

THE ROLE OF SOCIAL ENVIRONMENT IN SHAPING VOCAL
COMMUNICATION SYSTEMS IN WILD SONGBIRDS

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

Sara Christina Keen

May 2020

© 2020 Sara Christina Keen

THE ROLE OF SOCIAL ENVIRONMENT IN SHAPING VOCAL
COMMUNICATION SYSTEMS IN WILD SONGBIRDS

Sara Christina Keen, Ph. D.

Cornell University 2020

ABSTRACT

For many taxa, vocal communication is an essential means of navigating continuously changing social and ecological environments. Among passerine birds, vocal signals are critical to survival and play an important role in mate attraction, territory defense, and predator avoidance. In my dissertation, I study acoustic communication in a wild population of blue tits and great tits in order to investigate the effects of social environment on birds' responses to acoustic cues as well as their production of acoustic signals. In Chapter 1, I demonstrate that blue tits can learn to associate a novel acoustic cue with predation risk, and that the behavioral response to this cue can be socially transmitted to naïve great tits, despite their lack of first-hand experience. This study suggests that social learning of acoustic cues can occur between species. In the second chapter, I develop an unsupervised machine learning approach that can objectively measure similarity of vocal signals. I present this technique such that it can be broadly applied for the analysis of diverse acoustic datasets. Using this approach, I

then explore patterns of variation in great tit songs from multiple angles. In Chapter 3, I develop and test a mathematical model that describes the optimal levels of vocal similarity among neighbors under given social and ecological conditions. This model predicts that immigrant and resident birds will exhibit comparable levels of vocal similarity with neighbors, despite having different song learning opportunities in their natal environments. My empirical results agree with model predictions, showing that although immigrants more often use complex, unshared songs, they achieve high levels of vocal similarity with neighbors by using larger repertoires than residents. I further explore spatial and temporal variation of great tit songs in Chapter 4, and show that an individual's songs reflect both their immigration status and their breeding territory location. I also find that songs change between subsequent years, and that this is largely due to the appearance and disappearance of individual birds. Together, my findings suggest that individuals continuously adapt their vocal behavior to changing social environments, and that interactions with both conspecifics and heterospecifics may shape vocal communication systems.

BIOGRAPHICAL SKETCH

Sara Keen was born in the small coastal city of Melbourne, Florida. As a child, she spent much of her time outdoors and enjoyed many long afternoons exploring the woods beside her house and climbing trees. She attended college at the University of Florida, where she studied electrical engineering and was an active member of the machine intelligence laboratory and was encouraged, supported, and inspired by her dynamic undergraduate advisor, Dr. Eric Schwartz. Throughout college, she explored her interests in programming and machine intelligence by participating in an NSF REU summer program in robotics, being a teaching assistant for microprocessor classes, and leading robotics summer camps for elementary school students. Outside of the laboratory, she frequently ran, cycled, and explored the natural areas in Gainesville, Florida, and became fascinated with the natural world.

After finishing her undergraduate degree, she taught middle school English in Cali, Colombia, which offered more chances to explore different ecological and social landscapes, establish longstanding friendships, and develop fluency in Spanish. Following this, she worked as a technical assistant in a robotics laboratory at Yale University, and decided to return to the University of Florida to complete a masters degree in electrical engineering. She was worked with Dr. John Harris, who introduced her to the world of digital signal processing and who also demonstrated the essential skill of carrying out rigorous research while having as much fun as possible. During this period Sara worked as an intern at the Bioacoustics Research Program at the Cornell Lab of Ornithology, and became fascinated by the world of bioacoustics.

Her experience in the bioacoustics group led Sara to take a research position at the National University of Singapore for one year, and to then complete an MA in conservation biology with Dr. Dustin Rubenstein at Columbia University. During this period, Sara had the opportunity to apply acoustic analyses to study bird behavior, and she was inspired to continue this research as a PhD student in the Department of Neurobiology and Behavior at Cornell University. Sara was advised by Dr. Kern Reeve, who supported and encouraged her as she conducted field work, made numerous trips to her field site in England, and gained invaluable insights through long conversations about science at Ithaca Coffee Company. While in Oxford, she received essential guidance and encouragement from Drs. Ben Sheldon and Ella Cole during her studies in Wytham Woods. During this time, Sara carried developed and carried out her dissertation project, developed essential research skills, and made wonderful friendships in both Ithaca and Oxford. Sara plans to start a post-doctoral position in the Geology Department at Stanford University where she will continue to combine engineering and biology to better understand the natural world.

For my family

ACKNOWLEDGMENTS

I would like to thank everyone who supported me in completing this dissertation; it would not have been possible without their help. Foremost, I thank Kern Reeve, who was essential to all of this. Thank you for the time you took to discuss numerous concepts and models during our meetings, for shaping how I approach problems, and for demonstrating how to be a good scientist and human. Kern's enthusiasm for ecological research and for asking big questions helped me to thrive in NBB and to persist in the challenging times as well. His energy and support were crucial in completing this project. I also thank Ben Sheldon, whose guidance and encouragement made this study possible. Ben's curiosity about the world, his ability to identify and lead me to interesting questions, and his welcoming me to his lab group in Oxford were an incredible academic opportunity and a chance to form great friendships and collaborations. Working in Wytham Woods was a privilege and the days spent at the field station are among my favorite times during my PhD.

I am extremely grateful to the many other academic mentors who helped support and advise me during this process. Mike Sheehan and Mike Webster both gave invaluable feedback which helped to improve my project and experimental design, and through our conversations both helped me to become a stronger researcher by sharing their own experiences and insights. Thank you to Holger Klinck for offering an endless supply of energy, technical expertise, recording equipment, and reassurance that it was all good. I thank him especially for not letting me return to my previous job when I was discouraged in my second year, and for beginning CCB's weekly tradition of Friday evening gatherings. Thank you to Ella Cole for sharing her time and Wytham expertise and for helping me find my bearings in Oxford; to Keith McMahon for teaching me how to ring birds, helping me fix my mistakes during field

work, being available for critical phone calls from the field, and not taking the mickey every chance he got; to Lucy Aplin for showing me how to work in the aviary despite being in the midst of her own projects, and for her excellent advice on conducting experiments; to Josh Firth for sharing his expertise on Wytham birds, coding, and social networks, and for being a remarkable example of excellent time management; to Karan Odom, Marcelo Aray-Salas, Russ Charif, Wendy Erb, Maria Modanu, Liz Bergen, and Hailey Scofield for their help with developing my research as well as their friendship during the last five years.

I feel enormously fortunate to have found wonderful communities in both Ithaca and Oxford while leading parallel lives these last five years. Thank you to Emma Greig, Sarah Alexander, Aditi Sahasrabuddhe, Vannina Etti, Emily D'Angelo, Prantik Mazumder, Rohini Jalan, Kieron Guinemarde, Sarah Rugheimer, Freddy Hilleman, Ash Sendall-Price, Allison Roth, Benjamin Van Doren, Dena Clink, Ana Verarhami, Bobbi Estabrook, Yu Shiu, Liz and Joe Rowland, and Peter Wrege for your friendship and inspiration throughout this process. Most of all, I would like to thank my parents, Eric, and Emily, for always encouraging me and for each being role models of how to live a good life. I feel incredibly lucky to have ended up with such a remarkable family and for the support that each of you offered during this project.

Lastly, I would like to thank the agencies and organizations that helped to support this work, including Cornell Lab of Ornithology Athena Fund for enabling so many field seasons, the Center for Conservation Bioacoustics, the Cornell Lab of Ornithology and the fantastic oversight of students by Irby Lovette, Oxford's Edward Grey Institute for Field Ornithology, Sigma Xi, and the Cornell Department of Neurobiology and Behavior Animal Research grant.

TABLE OF CONTENTS

BIOGRAPHICAL SKETCH.....	v
ACNOWLEDGMENTS.....	viii
TABLE OF CONTENTS.....	x
CHAPTER 1.....	1
LITERATURE CITED.....	26
CHAPTER 2.....	30
LITERATURE CITED.....	61
CHAPTER 3.....	67
LITERATURE CITED.....	108
CHAPTER 4.....	113
LITERATURE CITED.....	142
APPENDIX A.....	148
APPENDIX B.....	151
APPENDIX C.....	164

CHAPTER 1

SOCIAL LEARNING OF ACOUSTIC ANTI-PREDATOR CUES OCCURS BETWEEN WILD BIRD SPECIES

Sara C. Keen^{1,2}, Ella F. Cole², Michael J. Sheehan¹, Ben C. Sheldon²

¹ Department of Neurobiology and Behavior, Cornell University,
Ithaca, NY 14850, USA

² Edward Grey Institute, Department of Zoology, University of Oxford,
Oxford, UK, OX1 3PS

ABSTRACT

In many species, individuals gather information about their environment both through direct experience and through information obtained from others. Social learning, or the acquisition of information from others, can occur both within and between species and may facilitate the rapid spread of antipredator behaviour. Within birds, acoustic signals are frequently used to alert others to the presence of predators, and individuals can quickly learn to associate novel acoustic cues with predation risk. However, few studies have addressed whether such learning occurs only through direct experience or whether it has a social component, nor whether such learning can occur between species. We investigate these questions in two sympatric species of Parids: blue tits (*Cyanistes caeruleus*) and great tits (*Parus major*). Using playbacks of unfamiliar bird vocalisations paired with a predator model in a controlled aviary setting, we

find that blue tits can learn to associate a novel sound with predation risk via direct experience, and that antipredator response to the sound can be socially transmitted to heterospecific observers, despite lack of first-hand experience. Our results suggest that social learning of acoustic cues can occur between species. Such interspecific social information transmission may help to mediate the formation of mixed-species aggregations.

INTRODUCTION

A central question in behavioural ecology is how learned traits spread through a population, and which individual characteristics may facilitate or impede their social transmission. Reflecting the increasing interest in this question is a growing body of literature which demonstrates the high adaptive value of social learning [1-4]. Unlike acquiring information directly, which requires a process of trial-and-error and often increases predation risk, acquisition of information from others can allow individuals to quickly learn about their surroundings and adjust to changing environments at a relatively lower cost [1-3]. This mechanism can enable rapid horizontal transmission of antipredator behaviour through a population, thereby directly impacting individual survival [5, 6]. Consequently, selection may act upon individuals' capabilities for social learning and social acquisition of traits [3, 4], making this area of research important in advancing our understanding of biological evolution and adaptation. Furthermore, because the acquisition of learned behaviours is an important mechanism in the establishment of animal culture, investigating this process could give insight into the emergence and persistence of novel traditions within a population [1, 2, 7].

In order to reduce uncertainty about the surroundings, information may be acquired from both con- and heterospecifics, though the amount of overlap in the ecological niches

occupied by the producer and receiver must be considered. For example, information acquired from heterospecifics that use comparable foraging strategies or experience similar predation risks is more useful than information gathered from species that rely on different food sources and/or are hunted by different predators, and may therefore be more likely to transmit across species boundaries [8-10]. In recent years, a number of studies have documented social learning both within and between species [11-15]. However, to date, much of the evidence of the spread of learned traits comes from conspicuous behaviours such as tool use in primates, propagation of foraging strategies in birds, and learned birdsong, e.g., [16-18]. Furthermore, experimental manipulations of social transmission of traits are few, and the best studied examples entail gathering information from conspecifics [e.g., 17, 19]. As we aim to better understand the spread of behavioural traits, the boundaries of social transmission must be examined from multiple angles, including a range of modalities and transmission between individuals with different phenotypes.

Many species of birds and mammals commonly use acoustic signals and cues to acquire information about predators [12, 13, 20-22], and a diverse array of alarm calling behaviours can be observed in different contexts, including calls directed at predators during mobbing events, distress calls made during predator attacks, calls produced whilst fleeing predators, and sentinel calls that alert nearby individuals to perceived risk levels [23]. Although information about predators is often obtained using acoustic signals produced by conspecifics, which may evolve through processes such as kin selection or reciprocal altruism [24,25], many birds and mammals commonly eavesdrop on signals intended for others [23]. Eavesdropping on heterospecifics may play an important role in the formation of mixed species assemblages [10], and response to heterospecific alarm calls can either be innate (e.g.,

if calls are acoustically similar among species [26-28]), or learned [24, 29], which may occur as early as the embryonic stage in birds [30]. In addition to learning heterospecific alarm calls, recent experimental evidence has shown that birds and mammals can learn to associate unfamiliar acoustic cues with perceived predation risk [14, 31, 32], adding to a growing body of literature suggesting that associative learning may be the mechanism underpinning the recognition of heterospecific alarm calls. Recent research has also shown that birds can learn to associate novel sounds with heterospecific alarm calls [15], suggesting that a behavioural response to an acoustic cue can be socially transmitted, even when the cue is not initially recognised as an alarm call. The possibility that this phenomenon can occur between species has been suggested [15], but not formally tested.

Here, we study birds captured from sympatric populations of blue tits (*Cyanistes caeruleus*) and great tits (*Parus major*), which spend the winter months foraging together in mixed-species flocks and use calls to alert others to the presence of predators such the Eurasian sparrowhawk (*Accipiter nisus*) [33]. This shared suite of natural history traits suggests that interspecific social learning is likely to occur (see [11] for a review of social learning between sympatric species), yet little experimental research investigating this question has been conducted. To address this question, in this study we investigate social learning of acoustic antipredator cues in two ecologically relevant contexts: within and between species. To test our hypotheses that intra- and interspecific social transmission occurs among blue tits and great tits, we carried out a two-stage experiment. First, using playbacks paired with a predator model, we trained groups of blue tit demonstrators to associate a novel acoustic cue with predation risk. We then introduced naïve blue tit and great tit observers and conducted multiple playbacks of the acoustic cue while demonstrators and observers were

housed together. Importantly, the predator model was not used during this stage of the experiment, ensuring that observers had access only to social information, but not private information, that could convey predation risk. We predicted that both conspecific and heterospecific observers would acquire an antipredator response to the acoustic cue despite having no direct exposure to the predator model, and independently tested observers to determine whether intra- and interspecific social transmission had occurred.

METHODS

Study site and species. The subjects for this experiment were eight great tits (*Parus major*) and 48 blue tits (*Cyanistes caeruleus*) captured using mist nets from a wild population at Wytham Woods, Oxfordshire, UK (51°46 N 1°20 W) between 29th December 2015 and 8th March 2016. Blue tits were used as both demonstrators and observers and thus more individuals of this species were included. All birds were fitted with a unique radio frequency identification (RFID) tag and metal BTO leg band as well as a temporary color band that was worn for the duration of the experiment. Upon catching, we determined the age (yearling or older) and sex of all birds based on plumage characteristics [33] (Sex: great tits: 6 males, 2 females; blue tits: 27 males, 15 females, and 3 individuals where sex could not be determined); (Age: great tits: 6 yearling and 2 older; blue tits: 33 yearling and 15 older). We randomly selected birds to use in this experiment from all individuals captured during mist netting, and did not take age or sex into consideration. For each replicate of our experiment, 6 blue tits and 1 great tit were captured together and kept in captivity for seven days before being released at the site of capture.

We conducted all experiments in an outdoor aviary at the John Krebs Field Station, Wytham, Oxfordshire, UK, between 29th December 2015 and 8th March 2016 (Fig. 1). Two cameras, an iPhone 5s and Logitech C920 HD Pro Webcam, were mounted on different walls such that the majority of the aviary space could be filmed. We placed a feeder station stocked with sunflower seeds and equipped with an RFID antenna and data logger in the center of the aviary which allowed for the time and individual identity of birds visiting the feeder to be recorded. Due to inconsistent wiring connections, RFID readers did not record some feeder visits. Therefore, for any feeder visit that was noted during video analysis but not recorded by the RFID logger, we determined identity using colored leg bands which could be seen in video footage.

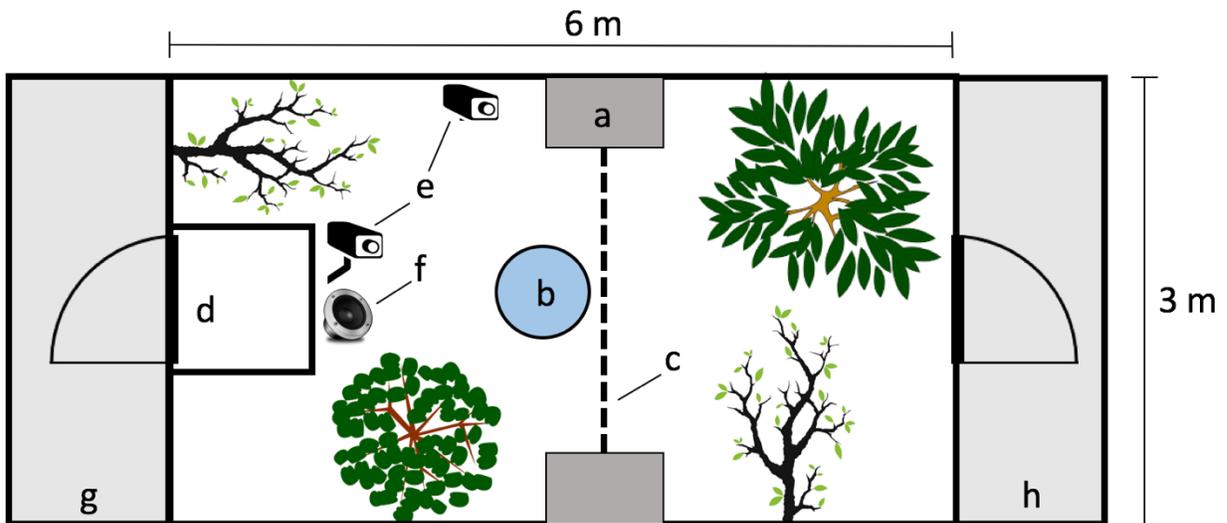


Figure 1. Diagram of outdoor aviary in which experiments took place. Labels refer to (a) box in which model sparrowhawk was positioned between training playbacks, (b) feeder station equipped with PIT tag reader, (c) zipline across which model sparrowhawk was flown, (d) booth with opaque walls in which the experimenter sat during playbacks, (e) cameras, (f), speaker, (g) adjacent buildings, (h) empty adjacent outdoor passageway.

