) found that song sparrows, *Melospiza melodia*, were initially unable to discriminate between matching song types from different individuals; a pattern that was later demonstrated more robustly by Beecher et al. (1994). Also, McGregor and Avery (1986) and McGregor (1989) have demonstrated that neighbor recognition in free-living great tits, *Parus major*, was impaired if new neighbors shared song types with prior neighbors. In

contrast, a much larger body of work suggests that individual recognition is generally common. Although earlier work suggested that extremely large song repertoires were likely to hinder individual recognition (Kroodsma 1976; Falls and d'Agincourt 1981; reviewed in Stoddard 1996), subsequent theoretical (Lambrechts and Dhont 1995) and empirical studies (Weary and Krebs 1992; Naguib and Todt 1998; Botero et al. 2007; Courvoisier and Aubin 2014) have demonstrated that recognition is not typically hindered by large repertoires since individual recognition in these species may relate to a number of individually distinctive vocal characteristics including vocal tract resonance effects and song syntax. In species with moderate vocal repertoires, neighbor song sharing will seldom be 100%, so even rarely delivered unshared song types may give up an individual's identity (Molles and Vehrencamp 2001; Skierczyński and Osiejuk 2010). Indeed, it has been proposed that individual recognition is the rule for species with fewer than 25 song types (Stoddard 1996). Even in species with single song-type repertoires individuals may be capable of individual recognition since song type renditions are rarely identical (Weeden and Falls 1959; Emlen 1971; Wunderle 1978). Therefore it seems likely that only in species with extremely accurate song copying is deceptive mimicry likely to be able to evolve (McGregor and Krebs 1984; but see Rainey and Grether 2007 pp. 2444-2445).

The ontogeny of vocal learning may also present a barrier to the evolution of deceptive mimicry if plastic song (which is by its nature imperfect) is produced near the intended operator(s) (i.e. the individuals that are the target of the deception). This is likely an important consideration since the models (i.e. future neighbors) of mimicry will be in very close proximity to intended operators (also future neighbors), and plastic song could easily reveal the identity of an intruder as such. Thus, for deceptive mimicry to evolve individuals must be capable of post-

dispersal learning to accurately copy future neighbors, but refrain from singing plastic song at the intended site of settlement.

The third major consideration regarding the evolution of vocal deceptive mimicry is that it must be effective in conferring net advantages to mimics, with respect to non-mimics. Here, empirical evidence is either lacking or equivocal. In many socially monogamous oscines, the potential advantage of deceiving neighbors about the residency status of the intruder is unlikely to overcome the constraints imposed by the strategic use of song in intra- and intersexual communication. A number of studies have demonstrated sexual selection for large repertoires (Catchpole 1986), modulation of singing diversity and switching rates (Botero and Vehrencamp 2007), and the strategic use of type and repertoire-matched (and unmatched) songs to escalate / de-escalate contests during countersinging (Beecher et al. 1996; Burt et al. 2001); behavioral patterns that are at odds with the use of song as a deceptive signal through mimicry (Molles and Vehrencamp 2001). A study by Payne (1982) demonstrated that yearling male indigo buntings, *Passerina cyanea*, that copied the songs of successful adults enjoyed increased fitness relative to males that did not match. However, subsequent work provided the most support for the conclusion that shared song was used in intrasexual communication rather than as a deceptive indicator of residential status (Payne 1983, Payne et al 1988). Deceptive mimicry, then, is expected to evolve only in cases in which communication systems do not constrain the use of song in a deceptive manner, and in which the benefit of deception is a large component of fitness, such as when suitable breeding habitat is severely limited (Fretwell 1972; Rohwer 1982).

Unsurprisingly, as no unambiguous examples of intrasexual vocal deceptive mimicry appear in the literature to date, this hypothesis has fallen out of favor as a plausible process by which song sharing might have evolved; nearly 15 years have passed since the last test of this

hypothesis (Wilson and Vehrencamp 2001). The seemingly sound conclusion that the deceptive mimicry hypothesis should be laid to rest, however, may have been premature. As I have discussed above, deceptive mimicry is only expected when 1) song copying is extremely accurate, 2) individuals learn their songs post-dispersal, 3) where the use of song is not constrained by complicated matching rules, and 4) where breeding opportunities are strongly controlled by access to limited territorial resources. Oddly, despite the existence of species possessing all of these characteristics, support for the deceptive mimicry hypothesis has been tested only in species that mismatch in at least one of these criteria; a phenomenon that is likely related to the extreme and widespread bias toward the study of vocal behavior in socially monogamous, territorial oscine passerines rather than a lack of availability of appropriate species to study (discussed in Falls and McNicholl 1979; Wilbrecht and Nottebohm 2003; Lovell and Lein 2004; Saranathan et al. 2007; Kroodsma et al. 2013; González and Ornelas 2014).

To test the deceptive mimicry hypothesis under conditions in which it is most likely to occur I therefore conducted a field playback experiment on a lekking hummingbird, the little hermit, *Phaethornis longuemareus*; a species meeting all four preconditions for the evolution of deceptive mimicry. First, little hermits exhibit a pattern of microgeographic dialects where neighbors often sing songs that are essentially indistinguishable from one another (Snow 1968). Second, as this species possesses a lek mating system, all learning necessarily occurs after juveniles have dispersed and have begun to settle on their future territories (Höglund and Alatalo 1995). Third, song in this species is extremely simple, consisting of a single song type repeated iteratively and comprising a fixed set of 2 – 7 individual elements (Wiley 1971). Singing males therefore do not modulate their songs in ways that might automatically give up their identities. Fourth, as is typical of lek mating systems, mating in little hermits does not appear to occur away

from leks (Kapoor, unpublished data), suggesting that male access to breeding opportunities is entirely restricted to individuals who have gained access to the lek (Höglund and Alatalo 1995). All four of these conditions set up a scenario in which a male attempting to join a lek and establish a territory might be expected to benefit from the deceptive mimicry of established lek residents.

In this study, I develop a new experimental design to test the deceptive mimicry hypothesis more robustly than has previously been possible. To do this I adopted experimental design features and novel statistical approaches to account for a number of confounding social factors that have plagued previous studies (discussed in Molles and Vehrencamp 2001; see Methods). To evaluate support for the hypothesis that song sharing in little hermits is the result of selection for deceptive mimicry, I tested for the relative ability of resident males to recognize simulated territorial intrusions (i.e. playback) by an unfamiliar individual whose song closely resembled that of an adjacent neighbor (hereafter, the mimic) to the recognition of neighbors with shared songs and strangers with unshared songs. Specifically, I predicted that 1) focal males would show reduced aggression to the playback of neighbors relative to strangers (i.e. they show a ‘dear enemy’ effect), 2) focal birds would be incapable of distinguishing between neighboring individuals sharing the same song (i.e. there is no individual recognition), and 3) the mimic stimulus would be treated more like a neighbor than a stranger (i.e. deception is successful), and the degree of similarity between the stimulus and the adjacent neighbor (i.e. the accuracy of mimicry) would predict the magnitude of this effect.

3.3 Materials and methods

3.3.1 General methods

3.3.1.1 Study population

The study was conducted in a 400 ha area of first and second-growth lower montane and montane rainforest surrounding the village of Brasso Seco in the Northern Range of Trinidad (10°45' N, 61°16' W) on a free-living population of little hermits that have been marked and monitored since 2008. During the breeding season (December to June), males aggregate on leks where they defend between one and five horizontal twigs that are used regularly as singing perches (Wiley 1971). Males sing persistently during daylight hours, but switch from song to an elaborate courtship display when either females or other rival males visit an occupied perch (Snow 1968; Wiley 1971). Visits by rival males often conclude in prolonged and extremely fast chases through the surrounding forest, and may involve physical fights (personal observation). Females visit leks for the sole purpose of mating, and raise their young solitarily.

3.3.1.2 Capture, marking, and age estimation

All capture and marking was conducted between January and May each year from 2008 to 2014 on the four leks on the field site (i.e. leks 5 – 7 and 12) using the methods described previously (see Chapter 2 for details); briefly, birds were captured with mistnets and marked with a metal band and colored plastic leg tag (see Kapoor 2012).

The ages of focal males in the playback experiment were calculated in one of two ways. First, some individuals were banded as juveniles on and off leks, based on the degree of striations on the base of the bill, which provides a rough estimate of age in hummingbirds under one year old (Ortiz-Crespo 1972). Second, in 2008 all adult males on leks 5 – 7 were banded, making it possible to estimate the age of new birds after this year to within 8 months, with high probability, and despite a lack of observable bill striations. In no case has a banded bird in his

second year or later been observed switching between leks, and all non-lekking males caught off leks on the field site after 2008 have been identified as juveniles.

3.3.1.3 Lek spatial surveys

To determine whether lek residents should be classified as neighbors or non-neighbors I conducted a spatial survey of all perches occupied by little hermits in 2014 using traditional land survey techniques employing an optical Kern DKM1 theodolite (Kern & Co. Ltd., Aarau, Switzerland). The details of this procedure will be provided elsewhere (Kapoor, in preparation). Because males generally defend as few as one song perch it is often impossible to estimate territorial boundaries by generating minimum convex polygons (which require a minimum of three points). Instead, I estimated boundaries by calculating Thiessen polygons (i.e. polygons within which any position is closer to the point that generated it than to any other point) for each song perch, with a maximum radius of 16.5 m (estimated using the methodology of Marten and Marler (1977), details in unpublished manuscript) representing the maximum distance that an average song is discriminable. The polygons of multiple song perches occupied by the same bird were merged to create a single polygon. A neighboring bird was defined as an individual that shared at least 15 % of the perimeter of a focal individual's territory.

3.3.1.4 Sound recordings

I obtained close-range recordings with minimal background noise of the unprompted songs of each lekking male on all leks in each year of the study for use as playback stimuli (for details of recording equipment, field methodology, and digitization procedures used see Chapter 2). Only the crystallized songs of adult males were included in analyses.

3.3.2 Playback experiment

3.3.2.1 Focal individuals

To be included as a focal male in the playback experiment each bird was required to be a regular lek resident, to possess at least two neighbors with a shared dialect, and to have recruited to the lek at least one season after a bird with a dialect shared with a current neighbor had disappeared (i.e. presumably died), and had been recorded. A majority of lek residents had at least two neighbors with a shared dialect, and due to extensive recording effort, all of these birds recruited to the lek one season or more after the disappearance of a male with a song matching the dialect of a neighbor of the focal male. These requirements ensured that a song stimulus shared with one or more of a focal's neighbors was available from two familiar (current neighbors) individuals and at least one unfamiliar (deceased) individual (discussed below).

All focal birds were of known age, ranging from three to five years after hatching at the time of playback. A total of 23 individuals were chosen as focal males for playback trials; 9 males from lek 5; 9 males from lek 6; and 5 males from lek 7.

3.3.2.2 Experimental setup

Although many studies test for neighbor-neighbor discrimination by comparing focal individual responses to playback of a neighbor from correct and incorrect territorial boundaries, they are unable to control for potential influences of 1) differences in the competitive history between a focal individual and his different neighbors, 2) differences in the value placed on particular territorial boundaries, and 3) the distance of a focal individual to different territory boundaries that could affect stimulus attenuation and response times (discussed in Wilson and Vehrencamp 2001). To control for these potential effects in my study, I held the location of the playback speaker constant: on the focal male's side of the boundary between the focal male and a randomly chosen neighbor (hereafter, the focal neighbor). Figure 3.1 provides a graphical

representation of an example playback trial setup, and illustrates the physical locations of the playback speaker and the territories of the focal male and of all males used to generate the stimulus set.

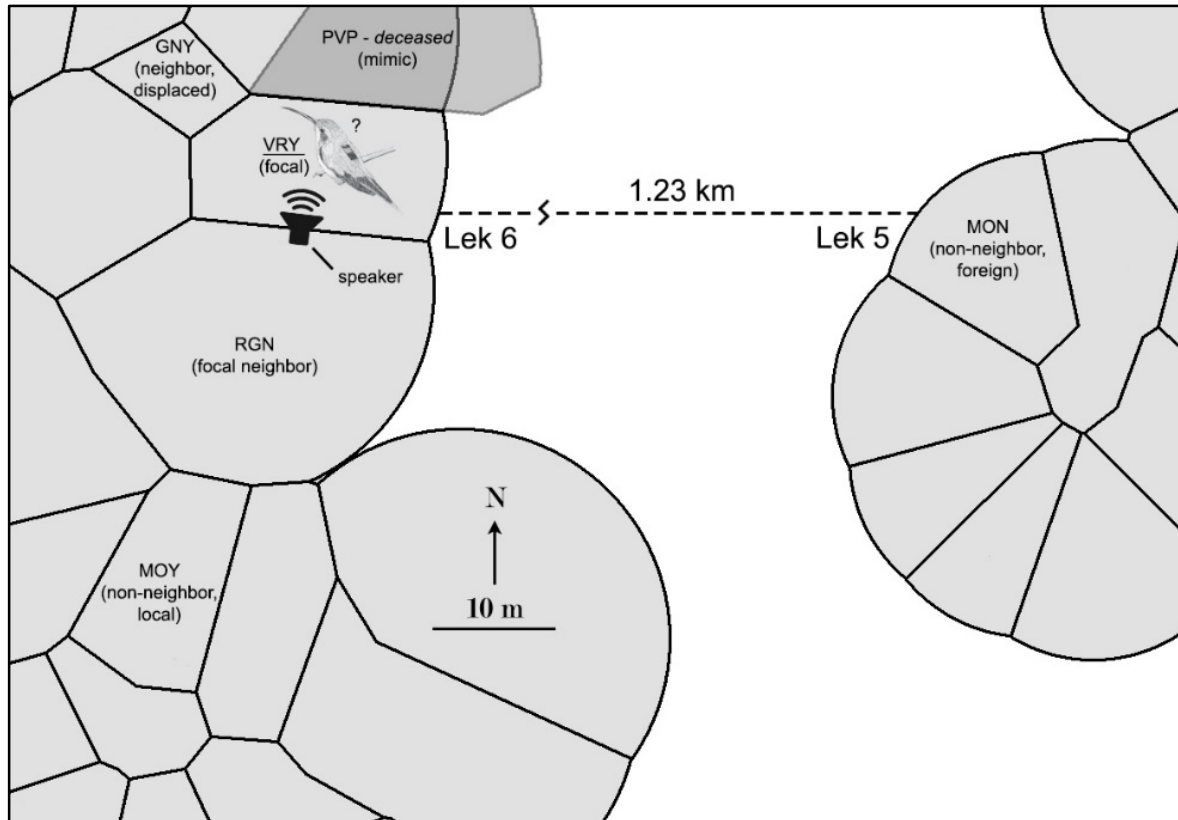


Figure 3.1. An example schematic overview of an experimental playback setup (i.e. the location of the playback speaker with respect to territorial boundary between the focal male and focal neighbor), and the territorial locations of the five song stimuli presented to the focal bird. Note that the scale bar applies only within the two leks depicted.

To avoid interference from neighbors during playback the playback speaker (XF-120-364, 10W, 4 Ohm, full range speaker, MCM Electronics) was placed inside a camouflaged parabolic baffle designed to direct sound towards the focal male and limit sound propagation to other individuals. Playback amplitude was standardized to match the average amplitude of singing little hermits (73.3 ± 5.3 dB at 1 m, $N=10$). In order to avoid a response due to startling the amplitude of the first three songs of the playback ramped from 33% to 66% to 100% of final playback volume, respectively. An artificial model of a male little hermit was placed 0.5 m in

front of the speaker to allow the focal individual the opportunity to attack or begin displaying upon approaching the speaker.

Each of the 23 focal males was presented with five different song stimulus treatments. (1) the focal neighbor's song played from a shared boundary with the focal (focal neighbor stimulus), (2) a different neighbor's song played from the same (i.e. unshared) boundary (dislocated neighbor stimulus), (3) an unfamiliar bird with a dialect shared with the focal neighbor (mimic stimulus), (4) a non-neighbor with an unshared dialect (local non-neighbor stimulus), and (5) a member of a different lek with an unshared dialect (foreign non-neighbor stimulus). Figure 3.2 shows the spectrograms of an example set of the five stimuli presented to one bird.

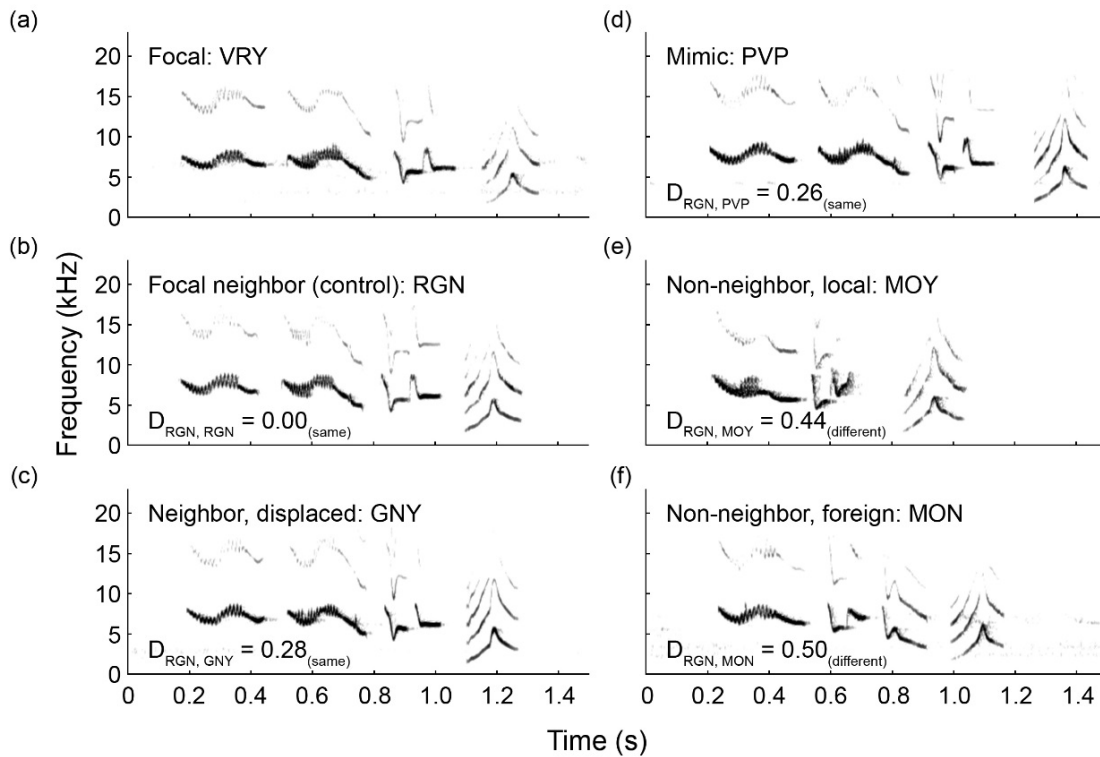


Figure 3.2. Spectrograms of a single song from an example focal bird and from males chosen to generate the five different stimuli. Letters refer to male IDs. In the lower left-hand corner of each panel the pair-wise song dissimilarity score between the focal neighbor and the stimulus is provided, along with the binary designation of whether the two songs shared a dialect (same) or not (different) (see Methods). The focal bird and stimuli in this figure match those of Figure 3.1.

Previous studies seeking to identify the link between recognition and song sharing have generated ‘mimic’ stimuli from purportedly unfamiliar birds by either selecting songs from a distant non-neighboring individual (Wilson and Vehrencamp 2001), or generating a ‘chimeric song’ by combining elements from one or more non-neighbors with those from a familiar bird (Molles and Vehrencamp 2001; Briefer et al. 2008b). Interpretation of responses to these stimuli can be complicated by 1) the possibility (albeit low) that the non-neighboring individuals chosen are truly unfamiliar given their concurrent existence in the population, and 2) the potential for subjects to perceive as unusual a stimulus containing the voices of multiple birds. The generation of a stimulus from a bird that shares the song of the focal neighbor but that died before the arrival of the focal into the social group ensures that there is no chance of familiarity between the focal and the mimic bird and provides a stimulus that does not differ from other stimuli in the number of constituent voices.

Song playback experiments were conducted between the 8th and 19th of May, 2014, a period when juveniles begin to settle on leks and prolonged male-male chases are common (personal observation). Trials were presented to each male on consecutive days from the same location and at approximately the same time between 0700 and 1600. Treatment order was balanced across subjects. To prevent habituation of focal birds to consecutive playbacks delivered to other focal individuals, all consecutive trials within a day were conducted on birds that were not in auditory contact with playback stimuli targeted at other focal males.

3.3.2.3 Stimulus selection and generation

Playback files were generated from digital recordings made at close range (< 1 m), which produced stimuli with high signal-to-noise ratios, no reverberation, and no detectable overlap with other animals (including other lekking little hermits). To minimize pseudoreplication issues

(McGregor 2000) each focal male received a unique playback stimulus for all treatments, whenever possible (of 112 playback trials 88 unique recordings were used). Each stimulus was standardized to a length of one minute, as the majority of responses to playback in preliminary trials occurred within the first few seconds if the bird responded at all. The playback stimuli were not altered to standardize song delivery rates, within-individual variation in song syntax, or total number of songs per playback because these features could carry information used in individual recognition, however, the one minute section of each source recording was chosen so that song delivery rates most closely approximated the mean song delivery rate of a randomly chosen subset of lekking males (i.e. 27.1 ± 8.9 songs/min, $N=36$). To minimize the effects of low-frequency environmental noise each audio file was highpass filtered (Butterworth type) at 1 kHz. The peak amplitude of all sound files was standardized to the same value to simplify the standardization of broadcast volume in the field.

I measured dialect sharing and a continuous measure of song dissimilarity between the focal neighbor and potential playback stimuli by performing an analysis of dialect structure on each lek, using the methodology described in detail in Chapter 2 (see Figures 3.2 and 3.3). Briefly, this procedure uses a sequence alignment algorithm known as dynamic time warping to compare song elements between birds, and to generate a matrix of song distances. An agglomerative hierarchical clustering approach is then used on this distance matrix to identify natural groups of individuals (i.e. those sharing the same dialect). In this way, I was able to identify dialect memberships and pairwise song dissimilarities for all individuals on each of the three experimental leks in 2014, and therefore also for playback stimuli involving males from those leks. For the playback treatments involving stimuli of males from a year before 2014 (mimic stimulus), or from outside the focal male's lek (foreign non-neighbor stimulus), I

repeated this basic dialect assignment process, again including all individuals from the focal male's lek in 2014, but also including the songs of single candidate males from other years or other leks. This procedure was used to select stimuli from shared dialects with the focal neighbor for treatments one through three and from an unshared dialect for treatments four and five.

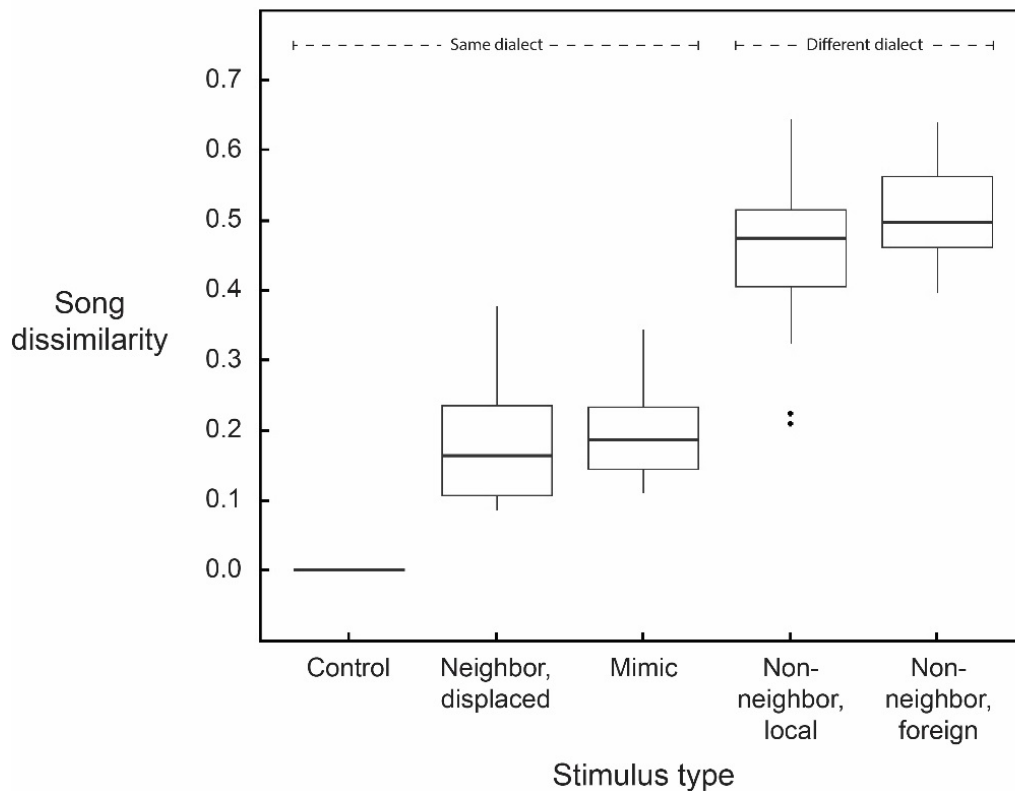


Figure 3.3. The relationship between stimulus type and the numeric score of pairwise similarity and dialect membership. The control treatment involved the comparison of the playback stimulus (the focal neighbor) to itself, therefore all song dissimilarity measures for this treatment are 0.