The experiment was replicated eight times, following the protocol summarized in Table 1. In total, we tested 40 blue tit demonstrators, eight blue tit observers, and eight great tit observers. Due to camera failure, one replicate of the pre-training tests (replicate 1) and two replicates of the post-training playback tests (replicates 1 and 8) of demonstrators had to be excluded. All post-training playback tests of observers were filmed. Thus, the final sample sizes were N=35 demonstrators for pre-training tests and N=30 for post-training tests and N=8 for post-training tests of both the blue tit and great tit observers. Demonstrator groups contained (mean \pm SE) 3 ± 0.3 males and 1.85 ± 0.4 females, and 2.8 ± 0.8 yearlings and 2.2 ± 0.8 older birds. Distributions of latency to resume feeding after playbacks within males and females were not significantly different in either pre-training or post-training playbacks (pre-training: t-test: $t = -1.21$, $df = 11.71$, $p = 0.25$; post-training: $t = -1.98$, $df = 20.84$, $p = 0.06$), nor were distributions of latency to resume feeding within yearlings and older birds (pre-training: t-test: $t = 0.27$, $df = 18.1$, $p = 0.79$, post-training: $t = -1.87$, $df = 31.2$, $p = 0.07$). For this reason, and because our sample size did not allow sufficient statistical power to include these factors in our analysis, all demonstrators from the same replicate were grouped together regardless of age and sex.

Table 1. Protocol for single replicate. The experiment was replicated eight times, each time using five blue tit demonstrators, one blue tit observer and one great tit observer.

Day	Protocol
1	<ul style="list-style-type: none"> ○ Birds captured from the wild and released into the aviary ○ Pre-training playback tests ○ Move observers to the indoor aviary
2	<ul style="list-style-type: none"> ○ Demonstrator training with predator model and playback (x 4)
3	<ul style="list-style-type: none"> ○ Demonstrator training with predator model and playback (x 4) ○ Demonstrator playback test for associative learning ○ Add observers into outdoor aviary with demonstrators
4	<ul style="list-style-type: none"> ○ Observer training with demonstrators and playback only (x 5) ○ Place all demonstrators and one observer indoors
5	<ul style="list-style-type: none"> ○ Playback test for social learning with observer 1
6	<ul style="list-style-type: none"> ○ Playback test for social learning with observer 2
7	<ul style="list-style-type: none"> ○ Release birds at site of capture

Experimental design. To test our hypotheses that intra- and interspecific social learning of acoustic cues can occur, this experiment necessarily comprised two stages: (1) training demonstrators to associate a novel sound with a predation event (*i.e.*, associative learning), and (2) exposing untrained conspecific and heterospecific observers to the trained demonstrators to test whether this behaviour is transferred horizontally (*i.e.*, social learning) (Fig. 2). To ensure that birds learned to associate the sound with predation and did not simply exhibit a neophobic response, we used acoustically similar “control” and “treatment” sounds as stimuli: recordings of songs from a Northern Cardinal (*Cardinalis cardinalis*) and an Eastern Whip-poor-will (*Antrostomus vociferus*). These signals occupy approximately the same frequency range as tits’ vocalizations (1.5 - 6 kHz), and are from North American species, and therefore unfamiliar to all birds used in the experiment. We downloaded both recordings from Xeno-canto [34] and normalized their amplitude using Audacity 2.1.1 [35] such that both recordings were of equal amplitude and eight seconds in duration (Fig. 3). We placed Dell AX210 speakers approximately 1m from the feeder station aiming towards the center of the aviary for playbacks, and adjusted the volume such that sounds played at an amplitude of approximately 65 dB at 10 m, the amplitude at which great tits sing in the wild [36], and within the range of amplitude at which great tits produce alarm calls [37]. Playbacks were always initiated when at least one bird was foraging at the feeder station, and this rule was used in tests of demonstrator groups as well as in tests of individual observers. In all stages of the experiment, we always separated playbacks by at least one hour. For each replicate, we alternated which sounds served as treatment and control stimuli, and, in order to minimize biases for factors such as motivation to feed, which may decrease throughout the

day, we alternated the order in which the control and treatment sounds were used and the order in which observers were tested (Table 2).

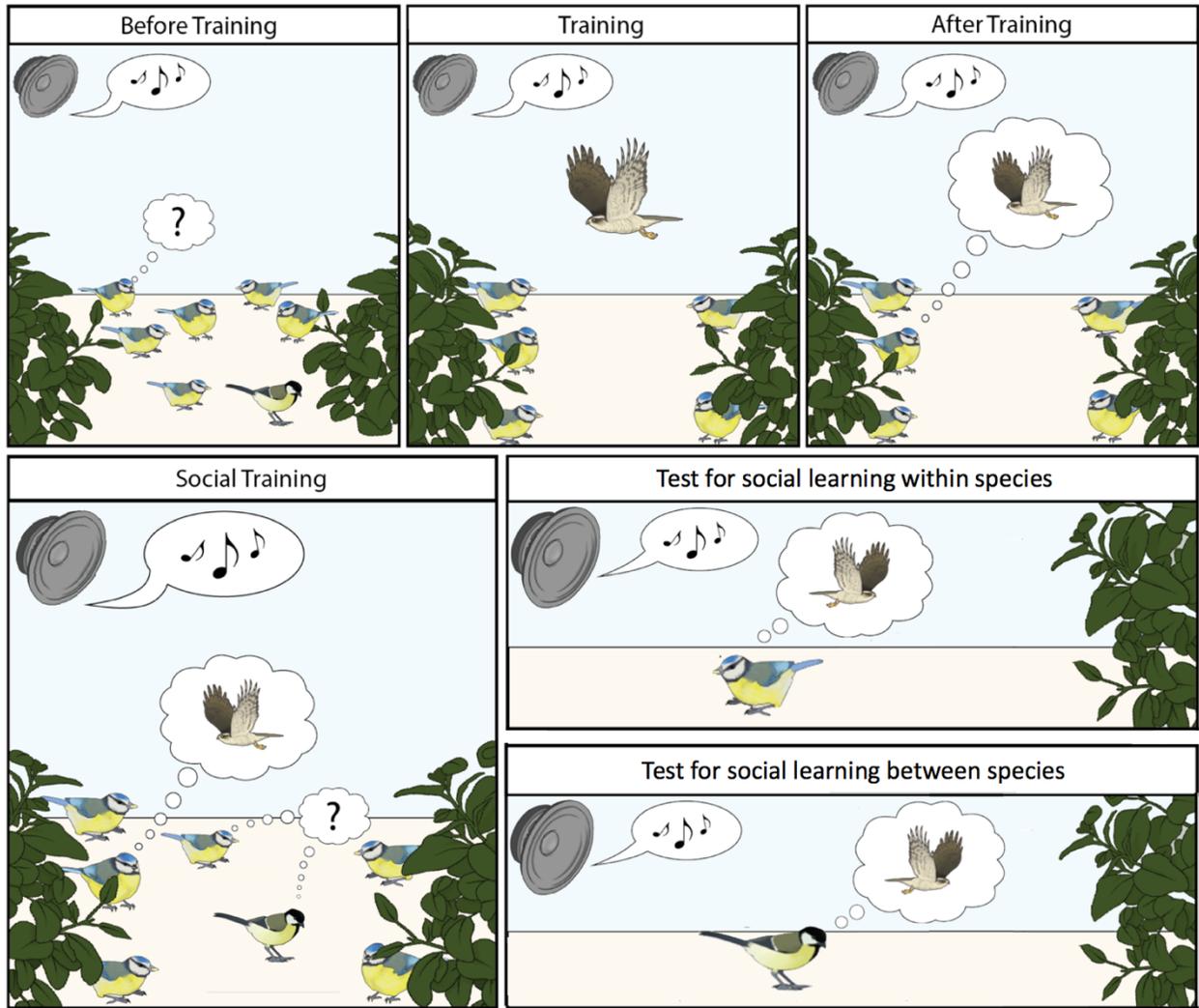


Figure 2. Graphical overview of experiment.

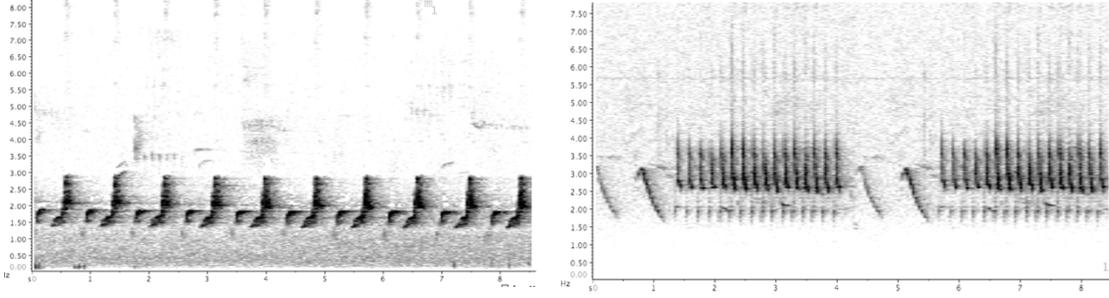


Figure 3. Spectrograms of sounds used for control and treatment playbacks plotted with Raven Pro 1.5 (www.birds.cornell.edu/raven) with 4095 point FFTs, Hann window, and 50% overlap. Sounds were downloaded from xeno-canto.org and amplitude-normalized and edited to 8 s duration. a) Northern Cardinal, b) Eastern Whip-poor-will.

Table 2. Order of playback stimuli and observer testing for the 8 replicate groups. In the first four replicates, recordings of and Eastern Whip-poor-will and Northern Cardinal were used as the treatment and control sounds, respectively; in the second four replicates this was reversed.

Replicate Group	Pre-training playback test		Direct learning playback test (5 BT demonstrators)		Social learning playback test (BT and GT observers)		Observer order		Play-back stimuli
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	
1	Trmt	Ctrl	Ctrl	Trmt	Ctrl	Trmt	GT	BT	Ctrl:
2	Ctrl	Trmt	Trmt	Ctrl	Trmt	Ctrl	GT	BT	NC
3	Trmt	Ctrl	Ctrl	Trmt	Ctrl	Trmt	BT	GT	Trmt:
4	Ctrl	Trmt	Trmt	Ctrl	Trmt	Ctrl	BT	GT	WPW
5	Trmt	Ctrl	Ctrl	Trmt	Ctrl	Trmt	GT	BT	Ctrl:
6	Ctrl	Trmt	Trmt	Ctrl	Trmt	Ctrl	GT	BT	WPW
7	Trmt	Ctrl	Ctrl	Trmt	Ctrl	Trmt	BT	GT	Trmt:
8	Ctrl	Trmt	Trmt	Ctrl	Trmt	Ctrl	BT	GT	NC

We fixed two large plastic boxes to the aviary ceiling with a cable running between them, upon which a sparrowhawk model could be flown across the 3m aviary width in under 0.5 s. Eurasian sparrowhawks (*Accipiter nisus*) are a primary cause of mortality among tits in Wytham Woods [38], and in previous experiments, great tits and blue tits have been shown to react to such models as they would live predators [39, 40]. The model was a plastic bird that was hand painted to closely resemble a sparrowhawk and was approximately the size of an adult male (length 350 mm, wingspan 560 mm). This model was also used in previous predator exposure experiments conducted using this population [40]. The openings of both boxes had plastic curtains such that the model was not visible when inside the box.

Experiment protocol. On the first day of an experiment, we caught six blue tits and one great tit before 0900 hr and placed all birds in the aviary within an hour of capture. After approximately one hour, we conducted pre-training playback tests using both the control and treatment sounds and filmed the group of seven birds for five minutes following each playback. Using this footage paired with RFID records from the feeder station, we measured the latency to resume feeding for all individuals. Latency was defined as the time from the end of the playback until first contact with the feeder. Approximately 30 minutes after playbacks were complete, we moved one blue tit and one great tit (hereafter referred to as observers) into the indoor aviary where they were housed together (see Appendix A for detailed description of indoor aviary). Five blue tits (hereafter referred to as demonstrators) remained in the outdoor aviary. The blue tit observer was selected as the first blue tit to fly into a mist net placed in the aviary.

Training the demonstrators. During the second and third days of an experiment, we trained the five demonstrators to associate the treatment sound with the presence of a predator by conducting eight repeat treatments (4 per day) during which a model Eurasian sparrowhawk was flown across the top of the aviary as the sound was broadcast over the speakers. All playbacks took place between 0900 and 1500 hr and were separated by at least one hour. At the end of the third day, to test whether the birds had learnt to associate the sound with the attempted predation event, we performed two additional playbacks using the treatment and control sounds, but not exposing birds to the predator model. The demonstrators were filmed for five minutes immediately following the final two playbacks; from this footage we extracted latency to resume feeding for all individuals as well as the number of alarm calls in order to compare pre- and post-training responses. Vocalizations matching descriptions of vocalizations produced by blue tits in response to predator presentations [41] were considered to be alarm calls. When analysing videos of playbacks, we counted all alarm calls and then calculated the average number of alarm calls per bird, as it was not possible to assign calls to individuals during trials. All video analyses were conducted in a blind manner. Following this test, the great tit and blue tit observers were returned to the outdoor aviary containing the demonstrators.

Training the observers. In the second stage of the experiment we tested our prediction that observers could socially learn to associate a sound with danger without having the direct experience of simultaneously seeing the predator model. To facilitate social transmission, on the fourth day of an experiment we conducted five playbacks of the treatment sound while the five demonstrators and the conspecific and heterospecific observers were in the aviary

together over the course of one day. We did not use the predator model during these tests, ensuring that any antipredator behaviors that the observers developed in response to the treatment sound were not due to direct experience of a potential predator. At the end of the fourth day, we moved the five demonstrators and one observer indoors.

Testing the observers. The next day, we conducted two playbacks with the observer that remained in the aviary (observer 1), once using the treatment sound and once using the control sound, and never exposing the observer to the predator model. Both playbacks were filmed; a blind observer used this footage to measure latency to resume feeding and number of alarm calls made in five minutes immediately following each playback. Vocalizations matching descriptions of blue tit and great tit alarm calls [41] were considered to be alarm calls produced by blue tit and great tit observers, respectively. At the end of day 5, we moved observer 1 indoors and placed observer 2 in the aviary, and performed identical playback tests the following day. Testing the observers separately ensured that they were responding only to the playback sound, rather than social cues from nearby birds. The next morning, we released all birds at the location where they were captured.

Quantification and statistical analysis. In pre-training tests with demonstrators (N=7 replicates), 17 of 35 birds returned to the feeders after playbacks of the control sound (mean \pm standard error: 2.83 ± 0.48 per replicate), and 13 of 35 birds returned after playbacks of the treatment sound (2.17 ± 0.4 per replicate). In post-training demonstrator tests (N=6) 20 of 30 birds returned after control playbacks (3.33 ± 0.49) and 20 of 30 returned after treatment playbacks (3.33 ± 0.42 per replicate).

To determine whether demonstrators learned to associate the treatment sound with a potentially dangerous event (i.e., whether associative learning had occurred), we conducted a survival analysis using a mixed effect Cox model to identify differences in latency to resume foraging between the pre- and post-training tests. We used two separate survival analyses, restricting the dataset first to pre- and then to post-training measurements, to assess whether demonstrators took significantly longer to resume foraging after the treatment versus control playbacks before and after training. We included stimulus (control or treatment sound) as a fixed binary effect, and individual bird identity and group number as random effects. By including bird identity in our model, we aimed to minimize the effects of noise in the latency measurements caused by variation between individuals in motivation to feed. In cases where an individual bird did not resume foraging within five minutes following a playback (60 of 130 demonstrator observations), latency times were censored. We also used paired t-tests to determine whether demonstrators made significantly more alarm calls within five minutes of control versus treatment playbacks after training, and used separate tests to compare demonstrators' response before and after training.

To test whether observers learned to associate the treatment sound with danger (i.e., whether social learning had occurred), we conducted separate survival analyses for blue tit observers and great tit observers with latency to resume foraging as a response variable, and used playback stimulus as a fixed effect and individual identity as a random effect. Latency values were censored when individual did not resume foraging within 5 minutes; this occurred in two of 16 trials of blue tit observers, and in two of 16 trials of great tit observers. We used paired t-tests to determine whether birds produced significantly more alarm calls following playback of the treatment sound as compared to the control sound. Analyses were performed

using the coxme and BSDA packages in R 3.4.1 [42-44]. See Appendix A information for further details of experimental procedures.

RESULTS

Associative learning of acoustic cues. Our results suggest that blue tit demonstrators learned to associate the novel cue with a predation threat. Before training, blue tit demonstrators showed no difference in latency to resume foraging after playbacks of the control or treatment sounds (mean \pm SE: control: 120.8 \pm 19.8 s; treatment: 104.51 \pm 18.8 s; Cox mixed effects model: $\chi^2 = 1.96$, df = 1, P = 0.161; Fig. 4a, c), showing that there is not an innate aversion or attraction to the sounds. After training, demonstrators took significantly longer to resume foraging after treatment playbacks compared to control playbacks (control: 90.2 \pm 12.53 s; treatment: 117.21 \pm 14.3 s; Cox mixed effects model: $\chi^2 = 5.81$, df = 1, P = 0.016; Fig. 4b, d). This suggests that the experimental training was successful in causing the demonstrators to associate the treatment sound with the presence of a predator. Both before and after training, demonstrator groups did not produce significantly more alarm calls in response to treatment vs. control playbacks (before: t = -0.42, df = 11.1, p = 0.68; after: t = 0.77, df = 8.94, p = 0.46).

Social transmission of antipredator response to acoustic cues. After exposure to trained demonstrators, great tit observers exhibited different behavioural responses to control versus treatment playbacks, whereas blue tit observers exhibited no detectable difference. Great tit observers took significantly longer to resume feeding after treatment playbacks (mean \pm SE: control: 48.3 \pm 17 s, treatment: 72.4 \pm 16.3 s, Cox mixed effects model: $\chi^2 = 7.88$, df = 1, p

=0.005, Fig. 5b, d) and made more alarm calls in the first five minutes after playbacks of the treatment sound, but this difference was not statistically significant (control: 12.5 ± 3.3 , treatment: 21.9 ± 6.8 , $t = -1.31$, $df = 7$, $p = 0.23$, Fig. 5f). Blue tit observers took longer to resume foraging and made more alarm calls after the playback compared to the control treatment, but neither effect was statistically significant (latency: mean \pm SE: control: 37.5 ± 15.3 s, treatment: 89.3 ± 42.7 s, Cox mixed effects model: $\chi^2 = 1.50$, $df = 1$, $p = 0.221$, Fig. 5a, c; alarm calls mean \pm SE: control: 2.86 ± 1.01 , treatment: 6.75 ± 1.93 , $t = -1.93$, $df = 7$, p -value = 0.09, Fig. 5e).

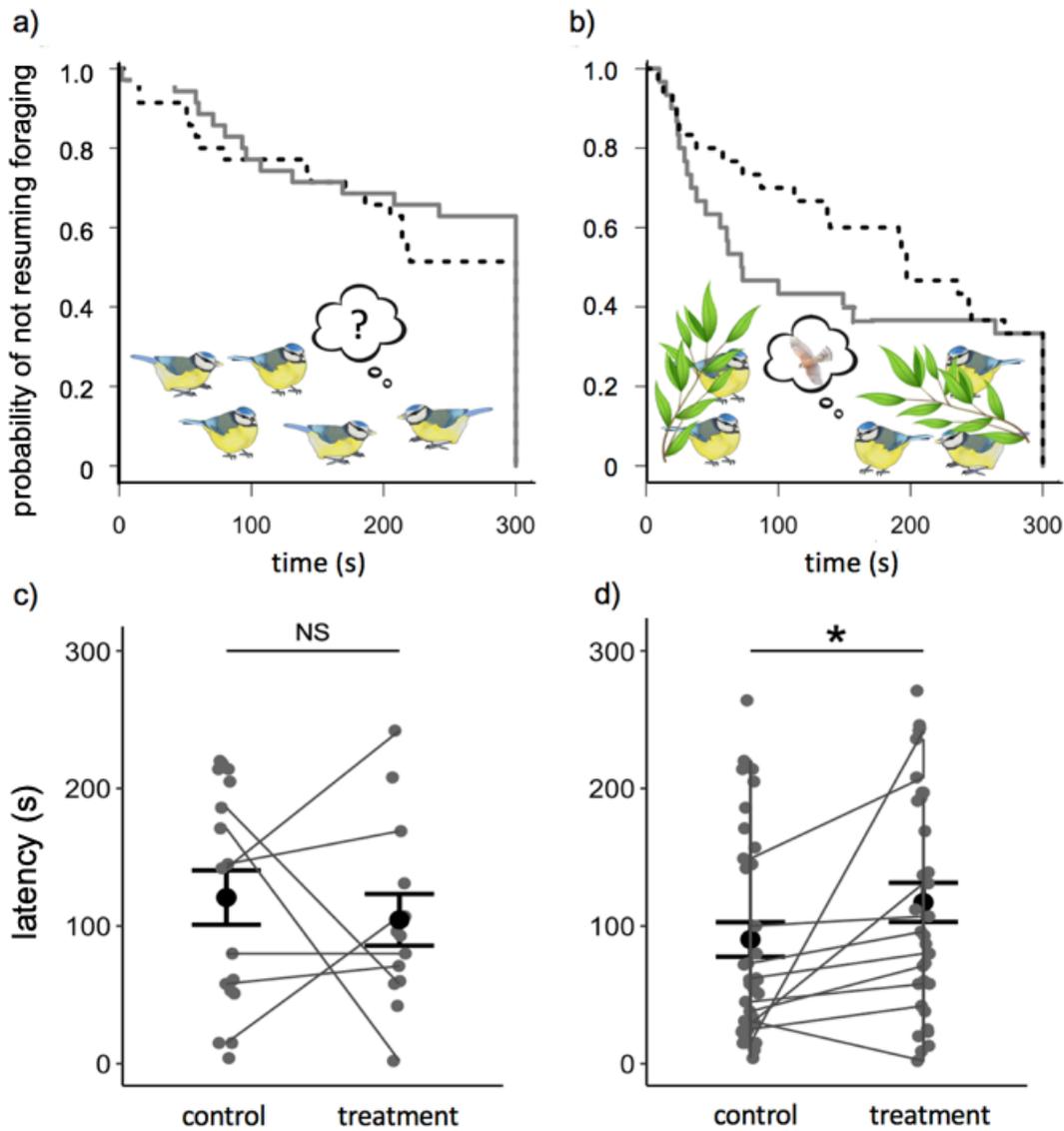
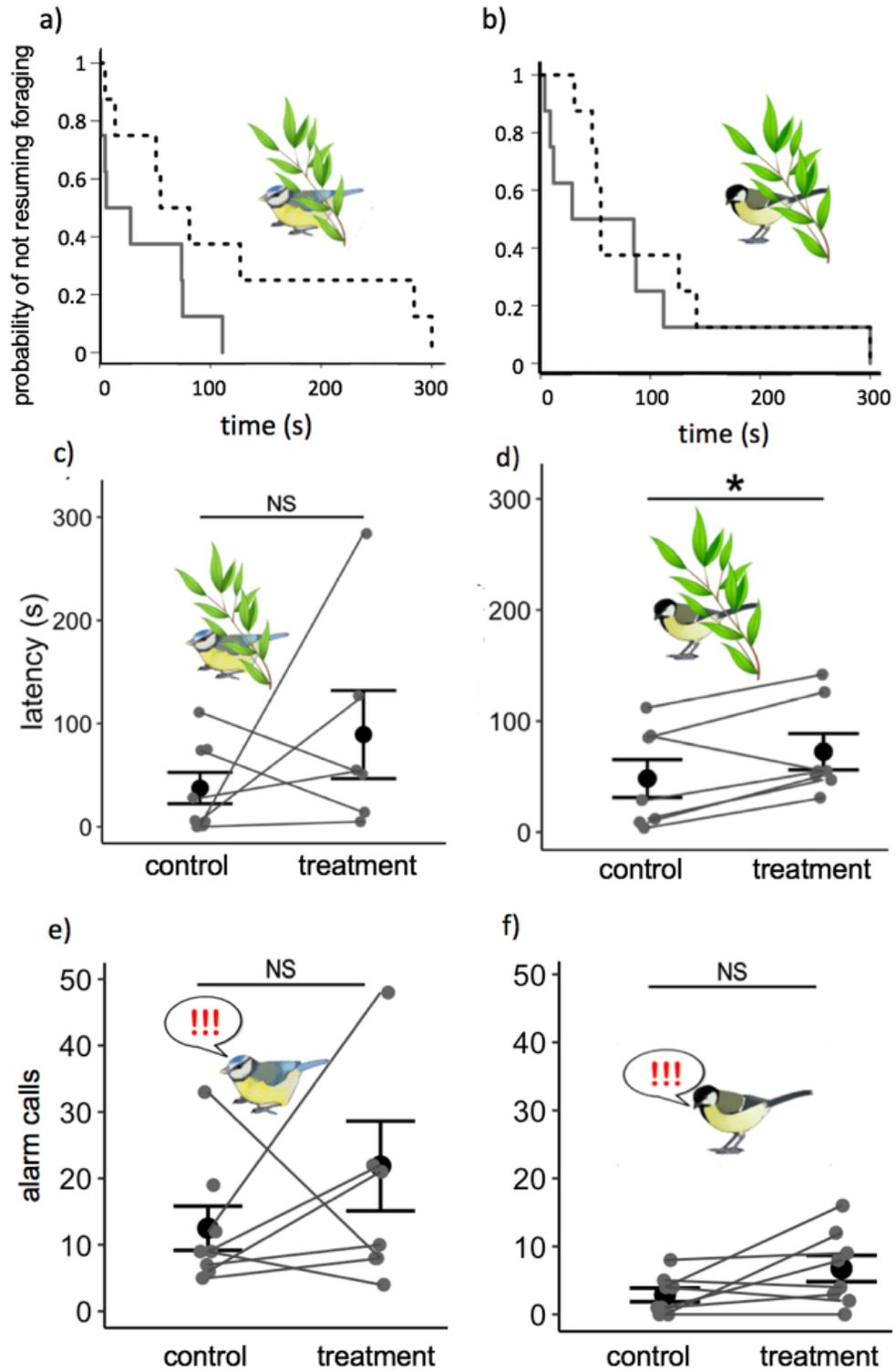


Figure 4. Associative learning of acoustic cues within demonstrator groups. a) Survival curves showing demonstrator latency to resume foraging before training. Demonstrators did not take significantly longer to resume foraging after playbacks of treatment sound (dashed line) versus control sound (solid line, see Results), b) Survival curves showing demonstrator latency to resume foraging after training. Demonstrators took significantly longer to resume foraging after playbacks of treatment sound (dashed line) than control sound playbacks (solid line, see Results). c) Demonstrator latency to resume foraging after treatment and control sound playbacks before training. d) Demonstrator latency to resume foraging after playbacks after training. Large black dots and bars represent means and standard error. Small grey dots represent individual birds and lines indicate paired samples from same individual within a single replicate. Asterisks correspond to $p < 0.05$, NS corresponds to $p \geq 0.05$. Censored birds (i.e., those that did not return within 300 seconds) are not shown here.

Figure 5. Tests for social transmission of antipredator response to heterospecific and conspecific observers after exposure to trained demonstrators. a) Survival curves showing conspecific observer latency to resume foraging after playbacks of treatment sound (dashed line) versus control sound (solid line) b) Survival curves showing heterospecific observer latency to resume foraging after social training, c) Latency of conspecific observers to resume foraging after treatment playbacks and control sound playbacks, d) latency of heterospecific observers to resume foraging after treatment playbacks and control sound playbacks, e) number of alarm calls made by conspecific observers after playbacks, f) number of alarm calls made by heterospecific observers after playbacks. Large black dots and bars represent means and standard error. Small grey dots represent individual birds and lines indicate paired samples from same individual within a single replicate. Note that individual birds that did not return within 300 s are not shown. Asterisks correspond to $p < 0.05$, NS corresponds to $p \geq 0.05$.



DISCUSSION

Evidence of interspecific social transmission of antipredator behaviour. Together, our results suggest that heterospecific observers can learn to associate a novel cue with predation threat without first-hand experience. Our results support findings from previous experimental work showing that antipredator behaviour can be acquired both through first-hand experience and secondary associations [14-15], and support the suggestion that flocking with heterospecifics gives greater access to social information that can enhance survival [45]. Social learning of predator avoidance may offer an adaptive advantage in dynamic environments; because our study population experiences strong spatial and temporal variation in food availability and predation risk, behavioural plasticity is likely under strong selection in these species [46]. Furthermore, as unfamiliar sounds were readily learnt, we suggest that both species exhibit an innate preparedness that increases likelihood of learning to associate any acoustic cue with predation.

Despite finding evidence that social transmission of antipredator information occurs between blue tits and great tits, we did not detect the same significant effect amongst blue tit demonstrators and observers. Specifically, while great tit observers increased latency to resume feeding after treatment playbacks, blue tit observers exhibited a non-significant increase in both alarm calling and latency to resume feeding. One possible explanation for the lack of observed conspecific social transmission is that Parid species differ in the manner in which they respond to predators. For example, blue tits have been shown to exhibit significantly more wing-flicking when presented with predator models that move and produce calls as compared to motionless, silent models [47], and perhaps great tits respond differently to such changes in predator model behaviour. It may also be the case that learning in the absence of a predator requires more

repetition within blue tits; previous work in which birds were trained without direct predator exposure included 10-12 training sessions [14,15], which is five to seven more than observers received in our experimental design. We also note that our sample sizes were relatively small (seven blue tit observers and seven great tit observers). Given that the non-significant responses of blue tit observers to the playbacks are in the expected direction, we cannot rule out the possibility that the absence of a detectable change in behaviour is due to lack of statistical power (see Appendix A). Although our results support the hypothesis that blue tits can learn to associate a novel acoustic cue with predation risk through direct experience with a predator, additional experiments that perhaps have longer training periods are needed to determine whether this behaviour can be socially transmitted between individuals in this species.

One issue that must be considered when interpreting our results is potential sensitization to the treatment sound due to repeated exposure during training. Because control sounds were presented only during test trials before and after training, whereas treatment sounds were presented multiple times, focal birds may have exhibited heightened responsiveness during treatment playbacks. However, if our results were caused by sensitization to the treatment sound, we would expect latency to feed after the sound was played to be significantly longer after repeated exposure. Rather, we saw a decrease in latency to return when birds were repeatedly exposed to the control sound (see Fig. 2). This suggests that, rather than birds becoming sensitized to the trained sound, they remained wary of the stimulus because it was paired with predator presentations and desensitized to the control sound, which was not associated with a threat. In future experiments, we advise that control sound playbacks that are not paired with a predator are conducted during the training period to enable testing of this alternative explanation [e.g. 14, 31].

Intriguingly, blue tit demonstrators did not produce significantly more alarm calls following exposure to treatment playbacks, suggesting that rather than learning from demonstrators' alarm calling, great tit observers learned from their behaviour. Although our findings cannot exclude the possibility that Parids also acquire anti-predator responses via acoustic association, our results present a different mechanism by which they may learn about predation risk. This therefore builds on recent work that has demonstrated that social learning can occur through acoustic association [15], and also suggests that there may be numerous ways in which individuals can acquire information about predators.

Level of perceived risk may encourage social learning. Interestingly, naïve observers adopted demonstrators' behaviour despite a lack of reinforcement during training, as the predator model was not presented after the initial demonstrator training. One possible explanation for this is that when costs of ignoring a cue are high, even unreliable social information is favoured over personal information [48]. Thus, as perceived risk increases, individuals are expected to copy rather than learn independently [49]. This tendency can enable extreme examples of cultural transmission of antipredator response to benign heterospecifics [50, 51], and can be used to train captive-bred animals before release [52]. Ultimately, learning strategy is likely determined by several factors, including the relative reliability of social and personal information, perceived cost of direct learning, degree of environmental variability, number of demonstrators, as well as observer and demonstrator identity.

Ecological and evolutionary implications. Two possible explanations for our results are that (1) the treatment sound is perceived as a vocalization produced by a novel predator, or (2) the treatment sound is perceived as an alarm call from a novel species. Neither can be ruled out within this experimental design; however, because sparrowhawks hunt primarily by surprise, as simulated in demonstrator training, the second alternative may be more likely. In either case, our results add to evidence that animals with complex vocal behaviours have evolved to efficiently process and use acoustic information, and that sympatric species may experience selection pressure to acquire acoustic information from both con- and heterospecifics. The ability to rapidly recognize and adjust behaviour in response to acoustic cues is expected to be adaptive for species that have evolved to efficiently encode and process sounds, such as most vertebrates [53], particularly passerine birds, which execute complex vocal communication tasks and acoustic environmental awareness [54]. These findings also suggest that within mixed-species communities, individuals may be predisposed to sharing and efficiently using social information from sympatric individuals, regardless of species. Our findings also add to research showing that social information transmission can facilitate recognition of novel predators [52], and suggest that social information acquired from heterospecifics may enable adaptation to dynamic environments [55].