3.3.2.4 Trial procedure

Playback trials were conducted only during periods of low wind and no rain, and were not started until all of the following pre-trial conditions were met: (1) at least five minutes had passed since all equipment had been set up and the observer was settled behind a blind, (2) the focal bird was present in the territory and singing, and (3) the focal neighbor was either not present on his territory or not singing. Once all three of these conditions were met, a two-minute

pre-playback observation period began in which the following variables were scored for the focal male's behavior: (1) time singing during the period, (2) closest approach distance to the speaker, (3) percent time spent moving through territory. These variables were measured so that baseline levels of singing, movement, and tendency to approach the playback area of the territory could be controlled for in the statistical analysis of the bird's response to the playback stimulus. If, at the end of this two-minute period, the focal male had left the territory, or moved to a location where its presence on the territory could not be confirmed the trial was aborted and restarted when the three pre-trial conditions had again been met. If the focal remained on the territory a one-minute period of song playback began. During preliminary trials, a focal male that responded would typically continue singing upon initiation of playback for some period of time, then stop singing and extend his neck in the direction of the playback before taking flight and slowly inspecting foliage and potential perches between the focal male's original location and the speaker. In most cases, the bird would return to his perch after inspecting the playback area, occasionally displaying briefly for the model before doing so. Based on these preliminary observations I chose to measure the following behavioral variables during the playback period: (1) movement duration, (2) duration singing, (3) duration displaying for the model, (4) duration on the territory, (5) minimum distance to the speaker, (6) latency to move, (7) latency to stop singing, and (8) latency to display. If, during the playback trial, any neighboring bird intruded into the focal's territory and interacted with the focal male, that trial was aborted. The display duration and latency measures were not considered further as displays for the model occurred in only a small proportion (9%) of playback trials. Similarly, the duration spent on the territory was omitted from further analysis since very few birds (8%) left their territories during playback.

3.3.3 Statistical analysis

3.3.3.1 Model construction

It has become standard practice to use variance reduction approaches for the analysis of playback responses due to the high levels of correlation between response variables typical of playback experiments (McGregor 1992). Here, this approach was not appropriate because the magnitudes of correlations between response variables were low ($|r| = 0.11 - 0.36$) for all pairwise comparisons excepting those involving minimum distance from the speaker ($|r| = 0.40 - 0.69$; this variable was removed from the analysis as it was effectively redundant with movement duration), and because techniques such as factor analyses resulted in only one factor with an eigenvalue ≥ 1.0 and which explained only a small proportion of overall variance in the original variables (the first factor explained only 31% of variance among response variables).

Furthermore, I intentionally measured different response variables to test specific aspects of male responses to simulated intrusions. In studies of recognition abilities, it is highly desirable to isolate initial responses (here, whether a decision to respond was made at all, and if so the latency to make that decision) from subsequent responses (here, the duration/intensity of the response itself), as the former can be more directly interpreted as a reflection of the difficulty of discrimination (Palmer et al. 2005; Botero et al. 2007), and the latter is likely to reflect both signal salience and interpretation of the level of threat represented by the stimulus. I therefore chose to analyze each of the four retained response variables (i.e. movement duration, song duration, latency to move, and latency to stop singing) in separate models.

For the two duration-type responses (i.e. time spent moving through the territory, and time spent singing) I used generalized linear mixed models (GLMMs) fit by the Laplace approximation using the R package *lme4* (Bates et al. 2015). Both response scores possessed an excess of zero values, and had non-normally distributed non-zero data, resulting in highly

overdispersed models when using a Poisson error distribution. Also, as my goal in this study was to test for recognition abilities of focal males, I wished to separate the effects of predictors on binary responses (i.e. whether recognition of a potential intruder occurred) from effects on the degree of response (i.e. interpretation of the level of threat a detected intruder may pose). To satisfy this goal and the issue of overdispersion I chose to use zero-altered negative binomial (ZANB) hurdle models, which are robust to overdispersion due to excess zeros, and allow for the separate analysis of the processes leading to response vs. non-response on the one hand and degree of response on the other (Zuur et al. 2009). Thus, each hurdle model was split into two parts: a binary response portion modelled with a binomial error distribution and logit link function, and a non-zero response portion modelled with a negative binomial error distribution.

The two latency responses required a different modelling approach because non-responses cannot be assigned arbitrary values (e.g. the duration of the playback trial) or censored (i.e. discarded) without altering the underlying distribution of the responses. Therefore, for the two latency scores (latency to begin moving and latency to stop singing) I used a survival analysis technique known as a Cox-proportional hazards mixed effects model (COXME) fit by the Laplace approximation, which is specifically designed to deal with this type of data (R package *coxme*, Therneau 2015). Whereas the hurdle models of duration responses allow for the analysis of whether an individual detected a simulated intruder at all, and if so, how strong a threat this intrusion represented, the separate analysis of response latencies enables assessment of the amount of time required before recognition (or the decision to respond) occurred.

In both the duration-type response and latency-type response global models the predictors included playback treatment, focal male age, and order of playback treatment as a categorical factor, with male ID nested within lek as random intercepts. Random slopes could not be

included due to model convergence issues with these models. To control for the pre-trial baseline behavior of each response type of interest each global model was fit with the pre-trial factor matching the response type being analyzed (e.g. pre-trial song duration was included for the song duration and latency models).

The inclusion of the categorical treatment factor for each model tests whether any of the six response measures of the focal male's behavior can be predicted from the relationship between the focal neighbor and the playback stimulus male. To assess whether the primary effect of this categorical playback treatment on male responses was due to the differences between the focal neighbor and playback stimulus in fine-scale acoustic structure of the songs, I reran each model, replacing the categorical treatment factor with a continuous measure of song dissimilarity between the focal neighbor and the playback stimulus (see the stimulus selection and generation section above). In the context of the experimental setup of this study, this factor is a continuous measure of the accuracy of mimicry of the focal neighbor's song by an intruding individual (i.e. the playback song). To these models I also added a covariate that measured the average song dissimilarity between the focal neighbor and all of the focal's other neighbors (hereafter, neighbor song dispersion), since previous studies have demonstrated that the degree of song sharing between neighbors can influence an individual's ability to detect acoustic differences in individuals recruiting to a breeding population (McGregor and Avery 1986; McGregor 1989). To assess whether models containing these two factors explain a greater proportion of variation in focal male responses than the categorical treatment models, I examined the evidence ratio for each pair of models (i.e. song dissimilarity versus categorical treatment) for each response type (Anderson and Burnham 2002; Burnham and Anderson 2002).

3.3.3.2 Model selection and averaging

I evaluated model support and effect sizes using an information theoretic approach (Burnham and Anderson 2002). Each global model was first evaluated for convergence, and when a model failed to converge (3 of the 12 models) I repeated the model fitting procedure using seven different optimizers to estimate the variance-covariance matrix of the random effects. I then assessed the agreement between coefficients calculated from these different optimizers and considered convergence warnings to be false positives when agreement among them was extremely similar (recommended by Bates et al. 2015). In all cases of non-convergence, inter-optimizer parameter agreement was high. I then standardized the predictors of all global models ($\mu = 0$, $sd = 0.5$ for each predictor) using the *standardize* function of the R package *arm* for the ZANB hurdle models, and manually for the COXME models, so that effect sizes were comparable between predictors (and models) and to allow for model averaging (see below; Schielzeth 2010; Grueber et al. 2011; Gelman and Su 2015). Next, to identify those factors that had the strongest support across all plausible models, I specified sets of submodels by generating all permutations of global model factor combinations, excepting interaction terms, using the *dredge* function of the R package *MuMIn* (Bartón 2015). I used this approach because I was particularly interested in estimating the relative importance of different factors hypothesized to be of potential ecological significance, but whose relative importance was unpredictable (Dochtermann and Jenkins 2011; see Duval and Kapoor 2015 for an example). All model comparison sets included a null model containing only the random effect. Last, for a given model set, I identified all candidate submodels within 2 AICc units of the best model (i.e. the model with the lowest AICc; Burnham and Anderson 2002) and averaged these submodels into a final model using the natural averages approach (so that factors with small but significant effects in some models could be retained in the final model).

I evaluated the fit of each final model in several ways. Influential outliers were identified with Cooks D (using the *influence* function) in the *influence.ME* R package. I evaluated overdispersion values for all 12 global models; none deviated significantly from 1. Residuals were evaluated visually to determine whether they deviated from normality for GLMMs. The assumption of proportional hazards for latency models was assessed visually from Kaplan-Meier plots. For compatibility with frequentist statistical inference, coefficient 95% confidence intervals that did not overlap with 0 are presented in boldface in tables and are “significant” at an alpha of 0.05. In evaluating support for factors in final averaged models, those that did not include a certain factor were interpreted as lacking support for that factor as an important contributor to the response (Grueber et al. 2011). To compare support for models containing the categorical treatment factor versus the continuous song dissimilarity factor I used evidence ratios based on the average of all pairwise comparisons between individual component models in the former versus latter model sets.

3.4 Results

3.4.1 Binary responses

I found strong support in the final model (and all component models) for an effect of categorical treatment on focal individuals’ likelihood to respond to playback by approaching the speaker (i.e. the movement response; Tables 3.1 and 3.2, Figure 3.4(a)). Focal individuals were significantly more likely to respond, by approaching the speaker, to playback of non-neighbors (both from the same lek and different leks) compared to the control treatment (i.e. playback of the focal neighbor’s song). These two playback treatments were also the only two in which dialects differed from the focal neighbor’s dialect. Responses to both the neighbor and mimic

treatments were slightly elevated with respect to controls, but 95% CIs overlapped 0 in both cases, suggesting little support for the importance of these trends. Releveling the treatment factor revealed no significant differences between any other pairs of treatment levels. Although focal male age was included in the final averaged model, there was little support for an effect of age on binary movement response, and had similarly negligible effects in all other models presented below.

Table 3.1. Component models used to generate the final model-averaged estimates in Table 3.2. The variables retained in various models are treatment (T), focal male age (A), movement during the pre-trial phase (M), trial order (O), dispersion between the focal neighbor and all the other focal's neighbors (D). df, degrees of freedom, $\log(L)$, log likelihood, Δ_i , AICc differences to top model, w_i , Akaike weight. Further details are included in Methods.

Type	Treatment = stimulus type (categorical)						Treatment = song dissimilarity (continuous)					
	Parameters	df	log(L)	AICc	Δ_i	w_i	Parameters	df	log(L)	AICc	Δ_i	w_i
Response (Y/N)	T + A	8	-67.48	152.35	0	0.56	T	4	-67.89	144.16	0	0.32
	T	7	-68.87	152.82	0.47	0.44	T + A	5	-66.82	144.22	0.05	0.31
Duration (>0)							D + A + T	6	-66.04	144.89	0.72	0.22
							D + T	5	-67.50	145.57	1.40	0.16
	M	5	-231.07	473.34	0	0.34	D + T	6	-226.11	465.94	0	0.57
Latency	(Null)	4	-232.32	473.42	0.08	0.33	D + M + T	7	-225.10	466.54	0.60	0.43
	T	8	-227.66	474.39	1.05	0.20						
	O	8	-228.09	475.24	1.90	0.13						
Latency	O + T	18	-226.12	498.00	0	0.68	D + O + T	16	-225.06	488.89	0	0.35
	A + O + T	19	-226.12	499.49	1.48	0.32	O + T	16	-224.90	489.06	0.17	0.32
							D + A + O + T	16	-225.12	490.22	1.33	0.18
						A + O + T	17	-224.91	490.48	1.59	0.16	

Table 3.2. Standardized model averaged predictor coefficients for all six movement response models. RI, relative importance; R^2 , likelihood ratio R^2 of Bartón (2015); ER, evidence ratio.

Treatment = stimulus type (categorical)		Treatment = song dissimilarity (continuous)													
Type	Parameters	Est.	SE	Z	95% CI	RI	R ²	Parameters	Est.	SE	Z	95% CI	RI	R ²	ER±SE
Response (Y / N)	(Intercept)	-1.42	0.56	2.52	-2.52 – -0.32	-	0.28	(Intercept)	0	0.27	0.01	-0.52 – 0.53	-	0.31	53.69±5.89
	T (Neighbor)	1.10	0.74	1.49	-0.34 – 2.55	1.00		Song distance	1.95	0.5	3.79	0.94 – 2.96	1.00		
	(Mimic)	1.05	0.74	1.43	-0.39 – 2.49	"		Age	-0.85	0.56	1.51	-1.96 – 0.25	0.53		
	(Dialect)	2.25	0.72	3.13	0.84 – 3.66	"		Song dispersion	-0.59	0.55	1.07	-1.66 – 0.49	0.38		
	(Lek)	2.33	0.75	3.10	0.86 – 3.79	"									
	Age	-0.85	0.52	1.64	-1.87 – 0.17	0.56									
Duration (> 0)	(Intercept)	3.02	0.38	8.02	2.28 – 3.76	-	0.21	(Intercept)	3.10	0.10	31.49	2.90 – 3.29	-	0.26	55.60±9.19
	Mov_b4	-0.40	0.23	1.74	-0.85 – 0.05	0.34		Song distance	0.64	0.20	3.17	0.24 – 1.03	1.00		
	T (Neighbor)	-0.05	0.45	0.11	-0.94 – 0.84	0.2		Song dispersion	-0.48	0.20	2.36	-0.88 – -0.08	1.00		
	(Mimic)	0.41	0.44	0.92	-0.46 – 1.27	"		Mov_b4	-0.15	0.23	0.65	-0.79 – 0.10	0.43		
	(Dialect)	0.63	0.39	1.61	-0.13 – 1.39	"									
	(Lek)	0.83	0.39	2.09	0.05 – 1.60	"									
Latency	Order (2)	-0.28	0.28	1.02	-0.83 – 0.26	0.13		Song distance	1.43	0.33	4.38	0.79 – 2.07	1.00		
	(3)	-0.31	0.29	1.07	-0.87 – 0.26	"		Order (2)	-0.52	0.38	1.36	-1.26 – 0.23	1.00		
	(4)	-0.33	0.30	1.10	-0.92 – 0.26	"		(3)	-0.70	0.40	1.73	-1.49 – 0.09	"		
	(5)	-1.20	0.40	3.01	-1.98 – -0.42	"		(4)	-0.80	0.43	1.88	-1.65 – 0.04	"		
	Order (2)	-0.61	0.38	1.60	-1.36 – 0.14	1.00		(5)	-1.79	0.54	3.33	-2.85 – -0.74	"		
	(3)	-0.83	0.42	1.98	-1.64 – -0.01	"		Song dispersion	-0.42	0.39	1.07	-1.19 – 0.35	0.53		
	(4)	-0.93	0.44	2.11	-1.80 – -0.07	"		Age	-0.09	0.42	0.22	-0.93 – 0.74	0.33		
	(5)	-1.88	0.55	3.44	-2.96 – -0.81	"									
	T (Neighbor)	0.46	0.60	0.76	-0.72 – 1.63	1.00									
	(Mimic)	0.80	0.59	1.36	-0.35 – 1.95	"									
	(Dialect)	1.63	0.52	3.11	0.61 – 2.66	"									
	(Lek)	1.68	0.53	3.15	0.63 – 2.72	"									
Age	-0.10	0.41	0.26	-0.90 – 0.69	0.32										

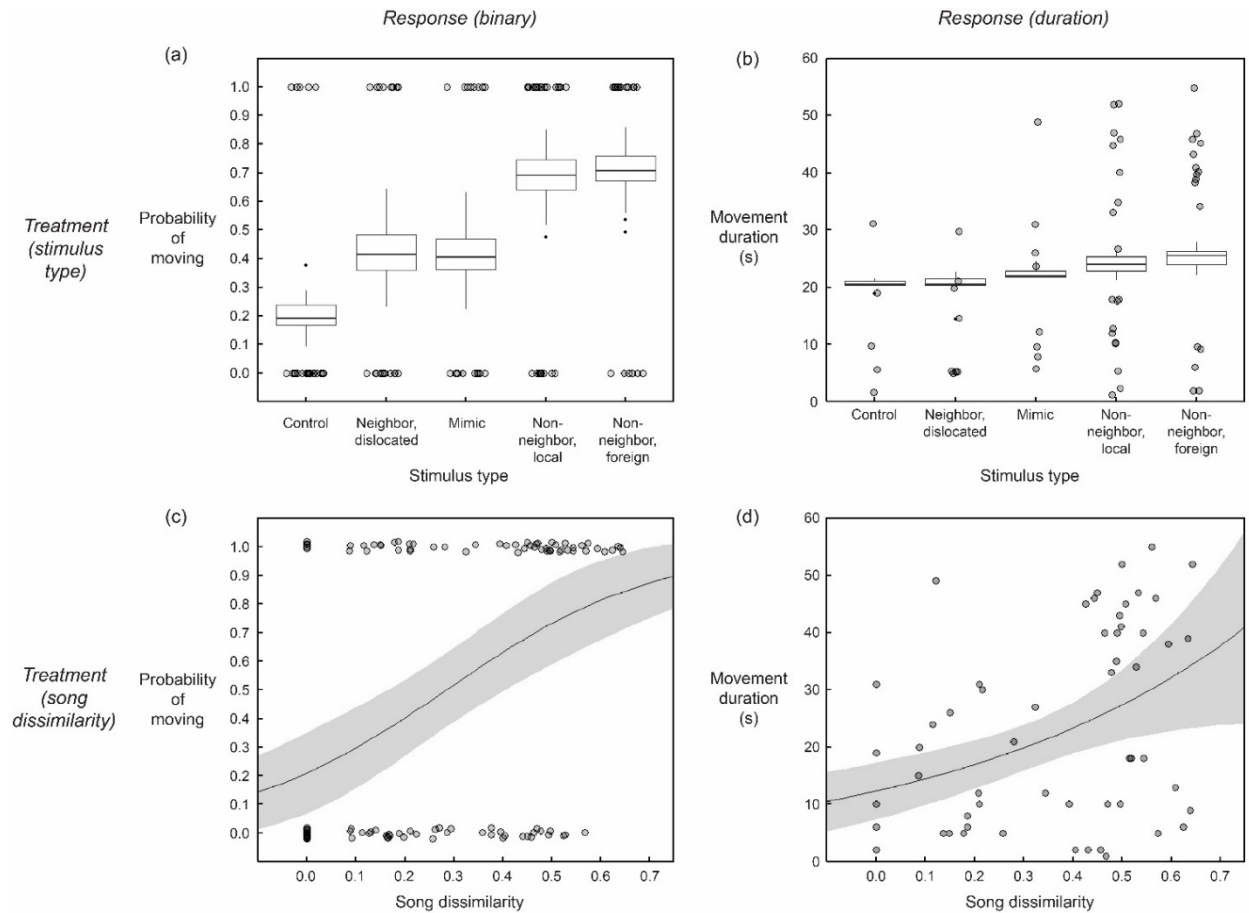


Figure 3.4. Binary and continuous movement responses for both the categorical treatment models and the continuous song dissimilarity models. Both local and foreign non-neighbors were more likely to elicit movement than other stimuli (a), whereas only foreign non-neighbors elicited increased response durations over the control (b). Similarly, stimulus-focal neighbor song dissimilarity strongly predicted the probability of a response (c), and predicted the duration of response to a lesser extent. Box-and-whisker plots (a) and (b) represent back-transformed model predictions, holding all numeric values at their means (i.e. 0 since all predictors are standardized) with raw data overlaid. The two scatter plots (c) and (d) show similarly back-transformed model predictions (solid lines), with 95% confidence intervals (CI) as shaded areas. All raw data points are semi-transparent and are jittered within categories to allow visualization of overlap.

In contrast to the movement response, the binary response to sing was unrelated to any playback treatment levels (Tables 3.3 and 3.4). The only factor that significantly influenced a focal male's likelihood to sing was the duration of singing in the pre-playback period: focal individuals that sang more in this period were more likely to sing during the playback phase.

Table 3.3. Component models used to generate the final model-averaged estimates in Table 3.4. Variable names and metrics follow the naming convention of Table 3.1; also, song duration in the pre-trial phase (V).

Type	Treatment = stimulus type (categorical)						Treatment = song dissimilarity (continuous)					
	Parameters	df	log(L)	AICc	Δ_i	w_i	Parameters	df	log(L)	AICc	Δ_i	w_i
Response (Y/N)												
	O + V	8	-50.48	118.35	0	1.00	O + V	8	-50.48	118.35	0	0.46
							O + T + V	9	-49.73	119.23	0.87	0.30
							T + V	5	-54.52	119.6	1.25	0.25
Duration (>0)												
	V	5	-290.46	591.85	0	1.00	T + V	6	-288.99	591.31	0	0.45
							V	5	-290.40	591.82	0.51	0.35
							T	5	-291.03	592.99	1.69	0.20
Latency												
	V	1	-260.99	524.02	0	0.69	V	1	-260.99	524.02	0	0.28
	A + V	2	-260.77	525.65	1.63	0.31	D + V	2	-260.14	524.40	0.38	0.23
							S + V	2	-260.32	524.75	0.73	0.19
							D + S + V	3	-259.37	524.97	0.95	0.17
							A + V	2	-260.77	525.65	1.63	0.12

Table 3.4. Standardized model averaged predictor coefficients for all six song response models. Metric naming scheme follows Table 3.2.

Type	Treatment = stimulus type (categorical)							Treatment = song dissimilarity (continuous)							
	Parameters	Est.	SE	Z	95% CI	RI	R ²	Parameters	Est.	SE	Z	95% CI	RI	R ²	ER±SE
Response (Y / N)	(Intercept)	0.81	0.55	1.46	-0.24 – 2.00	-	0.51	(Intercept)	0.81	0.51	1.60	-0.18 – 1.80	-	0.48	0.73±0.14
	Sing_b4	3.23	0.74	4.37	2.01 – 5.00	1.00		Sing_b4	3.14	0.74	4.23	1.68 – 4.59	1.00		
	Order (2)	-1.11	0.78	-1.41	-2.75 – 0.37	1.00		Order (2)	-1.11	0.79	1.40	-2.66 – 0.44	0.75		
	(3)	-0.75	0.77	-0.97	-2.33 – 0.75	"		(3)	-0.77	0.79	0.98	-2.32 – 0.77	"		
	(4)	0.61	0.77	0.80	-0.87 – 2.20	"		(4)	0.55	0.78	0.71	-0.98 – 2.09	"		
	(5)	1.46	0.93	1.57	-0.27 – 3.49	"		(5)	1.42	0.95	1.50	-0.44 – 3.27	"		
Duration (> 0)							0.20	Song distance	-0.75	0.55	1.37	-1.81 – 0.32	0.54		
	(Intercept)	3.11	0.13	24.82	2.84 – 3.36	-		(Intercept)	3.10	0.10	29.63	2.89 – 3.30	-	0.10	1.17±0.30
	Sing_b4	0.46	0.21	2.18	0.03 – 0.87	1.00		Song distance	-0.34	0.12	2.90	-0.58 – -0.11	0.80		
								Sing_b4	0.45	0.15	3.07	0.16 – 0.74	0.7		
Latency							0.23							0.21	1.17±0.19
	Sing_b4	1.24	0.30	4.20	0.66 – 1.82	1.00		Sing_b4	1.28	0.30	4.25	0.69 – 1.87	1.00		
	Age	0.17	0.25	0.68	-0.32 – 0.66	0.31		Song dispersion	-0.34	0.26	1.29	-0.86 – 0.18	0.40		
								Song distance	0.30	0.25	1.20	-0.19 – 0.79	0.37		
								Age	0.17	0.25	0.68	-0.32 – 0.66	0.12		

3.4.2 Duration responses

There was modest support for an effect of playback treatment on movement duration; only the foreign non-neighbor treatment had a significant effect (over the control response) on the response duration [Figure 3.4(b)]. Support for the importance of this effect was limited as the treatment factor had a relative importance of 0.2, and the candidate model set contained the null model as the second most highly-ranked model based on AICc [Table 3.1]. There was also modest support for the effect of trial order on movement response duration, with the final trial (5) having a moderate negative effect.

Similar to the binary song response model, there was no evidence of an effect of playback treatment on song duration in the playback phase. There was a mild, though significant, effect of time spent singing in the pre-trial period on time spent singing during the playback phase [Table 3.4].

3.4.3 Latency responses

Both of the non-neighbor treatments (local and foreign) had strong effects on the focal male's latency to begin to approach the playback speaker [Tables 3.1 and 3.2, Figure 3.5]. Here, latency coefficients represent the log change in the hazard function per unit change in standard deviation of the predictor variable (or change of level for a categorical factor), but when exponentiated (base e) can be interpreted directly as the multiplicative effect on the baseline rate of response. I have therefore presented the coefficients in the original scale in the data tables, but present the multiplicative effect in the text for interpretability. The non-neighbor, same lek and non-neighbor, different lek treatments evoked a 5.10x and 5.36x greater rate of initiation of approach over the control treatment, respectively. The neighbor and mimic treatments, for comparison, showed 1.58x and 2.22x greater rates of response, though neither of these rates were

significantly different from 1. In addition, trial order had strong negative effects on response rate, with the 5th presentation of a playback stimulus (across treatment types) reducing the rate of response to 0.15x that of the 1st stimulus presentation.