One constraint of this experiment was that a single exemplar of each sound was used; we were therefore unable to test whether receivers were able to recognize a general class of non-identical acoustic signals. In order to determine whether our findings extend more broadly to alarm calling in wild animals, further experiments in which the acoustic parameters and presentation of the signal are varied are required. Finally, we suggest that future experiments also videotape playbacks in a manner that allows for measuring individual hiding

and freezing behaviour. Although this was not feasible given the layout of the aviary in which we conducted this experiment, it may be an important behaviour used by Parids in response to model predator presentations.

Taken together, our results suggest that social transmission of predator avoidance behaviour occurs between species, and that using social information rather than private information may be favoured in the context of predator avoidance. Ultimately, our findings may help also to explain how species-level attributes and interspecific social learning could mediate the formation of mixed-species communities and the establishment of new traditions and cultures.

ACKNOWLEDGMENTS

We thank L.M. Aplin for her insightful suggestions during the development of this project; K. McMahon, F. Bell, D. Wilson, and N. Carlson for their valuable assistance during field work; H. Klinck and the Bioacoustic Research Program for technical advice and support; M.A. Pardo and E.L. Mudrak for assistance in statistical analysis; and H.K. Reeve, Russel Ligon, the Cornell Animal Behavior Lunch Bunch, and the Sheldon Lab group for their valuable feedback. All artwork in figures was created by Megan Bishop. This research was made possible by support to S.C.K. from the Cornell Lab of Ornithology and Department of Neurobiology and Behavior.

WORKS CITED

1. Danchin, É., Giraldeau, L. A., Valone, T. J., and Wagner, R. H. (2004). Public information: from nosy neighbors to cultural evolution. *Science* 305, 487-491.
2. Boyd, Robert, and Peter J. Richerson. (1985). *Culture and the evolutionary process* (Chicago: University of Chicago Press).
3. Laland, K. N. (2004). Social learning strategies. *Learn. Behav.* 32, 4-14.
4. Hoppitt, W., and Laland, K. N. (2013). *Social learning: an introduction to mechanisms, methods, and models* (Princeton: Princeton University Press).
5. Griffin, A. S. (2004). Social learning about predators: a review and prospectus. *Learn. Behav.* 32, 131-140.
6. Carthey, A. J., and Blumstein, D. T. (2017). Predicting Predator Recognition in a Changing World. *Trends Ecol. Evol.*
7. Whiten, A. (2000). Primate culture and social learning. *Cog. Sci.* 24, 477-508.
8. Slagsvold, T., and Wiebe, K. L. (2011). Social learning in birds and its role in shaping a foraging niche. *Proc. Biol. Sci.* 366, 969.
9. Seppänen, J. T., Forsman, J. T., Mönkkönen, M., and Thomson, R. L. (2007). Social information use is a process across time, space, and ecology, reaching heterospecifics. *Ecology* 88, 1622-1633.
10. Goodale, E., Beauchamp, G., Magrath, R. D., Nieh, J. C., and Ruxton, G. D. (2010). Interspecific information transfer influences animal community structure. *Trends Ecol. Evol.* 25, 354-361.
11. Avarguès-Weber, A., Dawson, E. H., and Chittka, L. (2013). Mechanisms of social learning across species boundaries. *J. Zool.* 290, 1-11.
12. Templeton, C. N., and Greene, E. (2007). Nuthatches eavesdrop on variations in heterospecific chickadee mobbing alarm calls. *P. Natl. A. Sci. USA.* 104, 5479-5482.
13. Magrath, Robert D., B. J. Pitcher, and J. L. Gardner. (2007). A mutual understanding? Interspecific responses by birds to each other's aerial alarm calls. *Behav. Ecol.* 18, 944-951.
14. Magrath, R.D., Haff, T.M., McLachlan, J.R., and Iqic, B. (2015). Wild birds learn to eavesdrop on heterospecific alarm calls. *Curr. Biol.* 25, 2047–2050.
15. Potvin, D. A., Ratnayake, C. P., Radford, A. N., and Magrath, R. D. (2018). Birds Learn Socially to Recognize Heterospecific Alarm Calls by Acoustic Association. *Curr. Biol.* 28, 2632-2637.

16. Nagell, K., Olguin, R. S., and Tomasello, M. (1993). Processes of social learning in the tool use of chimpanzees (*Pan troglodytes*) and human children (*Homo sapiens*). *J. Comp. Psychol.* *107*, 174.
17. Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., and Sheldon, B. C. (2015). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature* *518*, 538-541.
18. Catchpole, C. K., and Slater, P. J. (2003). *Bird song: biological themes and variations* (Cambridge: CUP).
19. Page, R. A., and Ryan, M. J. (2006). Social transmission of novel foraging behavior in bats: frog calls and their referents. *Curr. Biol.* *16*, 1201-1205.
20. Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* *210*, 801-803.
21. Blumstein, D. T., and Armitage, K. B. (1997). Alarm calling in yellow-bellied marmots: I. The meaning of situationally variable alarm calls. *Anim. Behav.* *53*, 143-171.
22. Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proc. Biol. Sci.* *268*, 2315-2324.
23. Magrath, R. D., Haff, T. M., Fallow, P. M., and Radford, A. N. (2015). Eavesdropping on heterospecific alarm calls: from mechanisms to consequences. *Biol. Rev.*, *90*, 560-586.
24. Smith, J. M. (1965). The evolution of alarm calls. *Am. Nat.* *99*, 59-63.
25. Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* *46*, 35-57
26. Fallow, P. M., Gardner, J. L., and Magrath, R. D. (2011). Sound familiar? Acoustic similarity provokes responses to unfamiliar heterospecific alarm calls. *Behav. Ecol.* *22*, 401-410.
27. Huang, X., Metzner, W., Zhang, K., Wang, Y., Luo, B., Sun, C., Tinglei, J., and Feng, J. (2018). Acoustic similarity elicits responses to heterospecific distress calls in bats (Mammalia: Chiroptera). *Anim. Behav.* *146*, 143-154.
28. Magrath, R. D., Pitcher, B. J. and Gardner, J. L. (2009b). Recognition of other species' aerial alarm calls: speaking the same language or learning another? *Proc. Biol. Sci.* *276*, 769-774.
29. Hollen, L. I., and Radford, A. N. (2009). The development of alarm call behaviour in mammals and birds. *Animal Behaviour*, *78*(4), 791-800.

30. Colombelli-Negrel, D., Hauber, M. E., Robertson, J., Sulloway, F. J., Hoi, H., Griggio, M. and Kleindorfer, S. (2012). Embryonic learning of vocal passwords in superb fairy-wrens reveals intruder cuckoo nestlings. *Curr. Biol.* 22, 2155–2160.
31. Dutour, M., Léna, J. P., Dumet, A., Gardette, V., Mondy, N., and Lengagne, T. (2019). The role of associative learning process on the response of fledgling great tits (*Parus major*) to mobbing calls. *Anim. Cogn.* 22, 1095-1103.
32. Wheeler, B. C., Fahy, M., and Tiddi, B. (2019). Experimental evidence for heterospecific alarm signal recognition via associative learning in wild capuchin monkeys. *Anim. Cogn.* 1-9.
33. Svensson, L. (1992). Identification guide to European passerines (BTO: Thetford, UK).
34. Xeno-canto. <https://www.xeno-canto.org>.
35. Audacity 2.1.1. The Audacity Team (2015). <http://audacityteam.org>.
36. Peake, T. M., Terry, A. M. R., McGregor, P. K., and Dabelsteen, T. (2002). Do great tits assess rivals by combining direct experience with information gathered by eavesdropping? *Proc. Biol. Sci.* 269, 1925-1929.
37. Templeton, C. N., Zollinger, S. A., and Brumm, H. (2016). Traffic noise drowns out great tit alarm calls. *Curr. Biol.* 26, 1173-1174.
38. Vedder, O., Bouwhuis, S., and Sheldon, B. C. (2014). The contribution of an avian top predator to selection in prey species. *J. Anim. Ecol.* 83, 99-106.
39. Gentle, L. K., and Gosler, A. G. (2001). Fat reserves and perceived predation risk in the great tit, *Parus major*. *Proc. Biol. Sci.* 268, 487-491.
40. Voelkl, B., Firth, J. A., and Sheldon, B. C. (2016). Nonlethal Predator effects on the turnover of wild bird flocks. *Sci. Rep.* 6, 33476.
41. Carlson, N. V., Healy, S. D., and Templeton, C. N. (2017). A comparative study of how British tits encode predator threat in their mobbing calls. *Anim. Behav.* 125, 77-92.
42. Therneau, T.M. (2018). *coxme: Mixed Effects Cox Models*. R package., 2.2-10 Edition.
43. Arnholt, A.M. and Evans, B. (2017). *BSDA: Basic Statistics and Data Analysis*, R package., 1.2-0 Edition.
44. R Core Team. (2018). *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing).
45. Krebs, J. R. (1973). Social learning and the significance of mixed-species flocks of chickadees (*Parus spp.*). *Can. J. Zool.* 51, 1275-1288.

46. Lima, S. L., and Dill, L. M. (1990). Behavioral decisions made under the risk of predation: a review and prospectus. *Can. J. Zool.* 68, 619-640.
47. Carlson, N. V., Pargeter, H. M., and Templeton, C. N. (2017). Sparrowhawk movement, calling, and presence of dead conspecifics differentially impact blue tit (*Cyanistes caeruleus*) vocal and behavioral mobbing responses. *Behav. Ecol. Sociobiol.* 71, 133.
48. Galef, B. G., and Laland, K. N. (2005). Social learning in animals: empirical studies and theoretical models. *AIBS Bull.* 55, 489-499.
49. Webster, M. M., and Laland, K. N. (2008). Social learning strategies and predation risk: minnows copy only when using private information would be costly. *Proc. Biol. Sci.* 275, 2869-2876.
50. Curio, E., Ernst, U., and Vieth, W. (1978). Cultural transmission of enemy recognition: one function of mobbing. *Science* 202, 899-901.
51. Vieth, W., Curio, E., and Ernst, U. (1980). The adaptive significance of avian mobbing. III. Cultural transmission of enemy recognition in blackbirds: cross-species tutoring and properties of learning. *Anim. Behav.* 28, 1217-1229.
52. Griffin, A. S., Blumstein, D. T., and Evans, C. S. (2000). Training captive-bred or translocated animals to avoid predators. *Conserv. Biol.* 14, 1317-1326.
53. Popper, N., and Fay, R. (1997). Evolution of the ear and hearing: issues and questions. *Brain Behav. Evol.* 50, 213-221.
54. Manley, G. A., and Gleich, O. (1992). Evolution and specialization of function in the avian auditory periphery. In *The Evolutionary Biology of Hearing*. (Springer: New York), pp. 561-580.
55. Farine, D. R., Aplin, L. M., Sheldon, B. C., and Hoppitt, W. (2015). Interspecific social networks promote information transmission in wild songbirds. *Proc. Biol. Sci.* 282,1803.

CHAPTER 2

A MACHINE LEARNING APPROACH FOR CLASSIFYING AND QUANTIFYING ACOUSTIC DIVERSITY

^{1,2,3}Sara Keen, ³Karan Odom, ⁴Marcelo Araya-Salas, ^{2,3}Mike Webster, ⁵Timothy F. Wright

¹Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Cornell University,
Ithaca, NY, 14850, USA.

²Department of Neurobiology and Behavior, Cornell University, Ithaca, NY, 14850, USA.

³Cornell Lab of Ornithology, Cornell University, Ithaca, NY, 14850, USA.

⁴. Sede del Sur, Universidad de Costa Rica, Golfito, 60701, Costa Rica

⁵Department of Biology, New Mexico State University, Las Cruces, NM 88003, USA.

ABSTRACT

1. Assessing diversity of discretely varying behavior is a classical ethological problem. In particular, the challenge of calculating an individuals' or species' repertoire size is often an important step in ecological and behavioral studies, but a reproducible and broadly applicable method for accomplishing this task is not currently available to researchers.
2. We offer a generalizable method to automate the calculation and quantification of acoustic diversity using an unsupervised random forest framework. We tested our method using natural and synthetic data sets of known repertoire sizes that exhibit

variation in common acoustic features and in recording quality, which allowed us to evaluate performance using signals with standardized variation. We tested two approaches to estimate acoustic diversity using the output from unsupervised random forest analyses: (i) cluster analysis to estimate the number of discrete acoustic signals (e.g., repertoire size) and (ii) an estimation of acoustic area in acoustic feature space, as a proxy for repertoire size.

3. Generally, we find that our unsupervised analyses classify acoustic structure with high accuracy. We also find that both approaches to estimate acoustic diversity offer robust means of estimating the number of discrete elements in scenarios when repertoire size is small to intermediate (5-20 unique elements). However, for larger data sets (20-100 unique elements), we find that calculating the size of the area occupied in acoustic space is a more reliable proxy for estimating repertoire size.
4. We conclude that our implementation of unsupervised random forest analysis offers a generalizable tool that researchers can apply to classify acoustic structure of diverse data sets. Additionally, using output from these analyses can be used to compare the distribution and diversity of signals in acoustic space, creating opportunities to quantify and compare the amount of acoustic variation among individuals, populations, or species in a standardized way.

INTRODUCTION

Many animals use vocal signals to transmit information and mediate a wide range of social behaviors, from resource competition to attracting mates (Payne et al. 1986, Kroodsma and

Miller 1996, Gerhardt and Huber 2002, Catchpole and Slater 2003, Janik 2009). Owing to the ubiquity and ecological importance of acoustic signaling, quantifying and comparing animal vocalizations is a major part of animal behavior and communication systems research. Data from several studies suggest that signals often fall into distinct categories based on their acoustic structure (e.g. birds, Kroodsma 1982; cetaceans, Janik 2009; primates, Owren et al. 1992). Such categories are often observed at the species level when conspecifics use a shared repertoire of distinct acoustic signals that are associated with different contexts (Marler 1982, Seyfarth and Cheney 2003). Distinct categories can also arise within a signal type, as when an individual uses several signal variants that have the same functional role (e.g., the song repertoires of many songbirds comprise multiple song types, Catchpole and Slater 2003).

Classifying or quantifying variation in animal signals is fundamental to many questions in animal communication. For example, metrics derived from measuring the number of unique elements or vocalizations produced by an individual, such as repertoire size and acoustic diversity, have been shown to correlate with indicators of quality, including territory size, cognitive ability, brain morphology, and levels of stress during early stages of development (Sewall et al. 2013, Devoogd et al. 1993, Podos et al. 2009). At the population level, differences in acoustic signals can facilitate species recognition (e.g., amphibians, Ryan 1985) and can play an important role in speciation by promoting isolation between sympatric groups (e.g., crickets, Mullen et al. 2007; birds, Mason et al. 2017). When assessing entire ecosystems, acoustic diversity, or the amount of variation within and among populations' or communities' vocal repertoires, serves as a commonly used metric for assessing ecosystem health or demographic aspects of communities (Seuer et al. 2008, Laiolo et al. 2008, Pijanowski et al. 2011). For these reasons, quantifying acoustic diversity is often an important

step in addressing questions and testing hypotheses regarding the social and ecological factors influencing signal function and evolution.

Classifying signals is often difficult or time consuming because acoustic variation across environments, individuals, or even different renditions of a signal by the same individual can be considerable. Furthermore, not all variation in acoustic structure is discrete, and therefore can be difficult to classify (Wadewitz et al. 2015). Within behavioral ecology, a common approach for quantifying variation among signals is to estimate repertoire size or element diversity. In this study, we consider diversity as the number of discrete vocalization types or elements used by an individual or species (this differs from ecological definition of diversity, which describes both the number and evenness of entities in the environment). While it is theoretically possible to count every discrete acoustic element in a data set of vocal elements, for animals with large repertoire sizes it is common to subsample a species repertoire and use either accumulation curves or a capture-recapture analysis to estimate repertoire size (Wildenthal 1965, Garamzegi et al. 2002, Catchpole and Slater 2003, Garamzegi et al. 2005). However, this approach requires first manually classifying elements or vocalizations, a process that can be subjective and vary among observers, and may become unwieldy or even nearly impossible for species with large repertoires or multispecies studies. Applying these approaches can also be complicated by the tendency of subsampling to result in biased measurements in some data sets (Botero et al. 2008). In recent years, several techniques have been developed which improve upon these methods (e.g., Peshek and Blumstein 2011; Kershenbaum et al. 2015), including approaches that use an information theory-based approach to quantify individuality of vocal signals (Beecher 1989, Freeberg and Kucas 2012, Linhart et al. 2019). Additionally, methods have been developed to help

distinguish among more graded element types (e.g., Wadewitz et al. 2015). Nevertheless, the general challenge of quantifying repertoire size still exists with many of these methods: human-based classification is both time intensive and unavoidably subjective.

In passive acoustic monitoring and quantification of soundscapes, there is an emphasis on creating fully automated approaches for classification and measurement of acoustic signals. One such approach, acoustic indices, has been used to quantify ecosystem-level to individual behavioral variation (Seuer et al. 2014). Studies suggest that ecosystem acoustic diversity indices may be indicators of biodiversity (Sueur et al. 2008a, Harris et al. 2016), degree of functional and/or phylogenetic diversity within a community (Gasc et al. 2013), and a proxy for local vocal activity (Pieretti et al. 2011). These metrics have become increasingly important to ecological assessment and monitoring (Gibb et al. 2019), however, they are often calculated at scales that are more appropriate to ecosystem or community ecology.