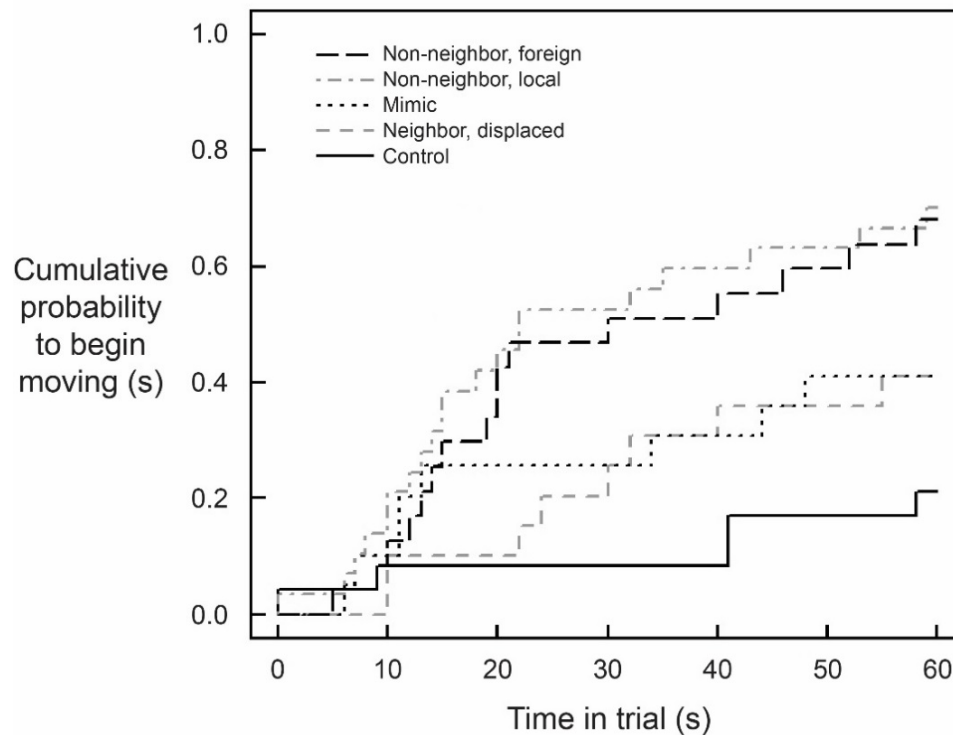


Figure 3.5. Kaplan-Meier cumulative failure plot for the latency of focal males to begin moving in response to the initiation of stimulus playback. Each line represents the cumulative percentage of subjects that had responded by that time point in the trial, for each of the five playback stimuli. The two non-neighbor stimuli had 95% CI that did not overlap with the control stimulus, whereas the displaced neighbor and mimic overlapped the control.

As in all other song response measures, only time singing in the pre-playback phase was positively related to the rate at which focal males stopped singing in response to the playback: a 1 SD increase in song duration in the pre-playback phase resulted in a 3.46x increase in the rate at which the bird stopped singing.

3.4.4 Song similarity and dispersion

In all cases where at least one treatment level had a significant effect on male response, the matching model replacing categorical treatment with focal neighbor – playback song

dissimilarity exhibited significant effects of this factor on response [Table 3.2, Figure 3.4(c) and (d)]. In only one case was song dissimilarity a significant factor where treatment effects were not or were not included in the paired model: that of the effect of song dissimilarity on the song duration response measure. In this case an increase in dissimilarity between focal neighbor and playback (i.e. a worse mimic) resulted in a shorter total duration of song during playback. Although the null model was not among the candidate models that were averaged in the final model, the R^2 for this model was extremely low (0.10; Table 3.4) indicating low overall explanatory power over the null model. The pairwise comparison of evidence ratios between models including continuous song dissimilarity and dispersion versus those including the categorical treatment revealed that, for all movement-type response models, the former factors resulted in much better fits than the categorical treatment models. In contrast, the evidence ratios for song response models involving continuous measures of song dissimilarity and dispersion for playback stimuli versus the categorical treatment models were either indistinguishable from 1 (indicating equal support of the two models) or less than 1 (indicating slightly stronger support for the model of categorical treatment).

In only one of the six models including neighbor song dispersion did this factor have a significant effect on focal male response; as the song dispersion of a focal male's neighbors increased the movement duration in response to playback decreased [Figure 3.6]. Although the upper bound of the coefficient 95% CI was close to 0, indicating tenuous support for this effect, song dispersion had equal relative importance to song dissimilarity in this model, and was a factor in both of the component models of the final averaged model. Note that whereas the matching model including treatment as a categorical factor included the null model as one of its

components, the song dissimilarity/dispersion model did not, and had a large evidence ratio, indicating greater support [Table 3.2].

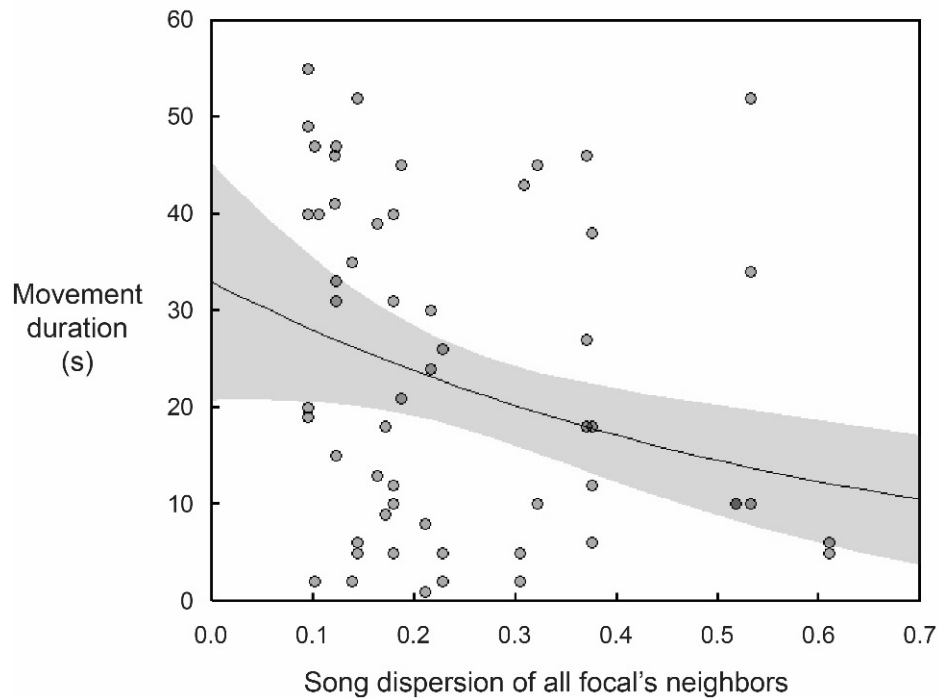


Figure 3.6. Neighbor song dispersion negatively predicted the duration of movement in response to the playback. The solid line represents the partial effect of neighbor song dispersion on the response variable (back-transformed from model predictions using the final averaged model), with 95% CI as shaded areas, and raw data overlaid as semi-transparent points.

3.5 Discussion

The results of this study provide support for all three predictions of the deceptive mimicry hypothesis for song sharing in little hermits:

1) Territorial males exhibit the ‘dear enemy’ effect on leks

Males responding to simulated intrusions (i.e. song playbacks) were markedly less likely to respond and responded more slowly to playback of familiar birds (neighbors) than to definitively unfamiliar birds (foreign non-neighbors), regardless of whether stimuli were played in the “correct” location for neighbors. The degree of aggression directed at the “intruder”, once a male responded, was also moderately elevated for foreign intruders over local intruders. All of

these responses suggest that there may be benefits (i.e. reduced likelihood, frequency, and intensity of aggression) to effective mimicry of local males.

2) Males are unable to recognize individuals based on their songs alone

The ability for territorial males to recognize the identity of established neighbors based upon individually distinctive voicing of songs, or consistent differences in syllable structure or syntax would negate any potential for a mimic to benefit from the ‘dear enemy’ effect, since mimics would immediately be recognized as such and treated as foreign birds. The finding that focal males responded similarly to the focal neighbor playback (where the stimulus matched the identity of the neighbor in the playback location) and to the displaced neighbor (where the stimulus mismatched the focal neighbor’s identity) suggests that either territorial males are incapable of individual recognition, or that they are capable of distinguishing among individuals but consider all familiar individuals to represent an equally low threat. This latter interpretation is unlikely, since the mimic stimulus would be attributed to an unfamiliar individual, and responded to aggressively; instead the response to this stimulus was indistinguishable from that of the displaced neighbor.

A final alternative scenario must be considered: one in which individual recognition occurs, but any individual (familiar or unfamiliar) with a shared song is considered to be a lower threat than a bird with an unshared song. Although this interpretation cannot be ruled out entirely, I consider it to be a less likely explanation because territorial takeovers frequently involve immediate neighbors that share songs with the usurped individual, indicating that territorial males capable of individual recognition should consider displaced neighbors to be as large a threat as an unknown individual (Temeles 1994); under this scenario, the displaced neighbor stimulus should elicit a markedly higher response than the focal neighbor stimulus,

however, all responses between these two stimulus types were indistinguishable. Therefore, the more parsimonious explanation appears to be that lekking little hermits are incapable of vocally-based individual recognition.

Interestingly, my results suggest that even in the absence of individual recognition males may distinguish local individuals with unshared songs from foreign individuals (i.e. from a different lek) and consider them to be less of a threat as evidenced by reduced movement response durations to the former stimulus compared to the latter. This suggests the possibility of familiarity with local song dialects despite the fact that local dialects are not more similar to one another than foreign dialects. Discrimination between local and foreign non-neighbors has been identified only in skylarks (Briefer et al. 2010), and suggests that there may be selection for individuals to attend to recruitment dynamics on the whole lek rather than just to that of their nearest neighbors; a possibility that deserves further attention.

3) Song sharing inhibits recognition of unfamiliar individuals as foreign intruders

The mimic stimulus, which effectively simulates intrusion of an unfamiliar bird that has copied the focal neighbor's song, received responses from focal birds that were indistinguishable from responses to familiar neighbors, providing strong support that shared vocal signatures inhibit recognition of intruders. It should be noted that the mean and variation of the accuracy of simulated song copying (measured as song dissimilarity) for the mimic treatment was indistinguishable from the dislocated neighbor stimulus (see Figure 3.3); indicating that low responses to the mimic stimulus were not due to unnaturally high levels of song similarity. Furthermore, males appeared to be sensitive to the accuracy of mimicry; the superiority of models treating song sharing as a continuous measure of song dissimilarity over those treating song as a categorical factor suggests that males evaluate the degree of similarity between an

intruder's songs and those of a resident, rather than rely solely upon a group signature (i.e. the song dialect).

Although males may pay attention to characteristics of the whole song when evaluating similarity, it is possible that particular features (e.g. syntax, certain syllable types, etc.) function as indicators of local song, and will be a focus of my future work in this system. Also, future work will be directed toward determining whether males attempting to gain access to a lek territory actually benefit from deception; i.e. whether more accurate mimics have elevated recruitment success and survival rates, and enjoy reduced aggression from neighbors.

3.5.1 Movement versus song responses

Contrary to initial expectations, only physical responses (i.e. responses involving movement toward the playback speaker, and inspection of the focal male's territory) related to the playback stimulus; vocal responses (i.e. the probability of singing, latency to stop singing, and duration of song) did not differ across different playback stimuli. In considering the purported function of song in this species, this result is unsurprising: whereas movement towards a potential rival represents willingness to engage in confrontation, often involving displays, chases, and physical fights, song is rarely used in confrontations and instead appears to be used in a manner consistent with long-range attraction of females. Therefore movement responses are likely to reflect recognition and estimation of the level of threat of intruders, whereas song appears to be unrelated to intrusion. For the only case in which treatment was a significant predictor of song behavior the relationship was negative; an indication that the focal male reduced vocal behavior to engage in more directed physical responses (e.g. inspection, display, and potentially attack).

3.5.2 Effects of song dispersion / cohesion on detection of intruders

I predicted that the average degree of song dissimilarity among an individual's neighbors (here, neighbor song dispersion) might negatively impact that individual's ability to recognize closely matched songs of a simulated intruder. An effect of this type could be due to two distinct processes: 1) the number of renditions of shared types among an individual's neighbors could negatively impact an individual's ability to recognize new versions of that type: an effect that has been attributed to pro-active memory interference in great tits, *Parus major* (McGregor and Avery 1986; McGregor 1989), and 2) as the songs of an individual's neighbors become more divergent, the more generalized must be the set of rules classifying all neighbors' songs as local, consequently increasing the likelihood of misclassification of a non-local mimic's song as local (analogous to arguments about the role of model phenotypic variance in the detectability of a batesian mimic; Oaten et al. 1975). Contrary to predictions, I found no relationship between neighbor song dispersion and response probability (binomial movement and song models) or response latency (movement and song latency models). One possible reason for this is that neighbor song dispersion levels may always be sufficiently low to prevent confusion in focal birds. This may be a distinct possibility as only three individuals in the playback experiment possessed neighbors whose songs were classified as belonging to a different dialect.

The predicted causal relationship between neighbor song dispersion and male response to simulated intruders also posits that effects should be greatest on the probability and difficulty of detection, but is unlikely to affect subsequent responses since these necessarily occur after any hurdle to recognition has been overcome. What I found in this study was precisely the opposite: although I found no evidence of an effect of song dispersion on detection probability (binary responses) or difficulty (latency to respond), I did find such a link in the continuous movement

response model. This result might indicate a reversal of the causal role of dispersion and response: males that direct less aggression towards intruders may be more likely to allow settlement of low-similarity singers than a more aggressive male might, thus creating a more acoustically dispersed song neighborhood.

3.5.3 Deception by intruders or a password of group membership?

An established hypothesis for the formation and maintenance of shared songs (i.e. dialects) in group living animals is that they may serve as “passwords” to help identify non-local intruders which might pose a threat to existing territory holders (Feekes 1977). Although this hypothesis makes many of the same predictions as the deceptive mimicry hypothesis the focus of this hypothesis is on selective pressures among existing group members to converge in their signals thus making the detection of intruders easier, as opposed to selection operating on intruders to deceive territorial males. Clearly these two hypotheses are related, and a possible outcome of a process like deceptive mimicry might be an arms race between territorial residents and intruders in which the former sing songs that represent difficult-to-fake passwords, and the latter develop increasingly accurate song copying strategies.

Several lines of evidence suggest that the password effect is unlikely in little hermits. First, a critical prediction of the password hypothesis is that neighbor- or colony-level song cohesion is expected to strongly influence the recognizability of intruders. As the results of my experimental playbacks demonstrate, cohesion/dispersion appears to be causally unrelated to intruder detection probability or difficulty. As stated previously, it is nevertheless possible that selection for song sharing has resulted in such unvaryingly high levels of song consistency among neighbors that there is no naturally observable relationship with detection rates despite strong selection. Second, since open-ended learning occurs in this species up to the second year

of life (Kapoor, unpublished data, this species; Araya-Salas and Wright 2013, in closely-related long-tailed hermits, *Phaethornis longirostris*) territory holders would be expected to track and match changes in the songs of neighbors, however, evidence of group level convergence is lacking.

3.5.4 Could deceptive mimicry be common on leks?

The past several decades of research on the adaptive significance of song sharing in birds has found general support for the conclusions that 1) song sharing only rarely hinders recognition, and when it does, does so only weakly or temporarily, and that 2) there is little, if any support for the conclusion that deceptive mimicry plays an important role in song sharing species. Below, I discuss why this conclusion may be a reflection of the omission of lek mating systems from research on this topic, and expand upon why certain species with lek mating systems and vocal learning may be particularly likely to provide support for the deceptive mimicry hypothesis.

3.5.4.1 There is high opportunity for deceptive mimicry to be possible on leks

An important consideration for the evolution of deceptive mimicry on leks is that the theory is plausible only for species with vocal learning capabilities. Aoki (1989) argues that the strong sexual selection typical of lek mating systems may have led to the evolution of imitative learning in a polygynous common ancestor to the oscines. Although Aoki did not model a specific link between imitative learning and intrasexual deception, and his model requires vertical transmission of song from father to son (which is implausible in most polygynous - especially lekking - species), the general argument that strong sexual selection may lead to imitative vocal behavior that maximizes mating success, an important precursor to deceptive mimicry, appears to be sound. Oddly, few studies of vocal learning in lekking species have been

conducted, however, evidence (ranging from suggestive to definitive) of imitative learning exists for suboscine screaming pihas, *Lipaugus vociferans* (Fitzsimmons et al. 2008), neotropical bellbirds, genus *Procnias* (Saranathan et al. 2007; Kroodsma et al. 2013), long-tailed manakins (Trainer et al. 2002), and for oscine golden bowerbirds, *Prionodura newtonia* (Westcott and Kroon 2002) and lyrebirds, genus *Menuridae* (Robinson and Curtis 1996), and many lekking hummingbirds (Snow 1968; Wiley 1971; Stiles and Wolf 1979; González and Ornelas 2009; Araya-Salas and Wright 2013).

Another consideration for the plausibility of deceptive mimicry on leks is whether the use of song in communication on leks might constrain its use in deception. A number of lekking species possess large vocal repertoires (e.g. the lyrebirds, Robinson and Curtis 1996) that could limit the use of song in mimicry. However, there are many species of lekking birds that possess relatively simple vocalizations, including many of the species noted above (i.e. pihas, bellbirds, manakins, hermit hummingbirds (see previous references for these species)), making these species especially good candidates for the investigation of deceptive mimicry. Even in species with large repertoires it may be possible to engage in deceptive mimicry if an individual can refrain from using an unshared song type under the correct circumstances (see Cockburn et al. 2009 for an example).

The constraint that vocal learning must occur post-dispersal in deceptive mimics is less likely to be an issue in lek mating systems since lekking vocal learners never hear conspecific song until they begin to arrive on leks; females raise offspring alone and away from leks (Höglund and Alatalo 1995). It is plausible, though not yet substantiated, that most lekking species that learn their vocalizations do so by listening to individuals near the point of future settlement. My observations in little hermits, suggest that this is the case; young males generally

spend several weeks sitting quietly near lek residents (presumably memorizing songs), then disappear for several more weeks as they practice their songs alone, and finally return to the same area once their songs have begun to crystallize and attempt to establish a territory (Kapoor, unpublished data). Interestingly, in little hermits, some individuals begin to sing on leks when their songs are still moderately variable, suggesting the possibility of a tradeoff between the timing of recruitment and the perfection of vocal mimicry.

3.5.4.2 Fitness benefits of mimicry are likely high on leks

Deceptive mimicry is more likely to evolve in species in which access to a breeding territory is severely limited (and therefore a dominant component of fitness), and where the risks associated with territorial aggression are high; both conditions exemplified by leks. In little hermits, as in many other lek mating systems, mating is unlikely to occur off leks, making territorial admittance to leks a large component of male fitness (Höglund and Alatalo 1995; DuVal and Kempenaers 2008). In this species, male-male aggression on leks is often extremely energetically costly and potentially risky, since fights between males can involve lengthy chases, bill fencing that can result in severe injuries as birds attempt to stab each other, and can disrupt courtship. In several cases I have observed prolonged fights in which male-male aggression continued unabated for weeks, and likely resulted in the death of one of the rivals. Competitive ‘challenges’ among neighbors are also common in lekking hermits, suggesting that birds are frequently attending to activities on neighbors’ territories and vigilant for potential intruders (Stiles and Wolf 1979); in several cases, playback resulted in responses within 1 – 2 s of the initiation of playback.

Perhaps the most convincing reason that deceptive mimicry may be likely to occur on leks, is that a dominant explanation for the evolution of male aggregation on leks is a hypothesis

invoking deception: the hotshot hypothesis (Beehler and Foster 1988). This hypothesis, posits that peripheral or subordinate males gain fitness by deceiving females about their dominance status, suggesting that deception, albeit directed at females, may be common in lekking species. Although the specific mechanisms of this deception have yet to be definitively identified, Cockburn et al. (2009) have found suggestive evidence of the use of deceptive vocal mimicry. In polygynous superb fairy-wrens subordinate males gather into ‘hidden leks’ (see Wagner 1998) in which they gain fitness by advertising in close proximity to dominants and limiting their songs to those shared with dominant individuals, which may lure females attempting to mate with dominants towards subordinate mimics. Mimicry in this system is targeted towards females as males aggregate only temporarily, and so are unlikely to exhibit ‘dear enemy’ effects that would make intra-sexual mimicry advantageous. Thus stable territorial boundaries, such as those in classic leks, are likely to be an important prerequisite for the evolution of intrasexual deception through vocal mimicry.

3.5.4.3 Evidence of deceptive mimicry in other lekking species

A review of the literature suggests that it is in species with lek-mating and lek-like (e.g. strongly polygynous) mating species that have garnered the greatest support for the deceptive mimicry hypothesis as an explanation of song sharing. Although direct evidence for deceptive mimicry, such as that presented here, is lacking, patterns of imitative vocal learning in these species are suggestive. For example, in yellow-rumped caciques, *Cacicus cela vitellinus* (Trainer 1989), males appear to track and match changes in the songs of colony-mates, hypothetically, in an attempt to increase their attractiveness to females. Similarly Payne (1985) demonstrated that in polygynous village indigobirds, *Vidua chalybeata*, individuals that take over existing sites are more successful when that bird copies local songs, and changes in song among successful males

result in mimicry by others. Although these behaviors do not necessarily entail deception, it sets up the conditions in which both intrasexual and intersexual deceptive mimicry might be favored. In lekking three-wattled bellbirds, *Procnias tricarunculata*, and bearded bellbirds, *Procnias averano*, males sing to attract females from stable clustered territories. Despite the strong indication that song is learned in all bellbirds, the majority of individuals within a population appear to share extremely similar structure in their vocalizations (Kroodsma et al. 2013; Kapoor, unpublished data). One of several possible explanations for such a pattern is a process in which low quality males precisely copied the songs of high quality males, either to lure females away from dominants (the hotshot phenomenon) or to gain access to lekking territories (intrasexual deceptive mimicry), resulting in broad-scale uniformity in vocal patterning. Interestingly, it is this vocal stereotypy that allowed the bellbirds to be overlooked as examples of suboscine learners for many decades, which suggests that vocal learning may actually be more common in polygynous non-oscines than is currently appreciated.

3.6 Conclusions

Here I have shown the first experimental evidence that the deceptive mimicry hypothesis may provide a plausible explanation for the evolution of song sharing in a polygynous song learner. Specifically, I have demonstrated that lekking little hermits 1) exhibit ‘dear enemy’ effects with respect to neighbors versus non-neighbors, 2) are unable to recognize individuals based on their songs, 3) treat mimic songs similarly to neighbor songs, and 4) are more likely to fail to recognize an intruder as similarity between the mimic and neighbor’s songs increases.

The combination of extreme selection for lek attendance, highly costly territorial disputes, and relatively simple use of song in sexual communication in the little hermit, and

potentially many other lekking species, makes the potential benefits of deceptive mimicry far greater in lekking systems than would be expected in territorial monogamous species where breeding sites are not as limited, territorial confrontations are typically less intense, and song is used in extremely varied ways. Additional work on other lekking song learners will reveal the generality of these conclusions.

Although much of the analysis I have presented here concerns the use of vocal mimicry in male-male deception in little hermits, a concurrent inter-sexual phenomenon targeted at deceiving females is possible (i.e. the hotshot phenomenon). This scenario may be especially likely if females exhibit site fidelity in their mating decisions and are therefore more susceptible to making mate selection errors (McDonald and Potts 1994; Höglund and Alatalo 1995).

Alternatively, vocal mimicry could serve a role in intersexual communication if females show preferences for high average song similarity among lekking males. In this case, song sharing might serve as an indicator of male quality (reviewed in Podos and Warren 2007), or could facilitate assessment of potential mates (Patricelli et al. 2011). Careful consideration of the link between male song sharing and female mating patterns is a focus of my current work in this system.

CHAPTER 4

COURTSHIP DIALECTS AND THEIR RELATIONSHIP WITH MALE REPRODUCTIVE SUCCESS IN A LEKKING HUMMINGBIRD

4.1 Abstract

Recent work investigating the relative reproductive fitness of individuals on leks has revealed a wide array of fine-scale mating tactics in both males and females. The interactions between these tactics can lead to stark phenotypic and spatial patterning on leks. One of the more striking patterns so far observed in lekking species is also the least well studied: vocal dialects between males within leks. The existence of vocal dialects within leks in hermit hummingbirds represents an evolutionary puzzle: why should competing males share songs with some individuals and not with others? Here, I test four hypotheses explaining the existence of dialects on leks due to female choice: 1) dialects are preferred by females as they permit easier comparisons of males, 2) dialects reveal the structure and relative positions of males queuing for dominance, 3) dialects serve as multi-male phenotypes preferred by females, and 4) dialects are the result of low quality males attempting to deceptively mimic the songs of attractive males. Using a combination of field observations, and acoustic and parentage analyses, I found the greatest support for the possibility that males are using deceptive song mimicry to intercept mate faithful females. The results of this hypothesis thus provide general support for the classic “hotshot” hypothesis, but extend it by identifying a potentially novel mechanism by which low quality males may deceive females on leks.

4.2 Introduction

The evolution of male mating aggregations (leks) has represented a major evolutionary puzzle in part because the intrinsic costs of social behavior demands the identification of its counterbalancing benefits (Alexander 1974; Hoogland and Sherman 1976). One of the primary goals of both early and ongoing studies of selection on leks has been to identify the benefits to both males and females of mating on leks versus dispersed territories (reviewed in Höglund and Alatalo 1995). The evolution of lekking behavior has been linked to a large number of individual benefits such as the avoidance of predation (Wittenberger 1978; Ryan et al. 1981) and female avoidance of courtship harassment (Clutton-Brock et al. 1992). In addition, the particular spatial placement of leks has been related to benefits of increased female encounter rates where female home ranges overlap (e.g. “hotspots”; Bradbury and Gibson 1983), improved access to resources important to females (Snow 1973; McNaughton 1988), and increased courtship signal propagation in particular habitats (Snow 1973). Furthermore, the size and composition of leks may play an important role in the evolution of lekking behavior such that female preferences for larger aggregations may lead to lek joining behavior (Alatalo et al. 1992; Hernandez et al. 1999) and males joining aggregations with kin may gain inclusive fitness benefits (Kokko and Lindstrom 1996). The majority of this earlier work, however, restricted its focus to selection at the level of the lek, leaving unexplored potential benefits accrued through fine-scale fitness maximizing tactics of males and females within leks.

Indeed, the lack of direct material benefits of female mate choice and typically high male mating skew in lekking species sets the stage for sexual selection to play a major role in shaping male and female mating tactics within leks (Höglund and Alatalo 1995; Patricelli et al. 2011). A shift in focus to within-lek processes, as well as novel techniques and the expansion of the

taxonomic scope of lekking species studied, has begun to uncover the complexity of male and female mating tactics on leks. For instance, males may increase their fitness on leks by spatially associating with attractive males and attempting to intercept females (e.g. the “hotshot” effect, Beehler and Foster 1988; “sneaker” strategies, Jukema and Piersma 2006) or by copying learned displays from successful neighbors (Payne 1985; Trainer 1989; Cockburn et al. 2009). Males of lekking species have also been reported manipulating the perceptual environment in which their displays are observed by females (Uy and Endler 2004; Endler et al. 2010) and negatively modifying the attractiveness of sexual signals of competitors (Borgia 1985). In addition, the advent of molecular genetic techniques has revealed remarkable patterns of kin structure among display sites within leks suggesting the possibility of kin selection on leks (Shorey et al. 2000; Krakauer 2005; Reynolds et al. 2008; Concannon et al. 2012). Some of the more striking examples of male reproductive tactics on leks involves the use of cooperative display and social coalitions with relatives (Foster 1981; van Rhijn 1983; McDonald and Potts 1994; DuVal 2007; Ryder et al. 2011) and nonrelatives (Petrie et al. 1999; Krakauer 2005) to attract mates and rise in social status or position in a mating queue (Díaz-Muñoz et al. 2014).

Males, of course, are not operating in a selective vacuum with respect to female mating tactics; female mate search behavior, cognitive processes, and courtship evaluation tactics undoubtedly coevolve with male mating tactics in complicated ways (Bateson and Healy 2005; Patricelli et al. 2011). As a single example, if females use comparative assessment of male displays, and can compare only a subset of males on the lek due to energetic or cognitive restrictions, low quality males may benefit from the avoidance of males of superior quality; a markedly different pattern from that expected under the “hotshot” hypothesis (Bateson and Healy 2005; Oh and Badyaev 2010; Patricelli et al. 2011).

The interplay between male and female mating tactics on leks is likely to lead to distinct patterning of male physiology (e.g. display traits), behaviors (e.g. aggression or cooperative display), and spatial relationships on leks that reflect the underlying selective processes at work, though in many cases intense scrutiny is required to reveal such patterns (for instance, fine-scale kin-associations among neighbors require molecular genetic analyses to reveal patterns; Petrie et al. 1999; Shorey et al. 2000; Krakauer 2005). Oddly, one of the most overt examples of within-lek phenotypic patterning is also one of the least well studied: discrete vocal dialects among neighboring males in lekking hermit (*Phaethornis*) hummingbirds. A great deal of work has been invested into understanding the forces leading to vocal dialects in animals (Baker and Cunningham 1985; Podos and Warren 2007), but nearly none of this work has sought to identify why it might occur on leks, and investigations of hermit dialectal patterns have been descriptive only (Snow 1968; Wiley 1971; Stiles and Wolf 1979; Ramjohn et al. 2003; González and Ornelas 2009). Whether previously described lek mating tactics can provide an explanation for the existence of dialects within leks remains an open question.

In previous work I have demonstrated that vocal dialects in the advertisement song of lekking little hermits (*Phaethornis longuemareus*) may result as an epiphenomenon of male-male competition for territorial vacancies on leks: males may copy the songs of territory holders and use them deceptively to gain entry onto the lek and benefit from reduced aggression from territorial neighbors (Chapter 3). Here, I address whether the existence of dialects can additionally be explained by patterns of female choice. Specifically, I test predictions of the following four non-exclusive hypotheses for the evolution of dialectal variation on leks via female choice: 1) females use the shared songs among dialect members to make accurate comparisons between male vocal performance, 2) females use a combination of dialect

membership and spatial information to identify central individuals that are high ranking in a social queue, 3) females use information contained in multi-male vocal displays (i.e. dialects) to find groups of males of high average quality, and 4) females use song structure to relocate previous preferred mates, and dialects reflect low-quality males attempting to deceive females into mating by copying the songs of these preferred males.

H1: females use shared songs to assist in accurate assessment of song performance.

In many species of songbird that have been studied, females show strong preferences for a number of traits related to song performance including: the use of large repertoires (Catchpole 1986), consistent songs (Byers 2007), and high quality renditions of challenging-to-produce syllables (Ballentine et al. 2004; Forstmeier and Schielzeth 2011). Although the use of individual performance in mate choice appears to be widespread, a small number of studies have reported negative results (Cardoso et al. 2012; Cramer 2012; Cramer 2013b). The authors of these studies have concluded that differences in song type repertoires may interact in complicated ways with performance estimates, negating the use of performance as a reliable signal of male quality by females. This logic suggests that the homogenization of song structure to a single “standardized” form (e.g. a song dialect) among a set of potential mates might allow females to compare performance in species where song otherwise varies between individuals. Under this scenario, little hermit dialects might serve as foci of female assessment, within which females make comparisons based on relative song performance. This situation makes the simple prediction that males that exhibit better song performance relative to other males in their dialect will have greater reproductive success than their lower-performance counterparts.

H2: Females use dialect structure to identify the positions of males that have queued for status within the dialect. Orderly and semi-orderly social queues for preferred central

territories or dominant social status have been identified in a number of lek mating species including *Chiroxiphia* manakins (McDonald 1993; DuVal 2007), black grouse, *Tetrao tetrix* (Kokko et al. 1998), and topi antelopes, *Damaliscus lunatus* (Bro-Jørgensen 2011). Females may show preferences for males on central territories when queuing results in older males or males with greater fighting abilities in more central locations on the lek (Kokko et al. 1998). Because male little hermits appear to learn their songs from immediate neighbors and tend to settle on the lek periphery (Chapter 3), males are expected to shift to more central territories within their dialects as older males die and new males arrive. Therefore, females may be able to use a combination of spatial and dialect membership queues to identify high quality or older males in the centers of dialects. This hypothesis therefore makes the prediction that males that are more central within their dialects (rather than the entire lek) are more likely to obtain matings than peripheral dialect members.

H3: Females use dialect structure as a direct indication of average male quality.

Although relatively rare, cooperative courtship display has been identified in a number of lekking species (see above), and aspects of the performance of this joint-male phenotype has been implicated in female choice in *Chiroxiphia* manakins (Trainer and McDonald 1995; Vanderbilt et al. 2015) which produce coordinated duet songs. In long-tailed manakins, *Chiroxiphia linearis*, the coordination of song frequency in duet songs improves over time and can serve as an indicator of age, and partnership stability to females, who prefer more coordinated duets (Trainer and McDonald 1995; Trainer et al. 2002). In little hermits, the coordination of song structure by neighboring males in a dialect may serve a similar role in female choice as in *Chiroxiphia*. Although song is learned early in life, the accuracy of vocal imitation and other average measures of performance in a dialect may serve as indicators of the

average quality of males within that dialect (Lachlan and Nowicki 2012). A testable prediction of this hypothesis is that the average reproductive success of males should correlate with a dialect level phenotype that indicates the quality of individuals in the group, namely, song copying accuracy, and average levels of song performance among group members.

H4: Females use song to relocate preferred males, dialects reflect deceptive mimicry of song by low quality males. A prevailing theory for the existence of lekking is that low quality males may surround attractive males and parasitize their attractiveness (i.e. the “hotshot” hypothesis; Beehler and Foster 1988). This hypothesis requires either that preferred males attract more females than they can mate with and peripheral males benefit from a “spill-over” effect or that females are constrained to use conservative mating tactics (rather than full assessment of all available males) that enable the use of deception by males. In little hermits, the former condition is unlikely as females visit the lek only rarely, however the latter condition may be especially likely in many lekking hummingbirds. The high metabolic rate of hummingbirds (Lasiewski 1963), combined with the limited supply of nectar resources (Kodric-Brown and Brown 1978) may present particularly high costs for mate searching among females, suggesting that conservative tactics that reduce the costs of mate searching may be common (Jennions and Petrie 1997). One way in which females may reduce the costs of searching for mates is by forgoing full assessment of potential mates and remaining faithful to previous mates, tactics which may be relatively common in lekking species (DuVal 2013). Females attempting to relocate previous mates may be expected to use a combination of behavioral (such as song) and spatial cues, especially in lekking species in which male territories are small and minor changes in habitat structure or territory boundaries may make reliance on spatial cues alone unreliable. Low quality males, then, might be expected to deceive females by copying these behavioral cues and

clustering near attractive males. Although this hypothesis has only rarely been tested, one study of superb fairy-wrens, *Malurus cyaneus* (Cockburn et al. 2009), has found compelling evidence that low quality males copy the songs of and cluster around attractive males, and thereby gain fitness by intercepting females attempting to mate with these attractive males. In the same way, low quality male little hermits may cluster around and copy the songs of attractive males and thereby deceive faithful females into mating. In this study I am able to test two important predictions of this hypothesis: 1) that females show a nonrandom tendency to remate with a male in the same dialect, and 2) at least some instances of remating on leks are with different males in the same dialect, indicating potential female assessment errors.

To test the predictions of these four hypotheses explaining the relationship between female choice and vocal dialects on leks I used a combination of novel sound analysis techniques, molecular genetics, and behavioral observations in a free-living population of lekking little hermits possessing song dialects.

4.3 Materials and methods

4.3.1 Field methods

4.3.1.1 Study population

I conducted this study on the population of little hermits that I have intensively marked and monitored since 2008 (see Chapters 2 and 3 for details on the field site and natural history of the study species). Briefly, little hermits are trap-lining rainforest understory hummingbirds endemic to Trinidad, Guyana, and NE Venezuela. Males form leks seasonally where they sing an advertisement song from one or a few song perches within a stable territory. Male-male aggression, in the form of chases, visual displays, and physical fights, is common during this

time. Females visit the lek to search for mates, and elicit complicated visual courtship displays from a territory owner when the female settles onto a display perch. After mating females raise offspring independently away from the lek. Nest predation is extremely high, and females may initiate five or more clutches during the breeding season, necessitating repeated visits to leks for mating.

4.3.1.2 Capture, marking, genetic samples

All capture and marking was conducted between January and May each year from 2008 to 2014 on the four core leks on the field site (i.e. leks 5 – 7 and 12). Males were captured using mistnets, measured for a number of physical traits (see below), and marked with a metal band and colored plastic leg tag, females, due to the use of their legs in nest building and incubation, were not given leg tags but instead marked with a colored mark on the tip of the tail with a permanent marker or given a colored plastic back tag (see Chapters 2, 3, and 5; Kapoor 2012 for details). Individuals were inspected for contour pin feathers, and if found, 5-10 feathers were collected for later genetic analysis. When pin feathers were not available a small blood sample (1-10 μ L) was obtained from a pin prick on the bird's foot and stored in lysis buffer.

4.3.1.3 Behavioral observations

I monitored individual male behavior on leks for all lekking males on leks 5, 6, and 7 in 2013. Behavioral observations consisted of 1-hour all-occurrence sampling (Altmann 2012) of individual male territories, and each male was observed for at least 10 hours across the season. Observers (n=11) recorded the identities and presence durations of all individuals that entered the focal territory as well as the durations of their song, display, chase, and foraging behavior. For the purposes of this study, only measures of the percent time a bird spent singing and percent time spent on the territory were analyzed (hereafter, male activity measures), as I considered

these measures (as well as the visual display) to be the most likely to relate to female assessment, and the analysis of visual display traits is beyond the scope of the study.

4.3.1.4 Sound recordings

I obtained recordings of all males on all four leks in 2013. Each recording was obtained by placing a microphone under the primary song perch of a male, then waiting for that male to return to this perch and begin singing (for details of recording equipment, field methodology, and digitization procedures used see Chapter 2). Only the crystallized songs of adult males were included in analyses.

4.3.1.5 Nest searching and genetic sampling of nestlings

Nests were located by systematic daily searches of understory vegetation, covering non-overlapping areas of the field site during each search. I supplemented visual nest searching by setting mistnets along active trap-line routes, fitting females with a small (0.11 g) custom VHF radiotag (Kapoor, manuscript in preparation), and tracking females to their nests. A total of 96 nests were found in 2013.

Once an active nest was located visually I set a mistnet along the female's preferred route of entry to / exit from the nest area to band, mark, and genetically sample the mother (nests identified by telemetry did not require recapturing the female). To prevent high nest predation rates from impacting the paternity dataset I modified the technique of Tori et al. (2006) as follows: If a nest was found in the egg stage I collected the clutch (all hummingbirds lay clutches of up to two eggs) and replaced each egg with a white plasticine model of a little hermit egg (all females readily accepted these eggs) and transported the clutch to a Brinsea® Mini Advance incubator (Brinsea Products Inc., Titusville, Florida). The nest was monitored once every three days to determine if predation had occurred. When predation occurred prior to hatching I

allowed the eggs to develop until candling revealed a well-developed embryo (Lokemoen and Koford 1996), then sacrificed the embryo and collected a tissue sample. If the nest was still active (i.e. had not been depredated) by the time the eggs hatched I took 2 – 3 down feather samples and cotton swabs of the interior of the eggshell membrane and placed them in lysis buffer for genetic analysis and returned the chicks to the nest (all females accepted chicks delivered to the nest in this manner). If the nest was in the nestling phase, I obtained pin feather samples of the chicks in the nest.

4.3.1.6 Lek spatial surveys

The close spacing of males on leks, combined with the topographically varied structure and extremely dense canopy cover of the habitat make conventional GPS mapping of display perches impossible (i.e. the error associated with waypoint locations frequently exceeded the distance between adjacent perches). The spatial structure of males on leks (i.e. the distances between the primary display perches of territorial males) was therefore obtained by using traditional land survey techniques employing an optical Kern DKM1 theodolite (Kern & Co. Ltd., Aarau, Switzerland). All angle and distance measurements were precise to the nearest one arc second and 0.001 m (respectively), and were calculated twice to minimize the effects of typographical and measurement errors.

4.3.2 Male physiology

To obtain measurements of putative male physical quality (hereafter, physiology measures) I measured the following traits related to male size, body condition, and age: body mass, head length, wing cord, tail length, length of the white tip on the two central rectrices (tail extension), ectoparasite number around the face and throat, and age. For cases in which multiple morphometric measurements were available for a given individual in 2013 I chose the

measurement closest to the middle of the breeding season (i.e. mid-March) to control for the potential for seasonal changes in male phenotypic characters. Because males were often captured returning from nectar foraging bouts estimates of mass fluctuated markedly and were therefore averaged for all captures of an individual in 2013. Body condition is typically measured as the residuals of a regression of mass on tarsus length (e.g. Cramer 2013b), however, as mass fluctuated with the amount of nectar a male was carrying in his crop, and tarsus length was extremely difficult to measure accurately, I did not include this measure of condition in the analysis. The ages of lekking males were calculated as described in Chapter 3.

4.3.3 Spatial analysis

I used ARCGIS version 10.1 to create a three-dimensional map of all measured perches, based on the angle and distance measures from the spatial survey. Because males generally defend as few as one song perch it is often impossible to estimate territorial boundaries by generating minimum convex polygons (which require a minimum of three points). Instead, I estimated boundaries by calculating Thiessen polygons (i.e. polygons within which any position is closer to the point that generated it than to any other point) for each song perch, with a maximum radius of 16.5 m (estimated using the methodology of Marten and Marler (1977), details in unpublished manuscript) representing the maximum distance that an average song is discriminable. The polygons of multiple song perches occupied by the same bird were merged to create a single polygon. These territory polygons were then used to estimate the area occupied by each dialect, as well as its density (i.e. the number of dialect members divided by the dialect area). I also calculated two measures of male spatial location (hereafter, spatial measures): the distance from the center of each individual's territory to 1) the center of that male's dialect, and 2) the center of the lek using this three-dimensional territorial map.

4.3.4 Song analysis

4.3.4.1 Dialect assignment

Each male was assigned to a dialect on his resident lek using the methods outlined in detail in Chapter 2, in this case using ten rather than fifteen songs per male in the analysis. In short, I extracted ten sequential songs per male from song recordings of each male in the analysis. I then used a dynamic time warping (DTW) algorithm to compare each individual song element from each song with all other elements in the ten selected songs. The algorithm searches for an optimal alignment of each pair of elements which minimizes the Euclidean distance between the two elements based on the fundamental frequency, change in frequency, and time features of both elements. For each individual, the elements were clustered into a subset of element types using an agglomerative hierarchical clustering algorithm, and the element closest to the group centroid (i.e. the “median element”) was chosen for subsequent comparisons between individuals. For comparing songs between-individuals I used the DTW algorithm again to measure pairwise element distances between all median elements of all birds’ songs in a lek (this time log transforming the two frequency measures to obey Weber’s law). Then, I used a sequence alignment algorithm to identify the most closely matched sets of element transitions (i.e. pairs of consecutive song elements) between individuals, and to create an average song dissimilarity score between all of each individuals’ songs and the songs of all other individuals in a lek. Finally, I used an agglomerative hierarchical clustering algorithm to establish the hierarchical relationships between individuals’ songs. The global silhouette index (a metric of the goodness of fit of any clustering solution; Rousseeuw 1987) was obtained to identify the best-supported set of dialect assignments for each male, and the significance of this clustering

solution from that expected by chance was calculated (all leks exhibited highly significant clustering).

4.3.4.2 Song performance measures

I scored four measures that describe individual song performance and dialect structure: song variability, acoustic centrality, dialect dispersion, and amplitude performance (hereafter, song performance measures; defined below).

Song variability - For each male I obtained a score of within-male song variability by first measuring the average acoustic distance of each recorded song element from the average element of its respective element cluster (as determined by the DTW algorithm). I then averaged these distance values for each element, and again across all element types produced by that male, to obtain a final song variability score [Figure 4.1(a)]. For relative song variability comparisons within dialects, I calculated a z-score for each individual based on the variability values of all males in his dialect.

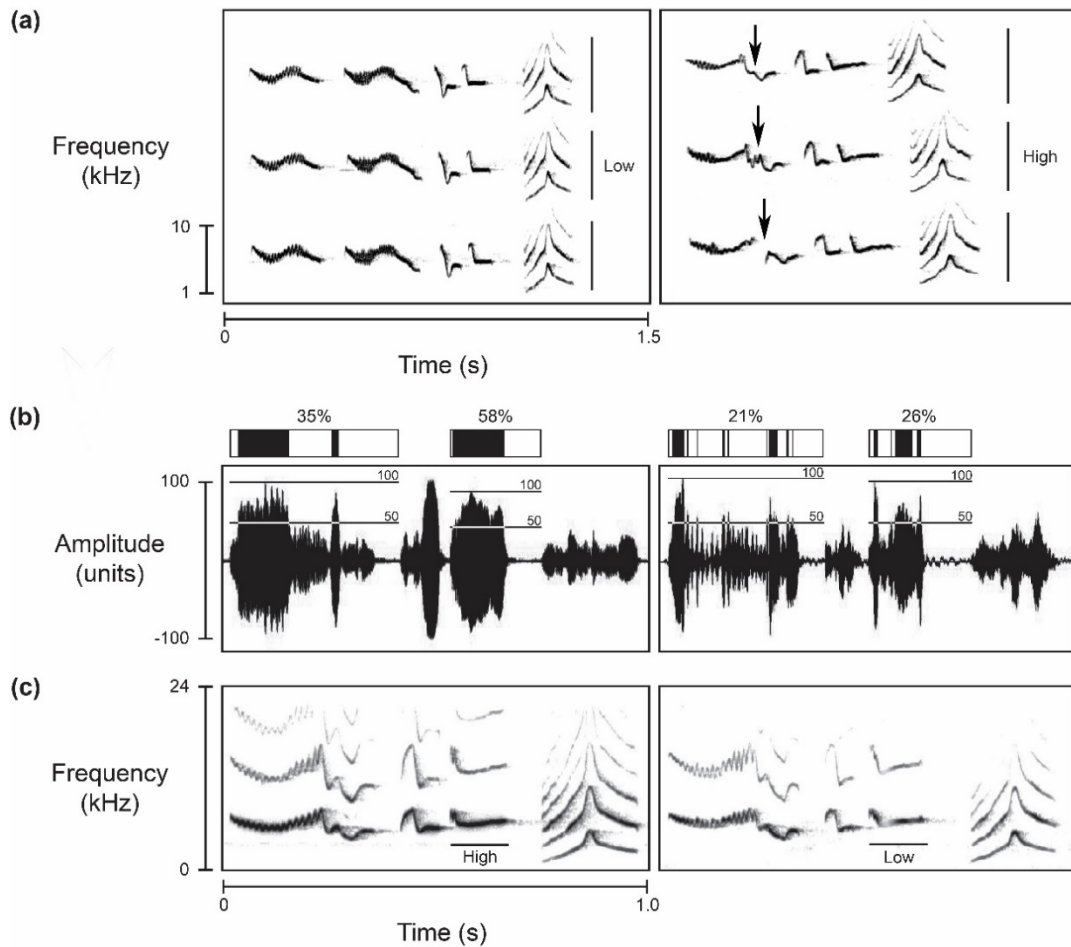


Figure 4.1. Example spectrograms and waveforms of songs exemplifying high and low measures of individual song variability (a), and percent peak performance used to calculate amplitude performance (b) and (c). Arrows in (a) denote the section of each of three repeated song elements that results in a high variability score. Panel (b) shows two sets of song elements with high peak performance (35% and 58% of each respective element with amplitude $\geq 50\%$ of maximum amplitude; left box) compared with the second set (21% and 26%, respectively; right box). Note that in both boxes the shorter of the two elements (second element) was sung with higher peak performance (58% and 26%) than the respective first, longer elements (35% and 21%). Panel (c) shows a spectrogram of song elements with high (left box) and low (right box) percent peak performance.

Acoustic centrality - To obtain a relative score of the similarity of a male's songs to an average song of his dialect, I first conducted a principal coordinates analysis of the pairwise SDI scores (see Chapter 2) for all males within a dialect, and then calculated the multivariate “acoustic centroid” of each dialect. Then, I calculated the Euclidean distance between each focal

male's point in acoustic space and that centroid. Lower values of this score indicate a male that sings more similarly to the average song of the dialect than a male with a higher score.

Song dispersion - To measure dialect-level song dispersion (essentially an inverse measure of the average acoustic centrality of males within a dialect), I measured the average pairwise SDI distance between all individuals within a dialect. Low values of song dispersion indicate a dialect in which the males share more similar songs than in a dialect with a higher dispersion value.

Amplitude performance - Across a number of study systems song performance, or the ability for an individual to produce a song with a combination of acoustic characteristics that are physiologically challenging, has been identified for its role in male-male competition and female choice. Two primary measures of song performance have received general support across avian taxa: 1) the interaction between the production of syllables with large frequency-bandwidths and fast repetition rates, also known as 'vocal deviation' (Podos 1997; Podos 2001; Ballentine et al. 2004; Riede et al. 2006; Cardoso et al. 2007), and 2) the interaction between the production of long-duration syllables with short inter-syllable intervals and high sound amplitude, here 'amplitude performance' (Hartley and Suthers 1987; Mota and Cardoso 2001; Forstmeier et al. 2002; Leadbeater et al. 2005; Holveck and Riebel 2007; Cardoso et al. 2012). In little hermits, vocal deviation is unlikely to play an important role as songs do not typically consist of broad bandwidth elements that are repeated rapidly. Amplitude performance, however, has potential to be an informative measure of performance across a wide range of element types. I therefore chose to investigate amplitude performance only.

A first step in measuring song performance involves establishing that a relationship exists between the acoustic characteristics that are hypothesized to trade-off with one another. In the

case of amplitude performance, individuals should be capable of producing songs with short duration elements that consist of both low and high amplitude elements, but should be less capable of producing long element songs with high amplitude; the resulting distribution of songs in the acoustic space defined by element duration and element amplitude is expected to be “triangular” (Podos 1997; Ballentine et al. 2004). To evaluate this potential relationship in little hermits I compared song element lengths for all males in the study with a measure of element amplitude: the percent time a song element’s amplitude meets or exceeds 50% of its peak amplitude, here ‘percent peak performance,’ (following Forstmeier et al. 2002; Figure 4.1(b) and (c)). A simple scatterplot comparing element lengths with their percent peak performance values, across all individuals in the study, reveals such a triangular relationship [Figure 4.2].

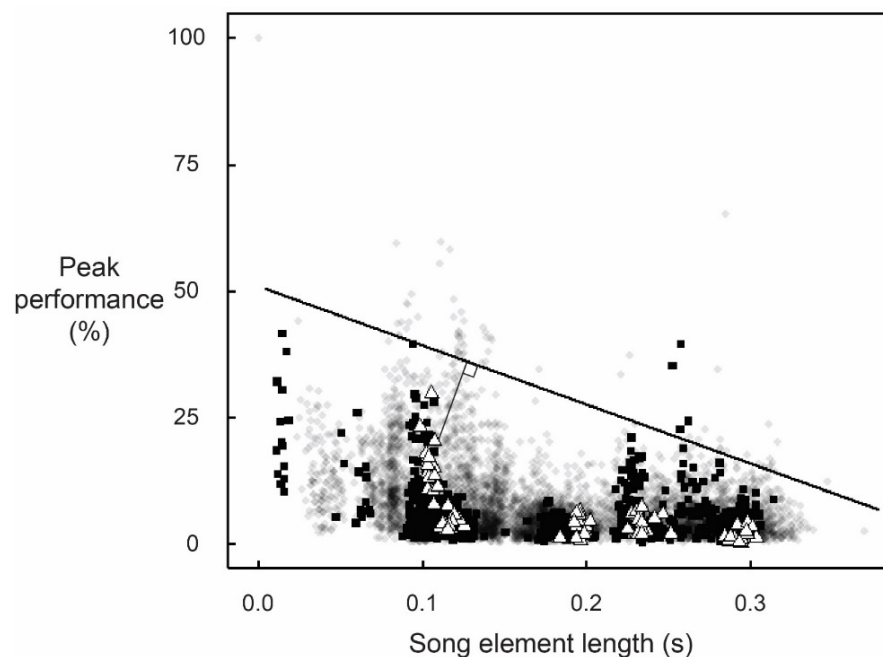


Figure 4.2. The calculation of amplitude performance for each song element. Small diamonds (semi-transparent) represent all song elements of all 108 candidate male sires in the population. Black squares depict the song elements of all males from one dialect, and those of a single individual from that same dialect (white triangles). The 90% quantile regression line used to calculate amplitude performance is calculated with all data points for individual measures, and restricted to a single dialect for relative measures (depicted here, thick line). An element’s amplitude performance value is calculated as the orthogonal distance from that element to the regression line (thin line).