Unlike soundscape analysis, measuring acoustic diversity on the species- or individual-level requires quantifying differences between discrete elements. Machine learning offers an automated and objective approach for such classification tasks, and is a powerful tool for detecting and distinguishing among vocal signals from different species (e.g., Acevedo et al. 2009, Briggs et al. 2013, Hershey et al. 2017, Stowell et al. 2019). In particular, unsupervised machine learning approaches offer several advantages that enhance their value for assessing behavioral diversity, namely in that they do not require a labeled training data set or a priori assumptions about the structure of data (Valletta et al. 2017). Unsupervised techniques can also determine which acoustic parameters contribute most to classification or splitting data into classes, therefore relieving researchers from the need to make potentially subjective choices about feature selection (Breiman 2001). Unsupervised

analyses have shown high performance in the classification of vocal signals to species as compared to other approaches (Keen et al. 2014), including in the case of large data sets (Stowell and Plumbley 2014), and there appears to be much promise in applying these techniques to evaluate acoustic diversity (Ulloa et al. 2018). However, a widely applicable tool for assessing acoustic diversity at the levels of individual, species, or communities is not readily available.

In this paper, we present and evaluate the use of unsupervised machine learning for classifying and quantifying acoustic diversity in animal signals. Specifically, we examine two approaches for estimating repertoire size: (1) a clustering method to identify discrete numbers of acoustic units and (2) an acoustic area calculation as a proxy for repertoire size. We evaluate the accuracy of these approaches on multiple data sets with known varying acoustic structure. Three unique aspects of our approach help ensure this method will be highly generalizable to diverse acoustic signals. First, we test algorithm performance using both field-recorded and synthesized acoustic data sets with known sample sizes and variation, making it possible to evaluate the usefulness of our method under a variety of conditions. Second, we incorporate several of the most commonly used acoustic parameters for characterizing signal structure. Third, we used test data sets with realistic distributions of variation and background noise, making it possible to evaluate the robustness of this approach to variable acoustic structures and across a range of recording scenarios. Together, these steps allow us to rigorously evaluate performance and provide recommendations about application in different scenarios. Based on our results, we believe this technique offers a powerful tool for researchers to quantify a diversity across taxa and communities.

METHODS

We estimated acoustic diversity for a collection of natural and synthetic acoustic signals using a machine learning approach (random forest) and evaluated the performance of this method following the workflow in Figure 1. This process involved creating sets of synthetic acoustic signals with known repertoire sizes and known amounts of structural variation, extracting acoustic features from these signals, running unsupervised random forest analyses to calculate pairwise distances between signals, and estimating repertoire size using both cluster analysis the size of the acoustic feature space (hereafter referred to as acoustic space). In addition, we evaluated how variation in repertoire size and acoustic structure affects the accuracy of supervised random forest.

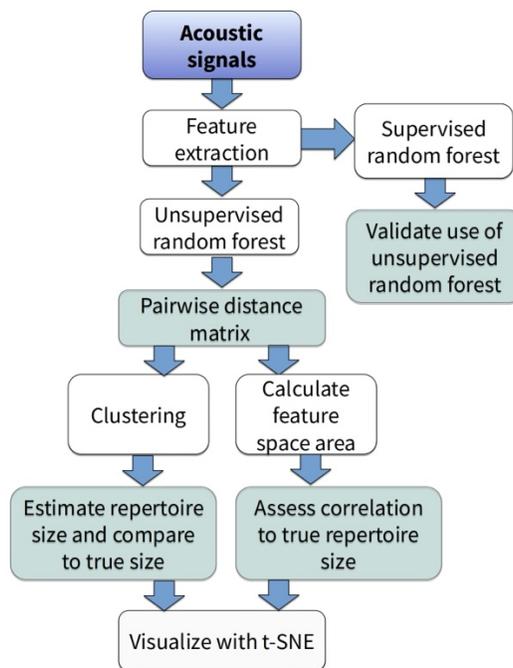


Figure 1. Flowchart of study design.

Using a random forest approach was integral to our workflow for several reasons. A key advantage of random forest is its ability to determine which feature measurements best

divide data into distinct categories; therefore, it is possible to use a large number of features and allow the algorithm to determine which are most useful for a given data set. Random forest also offers several additional advantages over other machine learning techniques: it is robust to collinearity, outliers and unbalanced data sets, is efficient even with large and highly multi-dimensional data sets, can be used in both a supervised and unsupervised manner, can handle non-monotonic relationships, ignores non-informative variables, produces low bias estimates, computes proximity of observations which can be used for representing trait spaces, and can be used to identify variables that contribute most to finding structure within a data set (Valletta et al. 2017). For these reasons, combining random forest with a large suite of automated acoustic feature measurements holds much promise as a generalizable tool for acoustic classification tasks.

Test data sets. We evaluated the performance of our proposed method using four data sets: annotated field recordings of long-billed hermits (*Phaethornis longirostris*), annotated lab recordings of budgerigars (*Melopsittacus undulatus*), and two collections of synthetic data sets that were modeled on natural vocalizations of these two species (see Table 1 for a summary of data sets and Figure 2 for sample spectrograms). This enabled us to assess performance using vocal signals collected from live birds that reflect the naturally occurring variation between individuals as well as with signals that have *a priori* defined discrete variation. The use of synthetic data sets as test cases also allowed us to conduct repeated tests of algorithm performance under different conditions.

Field recordings of long-billed hermits were collected from known individuals in wild populations at La Selva Biological Station, Costa Rica (10°, 25' N; 84°, 00' W), between 2008

and 2017. Males in this species live in territorial leks that exhibit local songs that are shared by sub groups of individuals (i.e. singing neighborhoods) within a lek (Araya-Salas and Wright 2013). For this study, we used songs recorded from 16 leks (mean \pm SE songs per group = 3.1 ± 0.51). Because the song types used by long-billed hermits change over time, it was possible to use songs recorded from the same lek in different years to compile a sample of 50 unique song types. We verified that song types exhibited distinct spectro-temporal structures using spectrograms created in the R package warbleR (Araya-Salas and Smith-Vidaurre 2017) (see Figure B1 for spectrograms). To create the test data set for this study, we identified the 50 song types had the most samples, and selected the 10 recordings with the highest signal-to-noise ratio for each type, yielding a data set of 500 signals.

Laboratory recordings of budgerigar contact calls were collected between July and November 2010 from a laboratory population originally acquired from a captive breeder. Individual budgerigars typically have repertoires of 2-5 acoustically distinct contact call types that are shared with some other individuals within their flock. Contact calls were recorded from 38 different individuals that were temporarily isolated from their flock mates in a homemade acoustic chamber constructed of an Igloo cooler lined with acoustic foam with a clear plexiglass door as described in Dahlin et al. (2014). In order to promote calling during recording sessions, we played recordings of unfamiliar budgerigar vocalizations at low amplitudes and also ensured that isolated individuals were in visual contact with the flock mates. Calls were recorded during 30 min sessions that occurred twice per week using a Audio-Technica Pro 37 microphone input to a Dell DHMPC running Syrinx 2.6 (Burt 2006) with a 22.05 kHz sampling rate. Calls were automatically partitioned and saved to separate wav files by Syrinx. Trained research assistants visually assessed spectrograms made from

wav files and assigned calls to classes using Raven 1.3 (Cornell Lab of Ornithology). Call classification was subsequently verified using a discriminant function analysis as described in Dahlin et al. (2014). To select the calls used in this study, we randomly selected 35 contact calls from each of 15 unique call types, resulting in a data set of 525 signals.

Synthetic data set creation. To create the synthesized song data sets used for testing, we first extracted the dominant frequency contours of the natural bird vocalizations (long-billed hermit songs and budgerigar calls). We then modeled these time series of frequency values using autoregressive moving average (ARMA) models. Briefly, these models find the maximum likelihood estimates of the parameters in a polynomial equation predicting the variation in time series. These parameters can be later used to simulate new time series, or, in our case, new dominant frequency contours for generating synthetic vocalizations. ARMA model parameters were estimated for each natural data set independently and later used to simulate frequency contours resembling those original data sets. These contours were converted into an audio clip using the R soundgen package (Anikin 2019). We allowed the synthetic sounds to vary in three features: duration (short: 150 ms; long: 300 ms), harmonic content (low and high) and background noise (low: 20 dB signal-to-noise ratio; high: 2 dB signal-to-noise ratio). Duration values were based on the observed variation in long-billed hermit and budgerigar data sets (mean \pm SE duration: long-billed hermit songs: 143.32 ± 17.5 ms, budgerigar calls: 138.23 ± 20.4 ms; histograms shown in Figure B2). The natural vocalizations used as template have very little harmonic content. Hence, harmonic content was simulated arbitrarily as frequency contours an octave (twice the frequency) and a fifth (2.5 times) above the dominant frequency contour. Variation in background noise was

generated by adding normally distributed noise (i.e. white noise) to each signal. In order to adequately test the ability of our method to estimate repertoire size and to determine whether this can be approximated by calculating the area occupied in acoustic space, we used this process to synthesize data sets with repertoire sizes of 5, 10, 15, 20, 50, or 100 unique elements. Each element type was represented by 10 examples. Variation within element types (i.e. between examples) was generated by adding randomly generated values to the simulated frequency contours drawn from a normal distribution with mean of 0 and a standard deviation equal to a tenth of the standard deviation in frequency for each contour. For each possible repertoire size, we used all possible combinations of duration, harmonic content, and background noise, resulting in 48 synthetic data sets for both long-billed hermit songs and budgerigar calls (see Table 1). See the Appendix B for further details of data synthesis. Sample spectrograms of signals from each data set are shown in Figure 2.

Table 1. Summary of test data sets used to evaluate performance.

Data description	Recording type	Number of data sets	Unique elements in repertoire	Examples of each element
Long billed hermit songs	Field	1	50	10
Budgerigar calls	Laboratory	1	15	35
			8 x 5	
			8 x 10	
Synthetic long-billed hermit songs	Synthetic	48	8 x 15	10
			8 x 20	
			8 x 50	
			8 x 100	
			8 x 5	
			8 x 10	
Synthetic budgerigar calls	Synthetic	48	8 x 15	10
			8 x 20	
			8 x 50	
			8 x 100	

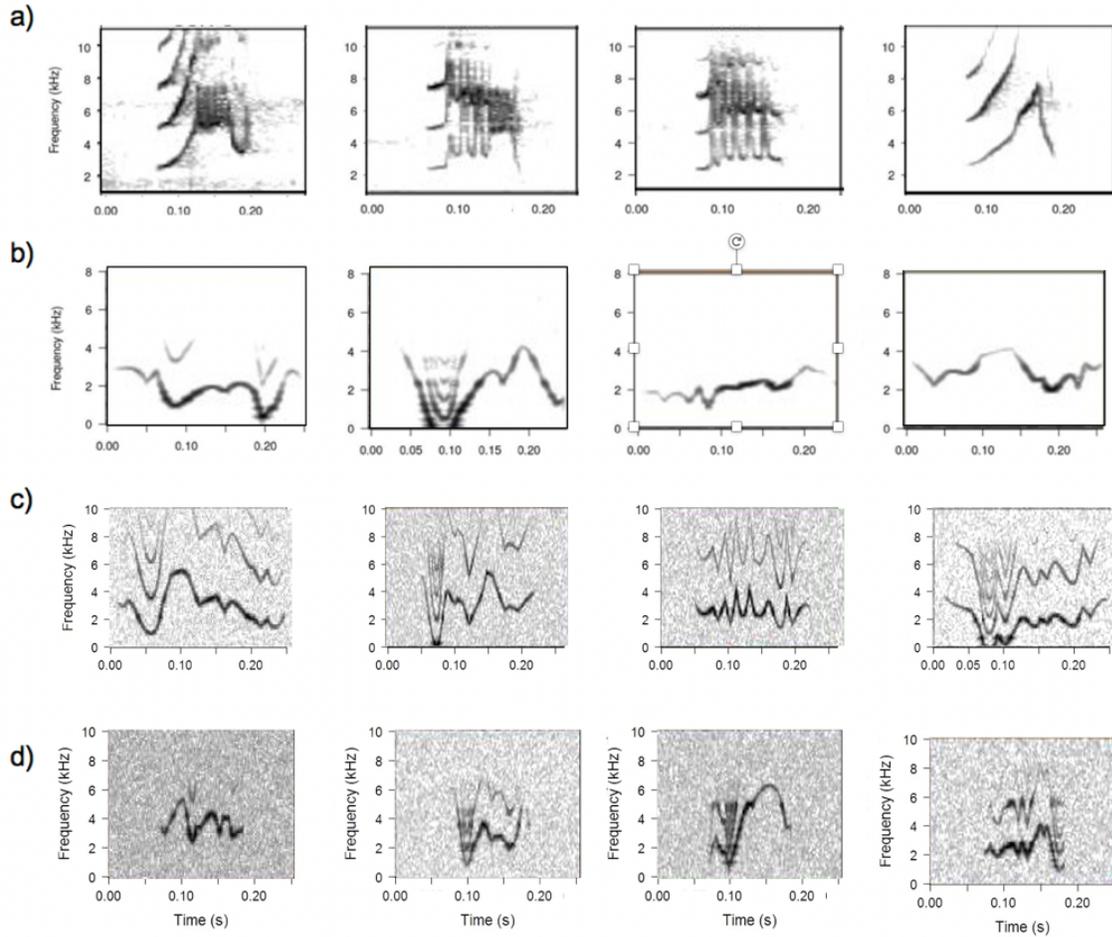


Figure 2. Spectrograms with examples from each data set. Example spectrograms from acoustic signals used to test algorithm performance from data sets including a) field recordings of long billed hermit songs, b) lab recordings of budgerigar songs, c) synthetic long billed hermit songs, d) synthetic budgerigar songs.

Acoustic feature measurements. We collected a suite of acoustic measurements from each audio clip in every test data set. We first applied a 500 Hz high pass filter to all audio clips to remove low frequency noise, and then created spectrograms for each sample clip using 300-point FFT with a Hann window and 90% overlap. We extracted several common acoustic feature measurements from each signal. These included 181 descriptive statistics of Mel Frequency cepstral coefficients (MFCCs; Lyon and Ordubadi 1982, *sensu* Salamon et al. 2014) and 28 acoustic parameters using the R packages warbler and sewave (Araya-Salas

and Smith-Vidaurre 2017, Sueur et al. 2008), which included commonly used acoustic measurements such as peak frequency, bandwidth, duration, as well as robust measurements based on energy distributions. We also calculated two pairwise distance matrices for every data set: one using spectrogram cross correlation (Clark et al. 1987) and one using dynamic time warping (Wolberg 1990). We then used classic multi-dimensional scaling (MDS) to translate the SPCC and DTW distance matrices into five-dimensional space, and used the axis coordinates for each sample as additional feature measurements (i.e., five SPCC MDS coordinates and five DTW MDS coordinates per sample). Together, this resulted in a vector of 219 feature measurements for each signal. The feature vectors for each audio clip were collated into a single matrix for each data set. We removed any collinear measurements from the matrix, used a Box-Cox transformation to improve normality, and scaled and centered all feature values. The resulting matrix was used as the input into the supervised and unsupervised random forest models.

Supervised random forest analyses. To evaluate the ability of random forest to classify signals into the correct categories, we used a supervised random forest created with the randomForest R package (Liaw and Weiner 2002), to classify the labeled signals in each data set into separate categories. Here, “supervised” denotes that the random forest model was created using a labeled data set. When using a supervised random forest approach, individual decision trees are constructed by splitting data into two classes at each node using a randomly selected feature measurement, with the goal of optimizing the split between labeled classes. Because all data sets were labeled by either human experts (field and lab recordings) or by software (synthetic data), we could then assess how well the supervised random forest models

were able to classify signals from the same category together using the out-of-bag error estimate (Breiman 2001). When using a supervised random forest, out-of-bag error is calculated by iteratively removing a single sample and building a random forest model with the remaining data, and then testing whether that sample is classified to the same category as other samples from the same class. These supervised random forest analyses served as a proof of concept, as it confirmed that models constructed from the selected acoustic features could accurately be assigned to the expected categories.

Unsupervised random forest analyses. To determine whether our method can be used to estimate repertoire size or acoustic diversity for unlabeled data, we created an unsupervised random forest model for each data set using the randomForest R package (Liaw and Weiner 2002). Unlike the supervised random forest approach, an unsupervised random forest uses unlabeled samples to create a collection of decision trees by optimally splitting the distribution of values for a randomly selected feature measurement at each node.

Unsupervised random forests are often used with the goal of finding underlying structure within data (Breiman 2001). This is possible with unlabeled data because decision trees assign all samples to end nodes, i.e. different classes, and one can then calculate the pairwise distance between samples within a data set as the proportion of times a pair of samples is classified in the same end node. For this study, each unsupervised random forest model was constructed using 10,000 decision trees that were built using the unlabeled feature measurements from each data set. We then used the output of each unsupervised model to obtain pairwise distances between all samples within each data set.