To obtain a measure of individual amplitude performance, I first performed a 90% quantile regression following Wilson et al. (2014), using the full complement of elements from the ten selected songs of each male included in the study (i.e. lekking males from leks 5-7, and 12 in 2013). I calculated the amplitude performance for each song element as the orthogonal distance between the duration – peak performance value and the regression line (following Podos 2001; Figure 4.2), then averaged these values for all the elements produced by a single male. In this study, positive values of amplitude performance indicate songs that are closer to the performance maximum. To score each male's amplitude performance relative to other males in his dialect I restricted the set of songs used to calculate the regression line to dialect members, calculated the corresponding amplitude performance value, then obtained a z-score of that male's average amplitude performance calculated with respect to the performance values of all other males in the dialect. I scored the dialect-level amplitude performance by averaging the amplitude performance scores, again based on an upper 90% quantile regression restricted to dialect members, for all males in that dialect.

4.3.5 Reproductive success

4.3.5.1 Laboratory procedures

Genomic DNA for later use in the assignment of paternity was extracted from blood and feather samples of all candidate sires (i.e. lekking males on core leks in 2013), chicks banded in nests or captured as fledglings in 2013, and all nesting or provisioning females, using the DNEasy Blood and Tissue Extraction Kit (Qiagen) following the manufacturer protocol. To develop polymerase chain reaction (PCR) primers for microsatellite markers for use in estimating pairwise relatedness, an enriched genomic DNA library was created using a protocol modified from (Hamilton et al. 1999). Briefly, genomic DNA was digested from 2 individuals,

and ligated into a SNX linker (IDT), and was then enriched by magnetic bead selection with a mixture of 3'-biotinylated oligonucleotides comprising 2 di-nucleotide, 8 tri-nucleotide, and 17 tetra-nucleotide repeats. Eluted genomic fragments were then amplified using PCR (protocol available upon request), screened for amplification, and then sequenced using a Titanium 454 sequencer. Primers were designed for candidate microsatellite regions that had a large tandem repeat number (>5), high quality sequences in both the forward and reverse directions, and that did not contain known transposable elements. Forward primers were designed with an Applied Biosystems fluorescent dye label on the 5' end (PET, VIC, and 6-FAM dyes), or (in the case of two loci that failed to amplify with fluorescent dyes: Pha05 and Pha22) were tagged with a 5' CGAGTTTCCAGTCACGAC sequence that worked in conjunction with a complimentary FAM-labeled universal M13(-21) primer to enable visualization of fragments under capillary electrophoresis (See Schuelke 2000 for detailed methodology). Reverse primers were unlabeled and designed with a 5' GTTCTT "pigtail" sequence to facilitate accurate genotyping by reducing allelic stutter (Brownstein et al. 1996). To facilitate multiplexing, primers were designed so that annealing temperatures were similar between forward and reverse primers, and across all loci. The 10 polymorphic loci used for this study were pooled into 2 multiplex panels each consisting of 5 primer pairs (or trios for loci using the universal tag), selected to avoid overlap in product sizes and to maximize the diversity of dye colors in each mix. I used a Type-it Microsatellite PCR Kit (Qiagen) both for initial primer screening and to amplify alleles for large-scale genotyping. Reactions had a final volume of 10 μ L, and consisted of 1-40 ng of genomic DNA (in 1 μ L AE buffer), 0.2 μ M of each primer in the multiplex panel, 1 μ L of water, and 5 μ L Multiplex PCR master mix (Qiagen). In the case of Pha05 and Pha22, the concentrations for the forward and universal primers were 0.05 and 0.15 μ M, respectively. The PCR was performed

under the following conditions: initial denaturation at 94°C for 5 min; a 7 cycle touch-down stage with an initial denaturation step at 94°C for 30 s, annealing at 61°C for 90 s (reduced by 1°C each cycle), and extension at 72°C for 30 s; 21 cycles at 94°C for 30 s, 57°C for 90 s, and 72°C for 30 s; and a final extension of 30 min at 60°C. All PCR products were resolved on an ABI 3730xl DNA Analyzer (Applied Biosystems) using 1 µL of a 1:10 dilution of PCR product to water, 18 µL formamide, and 0.2 µL size standard (GeneScan-500 LIZ). Fragment sizes were calculated with GENEMARKER version 2.4.0 (SoftGenetics, LLC.) semi-automatically and checked manually. For a list of all loci used in this study and associated information see Table 4.1.

Table 4.1. Descriptive statistics for microsatellite loci used in the paternity analysis.

Locus	Primer sequence (5'-3')	Dye	Repeat motif	Fragment size (bp)	Number of alleles	H_o	H_E	Null allele rate
<i>Pha01a</i>	F: TAGTCTCTGCCTGCACTACG R: CACTTCCTACCTGAAAGATAATCAG	PET	(ATT) ₉	107-131	7	0.61	0.62	0.0072
<i>Pha02</i>	F: CCTGTTCACAGTTTAGGGTAAGG R: AATGCCTTGGGCATGTACC	VIC	(ATAA) ₇	142-150	3	0.36	0.36	-0.0078
<i>Pha04</i>	F: ATTGGGCTCGGAGAGAGG R: TTTCTTCATAGACCTCCGAGAGC	6-FAM	(CAT) ₁₃	166-184	7	0.76	0.77	0.0079
<i>Pha05</i>	F: CCTGTCTTCCTTTGTGGAACC* R: GGTATCAAAAGTCTGAGTGTGTAGG	6-FAM	(TG) ₁₁	193-221	3	0.49	0.49	-0.0037
<i>Pha17</i>	F: CACTTGAGTGCACCTTAGAGAGC R: TTTGTTTCATGAGTTGAGGCTATG	PET	(TGA) ₁₁	214-229	5	0.54	0.57	0.0235
<i>Pha22</i>	F: AATGGCAAAACAAGCAGCTC* R: TGTGCGGACACAGATGATG	6-FAM	(CA) ₁₁	146-150	2	0.39	0.40	0.0135
<i>Pha25</i>	F: AGAGCAGTGGAAAGGAATGTG R: TGCTTGCAAAAGATAAAGCACTTAC	VIC	(AC) ₁₂	164-181	8	0.66	0.69	0.0196
<i>Pha26</i>	F: GCAGTTCAACAGCAGACAGG R: GCTATGCATTACCTTTTCACTGG	VIC	(AAGGC) ₁₇	268-331	13	0.86	0.90	0.0006
<i>Pha27</i>	F: AAGCTTTCCTGGCTTTGTCC R: GGAGCCAGGGTGTGTCC	6-FAM	(AGAT) ₁₃	217-253	8	0.76	0.77	0.0044
<i>Pha28</i>	F: ATTGCAGCAAAACCCACTTC R: GATCTTATTAGAGAGTTGATGCATGTG	VIC	(ACT) ₁₈	210-235	7	0.36	0.65	0.2862 [†]
$\mu \pm SE$		-	-	-	6.3 ± 1.0	0.58 ± 0.057	0.62 ± 0.054	0.035 ± 0.028

*Primers contained a (5'-CGAGTTTTTCCCGAGTCACGAC-3') sequence for use in 3-primer PCR protocol (see Methods). [†]Does not meet HWE assumptions at $p < 0.05$ due to linkage disequilibrium.

4.3.5.2 Paternity analysis

To validate the use of each locus in the paternity analysis (N=10) I calculated allele frequencies, observed and expected heterozygosity, and tested for linkage disequilibrium and departures from Hardy-Weinberg equilibrium (HWE) using the program GENEPOP version 4.2.2 (Rousset 2008) and including all individuals sampled on the field site (N=506). One locus, Pha28, was found to show evidence of a high frequency of null alleles (see Results and Table 4.1), so all downstream analyses were conducted with and without this locus to determine the effect of its inclusion on paternity estimates.

I estimated the genetic reproductive success of lekking males in 2013 only, the year for which the greatest number of offspring could be assigned to a sire. Although 13 and 15 offspring sampled in 2011 and 2012, respectively, could be assigned with high confidence to a lekking male sire in that year, this sample is likely inadequate to provide sufficient resolution of relative male success. To assign paternity I used the software package CERVUS version 3.0.3 (Kalinowski et al. 2007) which employs a maximum likelihood analysis approach, and then secondarily validated analysis results using genotypic exclusion (described below). I restricted paternity assignments of offspring to a sire to those calculated with a 95% confidence level. Assignment confidence is calculated by CERVUS through simulations of statistical power generated from estimates of population allele frequencies, candidate sire sampling completeness, and the number of candidate sires in the population. To generate paternity assignment confidence level estimates I simulated 100,000 genotypes, with 150 candidate sires, the proportion of candidate sires sampled equal to 0.95, a 0.01 genotyping error rate estimated from repeated genotyping of a subset of individuals, and a proportion of loci typed of 0.99. These simulations enable the

generation of the “delta score,” which provides an estimate of the likelihood that a particular male is an offspring’s father relative to the next most likely candidate sire.

I next conducted a genotypic exclusion step to ensure that offspring were assigned correctly without violating Mendelian inheritance patterns. For a given paternity assignment, to be included in the final dataset an assigned sire first had to be a perfect genotypic match to the offspring or to mismatch only at the Pha28 locus in a manner consistent with genotyping error due to a null allele, and to be the only otherwise perfect genotypic match to the offspring. Fifty-four of 67 offspring were assigned to a sire with high confidence using these criteria, and the inclusion of the Pha28 locus did not affect patterns of male reproductive success reported in this study.

4.3.6 Statistical analyses

4.3.6.1 Individual analysis

Before testing for associations between male reproductive success and dialect-level factors, I first tested for links between male physiology, activity, and spatial measures and song performance measures. I did this for two reasons; first, I predicted that putative male quality traits would positively correlate with song performance, and second, to ensure that any relationship between mating success and dialectal song patterns were not confounded by covariance with a third non-vocal male quality measure. To test for these associations I first tested for correlations between male quality measurements to assess the need for variance reduction techniques, and then ran separate general linear mixed models (LMM) using the lme function in R (Pinheiro et al. 2015), with a song performance measure (amplitude performance, or song variability) as the response, a single male quality measure as a predictor, and a random factor of dialect nested within lek. Song variability was log transformed to meet assumptions of

normality of the residuals. I chose to run separate models for each predictor variable, here and in subsequent analyses, due to the fact that a different complement of quality measurement variables were available for different males, and the separate analysis of each of these predictors enabled me to maximize the use of all available data.

Next, I tested whether any of the male physiology, activity, spatial, or song performance measures were related to individual reproductive success. As is typical of lek-mating systems, reproductive success among male little hermits is strongly skewed, which resulted in problematic non-normality of residuals in preliminary linear models. I therefore chose to examine the relationship between male reproductive success (number of chicks sired) and individual-level traits by running a generalized linear mixed model fit using the Laplace approximation with a negative binomial distribution to account for the strong positive skew and dialect nested within lek as a random factor using the `glmer.nb` function of R package `lme4` (Bates et al. 2015).

4.3.6.2 Females use dialects to compare male quality (H1)

To test the prediction that females base mate choice decisions on relative comparisons of song quality between males within a dialect, I ran a general linear model comparing relative mating success within dialects to relative measures of individual song performance with respect to dialect members. I ran separate models for each of three song measures related to dialect membership: relative amplitude performance, relative song variability, and acoustic centrality. For each model the response was a z-score of number of chicks sired with respect to other males in the dialect, and was transformed to meet normality assumptions as $\log(\text{z-score of chicks sired} + 1)$. No effects of individual male quality were significantly related to mating success or song characteristics after correction for multiple testing, however, two trends were apparent: a relationship between 1) male age and song variability, and 2) tail extension length and mating

success (see Results). I therefore chose to include both age and tail extension as covariates in this analysis as a conservative approach to controlling for possible effects of these factors on relative mating success.

4.3.6.3 Females use dialects to identify high ranking males (H2)

To test the prediction that females base mate choice decisions on the relative positions of males within a dialect, I ran a general linear model comparing relative mating success within dialects to each male's relative distance to the center of his dialect.

4.3.6.4 Females use dialect-level phenotypes to assess males (H3)

To assess whether female mate choice decisions are related to dialect-level acoustic features, I ran a general linear model for each dialect-level song trait (average variability, average amplitude performance, and dialect dispersion), with per capita number of chicks sired across the dialect as the response variable. To account for possible covariance with other dialect-level effects unrelated to shared song between dialect members I ran separate models testing for the effect of 1) dialect area, 2) density of territories in the dialect, 3) number of males in the dialect, and average proportion of time all males in a dialect spent 4) present on the lek, and 5) singing.

4.3.6.5 Dialects are the result of deceptive song mimicry (H4)

I was able to test two predictions of this hypothesis with my available data; namely, 1) that remating females are expected to exhibit dialect fidelity at a rate greater than expected by chance, and 2) that at least some mating decisions of remating females will be allocated to different males in the same dialect, representing female "mistakes." To do this I conducted a randomization test to compare the expected random distribution of female rematings with same-

dialect males, with the distribution observed on the leks. In 2013, I observed nine pairs of remating events (each from a different female) with males from the four leks on the field site. Of these nine pairs of rematings, five involved sires from the same dialect. To generate the expected number of dialect-faithful rematings under a scenario of random mating, I simulated two consecutive mate choice decisions by nine females among 108 potential mates assigned to the same dialects as the real males in the dataset. I then measured the number of cases out of these nine pairs of random draws in which a simulated female made consecutive choices of any male from the same dialect, as well as the number of times those choices involved different males from a given dialect. I repeated each permutation 999 times, and determined the test statistics (p-values) as 1) the number of permutations where the number of rematings with males in the same dialect was ≥ 5 (the observed number of dialect-faithful rematings; prediction 1), and 2) the number of permutations where same-dialect rematings that occurred with two different males was ≥ 3 (the observed number of potential “mistakes”, prediction 2), both divided by the number of permutations plus one. The simulations were run with a custom script in MATLAB.

4.3.6.6 General comments

I checked the residuals of all linear and generalized linear models for departures from normality using the Shapiro-Wilk test in R (R Core Team 2015). Overdispersion parameters for all generalized linear models ranged from 0.70 – 0.92. To correct for multiple testing I used the false discovery rate, applied table-wise, using the `fdrtool` function in the R package `fdrtool` (Klaus and Strimmer 2015). All reported p-values are un-corrected in data tables, but changes in significance after correction for multiple testing are reported for each table and in the Results section.

4.4 Results

4.4.1 Paternity

The ten microsatellite markers used to estimate paternity in this study amplified 2 – 13 alleles per locus (mean \pm SD = 6.3 ± 3.2 alleles, N = 506 individuals genotyped), and provided a combined second-parent non-exclusion probability of $P_{NE} = 0.00318$. Tests of departure from Hardy-Weinberg equilibrium (HWE) revealed a moderately high frequency of null alleles in one of the ten loci: $F_{Null} = 0.2862$ for Pha28 (Table 4.1). All other loci met HWE assumptions. The exclusion of the Pha28 locus in the paternity analysis did not alter the results reported here. I therefore chose to include Pha28 in paternity assignment, as its inclusion increased the number of chicks that could be assigned to a sire with strict (i.e. 95%) confidence. I was able to assign 80.1% of chicks sampled in 2013 (N = 54 of 67) to a known male in the population, with strict confidence. This corresponds to 84.8% (N = 28 of 33) of sampled nests and 96% (N = 24 of 25) of dams with at least one offspring assigned to a sire. Of the 108 males included in the pool of candidate males, 20.4% (N = 22) were assigned as a sire to at least one offspring. Of the 54 offspring assigned to a sire 44.4% (N = 24) mismatched a parent at one locus. In all of these cases the mismatch was compatible with a null allele at the Pha28 locus. For the 13 chicks that could not be assigned paternity with high confidence, chicks either matched more than one sire perfectly, or match no male with a perfect genotypic match, excluding the Pha28 locus.

4.4.2 Relationship between male quality and measures of song performance

Most of the male physiology, activity, and spatial measures were not significantly inter-correlated (data not shown), excepting the comparison of percent time males spent on their

territories and percent time spent singing ($r = 0.81$, $p < 0.001$). Variance inflation factors were all below 3.7, suggesting that multicollinearity among these predictors was not problematic.

Table 4.2. Two song performance measures in relation to metrics of individual male physiology, spatial location, and activity levels.

Predictor	Performance measure	N	Effect estimate \pm SE	R ² (marginal)	t _{df} (p)
Age	Amplitude	107	-0.001 \pm 0.002	0.001	t ₈₄ =-0.35 (0.73)
	Song variability	107	-0.062 \pm 0.023	0.049	t ₈₄ =-2.73 (0.01)
Weight	Amplitude	104	0.003 \pm 0.016	0.000	t ₈₁ =0.19 (0.85)
	Song variability	104	0.121 \pm 0.183	0.004	t ₈₁ =0.66 (0.51)
Wing chord	Amplitude	106	0.004 \pm 0.003	0.013	t ₈₃ =1.29 (0.20)
	Song variability	106	0.033 \pm 0.032	0.008	t ₈₃ =1.06 (0.29)
Head length	Amplitude	104	0.005 \pm 0.004	0.013	t ₈₁ =1.26 (0.21)
	Song variability	104	0.029 \pm 0.046	0.003	t ₈₁ =0.64 (0.52)
Tail length	Amplitude	99	0.003 \pm 0.001	0.026	t ₇₇ =1.77 (0.08)
	Song variability	99	0.015 \pm 0.016	0.006	t ₇₇ =0.91 (0.37)
Tail extension	Amplitude	97	0.002 \pm 0.002	0.014	t ₇₄ =1.26 (0.21)
	Song variability	97	0.032 \pm 0.017	0.028	t ₇₄ =1.89 (0.06)
Ectoparasites	Amplitude	104	0 \pm 0	0.000	t ₈₂ =0.19 (0.85)
	Song variability	104	0 \pm 0.001	0.001	t ₈₂ =-0.36 (0.72)
Distance to lek center	Amplitude	107	0 \pm 0	0.001	t ₈₄ =0.26 (0.8)
	Song variability	107	0.001 \pm 0.002	0.005	t ₈₄ =0.65 (0.52)
Distance to dialect center	Amplitude	107	0 \pm 0	0	t ₈₄ =-0.11 (0.91)
	Song variability	107	0.006 \pm 0.004	0.023	t ₈₄ =1.67 (0.1)
% time spent on territory	Amplitude	76	0 \pm 0	0.009	t ₅₇ =-0.95 (0.34)
	Song variability	76	0.003 \pm 0.002	0.039	t ₅₇ =1.91 (0.06)
% time spent singing	Amplitude	76	0 \pm 0	0	t ₅₇ =-0.11 (0.91)
	Song variability	76	0.002 \pm 0.002	0.021	t ₅₇ =1.42 (0.16)

Each predictor – response combination consisted of a separate general linear mixed model with dialect nested within lek as a random factor. Song variability was log transformed. R² (marginal) was calculated following recommendations by Nakagawa and Schielzeth (2013) and can be interpreted as the proportion of variance explained by the predictor. No effects remained significant after correction for multiple testing.

In comparisons between all putative male quality measures and song variability, none were significantly correlated with either individual song variability or amplitude performance (Table 4.2). Male age was weakly and negatively correlated with individual song variability, but the effect was not robust to correction for multiple comparisons [Table 4.2, Figure 4.3]. No male quality measures were significantly correlated with amplitude performance.

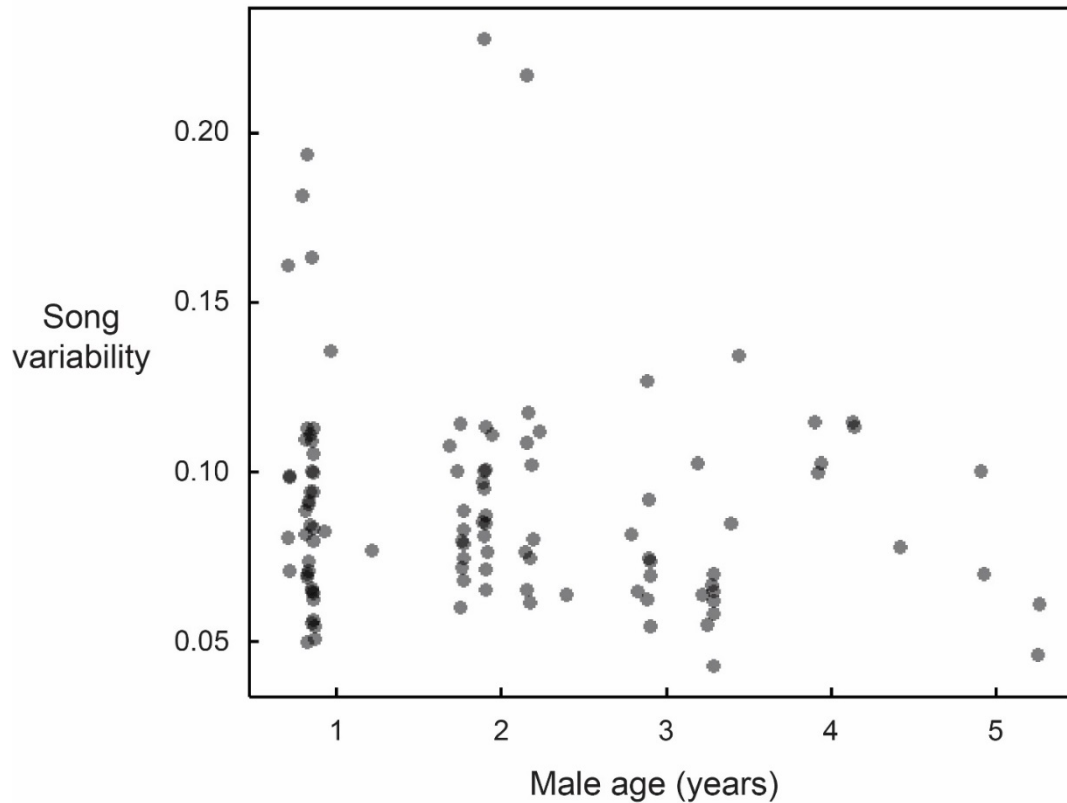


Figure 4.3. Relationship between male age and song variability. There was a weak negative relationship between male age and the average variability of his song (indicating that age is positively correlated with song consistency). The significance of this relationship was not robust to correction for multiple comparisons (see Results, Table 4.2).

4.4.3 Quality and song versus individual RS

All measures of male physiology, spatial location, activity patterns, and song performance were unrelated to male reproductive success (Table 4.3). One comparison, the relationship between mating success and the length of the male's tail extension, showed a weak (negative) correlation but significance was not robust to multiple testing [Figure 4.4, Table 4.3].

Table 4.3. The relationship between male physiology, spatial location, and activity patterns on individual reproductive success.

Predictor	N	Effect estimate \pm SE	LR R^2 (adjusted)	Z_{df} (p)
Amplitude performance	107	-0.246 \pm 9.815	0.056	$Z_{102}=-0.03$ (0.98)
Song variability	107	0.546 \pm 10.181	0.043	$Z_{102}=0.05$ (0.96)
Age	108	-0.001 \pm 0.209	0.036	$Z_{103}=-0.01$ (1)
Weight	105	-1.555 \pm 1.606	0.023	$Z_{100}=-0.97$ (0.33)
Wing chord	107	-0.051 \pm 0.273	0.005	$Z_{102}=-0.19$ (0.85)
Head length	105	-0.045 \pm 0.449	0.027	$Z_{100}=-0.1$ (0.92)
Tail length	100	-0.166 \pm 0.144	0.123	$Z_{95}=-1.16$ (0.25)
Tail extension	98	-0.398 \pm 0.169	0.180	$Z_{93}=-2.36$ (0.02)
Ectoparasites	105	0.012 \pm 0.014	0.110	$Z_{100}=0.86$ (0.39)
Distance to lek center	108	-0.006 \pm 0.015	0.002	$Z_{103}=-0.41$ (0.68)
Distance to dialect center	108	0.002 \pm 0.034	0.021	$Z_{103}=0.05$ (0.96)
% time spent on territory	76	0 \pm 0.020	0.493	$Z_{71}=0.02$ (0.98)
% time spent singing	76	-0.003 \pm 0.021	0.500	$Z_{71}=-0.12$ (0.90)

Each predictor – response combination consisted of a separate generalized linear mixed model with a negative binomial distribution, with dialect nested within lek as a random factor. LR R^2 (adjusted) is a maximum likelihood measure of the superiority of the fitted model to a null model including only the random effect, so is not directly interpretable as a proportion of variance explained by the predictor (Nakagawa and Schielzeth 2013). No effects remained significant after correction for multiple testing.

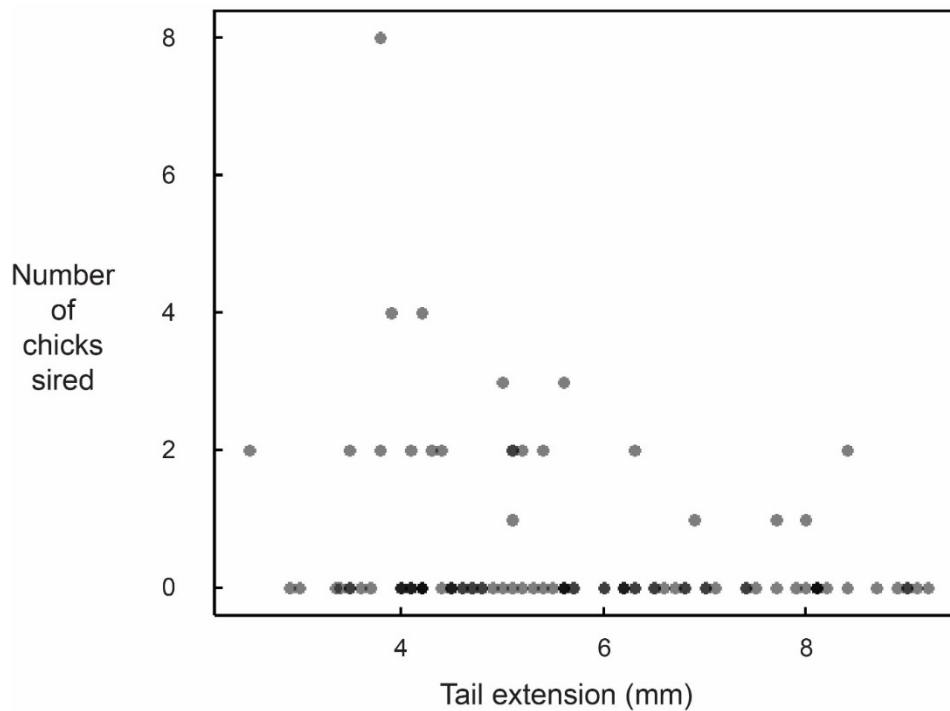


Figure 4.4. Relationship between tail extension and individual mating success. Male mating success (number of chicks sired) was weakly and negatively correlated with the length of the white tail extension. The significance of this relationship was not robust to corrections for multiple comparisons (see Results, Table 4.3).

4.4.4 Relative song performance and spatial centrality versus relative reproductive success (H1 and H2)

Mating success relative to other members of an individual's dialect was unrelated to relative measures of song variability and amplitude performance, and was also unrelated to that male's acoustic centrality (i.e. the similarity of that male's song to the average song of the dialect; Table 4.4).

Table 4.4. Male mating success relative to that of other dialect members and its relationship with three relative measures of song performance and one measure of dialect spatial centrality.

Predictor	N	Effect estimate \pm SE	Partial r	T _{df} (p)
Relative song variability	93	0.013 \pm 0.076	0	t ₈₉ =0.17 (0.87)
Relative amplitude performance	93	-0.440 \pm 0.280	0.025	t ₈₉ =-1.57 (0.12)
Acoustic centrality	93	0.158 \pm 2.179	0	t ₈₉ =0.07 (0.94)
Distance to dialect center	105	-0.018 \pm 0.0081	0.046	t ₈₉ =-2.15 (0.04)

Each predictor – response combination consisted of a separate general linear model with covariates of tail extension and age. Relative reproductive success was transformed as log(relative success +1). No effects were significant after false discovery rate corrections, or after exclusion of covariates from the model.

Relative mating success was also not significantly related to an individual's spatial centrality within his dialect. This comparison revealed a slight trend in the expected direction (i.e. individuals farther from the center of their dialects had relatively lower siring success), but the effect size was extremely small, and the effect was not robust to corrections for multiple testing (Table 4.4). The inclusion / exclusion of the age and tail extension covariates had no impact on significance of any of the factors.

4.4.5 Dialect-level performance versus per capita reproductive success (H3)

No measure of dialect-level song performance (average amplitude performance, average song variability, or dialect dispersion) related to per capita reproductive success of males in that dialect group (Table 4.5). Per capita success was also unrelated to measures of dialect size (area,

density, number of males) and activity levels (average % time singing and present on the dialect; Table 4.5).

Table 4.5. Relationship between per capita reproductive success among dialects and dialect-level song measures, spatial features, and activity patterns.

Predictor	N	Effect estimate \pm SE	Partial r	T _{df} (p)
Average variability	22	-0.647 \pm 6.157	0.001	t ₁₈ =-0.11 (0.92)
Average amplitude performance	19	14.212 \pm 7.936	0.115	t ₁₅ =1.79 (0.09)
Dialect dispersion	19	0.930 \pm 2.500	0	t ₁₅ =0.37 (0.72)
Dialect area	22	0 \pm 0	0.150	t ₁₈ =1.94 (0.07)
Dialect density	22	-30.374 \pm 18.433	0.114	t ₁₈ =-1.65 (0.12)
Dialect number	22	0.066 \pm 0.041	0.108	t ₁₈ =1.6 (0.13)
Average % time on territory	18	-0.003 \pm 0.014	0.023	t ₁₄ =-0.24 (0.81)
Average % time singing	18	-0.002 \pm 0.014	0.020	t ₁₄ =-0.12 (0.91)

Each predictor – response combination consisted of a separate general linear model with covariates of average tail extension and average age of males in the dialect. No effects were significant after corrections for multiple testing, or after exclusion of covariates from the model.

4.4.6 Female mate fidelity, and male deceptive song mimicry (H4)

In 2013 females were detected remating a total of nine times (16.7% of all offspring sampled). Of these cases five (55.6% of rematings, 9.2% of all matings) involved sires from the same dialect, and three of these (33.3% of rematings, 5.6% of all matings) involved different sires from the same dialect [Figure 4.5]. Both observed patterns of remating (i.e. proportion of dialect faithful matings, regardless of male ID, and proportion of dialect-faithful matings involving different males) were highly statistically significant. The remaining four cases of remating involved the choice of males from the same lek but different dialects (N=2) and from different leks (N=2).

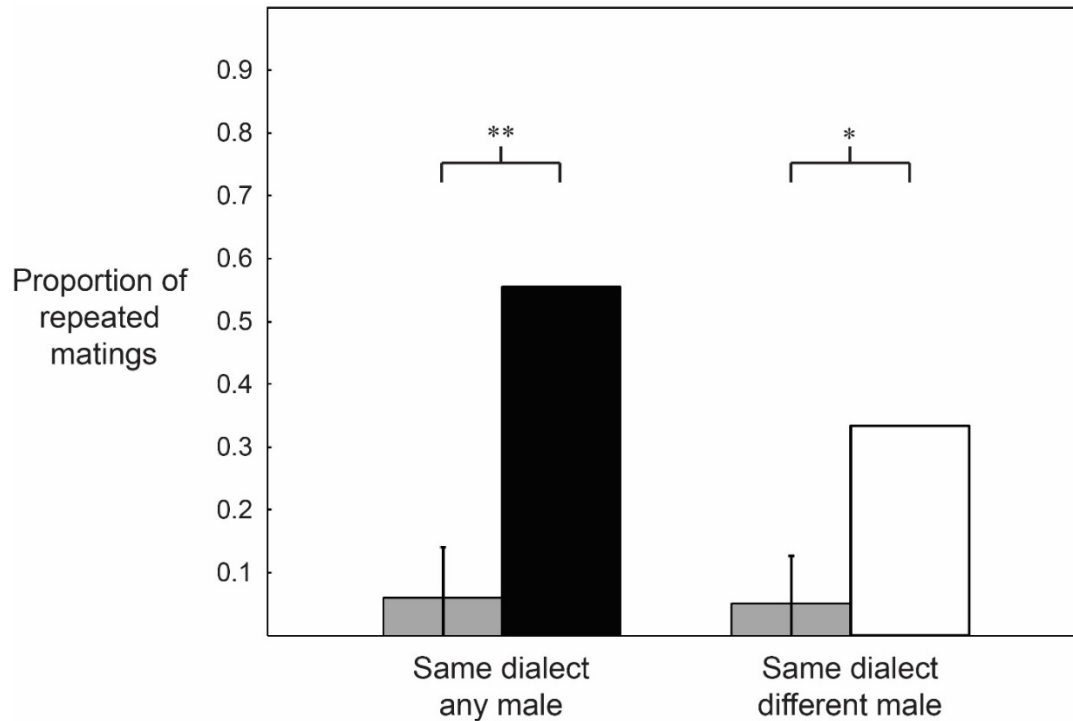


Figure 4.5. Expected (gray bars) and observed (black and white bars) proportion of repeated matings with any male (including the same male) in the same dialect and with only different males in the same dialect (representing potential cases of mistaken identity). * $p = 0.007$, ** $p = 0.001$

4.5 Discussion

4.5.1 The use of dialectal song variation in mate assessment

The results I have presented here indicate that dialectal structure in little hermit leks is not used by females in assessment of male quality (hypotheses 1 – 3). Below, I address possible interpretations for these results.

4.5.1.1 Statistical power

First, despite large sample sizes for the song, behavioral, spatial, and physiological datasets, the reproductive success data ($N=54$ chicks) was comparatively small, and likely impacted statistical power. It is probable that only a small proportion of nests were sampled from those initiated by females who mated with males on the field site, which likely resulted in the

incorrect assignment of candidate sires who actually sired some offspring to low (or zero) reproductive success status. Two factors, however, suggest that this may not be a major concern. First, the reproductive skew observed on leks was similar to skew values reported in other studies. For instance, on one lek a single male was assigned as a sire to eight offspring, one to three, three to two, three to one, and nineteen were assigned no offspring. This high skew is similar to skew patterns observed in a number of other lekking species (Alatalo et al. 1992; Höglund and Alatalo 1995; Ryder et al. 2009), and it is highly unlikely that this pattern of skew would emerge randomly (reviewed in Wiley 1991). Thus, although a number of offspring are likely to have been missed, the relative relationships between the reproductive success status of highly successful and less successful males is likely robust. Second, 16.7% of all offspring sampled were from renesting attempts of previously sampled females. Since nest searching (and female netting for radiotracking) was done systematically (i.e. not targeting particular females), the relatively high resampling rate of female nesting attempts suggests that a greater proportion of total nesting attempts was sampled than might otherwise be apparent. It is therefore more likely that the lack of support for the use of dialectal variation by females in assessment of male quality reflects biological reality rather than a statistical artifact.

4.5.1.2 Indicators of male quality

The strongest association between any measure of male quality and song performance (male age versus song variability) explained only 4.6% of the variation in performance [Table II, Figure 4.3], suggesting that none of the physiological, behavioral, or spatial measures of male quality were reliable predictors of performance (Table II). This conclusion is further supported by the lack of a relationship between song performance and male reproductive success (Table III). It is therefore unlikely that vocal performance would have evolved as a signal of quality in

little hermits, thus preventing its use in dialect-level assessment by females. I cannot entirely rule out the possibility that an unmeasured vocal performance measure would reveal the expected relationship with male quality, however, common measures such as repertoire size (Catchpole 1986), and vocal deviation (Podos 1997) are not relevant in species with single-song repertoires and without trilled vocalizations. Vocal performance, then appears to be a poor target for female assessment of male quality.

The context in which advertisement song versus courtship displays are initiated in little hermits may provide the best explanation for the lack of association between dialectal patterns of song performance and male reproductive success. Males in this species sing to defend territories and attract mates to their territories, but once a female arrives the male ceases singing and begins an elaborate visual display (described by Snow 1968; personal observation). Although female assessment of the vocal display is likely important for male fitness, in that it precedes the visual display, females may base their assessment of males on fine-scale aspects of the performance of the visual display, which is always given at close-range.

4.5.1.3 Dialects in relative quality assessment

Although the lack of a link between individual song performance and reproductive success suggests it is not used in female choice, if females are unable to compare performance between individuals in different dialects, this relationship might not be apparent. The results of this analysis, however, reject this possibility as no relative measure of song performance was linked to male reproductive success (Table IV).

There is potential that other sexually selected male traits, such as the visual display, that are used in assessment may reveal homogenizing effects by females to make comparisons easier. Low quality males, however, may destabilize this effect. If females prefer to assess males that

share similar displays because it makes assessment more accurate, low quality males may be selected to avoid standardizing their displays with high quality males, to avoid being easily distinguished from them. This may result in assortative segregation of males by attractiveness (“assortative lekking”), which could diminish or negate the benefits of relative comparisons (Bateson and Healy 2005; Oh and Badyaev 2010; Patricelli et al. 2011). It is unclear what kind of stable equilibrium point, if any would be reached in this scenario.

4.5.1.4 Dialects as indicators of queue membership and rank

A second hypothesis for the use of dialects by females in the assessment of male quality is that dialect membership may indicate the settlement history of individuals on a lek in which males queue for status. Leks in a number of species, including hermits, appear to have orderly queues for dominance status in which males from the periphery slowly move towards one or more foci on the lek as older central males die and younger males settle on the periphery (Stiles and Wolf 1979; McDonald 1993; Kokko et al. 1998; DuVal 2007; Bro-Jørgensen 2011). In little hermits, dialect membership could reveal that a male in the center of the lek may actually have only recently arrived if he is on the periphery of his dialect group. Despite potential selection for females to attend to this information, females do not appear to have preferences for central territories in dialect groups (although there is a trend in the expected direction; Table IV), nor on the lek as a whole (results not shown). This result might reflect the finding of Stiles and Wolf (1979) that central territories in lekking long-billed hermits, *Phaethornis longirostris*, have higher rates of courtship interruption and therefore potential for female harassment.

4.5.1.5 Dialects as multi-male phenotypes

Even in the absence of individual level associations between song performance and mating success, it is possible that females base initial decisions of which groups of males to visit

(and ultimately mate with) based on assessment of a multi-male vocal phenotype, e.g. dialect-level performance. The results I have presented here, however, suggest that this is not the case for little hermits.

Despite marked differences between dialects in the average amplitude performance and song variability of dialect members, as well as the acoustic cohesion of different dialects, these traits do not appear to relate to the success of males in the group. One potential reason for this is that similarity in acoustic structure of dialect members may be insufficient for assessment of group-level traits if individuals in the group are not sufficiently synchronized in time or clustered in space. Indeed, in species with cooperative displays used to attract mates, the degree of spatial clustering and temporal synchrony is often astounding (Trainer and McDonald 1995; Trainer et al. 2002; Vanderbilt et al. 2015). Although dialects in little hermits typically consist of males within auditory contact of one another (and therefore of females searching for mates), clustering and temporal synchrony does not appear to be higher than with neighbors with unshared dialects (Kapoor unpublished manuscript).

In addition, because song performance does not appear to change with age in little hermits, and there is constant turnover of dialect membership as new males arrive, the cohesion of dialect acoustic structure (as a whole) is unlikely to provide information about the age of males. Indeed, in *Chiroxiphia* manakins stable long-term male alliances, as well as age-graded changes in vocal structure are thought to convey honest information about male quality that make these displays useful in mate assessment (Trainer et al. 2002). On the other hand, male age may not be a target for selection in little hermits, as indicated by the results of the current study (Table III). A number of studies have demonstrated in several songbird species that developmental challenges can negatively impact later song copy accuracy in adulthood and that

females show preferences for more accurately copied songs (Nowicki et al. 1998; Nowicki et al. 2002; Searcy et al. 2010). Thus, highly cohesive acoustic structure within dialects might reflect high average male quality despite a mix of males of different ages. Further work is needed to determine whether nutritional or other stressors early in life negatively impact subsequent song structure in little hermits.

4.5.2 Dialects, mate fidelity, and the hotshot hypothesis

Although additional observational and experimental work is required, and sample size limitations demand caution in the interpretation of results, the finding that females may exhibit mate fidelity and that they remate at greater than expected rates with males in the same dialect provides suggestive evidence of the use of deceptive mimicry by males (thus leading to dialects) and is worthy of further attention [Figure 4.5].

One alternative explanation for this result – that females have stable preferences for different dialects – lacks theoretical support, though future work will be needed to rule out this possibility. In addition, I have no direct evidence that males actively use song to deceive females, or that females actually use song, rather than spatial location, to relocate preferred males. Future observational work will be required to determine whether females are more likely to mate with the neighbors of preferred males from the same rather than a different dialect.

There are several reasons this hypothesis deserves further consideration. First, as stated in the introduction, the costs of mate searching may be especially high in hummingbirds given their high metabolic rates and need for patchy and limited nectar resources. This may be especially the case in little hermits as courtship typically involves extremely rapid chases that undoubtedly require a great deal of energy to execute. Tactics that reduce the costs of mate searching, such as mate fidelity, may therefore be common in little hermits. Second, the advertisement song of little

hermits is delivered nearly incessantly during daylight hours during the breeding season, and its structure is typically highly stable among adults (Snow 1968). This, combined with the constantly changing vegetation of male territories (leks are located in areas with large amounts of dead and decaying branches used for perches), may select for the use of song, combined with spatial cues, in the relocation of preferred mates by females. Third, I have demonstrated in Chapter 3 that males settling on leks may be using shared song to deceive territory holders regarding their residency status. The use of shared song as a male-male deceptive signal supports its potential for use as a male-female deceptive signal as well.

4.6 Conclusions

In previous work I have demonstrated that within lek dialects relate to male-male competition. In this study I investigated the possibility that vocal dialects may also be important for female choice. In testing predictions of three hypotheses for the use of song in male quality assessment I found little evidence for tactical use of dialects by females, 1) in relative comparisons of male song performance, 2) as a signal of male rank in a dominance hierarchy, or 3) as a sexually selected multi-male phenotype. However, in the same way in which males may use songs tactically for deception in attaining territories, males may also be using them deceptively to attain matings from mate faithful females. Further work is needed to confirm these results using playbacks, as well as larger sample sizes. These findings provide important insights into the potential mechanisms by which the hotshot phenomenon might operate in other lek mating species.

CHAPTER 5

IMPROVED METHODS FOR COLOR-MARKING

HUMMINGBIRDS

5.1 Abstract

Individual color-marking is an essential tool for studying the behavior of free-living birds. Hummingbirds represent a particular challenge for traditional avian color-marking techniques because of their small size and short tarsi. Although several techniques have been used successfully, retention time of color-markers, and their safety and ease of construction could be improved. Here, I outline two new color-marking methodologies that I used for marking the little hermit (*Phaethornis longuemareus*): a plastic back tag constructed by fusing colored beads and a similar leg tag which is affixed to a metal band fitted around the bird's tarsus. Both tag designs were visible in field conditions, and neither appeared to adversely affect behavior. Back and leg tags had high retention rates within seasons, but the former had a poor retention rate between years.

5.2 Introduction

Marking techniques that minimally impact behavior have proven vital in studies of free-living animals (Craighead and Stockstad 1960; Hagler and Jackson 2001). For studies of hummingbirds, finding safe and effective color-markers has been problematic because their small size limits the number of possible marking techniques. Standard marking techniques, such

as colored leg bands, cannot be used because of the short tarsi of hummingbirds. Techniques that have been used to mark hummingbirds include: (1) brightly colored paints (Stiles and Wolf 1973) or colored fabrics glued to dorsal feathers (Baltosser 1978), and (2) colored streamers attached to a leg (Stiles and Wolf 1973). However, paint can be preened out by short-billed hummingbirds and requires restraining birds until the paint dries. Baltosser (1978) pieced together nylon fabric with glue to make tags that were glued to hummingbird dorsal feathers, but attachment required time for the adhesive to dry. In addition, back tags are lost when birds molt, so they are not suitable for long-term studies.

The first leg-marking method for hummingbirds involved the use of colored plastic streamers attached to a bird's leg (Ortiz-Crespo 1969). Stiles and Wolf (1973) increased the number of possible color combinations and the distinguishability of leg-markers by creating tags with strips of colored tape wrapped around a piece of acetate. The base of the tag was attached to the bird's leg by wrapping the acetate strip around the tarsus and then securing the end to the rest of the tag with glue. Stiles and Wolf's (1973) design was an improvement, but their leg-markers were difficult to apply quickly and attachment involved the use of adhesives near the leg, which could potentially be dangerous if a bird struggled during attachment.

Although these marking techniques have been used successfully, retention time of color-markers and their safety and ease of construction could be improved. Here, I describe two improved tag designs, and evaluate their visibility, durability, and effectiveness in field conditions. Finally, I discuss the advantages and disadvantages of these tags.

5.3 Methods

I studied Little Hermits (*Phaethornis longuemareus*) occupying dense sub-montane rainforest in the Northern Range of Trinidad (10°45' N, 61°16' W) from January to May 2010 and 2011. Birds were captured using mist-nets, marked with a numbered metal band and leg tag ($N = 97$), back tag ($N = 53$), or both ($N = 44$), then released. Only three females were marked; they received back markers and metal bands only. Most marked birds were captured on their territories at six different leks ($N = 163$), but 13 were captured along their foraging routes and 15 were captured at a lek that was not revisited. Tag retention and re-sighting rates were estimated only for individuals tagged on the six leks that were revisited. I subsequently identified 58 birds that lost their markers by their song structure and location, and all had been marked in 2010 with bands later determined to be too large (5.2 mm circumference). Values are presented as means \pm 1 SD. My marking techniques require colored plastic tubes known as Perler beads (Wilton Brands Inc., Woodridge, IL). These beads are available online and from craft supply stores. Beads can be obtained in large quantities for as little as US \$0.001 per bead and come in over 64 colors.

5.3.1 Back tag

Back tags can be constructed using a variable number of beads, but three or four colors are probably the most that can be distinguished at a distance in the field. To make a tag, use a sharp knife or slender pair of scissors to split a bead down the longitudinal axis and again down the cross-sectional axis, creating four equal pieces [Figure 5.1(a)]. Lay out three or four segments of the desired colors (long sides touching) on the ironing paper supplied with the beads [Figure 5.1(b)]. Next, fold the paper over the beads and gently apply a medium-hot iron to the top surface. This will quickly melt the beads together. Once beads have been flattened, lift the iron and allow the tag to cool and re-solidify for ~15 sec. After the tag has cooled, use a sharp

pair of scissors to cut the tag to the desired size and shape. Tags must be kept as small as is practical to avoid interference with flight and other behaviors. Back tags used in my study were 3.34 ± 0.50 mm wide and 9.90 ± 0.48 mm long ($N = 20$). This tag length represents ~11% of the body length of Little Hermits (~9 cm) and, for other species, tag dimensions should be scaled appropriately for larger or smaller birds. Tags can be stored out of contact with direct sunlight until needed.

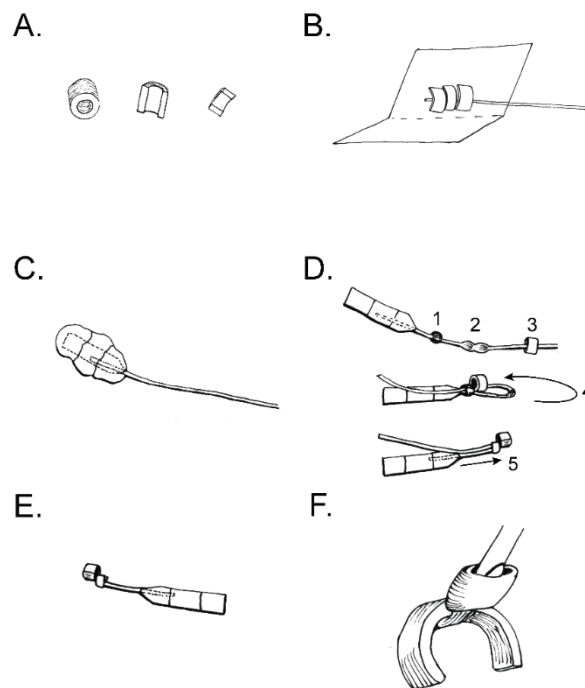


Figure 5.1. Design of the plastic back and leg tags: a) Perler beads showing uncut, longitudinally sliced, and both cross-sectionally and longitudinally sliced (quartered) tubes; b) Layout of beads for forming both leg and back tags using three quartered beads of different colors over fishing line and ironing paper (for back tags, the fishing line is omitted); c) Beads that have been successfully melted onto fishing line; the dotted line represents a cutting guide for forming a tag of the correct shape; d) Trimmed tag with crimping bead (1), flattened section of line (2), and metal band (3), looping the line through the crimping bead (4), and sliding the crimping bead up the line and crimping to tighten the connection around the band (5); e) Completed tag, with excess line trimmed; f) Close-up of a successfully flattened line-connection to the metal band (notice the lack of bulge where the line passes through the band), with crimping bead correctly positioned to allow placement on the bird's tarsus.

Back tags can be attached to hummingbirds by applying a small dab of quick-drying non-toxic cyanoacrylate glue to the tag, then quickly placing the tag on the back contour feathers posterior to the base of the wings and holding the tag in place until the glue dries (15-20 sec).

5.3.2 Leg tag

Leg tags used in my study are adapted from those first described by Stiles and Wolf (1973). The plastic section of the leg tags is made in much the same way as back tags, except that a short segment (10-12 cm) of stiff fishing line (hereafter, line; Berkley FireLine Fused Crystal, 20 lb test, 0.30 mm diameter, Pure Fishing Inc., Spirit Lake, IA) is placed under the color beads before ironing [Figure 5.1(b)]. Iron these beads flat to securely fuse the plastic beads to the line. Next, trim the tag so the leading end tapers with both sides of the tag sloping away from the line at a sharp angle for ~0.5 cm. This taper prevents the tag from snagging on vegetation. After the taper, trim the sides of the tag so they are straight and parallel with one another. Cut the trailing end of the tag perpendicular to the two sides [Figure 5.1(c)]. The appropriate dimensions for tags depend on the mass/size of the species to be marked. However, for a ~3 g hummingbird, my tags never exceeded 1 cm long and 3 mm wide. Next, thread a size #0 crimping bead (Beadalon, Coatesville, PA) onto the line [Figure 5.1(d-1)]. Flatten a small section of the line ~1 cm from the tapered tip of the plastic tag above the crimping bead using banding or needle-nose pliers [Figure 5.1(d-2)]; this keeps the line from bulging in the next step. Fold this flattened section in half and place a closed metal band (the size appropriate for the species + 0.1 mm) onto the line [Figure 5.1(d-3)], then loop the line back over itself and thread it back through the crimping bead [Figure 5.1(d-4)]. Move the metal band so it is held in place by the fold of the flattened section of line. Slide the crimping bead up toward the metal band, pull tight, and crimp the bead to secure the line with banding or needle-nose pliers [Figure 5.1(d-5)].