Performance evaluation. We used several metrics to evaluate how well our method could assign unlabeled signals into different classes. First, we assessed performance of each supervised random forest model by calculating out-of-bag error rates, which provided a misclassification rate for each data set. Using these values, we examined whether duration of audio clips (long vs. short), harmonic content (high vs. low), level of background noise (high vs. low), or number of discrete elements influenced the ability of models to assign signals to the correct class.

We evaluated how well the unsupervised analysis could measure acoustic diversity using two approaches: by estimating number of unique elements (i.e., repertoire size) in each data set and by calculating the area of the acoustic space occupied by all signals in a data set. To estimate repertoire size, we applied partitioning around medoids to the pairwise distance matrix returned by the unsupervised random forests for each data set (Kaufman and Rousseeuw 2009). For each data set, we calculated silhouette width to determine the optimal number of clusters, and then calculated the difference between this value (the estimated repertoire size) and the true repertoire size. For each data set, we also calculated the classification accuracy by assigning each cluster a label corresponding to the signal type that was most frequently placed in that cluster, and then dividing the total number of correctly assigned samples by the number of samples in the data set. We also calculated the adjusted Rand index for each data set, which is a metric of how often samples of the same type are assigned to the same cluster, and different types assigned to different clusters (Rand 1971).

To create the acoustic space, we first applied non-metric multidimensional scaling to the pairwise distance matrix produced by the unsupervised random forest created for each data set. We then calculated acoustic area as the 95% minimum convex polygon (i.e.

excluding the proportion of outliers above 95%) of these points. We then used Spearman's rank correlation to test whether acoustic area increased with true repertoire size.

We ran these analyses on the four collections of data sets described above. Lastly, in order to visualize how well the unsupervised analyses clustered distinct signal types in our data sets, we used the t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction technique to display all samples in two dimensions (Maatan and Hinton 2008). All statistical analyses were conducted using the R packages cluster, tsne, MASS, and adehabitatHR (Maechler et al. 2019, Donaldson 2016, Venables and Ripley 2002, Calenge 2006).

RESULTS

Supervised random forest performance. Out-of-bag error was below what would be expected by chance for all supervised random forest models: field recordings of long-billed hermits: 0.04, lab recordings of budgerigars: 0.053; synthetic long-billed hermit data sets (mean \pm SE): 0.02 ± 0.043 ; synthetic budgerigar data sets: 0.049 ± 0.017 (see Appendix B for further details). However, we observed that certain signal characteristics in our synthetic calls sets influenced error rates. Namely, synthetic long billed hermit songs that have low harmonic content or high background noise have higher out-of-bag error rates, and typically error rates were higher in long billed hermits than in budgerigars. Synthetic data sets that had higher numbers of discrete element types also had higher out-out bag error rates (Figure 3). Variable importance rankings indicating which feature measurements were most useful in splitting data into distinct classes were different for each of the four data set types used for testing (Table B1).

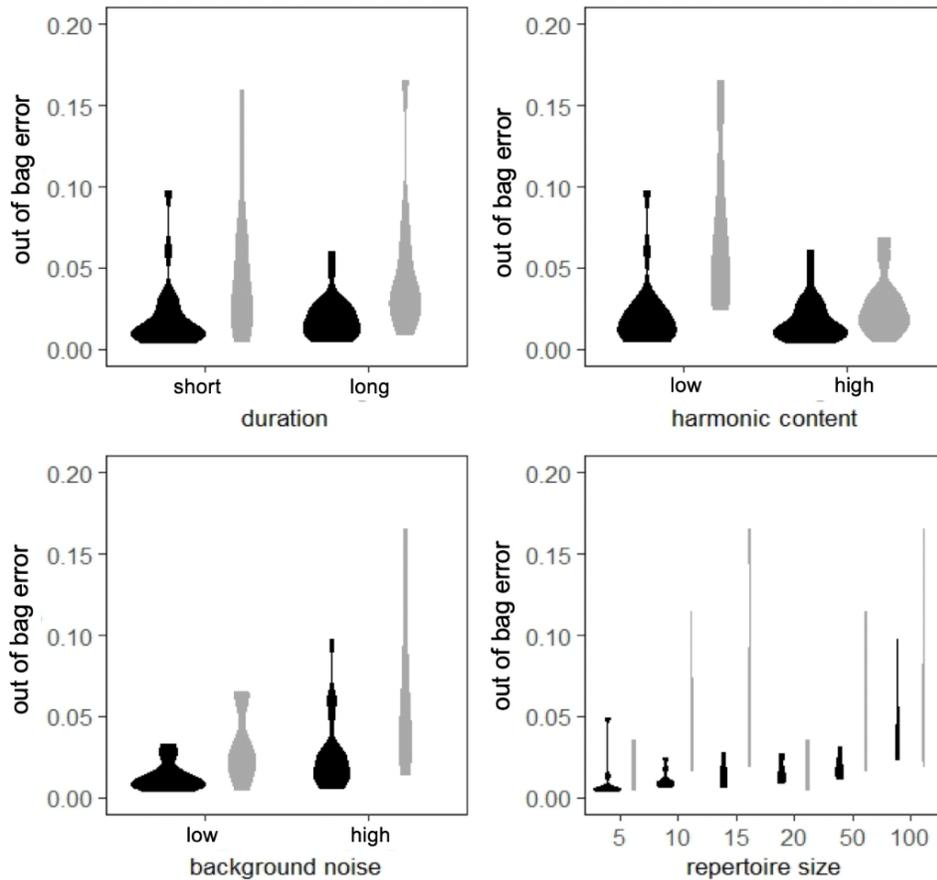


Figure 3. Out-of-bag error rates for supervised random forest models created for synthetic data sets with varying a) duration, b) harmonic content, c) levels of background noise. Black violin plots show results for synthetic budgerigar and gray plots results for synthetic long billed hermit data sets.

Unsupervised random forest performance and calculating acoustic diversity. Using cluster analysis to evaluate repertoire size, we observed that our estimates of repertoire size were most accurate for synthetic data sets that contained 20 or fewer unique elements (Figure 4a). Classification accuracy was often above 90% for data sets with five unique elements, and decreased as the true number of discrete elements in a data set increased, reaching around 60% for data sets with 100 unique elements (Figure 4b). Similarly, adjusted Rand indices were relatively high for synthetic data sets with small numbers of unique elements, and

decreased among data sets as the number of unique elements increased (Figure 4c). An exception to this pattern was the synthetic budgerigar data sets with five unique elements, which had lower adjusted Rand indices because data were often clustered into less than five classes. The scatter plots in Figure 5a-d illustrate the ability of the unsupervised analysis to cluster synthetic signals of the same class together.

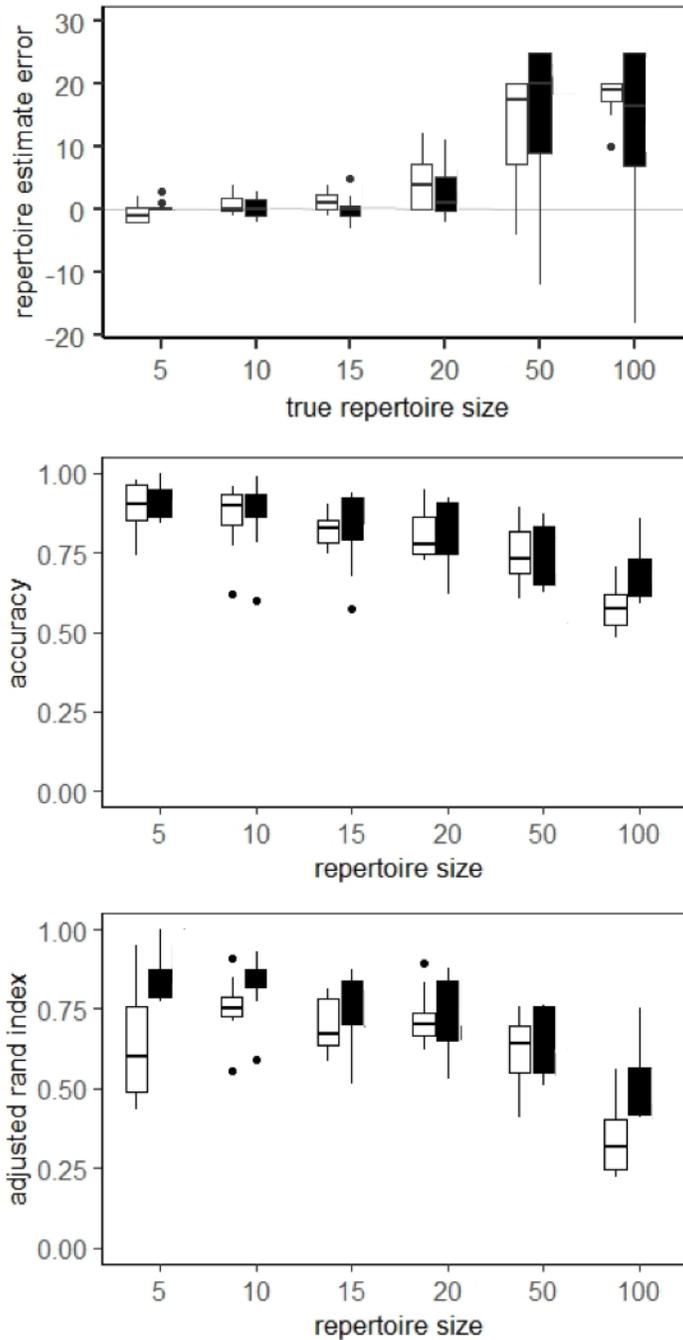
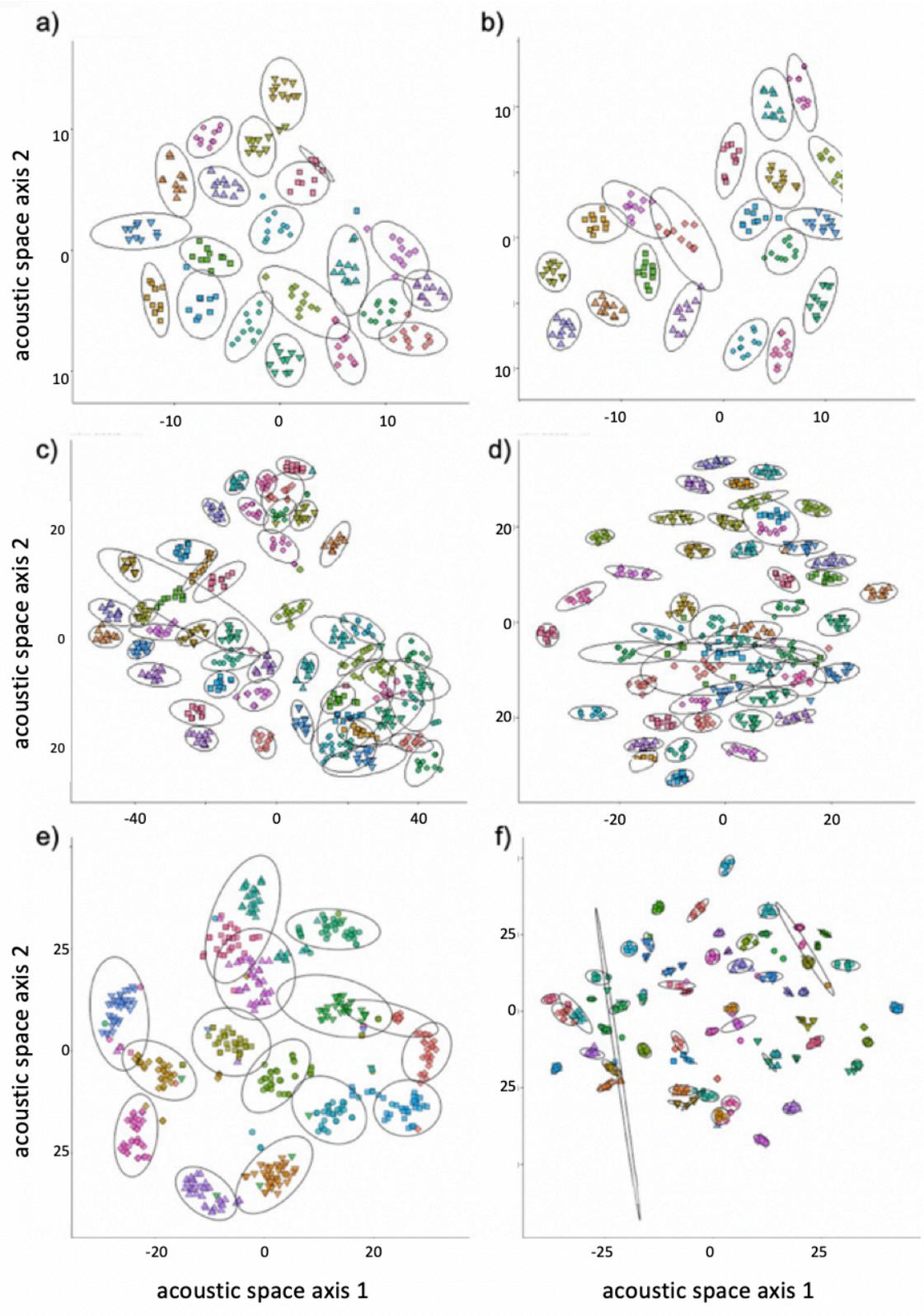


Figure 4. Unsupervised performance varies with number of unique elements in synthetic data sets. Plots of results from cluster analysis of unsupervised random forest output showing a) estimated repertoire size, b) classification accuracy, c) adjusted Rand index versus true repertoire size. White and black boxes represent results from synthetic budgerigar calls and synthetic long billed hermit songs, respectively.

The unsupervised analysis of live budgerigar calls using cluster analysis correctly estimated that there were 15 unique signal types in the data set. However, all calls from the same truth class were not always assigned to the same cluster (Figure 4c), which is reflected by the classification accuracy of 79.0 % and adjusted rand index of 0.602. The unsupervised analysis of field-recorded long-billed hermit songs incorrectly estimated 75 unique signal types in the data set, which was the maximum allowed number of clusters during our testing, rather than the true number of 50 unique signal types. However, the classification accuracy for this data set was 78.2 %, and the adjusted rand index was 0.776, indicating that signals of the same class were often clustered together. Scatter plots showing the unsupervised clustering of both live bird data sets are shown in Figure 5e,f.

Figure 5. Example scatter plots from unsupervised clustering of data sets. a) synthetic budgerigar data set with 20 unique elements, short duration, low harmonic content, and low background noise (clustered into 21 groups), b) synthetic long-billed hermit data set with 20 unique elements, short duration, high harmonic content, and low background noise (clustered into 20 groups), c) synthetic budgerigar data set with 50 unique elements, long duration, low harmonic content, and low background noise (clustered into 47 groups), d) synthetic long billed hermit data set with 50 unique elements, short duration, high harmonic content, and low background noise (clustered into 47 groups), e) lab data set of budgerigar calls with 15 unique elements (clustered into 15 groups), f) field data set of long billed hermit songs with 50 unique elements (clustered into 75 groups). We used t-sne dimensionality reduction technique to display all data points in two dimensions. Axes represent acoustic space, points represent single audio samples, and point colors and shapes represent samples of the same element type.



When acoustic area as used to estimate repertoire size, we observed a significant, positive correlation between acoustic area and the number of discrete elements. In addition, the acoustic area metric estimated repertoire size with similar accuracy across all values of true repertoire size (Figure 6). We observed this same pattern for synthetic data sets of long-billed hermit songs and budgerigar calls (Spearman correlation: budgerigars: $r = 0.91$, $N = 99$, $p < 0.0001$, long-billed hermits: $r = 0.95$, $N = 99$, $p < 0.0001$; Figure 6).

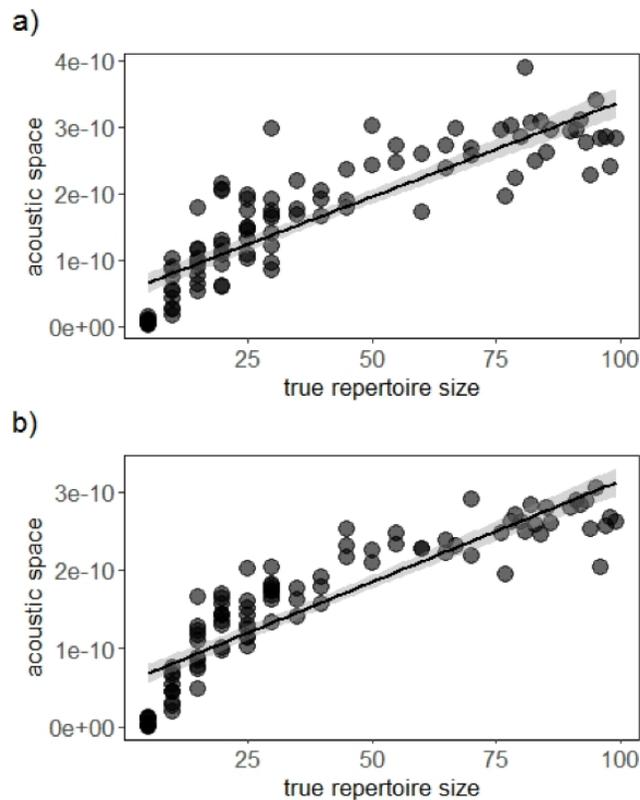


Figure 6. Datasets with more discrete elements have larger distributions in acoustic space. As repertoire size increases, the distribution of samples in acoustic space occupies a larger area for a) synthetic budgerigar calls, b) synthetic long-billed hermit songs. Acoustic space values have been squared to better illustrate differences between values on a small scale.