Cut the trailing end of the line as close to the crimping bead as possible [Figure 5.1(e)]. The portion of the flattened line on the inner surface of the band should be further smoothed by drawing a tapestry needle thicker than the inner diameter of the band through the band. The band will open, but can be reclosed after the line has been sufficiently flattened.

Pre-formed metal bands are not available for use on hummingbirds. However the United States Bird Banding Laboratory (United States Geological Survey) and Canadian Bird Banding Office (Canadian Wildlife Service) provide thin metal sheets for forming hummingbird bands for banders authorized to band hummingbirds in these countries. Outside the United States and Canada, sheets of aluminum 1100 alloy with photolithographed numbers and band-cutting guidelines must be custom ordered (Oniki 1996). Detailed methods for forming and applying hummingbird bands are available in Russell and Russell (2001). Band size calculations should be made for each species, but Oniki (1996) provides a good metric for appropriate band sizes for 31 species of South American hummingbirds.

To make it possible to close the leg tag's band around the tarsus of a hummingbird, it is important to move the crimping bead away from the center of the band (so there is room for the pliers to close). The best way to do this is to slide the crimping bead up or down along the wall of the metal band so it sits against one edge of the band [Figure 5.1(f)]. Leg tags can be stored like this until needed.

To place a tag on a hummingbird's tarsus, open the band by sliding a tapestry needle through it, grasp the band with banding pliers and place it over the tarsus of the bird (the crimping bead should be on the side of the band proximal to the body), and then close the band. Holding the band in place with tweezers, position the crimping bead so that the tag projects perpendicularly away from the leg and downward, i.e., towards the ground when the bird is perched. Check that the band rotates freely on the tarsus, but does not slide easily over the foot; this can be particularly difficult with hermits because their closed feet are nearly the same diameter as their tarsi (see Figure 5.2 and insert).

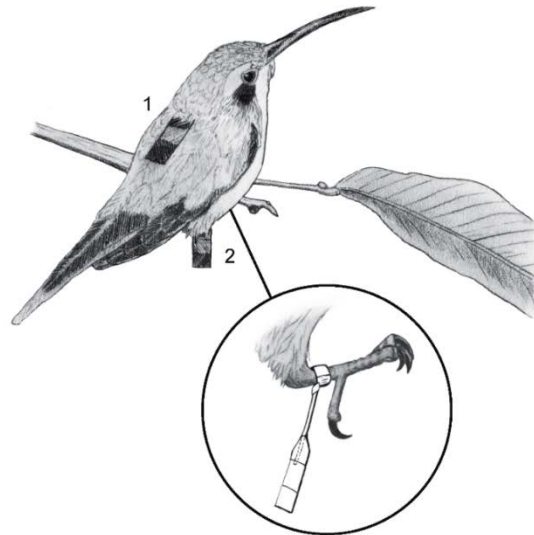


Figure 5.2. Correctly positioned back tag (1) and leg tag (2) on a male little hermit. Insert shows correctly fitted leg tag with crimping bead angled away from the surface of the leg and on the proximal side of the metal band.

5.4 Results

I observed birds using 10 x 42 mm binoculars and was usually able to distinguish the 13 colors used for back and leg tags on both flying and perched birds that were up to ~9 m distant.

Visibility of back tags was restricted to angles allowing a view of the back of the hummingbird,

but leg tags were often visible from both dorsal and ventral angles. When tags were observed in low-light conditions, adjacent similar colors (e.g., light green and toothpaste or pastel lavender and purple) were often difficult to distinguish. Most tag colors showed little sign of fading over a 2-yr period, but several colors faded noticeably [Table 5.1].

Table 5.1. Perler beads used in my study of little hermits and whether bead color faded over a 2-year period.

Color	Faded (Y/N)?	Final color (if different)
Black	N	-
Brown	N	-
Dark blue	N	-
Dark green	N	-
Grey	N	-
Kiwi lime	N	-
Orange	Y	Clear
Pastel lavender	N	-
Purple	Y	Pale lavender
Red	Y	Pale orange
Toothpaste	N	-
White	N	-
Yellow	N	-

5.4.1 Back tag

The average mass of back tags with three quartered beads was 17.1 ± 3.9 mg ($N = 10$). The mean mass of hummingbirds was 2.87 ± 0.10 g ($N = 445$), making the tags 0.6% of their body weight.

Birds fitted with back tags typically flew a short distance from the marking site to a branch and remain perched 5 to 10 min after release. During the end of this period birds would direct preening behavior towards the tags. In every case the bird flew off before 20 min had elapsed. Preening behavior was rarely directed at the tags after this period. In most cases the departing bird foraged on nearby flowers before flying out of view. On several occasions hummingbirds resumed singing only seconds after being released with a new back tag.

Of 94 territorial birds fitted with back tags, 83 (88%) still had tags during the last month of my study (i.e., up to 4 months after initial marking). Nine birds (9%) recaptured and identified by their metal band numbers lost their back tags before the end of the field season (all were re-tagged), and two marked birds (2%) were neither re-sighted nor recaptured during the same period. I either re-sighted or recaptured 48 of 91 birds (53%) with back tags by the end of the subsequent field season, and only one (2%) had retained its tag. Because most back tags were lost between field seasons, I was unable to estimate average retention time.

5.4.2 Leg tag

The mean mass of leg tags with three quartered beads was 25.4 ± 5.1 mg; leg tags were 0.89% of average total body weight.

Birds banded with leg tags reacted similarly to those fitted with back tags, however, preening behavior was directed at the tag's metal band and the bird would occasionally shake the tagged leg vigorously for short periods (~1 sec). Although leg-directed preening behavior lasted up to 20 min post-marking, birds almost entirely ignored the tags after this period.

I fitted 113 territorial birds with leg tags. Some birds were recaptured within and between years to determine if tag-related injuries had occurred ($N = 5$) and to clean ($N = 22$) or replace lost tags ($N = 5$). No recaptured birds showed signs of injury or irritation. Of 113 tagged birds, 103 (91%) retained tags within a season, five (4%) lost them, and five (4%) were not re-sighted or recaptured. I re-sighted or recaptured 43 birds (38%) in the year after marking, with 38 (88%) retaining tags and five (12%) losing tags.

5.5 Discussion

5.5.1 Back tag

The back tags I used can be made in seconds, the material used is durable and light, and birds being marked must be restrained for only a few seconds to allow the cyanoacrylate glue to dry.

Tag retention was high during the first 4 months after marking, but was near zero between years. Back tags are non-permanent, but may be useful for studies where markers are needed for only a few months (e.g., a breeding season).

A concern of any color-marking technique is that markers may increase predation risk or affect social interactions (e.g., Kessler 1964; Cuthill et al. 1997). I did not compare the survivorship and behavior of marked and unmarked birds so cannot address the possibility that back tags increase mortality risk or alter social interactions.

However, back tag weight was far less than the commonly accepted auxiliary marker weight thresholds for marking vertebrates, i.e., 3% of total body weight (Bird Banding Laboratory), and 5% (Kenward 2001, p. 127). In addition, anecdotal evidence suggests that back tags did not affect behavior beyond the first 30 min post-release. The back tags on nesting females ($N = 3$) were completely obscured when they were incubating because hermits draw their tails over their backs when on nests. Therefore, it is unlikely that back tags would increase predation risk of incubating female hermits. Typical hummingbirds (*Trochilinae*), however, do not sit on the nest like hermits and would likely be more conspicuous with back tags than without. Also, back tags may draw a predator's attention to a nest as females fly to and from it, and may make both males and females more conspicuous to predators in general (but see Cresswell et al. 2007; Roche et al. 2010). Clearly, more information is needed before it can be concluded that back tags do not affect survivorship.

5.5.2 Leg tag

The leg tags I used were cheap, easy to apply, durable and long lasting. Whereas Stiles and Wolfs' (1973) acetate leg tag design is "tricky to master" and requires completing the tag while the bird is being restrained, attaching the leg tag described here takes under five sec and requires only the application of a metal band to the tarsus. The tag also avoids the potential dangers associated with using adhesives near the feet, and it eliminates the rough edges of plastic colored tape, which might catch on vegetation.

Previous versions of the leg tag used in this study, which did not use a blank metal band, dropped off the tarsi too easily, were not durable, or irritated the legs of the birds and had to be removed. The addition of a metal band to hold the tag, as opposed to directly attaching the line of the tag to the bird's leg, provides a rigid form that prevents the line from kinking and cutting into the leg.

Although tags remained attached to most birds marked, colored plastic pieces were not always clearly distinguishable due to fading of some colors and to deposition of grime on tags over a year old. In the latter cases, tags had to be cleaned between years. Testing a broader range of bead colors for UV resistance or using a UV protective varnish may provide a solution to color fading.

Retention rates of leg tags were high both within (91%) and between (88%) seasons. However, because identification of individuals was limited to birds that retained at least their numbered aluminum band, birds that lost all markers were not factored into these retention rates. Loss of leg tags by these birds was likely a result of incorrect band sizing (because both aluminum bands were shed) rather than a flaw in tag design. Correct band sizes are vital to the success of these tags because bands that are too tight put birds at risk for leg injuries and tags that are too loose are likely to fall off or slide over the foot.

Leg tags weighed less than the 2% threshold adopted by the Bird Banding Laboratory for auxiliary markers on the legs of birds and did not appear to affect behavior beyond the period immediately following release. However, because female hummingbirds use their feet to move eggs and shape nests, leg tags might impair their ability to build and maintain nests (Waser and Calder 1975). I did not place leg tags on females in my study, and caution against use of leg tags on female hummingbirds for the reasons described by Waser and Calder (1975).

Leg tags are clearly more useful than back tags for studies requiring permanent color-marking of hummingbirds. Although the marking techniques used in my study were designed for and tested on one species of small hummingbird, they would likely be useful with other species of hummingbirds.

BIBLIOGRAPHY

- Alatalo, R. V., J. Höglund, A. Lundberg, and W. J. Sutherland. 1992. Evolution of black grouse leks: female preferences benefit males in larger leks. *Behavioral Ecology* 3:53–59.
- Alexander, R. D. 1974. The evolution of social behavior. *Annual Review of Ecology, Evolution, and Systematics* 5:325–383.
- Almeida, H., D. Guedes, W. Meira Jr, and M. Zaki. 2011. Is there a best quality metric for graph clusters? In G. Dimitrios, T. Hofmann, D. Malerba, & V. Michalis, eds., *Machine Learning and Knowledge Discovery in Databases*, volume 6911 of *Lecture Notes in Computer Science* (pp. 44–59). Springer, Berlin.
- Altmann, J. 2012. Observational study of behavior: sampling methods. *Behaviour* 49:227–267.
- Anderson, D. R., and K. P. Burnham. 2002. Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management* 66:912–918.
- Aoki, K. 1989. A sexual-selection model for the evolution of imitative learning of song in polygynous birds. *The American Naturalist* 134:599–612.
- Araya-Salas, M., and T. Wright. 2013. Open-ended song learning in a hummingbird. *Biology Letters* 9:1–5.
- Auger, F., and P. Flandrin. 1995. Improving the readability of time-frequency and time-scale representations by the reassignment method. *IEEE Transactions on Signal Processing* 43:1068–1089.
- Badyaev, A. V, R. L. Young, K. P. Oh, and C. Addison. 2008. Evolution on a local scale: developmental, functional, and genetic bases of divergence in bill form and associated changes in song structure between adjacent habitats. *Evolution* 62:1951–64.
- Baker, M. C. 1982. Genetic population structure and vocal dialects in *Zonotrichia* (Emberizidae). In D. E. Kroodsma & E. Miller, eds., *Acoustic Communication in Birds* (pp. 209–232). Comstock Publishing Associates, Ithaca, NY.

- Baker, M. C., and M. A. Cunningham. 1985. The biology of bird-song dialects. *Behavioral and Brain Sciences* 8:85–133.
- Ballentine, B., J. Hyman, and S. Nowicki. 2004. Vocal performance influences female response to male bird song: an experimental test. *Behavioral Ecology* 15:163–168.
- Balsby, T. J. S., and J. W. Bradbury. 2009. Vocal matching by orange-fronted conures (*Aratinga canicularis*). *Behavioural Processes* 82:133–139.
- Baltosser, W. 1978. New and modified methods for color-marking hummingbirds. *Bird Banding* 49:47–49.
- Baptista, L. 1985. The functional significance of song sharing in the white-crowned sparrow. *Canadian Journal of Zoology* 63:1741–1752.
- Bartón, K. 2015. MuMIn: Multi-Model Inference.
- Bates, D., M. Maechler, B. Bolker, and S. Walker. 2015. lme4: Linear mixed-effects models using Eigen and S4.
- Bateson, M., and S. D. Healy. 2005. Comparative evaluation and its implications for mate choice. *Trends in Ecology and Evolution* 20:659–64.
- Beecher, M. D., S. E. Campbell, and J. M. Burt. 1994. Song perception in the song sparrow: birds classify by song type but not by singer. *Animal Behaviour* 47:1343–1351.
- Beecher, M. D., P. K. Stoddard, S. E. Campbell, and C. L. Horning. 1996. Repertoire matching between neighbouring song sparrows. *Animal Behaviour* 51:917–923.
- Beehler, B. B. M., and M. S. M. Foster. 1988. Hotshots, hotspots, and female preference in the organization of lek mating systems. *The American Naturalist* 131:203–219.

- Borgia, G. 1985. Bower destruction and sexual competition in the satin bowerbird (*Ptilonorhynchus violaceus*). *Behavioral Ecology and Sociobiology* 18:91–100.
- Botero, C. A., J. M. Riveros, and S. L. Vehrencamp. 2007. Relative threat and recognition ability in the responses of tropical mockingbirds to song playback. *Animal behaviour* 73:661–669.
- Botero, C. A., and S. L. Vehrencamp. 2007. Responses of male tropical mockingbirds to variation in within-song and between-song versatility. *The Auk* 124:185–196.
- Boughman, J. W. 1997. Greater spear-nosed bats give group-distinctive calls. *Behavioral Ecology and Sociobiology* 40:61–70.
- Bradbury, J. W., and R. M. Gibson. 1983. Leks and mate choice. In P. Bateson, ed., *Mate Choice* (pp. 109–138). Cambridge University Press, Cambridge, MA.
- Bradbury, J. W., S. L. Vehrencamp, and R. M. Gibson. 1985. Leks and the unanimity of female choice. In P. J. Greenwood, P. H. Harvey, & M. Slatkin, eds., *Evolution: essays in honour of John Maynard Smith* (pp. 301–314). Cambridge University Press, Cambridge, MA.
- Bretagnolle, V. 1996. Acoustic communication in a group of nonpasserine birds, the petrels. In D. E. Kroodsma & E. H. Miller, eds., *Ecology and Evolution of Acoustic Communication in Birds* (pp. 160–178). Comstock Publishing Associates, Ithaca, NY.
- Briefer, E., T. Aubin, K. Lehongre, and F. Rybak. 2008a. How to identify dear enemies: the group signature in the complex song of the skylark *Alauda arvensis*. *The Journal of Experimental Biology* 211:317–326.
- Briefer, E. F., F. Rybak, and T. Aubin. 2013. Does true syntax or simple auditory object support the role of skylark song dialect? *Animal Behaviour* 86:1131–1137.
- Briefer, E., F. Rybak, and T. Aubin. 2008b. When to be a dear enemy: flexible acoustic relationships of neighbouring skylarks, *Alauda arvensis*. *Animal Behaviour* 76:1319–1325.

- Briefer, E., F. Rybak, and T. Aubin. 2010. Are unfamiliar neighbours considered to be dear-enemies? *PLoS One* 5:1–7.
- Brindley, E. L. 1991. Response of European robins to playback of song: neighbour recognition and overlapping. *Animal Behaviour* 41:503–512.
- Bro-Jørgensen, J. 2011. Queuing in space and time reduces the lek paradox on an antelope lek. *Evolutionary Ecology* 25:1385–1395.
- Brownstein, M., J. Carpten, and J. Smith. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20:1004–1010.
- Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach (2nd ed.). Springer-Verlag, New York, NY.
- Burt, J. M., S. E. Campbell, and M. D. Beecher. 2001. Song type matching as threat: a test using interactive playback. *Animal Behaviour* 62:1163–1170.
- Byers, B. E. 2007. Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behavioral Ecology* 18:130–136.
- Campbell, S. E., M. D. Beecher, P. K. Stoddard, P. Loesche, M. D. Beecher, P. Loesche, and S. E. Campbell. 1992. Memory does not constrain individual recognition in a bird with song repertoires. *Behaviour* 122:274–287.
- Cardoso, G. C. 2013. Using frequency ratios to study vocal communication. *Animal Behaviour* 85:1529–1532.
- Cardoso, G. C., J. W. Atwell, Y. Hu, E. D. Ketterson, and T. D. Price. 2012. No correlation between three selected trade-offs in birdsong performance and male quality for a species with song repertoires. *Ethology* 118:584–593.

- Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2007. Inferring performance in the songs of dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology* 18:1051–1057.
- Cardoso, G. C., J. W. Atwell, E. D. Ketterson, and T. D. Price. 2009. Song types, song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology* 20:901–907.
- Catchpole, C. K. 1986. Song repertoires and reproductive success in the great reed warbler *Acrocephalus arundinaceus*. *Behavioral Ecology and Sociobiology* 19:439–445.
- Christie, P. J., D. J. Mennill, and L. M. Ratcliffe. 2004. Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. *Behavioral Ecology and Sociobiology* 55:341–348.
- Clark, C. W., P. Marler, and K. Beeman. 1987. Quantitative analysis of animal vocal phonology. *Ethology* 76:101–115.
- Clutton-Brock, T. H., O. F. Price, and a D. C. Maccoll. 1992. Mate retention, harassment, and the evolution of ungulate leks. *Behavioral Ecology* 3:234–242.
- Cockburn, A., A. H. Dalziel, C. J. Blackmore, M. C. Double, H. Kokko, H. L. Osmond, N. R. Beck, et al. 2009. Superb fairy-wren males aggregate into hidden leks to solicit extragroup fertilizations before dawn. *Behavioral Ecology* 20:501–510.
- Concannon, M. R., A. C. Stein, and J. A. C. Uy. 2012. Kin selection may contribute to lek evolution and trait introgression across an avian hybrid zone. *Molecular Ecology* 21:1477–1486.
- Courvoisier, H., and T. Aubin. 2014. When neighbours are not “dear enemies”: a study in the winter wren, *Troglodytes troglodytes*. *Animal Behaviour* 90:229–235.
- Craighead, J. J., and D. S. Stockstad. 1960. Color marker for big game. *The Journal of Wildlife Management* 24:435–438.

- Cramer, E. R. A. 2012. Vocal deviation and trill consistency do not affect male response to playback in house wrens. *Behavioral Ecology* 24:412–420.
- Cramer, E. R. A. 2013a. Measuring consistency: spectrogram cross-correlation versus targeted acoustic parameters. *Bioacoustics* 22:247–257.
- Cramer, E. R. A. 2013b. Physically challenging song traits, male quality, and reproductive success in house wrens. *PLoS One* 8:e59208.
- Cresswell, W., J. Lind, J. L. Quinn, J. Minderman, and D. P. Whitfield. 2007. Ringing or colour-banding does not increase predation mortality in redshanks *Tringa totanus*. *Journal of Avian Biology* 38:309–316.
- Cuthill, I. C., S. Hunt, C. Cleary, and C. Clark. 1997. Colour bands, dominance, and body mass regulation in male zebra finches (*Taeniopygia guttata*). *Proceedings of the Royal Society B: Biological Sciences* 264:1093–1099.
- Cynx, J., and M. Shapiro. 1986. Perception of missing fundamental by a species of songbird (*Sturnus vulgaris*). *Journal of Comparative Psychology* 100:356–360.
- Derryberry, E. P. 2009. Ecology shapes birdsong evolution: variation in morphology and habitat explains variation in white-crowned sparrow song. *The American Naturalist* 174:24–33.
- Díaz-Muñoz, S. L., E. H. DuVal, A. H. Krakauer, and E. A. Lacey. 2014. Cooperating to compete: altruism, sexual selection and causes of male reproductive cooperation. *Animal Behaviour* 88:67–78.
- Dochtermann, N. A., and S. H. Jenkins. 2011. Developing multiple hypotheses in behavioral ecology. *Behavioral Ecology and Sociobiology* 65:37–45.
- Dooling, R. 1982. Auditory perception in birds. In D. Kroodsma, E. Miller, & H. Ouellet, eds., *Acoustic Communication in Birds*, volume 1 (pp. 95–130). Academic Press, Inc., New York, NY.

- Dooling, R. J., and R. J. Haskell. 1978. Auditory duration discrimination in the parakeet (*Melopsittacus undulatus*). *The Journal of the Acoustical Society of America* 63:1640–1642.
- Dooling, R., B. Lohr, and M. Dent. 2000. Hearing in birds and reptiles. In R. Dooling, R. Fay, & A. Popper, eds., *Comparative Hearing: Birds and Reptiles* (p. 380). Springer-Verlag, New York.
- DuVal, E. H. 2007. Adaptive advantages of cooperative courtship for subordinate male lance-tailed manakins. *The American Naturalist* 169:423–432.
- DuVal, E. H. 2013. Female mate fidelity in a Lek mating system and its implications for the evolution of cooperative lekking behavior. *The American Naturalist* 181:213–222.
- Duval, E. H., and J. A. Kapoor. 2015. Causes and consequences of variation in female mate search investment in a lekking bird. *Behavioral Ecology* 26:1537–1547.
- DuVal, E. H., and B. Kempenaers. 2008. Sexual selection in a lekking bird: the relative opportunity for selection by female choice and male competition. *Proceedings of the Royal Society B: Biological Sciences* 275:1995–2003.
- Edwards, S. V., S. B. Kingan, J. D. Calkins, C. N. Balakrishnan, W. B. Jennings, W. J. Swanson, and M. D. Sorenson. 2005. Speciation in birds: genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences of the United States of America* 102:6550–6557.
- Emlen, S. T. 1971. The role of song in individual recognition in the Indigo Bunting. *Zeitschrift fuer Tierpsychologie* 28:241–246.
- Endler, J. A., L. C. Endler, and N. R. Doerr. 2010. Great bowerbirds create theaters with forced perspective when seen by their audience. *Current Biology* 20:1–6.
- Falls, J. B., and L. G. d'Agincourt. 1981. A comparison of neighbour–stranger discrimination in eastern and western meadowlarks. *Canadian Journal of Zoology* 59:2380–2385.

- Falls, J. B., and M. K. McNicholl. 1979. Neighbor–stranger discrimination by song in male blue grouse. *Canadian Journal of Zoology* 57:457–462.
- Feekes, F. 1977. Colony specific song in *Cacicus cela* (Icteridae, Aves): the password hypothesis. *Ardea* 65:197–202.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist*.
- Fisher, R. A., and J. B. Fisher. 1954. Evolution and bird sociality. In J. Huxley, A. C. Hardy, & E. B. Ford, eds., *Evolution as a Process* (pp. 71–83.). Allen & Unwin, London.
- Fitzsimmons, L. P., N. K. Barker, and D. J. Mennill. 2008. Individual variation and lek-based vocal distinctiveness in songs of the screaming piha. *The Auk* 125:908–914.
- Forstmeier, W., B. Kempenaers, A. Meyer, and B. Leisler. 2002. A novel song parameter correlates with extra-pair paternity and reflects male longevity. *Proceedings of the Royal Society B: Biological Sciences* 269:1479–85.
- Forstmeier, W., and H. Schielzeth. 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner’s curse. *Behavioral Ecology and Sociobiology* 65:47–55.
- Foster, M. S. 1981. Cooperative behavior and social organization of the swallow-tailed manakin (*Chiroxiphia caudata*). *Behavioral Ecology and Sociobiology* 9:167–177.
- Fretwell, S. D. 1972. Populations in a seasonal environment. *Monographs in Population Biology* 5:1–217.
- Gahr, M. 2000. Neural song control system of hummingbirds: comparison to swifts, vocal learning (songbirds) and nonlearning (suboscines) passerines, and vocal learning (budgerigars) and nonlearning (dove, owl, gull, quail, chicken) nonpasserines. *The Journal of comparative neurology* 426:182–96.

- Gelman, A., and Y.-S. Su. 2015. *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*.
- González, C., and J. F. Ornelas. 2009. Song variation and persistence of song neighborhoods in a Lekking hummingbird. *The Condor* 111:633–640.
- González, C., and J. F. Ornelas. 2014. Acoustic divergence with gene flow in a lekking hummingbird with complex songs. *PLoS One* 9:e109241.
- Grant, B. R., and P. R. P. Grant. 1996. Cultural inheritance of song and its role in the evolution of Darwin's finches. *Evolution* 50:2471–2487.
- Grimsley, J. M. S., M. A. Gadziola, and J. J. Wenstrup. 2012. Automated classification of mouse pup isolation syllables: from cluster analysis to an Excel-based “mouse pup syllable classification calculator”. *Frontiers in Behavioral Neuroscience* 6:1–12.
- Grueber, C. E., S. Nakagawa, R. J. Laws, and I. G. Jamieson. 2011. Multimodel inference in ecology and evolution: challenges and solutions. *Journal of Evolutionary Biology* 24:699–711.
- Hagler, J. R., and C. G. Jackson. 2001. Methods for marking insects: current techniques and future prospects. *Annual Review of Entomology* 46:511–543.
- Hall, M. L., S. A. Kingma, and A. Peters. 2013. Male songbird indicates body size with low-pitched advertising songs. *PloS One* 8:e56717.
- Hamilton, M., E. Pincus, A. Di Fiore, and R. Fleischer. 1999. Universal linker and ligation procedures for construction of genomic DNA libraries enriched for microsatellites. *Biotechniques* 27:500–507.
- Handl, J., J. Knowles, and D. B. Kell. 2005. Computational cluster validation in post-genomic data analysis. *Bioinformatics* 21:3201–3212.

- Hartley, D. J., and R. A. Suthers. 1987. The sound emission pattern and the acoustical role of the noseleaf in the echolocating bat, *Carollia perspicillata*. *The Journal of the Acoustical Society of America* 82:1892–1900.
- Hernandez, M. L., A. I. Houston, and J. M. McNamara. 1999. Male rank and optimal lek size. *Behavioral Ecology* 10:73–79.
- Ho, C. E., B. Pesaran, M. S. Fee, and P. P. Mitra. 1998. Characterization of the structure and variability of zebra finch song elements. *Proceedings of the Joint Symposium on Neural Computation* 5:76–83.
- Höglund, J., and R. V. Alatalo. 1995. *Leks*. Princeton University Press, Princeton.
- Holveck, M.-J., and K. Riebel. 2007. Preferred songs predict preferred males: consistency and repeatability of zebra finch females across three test contexts. *Animal Behaviour* 74:297–309.
- Hoogland, J. L., and P. W. Sherman. 1976. Advantages and disadvantages of bank swallow (*Riparia riparia*) coloniality. *Ecological Monographs* 46:33–58.
- Hulse, S., and J. Cynx. 1985. Relative pitch perception is constrained by absolute pitch in songbirds (*Mimus*, *Molothrus*, and *Sturnus*). *Journal of Comparative Psychology* 99:176–196.
- Hurly, T., L. Ratcliffe, and R. Weisman. 1990. Relative pitch recognition in white-throated sparrows, *Zonotrichia albicollis*. *Animal Behaviour* 40:176–181.
- Husak, J. F., and S. F. Fox. 2003. Adult male collared lizards, *Crotaphytus collaris*, increase aggression towards displaced neighbours. *Animal Behaviour* 65:391–396.
- Irwin, D., S. Bensch, and T. Price. 2001. Speciation in a ring. *Nature* 409:333–337.
- Itakura, F. 1975. Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech and Signal Processing* 1:67–72.

- Ito, K., and K. Mori. 1999. Dynamic programming matching as a simulation of budgerigar contact-call discrimination. *Journal of the Acoustical Society of America* 105:552–559.
- Janik, V. 1999. Pitfalls in the categorization of behaviour: a comparison of dolphin whistle classification methods. *Animal behaviour* 57:133–143.
- Jennions, M. D., and M. Petrie. 1997. Variation in mate choice and mating preferences: a review of causes and consequences. *Biological Reviews of the Cambridge Philosophical Society* 72:283–327.
- Jones, A. E., C. Ten Cate, and C. C. Bijleveld. 2001. The interobserver reliability of scoring sonograms by eye: a study on methods, illustrated on zebra finch songs. *Animal Behaviour* 62:791–801.
- Jukema, J., and T. Piersma. 2006. Permanent female mimics in a lekking shorebird. *Biology Letters* 2:161–164.
- Kaib, M., S. Franke, W. Francke, and R. Brandl. 2002. Cuticular hydrocarbons in a termite: phenotypes and a neighbour-stranger effect. *Physiological Entomology* 27:189–198.
- Kalinowski, S. T., M. L. Taper, and T. C. Marshall. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16:1099–1106.
- Kapoor, J. A. 2012. Improved methods for color-marking hummingbirds. *Journal of Field Ornithology* 83:186–191.
- Kenward, R. E. 2001. *A manual for wildlife radio tagging*. Academic Press, Inc., San Diego.
- Keogh, E., and M. Pazzani. 2001. Derivative dynamic time warping. *First SIAM International Conference on Data Mining* (pp. 1–11). Chicago, USA.
- Kessler, F. W. 1964. Avian predation on pheasants wearing differently colored plastic markers. *The Ohio Journal of Science* 64:401–402.

- Klaus, B., and K. Strimmer. 2015. fdrtool: estimation of (local) false discovery rates and higher criticism.
- Kodric-Brown, A., and J. H. Brown. 1978. Influence of economics, interspecific competition, and sexual dimorphism on territoriality of migrant rufous hummingbirds. *Ecology* 59:285–296.
- Kogan, J., and D. Margoliash. 1998. Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: A comparative study. *The Journal of the Acoustical Society of America* 103:2185–2196.
- Kokko, H., and J. Lindstrom. 1996. Kin selection and the evolution of leks: whose success do young males maximize? *Proceedings of the Royal Society B: Biological Sciences* 263:919–923.
- Kokko, H., J. Lindström, R. V. R. Alatalo, and P. T. P. Rintamäki. 1998. Queuing for territory positions in the lekking black grouse (*Tetrao tetrix*). *Behavioral Ecology* 9:376–383.
- Krakauer, A. 2005. Kin selection and cooperative courtship in wild turkeys. *Nature* 434:69–72.
- Kroodsma, D. E. 1976. The effect of large song repertoires on neighbor “recognition” in male song sparrows. *The Condor* 78:97–99.
- Kroodsma, D. E. 2005. *The Singing Life of Birds*. Houghton Mifflin Press, Boston, MA.
- Kroodsma, D., D. Hamilton, J. E. Sánchez, B. E. Byers, H. Fandiño-Mariño, D. W. Stemple, J. M. Trainer, et al. 2013. Behavioral evidence for song learning in the suboscine bellbirds (*Procnias* spp.; Cotingidae). *The Wilson Journal of Ornithology* 125:1–14.
- Lachlan, R. F., and S. Nowicki. 2012. How reliable is song learning accuracy as a signal of male early condition? *The American Naturalist* 180:751–61.

- Lachlan, R. F., L. Verhagen, S. Peters, and C. Ten Cate. 2010. Are there species-universal categories in bird song phonology and syntax? A comparative study of chaffinches (*Fringilla coelebs*), zebra finches (*Taenopygia guttata*), and swamp sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology* 124:92–108.
- Lachlan, R. F., M. N. Verzijden, C. S. Bernard, P.-P. Jonker, B. Koese, S. Jaarsma, W. Spoor, et al. 2013. The progressive loss of syntactical structure in bird song along an island colonization chain. *Current Biology* 23:1896–1901.
- Lambrechts, M. M., and A. A. Dhont. 1995. Individual voice discrimination in birds. *Current Ornithology* 12:115–139.
- Lank, D. B., and C. M. Smith. 1987. Conditional lekking in ruff (*Philomachus pugnax*). *Behavioral Ecology and Sociobiology* 20:137–145.
- Lasiewski, R. C. 1963. Oxygen consumption of torpid, resting, active and flying hummingbirds. *Physiological Zoology* 36:122–140.
- Leadbeater, E., F. Goller, and K. Riebel. 2005. Unusual phonation, covarying song characteristics and song preferences in female zebra finches. *Animal Behaviour* 70:909–919.
- Leader, N., J. Wright, and Y. Yom-Tov. 2000. Microgeographic song dialects in the orange-tufted sunbird (*Nectarinia osea*). *Behaviour* 137:1613–1627.
- Legendre, P., and L. Legendre. 2012. *Numerical Ecology*. Elsevier Ltd, Amsterdam.
- Leiser, J., and M. Itzkowitz. 1999. The benefits of dear enemy recognition in three-contender convict cichlid (*Cichlasoma nigrofasciatum*) contests. *Behaviour* 136:983–1003.
- Lesbarrères, D., and T. Lodé. 2002. Variations in male calls and responses to an unfamiliar advertisement call in a territorial breeding anuran, *Rana dalmatina*: evidence for a “dear enemy” effect. *Ethology Ecology & Evolution* 14:287–295.

- Lohr, B., T. F. Wright, and R. J. Dooling. 2003. Detection and discrimination of natural calls in masking noise by birds: estimating the active space of a signal. *Animal Behaviour* 65:763–777.
- Lokemoen, J. T., and R. R. Koford. 1996. Using candlers to determine the incubation stage of passerine eggs. *Journal of Field Ornithology* 67:660–668.
- Lovell, S. F., and M. R. Lein. 2004. Neighbor-stranger discrimination by song in a suboscine bird, the alder flycatcher, *Empidonax alnorum*. *Behavioral Ecology* 15:799–804.
- Macdougall-shackleton, A. E. A., and S. A. Macdougall-shackleton. 2001. Cultural and genetic evolution in mountain white-crowned sparrows: song dialects are associated with population structure. *Evolution* 55:2568–2575.
- Marten, K., and P. Marler. 1977. Sound transmission and its significance for animal vocalization: I. temperate habitats. *Behavioral Ecology and Sociobiology* 2:271–290.
- McDonald, D. B. 1993. Delayed plumage maturation and orderly queues for status: a manakin mannequin experiment. *Ethology* 94:31–45.
- McDonald, D. B. 2009. Young-boy networks without kin clusters in a lek-mating manakin. *Behavioral Ecology and Sociobiology* 63:1029–1034.
- McDonald, D. B., and W. K. Potts. 1994. Cooperative display and relatedness among males in a lek-mating bird. *Science* 266:1030–1032.
- McGregor, P. K. 1989. Pro-active memory interference in neighbour recognition by a songbird. *Proceedings of the 19th International Ornithological Congress* 19:1391–1397.
- McGregor, P. K. 1992. Quantifying responses to playback: one, many or multivariate composite measures? In P. K. McGregor, ed., *Playback and Studies of Animal Communication* (pp. 79–96). Plenum Press, New York, NY.
- McGregor, P. K. 2000. Playback experiments: design and analysis. *Acta Ethologica* 3:3–8.

- McGregor, P. K., and M. I. Avery. 1986. The unsung songs of great tits (*Parus major*): learning neighbours' songs for discrimination. *Behavioral Ecology and Sociobiology* 18:311–316.
- McGregor, P. K., and J. R. Krebs. 1984. Song learning and deceptive mimicry. *Animal Behaviour* 32:280–287.
- McGuire, J. A., C. C. Witt, J. V. Remsen, R. Dudley, and D. L. Altshuler. 2009. A higher-level taxonomy for hummingbirds. *Journal of Ornithology* 150:155–165.
- McNaughton, S. J. 1988. Mineral nutrition and spatial concentrations of African ungulates. *Nature* 334:343–345.
- McShane, L. M., M. D. Radmacher, B. Freidlin, R. Yu, M.-C. Li, and R. Simon. 2002. Methods for assessing reproducibility of clustering patterns observed in analyses of microarray data. *Bioinformatics* 18:1462–1469.
- Meliza, C. D., S. C. Keen, and D. R. Rubenstein. 2013. Pitch- and spectral-based dynamic time warping methods for comparing field recordings of harmonic avian vocalizations. *The Journal of the Acoustical Society of America* 134:1407–15.
- Molles, L. E., and S. L. Vehrencamp. 2001. Neighbour recognition by resident males in the banded wren, *Thryothorus pleurostictus*, a tropical songbird with high song type sharing. *Animal Behaviour* 61:119–127.
- Mota, P. G., and G. C. Cardoso. 2001. Song organisation and patterns of variation in the serin (*Serinus serinus*). *Acta Ethologica* 3:141–150.
- Mundinger, P. C. 1982. Microgeographic and macrogeographic variation in the acquired vocalizations of birds. In D. E. Kroodsma & E. H. Miller, eds., *Acoustic Communication in Birds*, volume 2 (pp. 147–208). Academic Press, New York, NY.
- Naguib, M., and D. Todt. 1998. Recognition of neighbors' song in a species with large and complex song repertoires: the thrush nightingale. *Journal of Avian Biology* 29:155–160.

- Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4:133–142.
- Nelson, D. A., K. I. Hallberg, and J. A. Soha. 2004. Cultural evolution of Puget Sound white-crowned sparrow song dialects. *Ethology* 110:879–908.
- Nowicki, S., S. Peters, and J. Podos. 1998. Song learning, early nutrition and sexual selection in songbirds. *Integrative and Comparative Biology* 38:179–190.
- Nowicki, S., W. A. Searcy, and S. Peters. 2002. Brain development, song learning and mate choice in birds: a review and experimental test of the “nutritional stress hypothesis.” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology* 188:1003–1014.
- Oaten, a, C. E. M. Pearce, and M. E. B. Smyth. 1975. Batesian mimicry and signal detection theory. *Bulletin of Mathematical Biology* 37:367–387.
- Oh, K. P., and A. V Badyaev. 2010. Structure of social networks in a passerine bird: consequences for sexual selection and the evolution of mating strategies. *The American Naturalist* 176:E80–E89.
- Oniki, Y. 1996. Band sizes of southeastern Brazilian hummingbirds. *Journal of Field Ornithology* 67:387–391.
- Ortiz-Crespo, F. 1972. A new method to separate immature and adult hummingbirds. *The Auk* 89:851–857.
- Ortiz-Crespo, F. I. 1969. *Interactions in three sympatric species of Californian hummingbirds in a planted area containing melliferous flowering vegetation*. University of California at Berkeley.
- Osiejuk, T. S., K. Ratyńska, J. P. Cygan, and S. Dale. 2005. Frequency shift in homologue syllables of the ortolan bunting *Emberiza hortulana*. *Behavioural Processes* 68:69–83.

- Pakkanen, J., J. Iivarinen, and E. Oja. 2006. The evolving tree--analysis and applications. *IEEE Transactions on Neural Networks* 17:591–603.
- Palmer, J., A. C. Huk, and M. N. Shadlen. 2005. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision* 5:376–404.
- Patricelli, G., A. Krakauer, and R. McElreath. 2011. Assets and tactics in a mating market: economic models of negotiation offer insights into animal courtship dynamics on the lek. *Current Zoology* 57:225–237.
- Payne, R. B. 1982. Ecological consequences of song matching: breeding success and transpecific song mimicry in indigo buntings. *Ecology* 63:401–411.
- Payne, R. B. 1985. Behavioral continuity and change in local song populations of village indigobirds, *Vidua chalybeata*. *Zeitschrift für Tierpsychologie* 70:1–44.
- Payne, R., and P. Budde. 1979. Song differences and map distances in a population of acadian flycatchers. *The Wilson Bulletin* 91:29–41.
- Petrie, M., A. Krupa, and T. Burke. 1999. Peacocks lek with relatives even in the absence of social and environmental cues. *Nature* 401:1998–2000.
- Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, and R Core Team. 2015. nlme: linear and nonlinear mixed effects models.
- Podos, J. 1997. A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution* 51:537–551.
- Podos, J. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature* 409:185–188.
- Podos, J., and P. Warren. 2007. The evolution of geographic variation in birdsong. *Advances in the Study of Behavior* 37:403–458.

- Price, T. D. 2008. Speciation in Birds. Roberts and Company Publishers, Greenwood Village, CO.
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rainey, M. M., and G. F. Grether. 2007. Competitive mimicry: Synthesis of a neglected class of mimetic relationships. *Ecology* 88:2440–2448.
- Ramjohn, C. L., F. B. Lucas, F. E. Hayes, S. T. Ballah, N. C. Johnson, and K. M. Garcia. 2003. Lek mating behavior of the sooty-capped hermit (*Phaethornis augusti*) in the Paria Peninsula of Venezuela. *Journal of Field Ornithology* 74:205–209.
- Ranjar, L., and H. A. Ross. 2008. Unsupervised bird song syllable classification using evolving neural networks. *The Journal of the Acoustical Society of America* 123:4358–68.
- Reynolds, S. M., M. C. Christman, J. a. C. Uy, G. L. Patricelli, M. J. Braun, and G. Borgia. 2008. Lekking satin bowerbird males aggregate with relatives to mitigate aggression. *Behavioral Ecology* 20:410–415.
- Riede, T., R. A. Suthers, N. H. Fletcher, and W. E. Blevins. 2006. Songbirds tune their vocal tract to the fundamental frequency of their song. *Proceedings of the National Academy of Sciences of the United States of America* 103:5543–5548.
- Robinson, F., and H. Curtis. 1996. The vocal displays of the lyrebirds (Menuridae). *Emu* 96:258.
- Roche, E. A., T. W. Arnold, J. H. Stucker, and F. J. Cuthbert. 2010. Colored plastic and metal leg bands do not affect survival of piping plover chicks. *Journal of Field Ornithology* 81:317–324.
- Rohwer, S. 1982. The evolution of reliable and unreliable badges of fighting ability. *Integrative and Comparative Biology* 22:531–546.

- Rousseeuw, P. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65.
- Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* 8:103–106.
- Russell, S. M., and R. O. Russell. 2001. *The North American Bander's Manual for Banding Hummingbirds*. North American Banding Council, Point Reyes Station, CA.
- Ryan, M. J., M. D. Tuttle, and L. K. Taft. 1981. The costs and benefits of frog chorusing behavior. *Behavioral Ecology and Sociobiology* 8:273–278.
- Ryder, T. B., J. G. Blake, P. G. Parker, and B. A. Loiselle. 2011. The composition, stability, and kinship of reproductive coalitions in a lekking bird. *Behavioral Ecology* 22:282–290.
- Ryder, T. B., P. G. Parker, J. G. Blake, and B. A. Loiselle. 2009. It takes two to tango: reproductive skew and social correlates of male mating success in a lek-breeding bird. *Proceedings of the Royal Society B: Biological Sciences* 276:2377–84.
- Ryder, T., D. McDonald, J. Blake, P. Parker, and B. Loiselle. 2008. Social networks in the lek-mating wire-tailed manakin (*Pipra filicauda*). *Proceedings of the Royal Society B: Biological Sciences* 275:1367–1374.
- Sakoe, H., and S. Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26:43–49.
- Sankoff, D., and J. B. Kruskal. 1983. Time warps, string edits, and macromolecules: the theory and practice of sequence comparison. (D. Sankoff & J. B. Kruskal, eds.). Addison-Wesley, Inc., Reading, MA.
- Saranathan, V., D. Hamilton, G. V. N. Powell, D. E. Kroodsma, and R. O. Prum. 2007. Genetic evidence supports song learning in the three-wattled bellbird *Procnias tricarunculata* (Cotingidae). *Molecular Ecology* 16:3689–3702.

- Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution* 1:103–113.
- Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18:1–2.
- Searcy, W. A., S. Peters, S. Kipper, and S. Nowicki. 2010. Female response to song reflects male developmental history in swamp sparrows. *Behavioral Ecology and Sociobiology* 64:1343–1349.
- Seddon, N. 2005. Ecological adaptation and species recognition drives vocal evolution in neotropical suboscine birds. *Evolution* 59:200–215.
- Shofner, W. P. 2005. Comparative aspects of pitch perception. In C. Plack, R. Fay, A. Oxenham, & A. Popper, eds., *Pitch: Neural Coding and Perception* (pp. 56–98). Springer-Verlag, New York.
- Shorey, L., S. Pierniey, J. Stone, and J. Höglund. 2000. Fine-scale genetic structuring on *Manacus manacus* leks. *Nature* 408:352–353.
- Sinervo, B., and C. M. Lively. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature* 380:240–243.
- Skierczyński, M., and T. Osiejuk. 2010. Sharing songs within a local dialect does not hinder neighbour–stranger discrimination in ortolan bunting (*Emberiza hortulana*) males. *Behaviour* 147:333–351.
- Slabbekoorn, H., and T. B. Smith. 2002a. Habitat-dependent song divergence in the little greenbul: an analysis of environmental selection pressures on acoustic signals. *Evolution* 56:1849–58.
- Slabbekoorn, H., and T. B. Smith. 2002b. Bird song, ecology and speciation. *Philosophical Transactions of the Royal Society of London B: Biological sciences* 357:493–503.

- Slater, P. 1989. Bird song learning: causes and consequences. *Ethology Ecology & Evolution* 1:19–46.
- Snow, D. 1968. The singing assemblies of Little Hermits. *Living Bird* 7:47–55.
- Snow, D. 1973. Distribution, ecology and evolution of the bellbirds (Procnias, Cotingidae). *Bulletin of the British Museum (Natural History)* 25:369–391.
- Sokal, R., and F. Rohlf. 1962. The comparison of dendrograms by objective methods. *Taxon* 11:33–40.
- Stiles, F. G., and L. L. Wolf. 1973. Techniques for color-marking hummingbirds. *The Condor* 75:244–245.
- Stiles, F., and L. Wolf. 1979. Ecology and evolution of lek mating behavior in the long-tailed hermit hummingbird. *Ornithological Monographs* 27:1–78.
- Stoddard, P. K. 1996. Vocal recognition of neighbors by territorial passerines. In D. E. Kroodsma & E. H. Miller, eds., *Ecology and Evolution of Acoustic Communication in Birds* (pp. 356–374). Cornell University Press, Ithaca, New York.
- Stoddard, P. K., M. D. Beecher, C. L. Horning, and M. S. Willis. 1990. Strong neighbor-stranger discrimination in song sparrows. *Condor* 92:1051–1056.
- Tchernichovski, O., F. Nottebohm, C. Ho, B. Pesaran, and P. Mitra. 2000. A procedure for an automated measurement of song similarity. *Animal Behaviour* 59:1167–1176.
- Temeles, E. 1994. The role of neighbours in territorial systems: when are they “dear enemies”? *Animal Behaviour* 47:339–350.
- Therneau, T. M. 2015. *Mixed Effects Cox Models*. Mayo Clinic.

- Thorpe, W. 1958. The learning of song patterns by birds, with especial reference to the song of the chaffinch *Fringilla coelebs*. *Ibis* 100:535–570.
- Tori, W. P., T. B. Ryder, R. Duraes, J. R. Hidalgo, B. A. Loiselle, and J. G. Blake. 2006. Obtaining offspring genetic material: a new method for species with high nest predation rates. *Condor* 108:948–952.
- Trainer, J. 1989. Cultural evolution in song dialects of yellow-rumped caciques in Panama. *Ethology* 80:190–204.
- Trainer, J. M., and D. B. McDonald. 1995. Singing performance, frequency matching and courtship success of long-tailed manakins (*Chiroxiphia linearis*). *Behavioral Ecology and Sociobiology* 37:249–254.
- Trainer, J. M., D. B. McDonald, and W. A. Learn. 2002. The development of coordinated singing in cooperatively displaying long-tailed manakins. *Behavioral Ecology* 13:65–69.
- Uy, J. A. C., and J. A. Endler. 2004. Modification of the visual background increases the conspicuousness of golden-collared manakin displays. *Behavioral Ecology* 15:1003–1010.
- van Rhijn, J. G. 1983. On the maintenance and origin of alternative strategies in the ruff *Philomachus pugnax*. *Ibis* 125:482–498.
- Vanderbilt, C. C., J. P. Kelley, and E. H. DuVal. 2015. Variation in the performance of cross-contextual displays suggests selection on dual-male phenotypes in a lekking bird. *Animal Behaviour* 107:213–219.
- Wagner, R. H. 1998. Hidden leks: sexual selection and the clustering of avian territories. *Ornithological Monographs* 49:123–145.
- Wang, C., and S. Seneff. 2000. Robust pitch tracking for prosodic modeling in telephone speech. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* 3:1343–1346.

- Ward, J. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58:236–244.
- Waser, N. M., and W. A. Calder. 1975. Possible impairment of nest-building of hummingbirds by acetate leg tags. *The Condor* 77:361.
- Weary, D. M., and J. R. Krebs. 1992. Great tits classify songs by individual voice characteristics. *Animal Behaviour* 43:283–287.
- Weeden, J. S., and B. J. Falls. 1959. Differential responses of male ovenbirds to recorded songs of neighboring and more distant individuals. *The Auk* 76:343–351.
- Westcott, D., and F. Kroon. 2002. Geographic song variation and its consequences in the golden bowerbird. *The Condor* 104:750–760.
- Wilbrecht, L., and F. Nottebohm. 2003. Vocal learning in birds and humans. *Mental Retardation and Developmental Disabilities Research Reviews* 9:135–148.
- Wiley, R. 1971. Song groups in a singing assembly of little hermits. *The Condor* 73:28–35.
- Wiley, R. H. 1991. Lekking in birds and mammals: behavioural and evolutionary issues. *Advances in the Study of Behavior* 20:201–291.
- Wilkins, M. R., N. Seddon, and R. J. Safran. 2013. Evolutionary divergence in acoustic signals: causes and consequences. *Trends in Ecology and Evolution* 28:156–166.
- Wilson, D. R., P.-P. Bitton, J. Podos, and D. J. Mennill. 2014. Uneven sampling and the analysis of vocal performance constraints. *The American Naturalist* 183:214–228.
- Wilson, P. L., and S. L. Vehrencamp. 2001. A test of the deceptive mimicry hypothesis in song-sharing song sparrows. *Animal Behaviour* 62:1197–1205.
- Wittenberger, J. F. 1978. The evolution of mating systems in grouse. *The Condor* 80:126–137.

- Wright, T. F., C. R. Dahlin, and A. Salinas-Melgoza. 2008. Stability and change in vocal dialects of the yellow-naped amazon. *Animal Behaviour* 76:1017–1027.
- Wunderle, J. M. J. 1978. Differential Response of Territorial Yellowthroats To the Songs of of Neighbors and Non-Neighbors. *The Auk* 95:389–395.
- Xiao, J., and P. Flandrin. 2007. Multitaper time-frequency reassignment for nonstationary spectrum estimation and chirp enhancement. *IEEE Transactions on Signal Processing* 55:2851–2860.
- Zahn, C. T. 1971. Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Transactions on Computers* C-20:68–86.
- Zuur, A. F., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R*. (M. Gail, K. Krickeberg, J. M. Samet, A. Tsiatis, & W. Wong, eds.). Springer, New York.