DISCUSSION

Our goal was to provide researchers with a flexible, unsupervised method for quantifying diversity in acoustic signals, a general problem encountered when evaluating the vocal repertoires of individuals, populations, or species. We aimed to replicate the process researchers might use when assessing variation in their own unlabeled data sets. We find that unsupervised learning paired with either cluster analyses or acoustic area calculations can approximate small and intermediate sample sizes well. In cases in which the number of discrete elements in a data set are large, however, quantifying the size of the area occupied in acoustic space may offer a more accurate alternative to estimating repertoire size than with cluster analyses. Below, we make specific recommendations about which signal characteristics might influence the accuracy of estimating acoustic diversity under different conditions, repertoire sizes, and acoustic features.

Supervised random forest performance. Supervised random forest analyses allowed us to verify that random forest analysis can accurately identify underlying patterns in acoustic data. We assessed the efficacy of this process and confirmed that our test data sets had the expected structure. Our results suggest that signal duration (short vs. long) and harmonic content (low vs. high) largely do not affect classification accuracy in most cases (Figure 3). Interestingly, synthetic long-billed hermit songs that have low harmonic content or high background noise suffered from higher out-of-bag error. Additionally, in almost all cases, synthetic long-billed hermit songs exhibited higher out-of-bag error rates than synthetic budgerigar songs. A likely explanation is that the harmonic content of natural long-billed hermit songs provides physical acoustic structure that aids in classification among element types, and low power content in

harmonic bands of our synthetic songs, or high background noise may mask this helpful feature. Harmonic structure is known to encode individual identity in some species' vocalizations (e.g., penguins, Aubin et al. 2000; humans, Imperl et al. 1997). The energy distribution of songs may be a salient feature that allows both conspecifics and automated approaches to better discern fine differences in signal structure. Therefore, harmonic structure is likely an important feature to capture in field recordings and feature measurements when it exists in natural vocalizations. As for the higher classification error for hermit elements in general, it is possible that the feature measurements we used might not be as effective at identifying the spectrotemporal variation for this species compared to budgerigars. Alternatively, focusing on frequency contours for representing variation in signal structure might miss other important features that help to distinguish between types, as the subtle variation in harmonic structure. Hence, it is likely that our simulation underestimated the overall discriminatory power of the methods. For both classes of synthetic data sets, we observed that error rates increased with true repertoire size, suggesting that the method is less effective at finding structure in data when there are large numbers of discrete elements. This decrease in discriminatory power with increasing repertoire size might be due to a saturation of the acoustic space

Unsupervised random forest performance. Cluster analysis using output from unsupervised random forest models showed that it was possible to estimate the true number of discrete elements in synthetic data sets with little error when the number of discrete elements was equal to or less than 20 (Figure 4a). For data sets that had 50 or 100 discrete elements, the unsupervised clustering technique often estimated repertoire size as being much higher than

its true value. One possible reason for this may be overfitting during clustering, i.e., when subsets of samples of the same signal type are assigned to separate clusters, which can occur when there is high similarity among a subset of samples in a class. Additionally, higher inaccuracy is expected as more unique elements are introduced when the acoustic space becomes saturated. Classification accuracy and adjusted rand indices were also higher for data sets with few discrete elements, and both metrics were consistently slightly higher for synthetic long-billed hermit data sets relative to synthetic budgerigar data sets (Figure 4b,d). This might be explained by the fact that the synthetic long-billed hermit exhibit more pronounced differences between classes than the synthetic budgerigar calls (Figure B1) which might allow for classes to be more easily distinguished.

Our second approach of quantifying acoustic diversity by calculating the size of the acoustic area occupied in acoustic space avoids the issue of needing to assign signals to discrete classes. For synthetic budgerigar and long-billed hermit data sets, acoustic area was positively correlated with the number of discrete elements in a data set (Figure 6). Additionally, unlike the clustering approach, acoustic area estimates were robust to large repertoire sizes. We suggest that this may be a useful technique for quantifying diversity in species anticipated to have large repertoires or high element diversity, as it precludes the need for defining discrete categories which may be difficult to define statistically in a crowded acoustic space. We note, however, that making relative comparisons between different data sets requires that all data points are analyzed concurrently; acoustic area has no value or meaning on its own. Recently, researchers have suggested that using latent acoustic space created by compressing data into fewer dimensions could be a powerful way to cluster

discrete vocal signals (Goffinet et al. 2019, Sainburg et al. 2019), but to our knowledge no previous studies have assessed signal diversity by evaluating acoustic space occupancy.

For the natural field and lab recorded data sets, we also observed limitations of the clustering method. Although cluster analysis accurately estimated small repertoire sizes with the synthetic data, for the lab-recorded budgerigar data set, which included only 15 unique element types, signals of the same class were sometimes placed in separate clusters. This could be one shortcoming of using clustering, as the algorithm may not assign the correct labels to every signal in a data set, although we observed that classification accuracy was rather high overall (79%). As with the synthetic data sets that had 50 unique elements, the unsupervised analysis overestimated the repertoire size of the field-recorded long-billed hermit data set of 50 elements, indicating that there were 75 unique elements present. Our results indicate that evaluating acoustic area is a more robust means of assessing acoustic diversity in such scenarios.

The feature measurements that were most useful in the unsupervised random forest approach varied among test data sets, presumably because different signal types were best distinguished by different features (Table B1). The ability for the analysis to detect this latent variation without requiring us to specify a priori which features we expected to vary exemplifies one of the primary strengths of random forest analysis. For this reason alone, we expect this approach may permit a high degree of adaptability to diverse acoustic data sets. Overall, given the relatively low out-of-bag error rates, we were confident that constructing random forest models in an unsupervised manner would be a useful tool for assessing acoustic diversity.

Potential Uses. Both methods we tested allowed for accurate estimates of repertoire size, however, we see promising attributes and limitations of both approaches. As we pointed out, the cluster analysis was particularly useful for assessing small or intermediate repertoire sizes. Interestingly, previous work has shown that parrot repertoires often contain 10-15 elements (Bradbury, *in press*) and that most songbird repertoires typically include below 20 elements or song types (MacDougall-Shackelton 1997, Byers and Kroodsma 2009, Snyder and Creanza 2018). Repertoires can refer both to total signal repertoire in a species (signal ethogram), and total number of signals of a certain type within an individual (song repertoire or call repertoire). Here we evaluated performance with individual vocal elements, however, our proposed approach can potentially be applied in both scenarios.

We suggest that both approaches can also be applied to address several ecological questions. Comparisons among species suggest acoustic diversity may correspond to a number of ecological characteristics, including viability of populations (Lailo et al. 2008), local habitat structure (Morton 1975, Boncoraglio and Saino 2007), as well as social system structure and complexity (Dunbar 1998, Freeburg 2006, elephants, Leighton 2017). Additionally, acoustic diversity within an individual, population, or species is also a key characteristic of animal vocal behavior that has been evaluated in terms of its role in social and sexual signaling (Tobias and Seddon 2009, Wilkins et al. 2013).

We envision that acoustic space is an especially promising method to estimate and compare acoustic diversity across individuals, populations, or species. This method is especially well-suited for large comparative analyses in which little might be known ahead of time about repertoire sizes for individual species and whether they surpass the limit appropriate for cluster analysis. In addition, all species or individuals can be compared in the

same acoustic space, allowing for comparable estimates of acoustic area for all species.

Lastly, automated procedures are especially beneficial for efficiency and reliability when comparing large numbers of species. Although the analyses presented here were conducted in a two-dimensional acoustic space, future analyses could calculate multi-dimensional acoustic volumes (as opposed to 2-D acoustic areas), although this is more computationally intensive.

Challenges and limitations. The broader challenge of assigning signals to categories is expected to scale in difficulty as the number of classes increase and the acoustic space becomes saturated. This inherent challenge cannot be entirely avoided, but certain aspects of our technique can help to mitigate this issue, namely by considering how individual vocalizations occupy acoustic space rather than estimating repertoire size. Acoustic space may not linearly correlate with the number of discrete elements in a data set, but we can use this approach to capture differences between large versus small repertoires. Additionally, assessing acoustic space may also allow researchers to avoid evaluating acoustic niche in a manner that is not meaningful for a study species. Lastly, when using this approach, we recommend that researchers take care in collecting high quality recordings, make sufficient sampling effort to capture the full repertoire to be analyzed, and select features that can adequately capture variation in their data.

Conclusions. We build upon previous work that has demonstrated the utility of unsupervised analyses for classifying acoustic signals and propose a novel combination of techniques for quantifying vocal diversity and/or measuring differences among individuals, species, and ecosystems. We propose that this method can be applied to estimate repertoire size and

calculating acoustic space occupancy, and both may be used to characterize vocalizations. By testing this method under diverse conditions and facilitating testing using synthetic data, we hope to offer researchers a robust and generalizable method for analyses of vocalizations.

ACKNOWLEDGEMENTS

We thank Holger Klinck, Chris Pelkie, the Cornell Center for Conservation Bioacoustics, and the Cornell Lab of Ornithology for essential support and technical advising while carrying out this project. This work was supported by funding from the Cornell Lab of Ornithology Athena Award, Cornell Sigma Xi research grants, and the Cornell Department for Neurobiology and Behavior.

WORKS CITED

- Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., and Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics*, 4: 206-214.
- Anikin, A. (2019). Soundgen: An open-source tool for synthesizing nonverbal vocalizations. *Behavior research methods*, 51: 778-792.
- Araya-Salas, M., and Smith-Vidaurre, G. (2017). warbleR: an R package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution*, 8: 184-191.
- Aubin, T., Jouventin, P., and Hildebrand, C. (2000). Penguins use the two-voice system to recognize each other. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267:1081-1087.
- Beecher, M. D. (1989). Signaling systems for individual recognition - An information-theory approach. *Animal Behaviour*, 38: 248-261.
- Boncoraglio, G., and Saino, N. (2007). Habitat structure and the evolution of bird song: a meta-analysis of the evidence for the acoustic adaptation hypothesis. *Functional Ecology*, 21: 134-142.
- Bormpoudakis, D., Sueur, J., and Pantis, J. D. (2013). Spatial heterogeneity of ambient sound at the habitat type level: ecological implications and applications. *Landscape Ecology*, 28: 495-506.
- Botero, C. A., Mudge, A. E., Koltz, A. M., Hochachka, W. M., and Vehrencamp, S. L. (2008). How reliable are the methods for estimating repertoire size? *Ethology*, 114: 1227-1238.
- Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X. Z., Raich, R., Hadley, S. J., Hadley, A. S., and Betts, M. G. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. *The Journal of the Acoustical Society of America*, 131: 4640-4650.
- Briggs, F., Huang, Y., Raich, R., Eftaxias, K., Lei, Z., Cukierski, W., and Irvine, J. (2013). The 9th annual MLSP competition: new methods for acoustic classification of multiple simultaneous bird species in a noisy environment. In *2013 IEEE international workshop on machine learning for signal processing (MLSP)*. IEEE.
- Catchpole, C. K., and Slater, P. J. (2003). *Bird song: biological themes and variations*. Cambridge university press.
- Clark, C. W., Marler, P., and Beeman, K. (1987). Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology*, 76: 101-115.

Costa, B., Taylor, J. C., Kracker, L., Battista, T., and Pittman, S. (2014). Mapping reef fish and the seascape: using acoustics and spatial modeling to guide coastal management. *PLoS One*, 9: e85555.

Dahlin, C. R., Young, A. M., Cordier, B., Mundry, R., and Wright, T. F. (2014). A test of multiple hypotheses for the function of call sharing in female budgerigars, *Melopsittacus undulatus*. *Behavioral ecology and sociobiology*, 68(1), 145-161.

Devoogd, T. J., Krebs, J. R., Healy, S. D., and Purvis, A. (1993). Relations between song repertoire size and the volume of brain nuclei related to song: comparative evolutionary analyses amongst oscine birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 254(1340), 75-82.

Farabaugh, S. M., Linzenbold, A., and Dooling, R. J. (1994). Vocal plasticity in Budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108(1), 81.

Freeberg, T. M., and Lucas, J. R. (2012). Information theoretical approaches to chick-a-dee calls of Carolina chickadees (*Poecile carolinensis*). *Journal of Comparative Psychology*, 126: 68.

Gasc, A., Sueur, J., Jiguet, F., Devictor, V., Grandcolas, P., Burrow, C., and Pavoine, S. (2013). Assessing biodiversity with sound: Do acoustic diversity indices reflect phylogenetic and functional diversities of bird communities? *Ecological Indicators*, 25, 279-287.

Garamszegi, L. Z., Boulinier, T., Moller, A. P., Torok, J., Michl, G. and Nichols, J. D. (2002). The estimation of size and change in composition of avian song repertoires. *Animal Behavior*, 63, 623-630.

Garamszegi, L. Z., Balsby, T. J. S., Bell, B. D., Borowiec, M., Byers, B. E., Draganoiu, T., Eens, M., Forstmeier, W., Galeotti, P., Gil, D., Gorissen, L., Hansen, P., Lampe, H. M., Leitner, S., Lontkowski, J., Nagle, L., Nemeth, E., Pinxten, R., Rossi, J. M., Saino, N., Tanvez, A., Titus, R., Torok, J., Van Duyse, E. and Muller, A. P. (2005). Estimating the complexity of bird song by using capture-recapture approaches from community ecology. *Behavioral ecology and sociobiology*, 57, 305—317.

Gerhardt, H. C., and Huber, F. (2002). *Acoustic communication in insects and anurans: common problems and diverse solutions*. University of Chicago Press.

Gibb, R., Browning, E., Glover-Kapfer, P., and Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10: 169-185

Goffinet, J., Mooney, R., and Pearson, J. (2019). Inferring low-dimensional latent descriptions of animal 740 vocalizations. *bioRxiv*, 811661.

- Harris, S. A., Shears, N. T., and Radford, C. A. (2016). Ecoacoustic indices as proxies for biodiversity on temperate reefs. *Methods in Ecology and Evolution*, 7: 713-724.
- Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., and Slaney, M. (2017, March). CNN architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE.
- Imperl, B., Kačič, Z., and Horvat, B. (1997). A study of harmonic features for the speaker recognition. *Speech communication*, 22: 385-402.
- Janik, V. M. (2009). Acoustic communication in delphinids. *Advances in the Study of Behavior*, 40: 123-157.
- Kaufman, L., and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley and Sons.
- Keen, S., Ross, J. C., Griffiths, E. T., Lanzone, M., and Farnsworth, A. (2014). A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). *Ecological Informatics*, 21: 25-33.
- Kershenbaum, A., Freeberg, T. M., and Gammon, D. E. (2015). Estimating vocal repertoire size is like collecting coupons: a theoretical framework with heterogeneity in signal abundance. *Journal of theoretical biology*, 373: 1-11.
- Kroodsma, D. E., Miller, E. H., and Ouellet, H. (Eds.). (1982). *Acoustic Communication in Birds: Song learning and its consequences* (Vol. 2). Academic press.
- Kroodsma, D. E., and Miller, E. H. (Eds.). (1996). *Ecology and evolution of acoustic communication in birds*. Comstock Publishing.
- Laiolo, P., Vögeli, M., Serrano, D., and Tella, J. L. (2008). Song diversity predicts the viability of fragmented bird populations. *PLoS One*, 3: e1822.
- Langbauer, W. R., Payne, K. B., Charif, R. A., Rapaport, L., and Osborn, F. (1991). African elephants respond to distant playbacks of low-frequency conspecific calls. *Journal of Experimental Biology*, 157: 35-46.
- Leighton, G. M. (2017). Cooperative breeding influences the number and type of vocalizations in avian lineages. *Proceedings of the Royal Society B: Biological Sciences*, 284: 20171508.
- Linhart, P., Osiejuk, T. S., Budka, M., Šálek, M., Špinka, M., Policht, R., and Blumstein, D. T. (2019). Measuring individual identity information in animal signals: Overview and performance of available identity metrics. *Methods in Ecology and Evolution*, 10: 1558-1570.

- Lyon, R. H., and Ordubadi, A. (1982). Use of cepstra in acoustical signal analysis. *Journal of Mechanical Design*, 104: 303-306.
- Maaten, L. V. D., and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9: 2579-2605.
- Marler, P. R. (1982). Avian and primate communication: The problem of natural categories. *Neuroscience and Biobehavioral Reviews*, 6: 87-94.
- Mason, N. A., Burns, K. J., Tobias, J. A., Claramunt, S., Seddon, N., and Derryberry, E. P. (2017). Song evolution, speciation, and vocal learning in passerine birds. *Evolution*, 71: 786-796.
- Michie, D., Spiegelhalter, D. J., and Taylor, C. C. (1994). Machine learning. *Neural and Statistical Classification*, 13.
- Mullen, S. P., Mendelson, T. C., Schal, C., and Shaw, K. L. (2007). Rapid evolution of cuticular hydrocarbons in a species radiation of acoustically diverse Hawaiian crickets (Gryllidae: Trigonidiinae: Laupala). *Evolution*, 61: 223-231.
- Owren, M. J., Seyfarth, R. M., and Hopp, S. L. (1992). Categorical vocal signaling in nonhuman primates. *Studies in emotion and social interaction. Nonverbal vocal communication: Comparative and developmental approaches*, 102-122
- Payne, R. B. (1986). Bird songs and avian systematics. In *Current ornithology*. Springer, Boston, MA.
- Peshek, K. R., and Blumstein, D. T. (2011). Can rarefaction be used to estimate song repertoire size in birds? *Current Zoology*, 57: 300-306.
- Pieretti, N., Farina, A., and Morri, D. (2011). A new methodology to infer the singing activity of an avian community: The Acoustic Complexity Index (ACI). *Ecological Indicators*, 11: 868-873.
- Pijanowski, B. C., Villanueva-Rivera, L. J., Dumyahn, S. L., Farina, A., Krause, B. L., Napoletano, B. M., and Pieretti, N. (2011). Soundscape ecology: the science of sound in the landscape. *BioScience*, 61: 203-216.
- Podos, J., Lahti, D. C., and Moseley, D. L. (2009). Vocal performance and sensorimotor learning in songbirds. *Advances in the Study of Behavior*, 40: 159-195.
- Rand, W.M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66: 846-850.
- Ryan, M. J. (1985). *The tungara frog: a study in sexual selection and communication*. University of Chicago Press.

Sainburg, T., Thielk, M., and Gentner, T. Q. (2019). Latent space visualization, characterization, and generation of diverse vocal communication signals. *bioRxiv*, 870311.

Salamon, J., Jacoby, C., and Bello, J. P. (2014). A data set and taxonomy for urban sound research. In *Proceedings of the 22nd ACM international conference on Multimedia*, 1041-1044.

Sewall, K. B., Soha, J. A., Peters, S., and Nowicki, S. (2013). Potential trade-off between vocal ornamentation and spatial ability in a songbird. *Biology Letters*, 9: 20130344.

Seyfarth, R. M., and Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual review of psychology*, 54, 145-173.

Smith-Vidaurre, G., Araya-Salas, M., and Wright, T. F. (2019). Individual signatures outweigh social group identity in contact calls of a communally nesting parrot. *Behavioral Ecology*, 31: 448-458.

Stowell, D., and Plumbley, M. D. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ*, 2: e488.

Stowell, D., Wood, M. D., Pamuła, H., Stylianou, Y., and Glotin, H. (2019). Automatic acoustic detection of birds through deep learning: the first Bird Audio Detection challenge. *Methods in Ecology and Evolution*, 10: 368-380.

Sueur, J., Pavoine, S., Hamerlynck, O., Duvail, S., (2008a). Rapid acoustic survey for biodiversity appraisal. *PLoS One*, 3: e4065.

Sueur, J., Aubin, T., and Simonis, C. (2008b). Seewave, a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18: 213-226.

Sueur, J., Farina, A., Gasc, A., Pieretti, N., and Pavoine, S. (2014). Acoustic indices for biodiversity assessment and landscape investigation. *Acta Acustica united with Acustica*, 100: 772-781.

Sullivan-Beckers, L. and Cocroft, R.B. (2010) The importance of female choice, male-male competition, and signal transmission as causes of selection on male mating signals. *Evolution*, 64: 3158–3171

Tobias, J.A. and Seddon, N. (2009) Signal design and perception in *Hypocnemis* antbirds: evidence for convergent evolution via social selection. *Evolution*, 63: 3168–3189

Ulloa, J. S., Aubin, T., Llusia, D., Bouveyron, C., and Sueur, J. (2018). Estimating animal acoustic diversity in tropical environments using unsupervised multiresolution analysis. *Ecological Indicators*, 90: 346-355.

Valletta, J. J., Torney, C., Kings, M., Thornton, A., and Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124: 203-220.

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., and Fischer, J. (2015). Characterizing vocal repertoires—Hard vs. soft classification approaches. *PloS one*, 10: e0125785.

Wildenthal, J. L. 1965: Structure in primary song of the mockingbird (*Mimus polyglottos*). *The Auk* 82: 161-189.

Wilkins, M. R., Seddon, N., and Safran, R. J. (2013). Evolutionary divergence in acoustic signals: causes and consequences. *Trends in ecology and evolution*, 28: 156-166.

Wolberg, G. (1990). *Digital image warping*, Vol. 10662. Los Alamitos, CA: IEEE computer society press.

Wrege, P. H., Rowland, E. D., Keen, S., and Shiu, Y. (2017). Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods in Ecology and Evolution*, 8: 1292-1301.

CHAPTER 3

PATTERNS OF VOCAL CONVERGENCE ARE SHAPED BY OPPOSING FORCES: EVIDENCE FROM WILD SONGBIRDS

Sara Keen^{1,2}

¹ Department of Neurobiology and Behavior, Cornell University, Ithaca, NY

² Cornell Lab of Ornithology, 159 Sapsucker Woods Rd, Ithaca, NY

ABSTRACT

Vocal convergence is a widespread phenomenon that occurs in many species. To date, a universal set of factors or conditions that lead to convergence has not been identified. A central challenge in predicting when vocal convergence is expected is that both ecological and social environment can influence the benefits an individual may gain from convergence. Here, I develop and test a series of models that incorporate ecological and social context in order to predict whether signal convergence or divergence is expected in a given system, as well as the optimal level of convergence or divergence an individual will exhibit. I also consider the specific case of movement between populations in order to predict the optimal amount of convergence effort for immigrants and residents. I test these predictions using empirical data collected from a wild population of great tits, *Parus major*, a species known to exhibit song sharing among neighbors. The convergence model predicts that that there will be partial, but

not full, vocal convergence among neighbors, and that immigrants, not residents, should make effort to converge with neighbors' songs. This is premised upon the hypothesis that, immigrants, who initially exhibit low vocal similarity with residents, can significantly increase their fitness by converging with residents, whereas residents can only marginally increase their fitness by converging with a newly arrived immigrant. We found that levels of acoustic similarity were higher among neighbors sharing territory boundaries, but that neighbors did not exhibit complete convergence, supporting model predictions. Using playbacks of non-local songs to simulate the arrival of immigrants, we found that residents did not exert convergence effort, as predicted by the model. We consider how our model might help to explain patterns of signal convergence in many species, and suggest that vocal convergence may be a dynamic, context-dependent trait which can be shaped by an individual's current environment.

INTRODUCTION

In many animal societies, cooperative and competitive interactions are mediated by an individual's ability to signal group membership or identity (Sherman et al. 1997, Tibbetts and Dale 2007, Bradbury and Verhencamp 2011). In many taxa, vocal signals encode information about population or group membership, which is often conveyed by highly similar signal structure among group members (e.g., parrots, Wright 1996; songbirds, Catchpole and Slater 2008; hummingbirds, Gaunt et al. 1994; cetaceans, Deecke et al. 2000, Garland et al. 2011; bats, Boughman 1997). The phenomenon of nearby conspecifics converging upon similar vocalizations has received considerable attention from researchers (Wright and Dahlin 2018), and several hypotheses have been proposed to explain the emergence and maintenance of

such geographical variation (Payne 1982, Baker and Cunningham 1985, Podos and Warren 2007). Although vocal convergence is widespread, the observed degree of convergence and the spatial scale on which it occurs varies among species, and no single suite of ecological or social factors have been found to explain its presence or absence (Podos and Warren 2007, Wright and Dahlin 2018).

Bird song is a useful trait in which to explore vocal convergence as songs and calls often exhibit within-species geographic variation (Catchpole and Slater 2008). Dialects, in which nearby conspecifics converge upon acoustically similar signals that vary widely among locations, have been observed in dozens of species since first documented in white-crowned sparrows by Marler and Tamura (1964), and have prompted a number of researchers to posit explanations for the evolution of vocal convergence (Baker and Cunningham 1985, reviewed by Podos and Warren 2007 and Wright and Dahlin 2018). These hypotheses aim at explaining why vocal convergence is favored by selection, and fall into three, non-mutually exclusive classes. The first class can broadly be described as the social benefits hypothesis, which suggests that individuals benefit from using vocalizations similar to those of nearby birds because the social costs of using nonlocal signals are high. This encompasses the colony password hypothesis (Feekees 1977), which suggests that dialects signal social group membership and serve to identify nonlocal intruders, as well as the deceptive mimicry hypothesis (Payne 1982), which proposes that less dominant or nonlocal individuals adopt local songs used by dominant birds in order to gain acceptance and deter competitors. The social benefits hypothesis suggests that, in addition to helping individuals gain group acceptance, using local signal types can also facilitate cooperation, group cohesion, or coordinated behavior with particular individuals (Wilkinson and Boughman 1998,

Vehrencamp et al. 2003, Janik and Slater 1998). The second class of hypotheses suggests that convergence could be shaped by sexual selection acting through mate choice. In this scenario, potential mates would exhibit a preference for vocalizations used in their population, as this could signal that individuals were born nearby and may therefore be better adapted to the local environment (i.e., coadapted gene complex; Baker 1982). Therefore, using local signal types could offer a selective advantage through enhanced mate attraction, as has been shown in several species in which females prefer local signal types (e.g., Searcy et al. 2002, Lachlan et al. 2014). The third class of hypotheses suggest that vocal convergence may arise because particular signals are better adapted to the local environment (i.e., the acoustic adaptation hypothesis, Morton 1975). In this case, local signal types confer a selective advantage through more effective transmission to receivers. Levels of transmission loss are known to vary with habitat characteristics (Bradbury and Vehrencamp 2011), and several studies of bird song in urban areas demonstrate that signal structure is under selection to best fit the environment (e.g., Slabberkoorn and Peet 2003, Luther and Derryberry 2012). A final possibility that could serve as a null hypothesis is that geographical variation arises through cultural drift (i.e., the epiphenomenon hypothesis, Wiens 1982). For example, copying errors can give rise to distinct song variants in different populations, and a particular variant may be adopted by a population not because it offers some selective advantage, but because it is the model to which many juveniles are exposed. In this case, convergence offers no selective advantage and is analogous to random genetic drift with local fixation.

The many hypotheses proposed to explain vocal convergence reflect not only the sustained interest in this phenomenon, but also the large amount of variation in patterns of vocal convergence observed among species. For example, vocal convergence can take the

form of sharing a single call type as well as partial or full overlap of song repertoires, and the timescale on which vocal convergence persists can vary between weeks to decades (Kroodsma 2004). Furthermore, although many of the hypotheses above specify the mechanisms by which convergence would be expected to arise, the exact predictions made by each will depend on species characteristics, such as dispersal distance, repertoire size, and spatial distribution of breeding territories (Baker and Cunningham 1985, Podos and Warren 2007). Many past studies offer support for the social benefits hypothesis (Wright and Dahlin 2018), though at present there appears to be no single hypothesis that provides a universal explanation for vocal convergence. Although this broad phenomenon may not be described by a single theory, researchers agree that vocal convergence is shaped by a complex suite of social and ecological factors. Here, we suggest that by considering these factors as parameters in an optimality model, we can estimate the expected amount of vocal convergence within a given population.

In order to make predictions about the conditions under which vocal convergence is expected, it is important to differentiate vocal convergence from counter singing. The former is a population-level process that persists over a period of time on the order of weeks or longer. In contrast, counter singing, or vocal matching, is an interactive display that often occurs between two individuals, typically during an agonistic encounter such as the negotiation or defense of territory boundaries (McGregor et al. 1992, Vehrencamp 2001, King and McGregor 2016). Like vocal convergence, counter singing entails matching the signals of others, and indeed the need to counter sing with competitors may be among the suite of factors that gives rise to vocal convergence or song sharing (Macdougall-Shackelton 1997, Nelson 2000). However, because vocal matching occurs in the context of aggressive

interactions, whereas vocal convergence is the resulting pattern observed across a group or population, these processes are functionally dissimilar and therefore yield different predictions. In other words, vocal convergence is an ontogenetic process, and counter singing serves an adaptive function in the moment.

Central to understanding the processes that shape vocal convergence is determining the fitness advantage an individual stands to gain by resembling nearby conspecifics. In addition to the hypotheses above, which suggest that individuals may increase their fitness via vocal convergence, we must also consider scenarios in which individuals benefit from being dissimilar to nearby conspecifics. Dissimilar signals are known to be advantageous in social systems in which individual recognition is important for avoiding aggression that is intended for others (Dale et al. 2001, Sheehan and Tibbetts 2009) and has been suggested to enhance offspring recognition (Medvin et al. 1993). Signal divergence might also be favored by sexual selection, for example when signaling individual identity offers a selective advantage in mate attraction (Thom et al. 2012). In this case, individuals are subject to some penalty if they do not sufficiently differentiate themselves from other members of the population. Additionally, sexual selection might lead to divergence when signals serve as honest indicators of quality, leading to divergent signals among signalers of varying quality. Examples of this may include cases in which using more elaborate songs or larger repertoires increases access to mating opportunities (e.g., Read and Weary 2002, Snell-Rood and Badyaev 2008).

To evaluate the costs and benefits of using vocal signals that are similar or dissimilar to those of nearby conspecifics under different conditions, we first consider two models that represent alternative scenarios: selection for signal convergence and selection for signal divergence. Here, the term divergence refers to signal dissimilarity, and thus selection for

divergence would result in greater overall diversity within a population. For each case, we estimate the amount of effort that a focal individual should invest in converging with or diverging from nearby individuals in the same group or population (hereafter neighbors) in order to maximize their fitness, and list the specific hypotheses generated from model in Table 1. We also present a combined model, in which there is opposing selection for both convergence and divergence, that can be used to predict which scenario (convergence or divergence) is expected given the characteristics of a particular system. Regardless of whether selection favors convergence or divergence, the predicted degree of signal similarity is expected to be mediated by social and ecological factors, i.e., to be context-specific. Context-specific tuning of signal similarity is a new way in which social recognition systems can be seen as flexible, just as receiver acceptance thresholds for recognition signals have been shown vary widely among interaction contexts (Reeve 1989, Johnstone 1997, Sheehan and Reeve 2020). The models below aim to describe the expected degree of vocal convergence or divergence in each scenario given the particular social and ecological context in which signaling occurs.

Table 1. Summary of model hypotheses and predictions for our study system.

Model	Hypotheses	Predictions
Convergence	Individual fitness costs are incurred when a focal individual is very different from neighbors, and therefore non-local individuals newly arrived in a population will bear the costs of vocal convergence.	<ol style="list-style-type: none"> 1. Males breeding nearby one another converge upon similar vocal signals. 2. Immigrants will change their signals to converge with those of neighbors. 3. Local birds will exert no effort to adopt nonlocal playback songs into their repertoires.
Divergence	Individuals increase their fitness by distinguishing themselves from neighbors. Non-local individuals that use very dissimilar signal incur no fitness costs for not matching the local population.	<ol style="list-style-type: none"> 1. Males breeding nearby one another will not have higher vocal similarity than other birds within the population. 2. Immigrants will exhibit less vocal similarity with neighbors than resident males

Table 1 (Continued)

Model	Hypotheses	Predictions
Combined	Individuals experience opposing selective forces.	The unique system characteristics will determine if predictions match those of the convergence or divergence model. This is decided by relative values of the convergence and divergence parameters, <i>cc</i> and <i>cd</i> , respectively.

I test the predictions from each model using a dataset of songs collected from a wild population great tits (*Parus major*). Because males of this species disperse from natal territories to establish breeding territories during their first year (Krebs 1981), it was possible to assess whether levels of pairwise acoustic similarity between males correlates with geographic distance between their natal nests and/or breeding territories, which may give insight into whether there is selection for vocal convergence when individuals establish territories. I also evaluated whether birds were more acoustically similar to neighbors than non-neighbors, and whether levels of acoustically similarity varied among resident birds born locally and immigrant birds born outside of the study system. Lastly, I used playbacks to test whether simulating the arrival of a nonlocal male on a neighboring territory prompted birds to incorporate nonlocal songs into their repertoires.

My results below support the predictions of the convergence model: neighboring males used more similar songs than expected by chance, and their songs are more similar to songs of current neighbors than to songs of natal neighbors. I also observed that immigrant birds exhibited the same levels of vocal convergence as residents, and that birds did not adopt nonlocal songs used in playbacks. Together, my results suggest that individuals benefit from matching neighbors' signals and that immigrants show higher likelihood of investment in vocal convergence than residents. These findings also suggest that the observed levels of vocal convergence may reflect the tradeoff between benefits of convergence and the costs of acquiring new signals. The models we present here may help to explain patterns of vocal convergence found in different systems and to unify observations under a general framework.

METHODS

Models. I propose three models describing the expected level of vocal similarity between a focal individual and their neighbors under selection for convergence, divergence, or opposing selection for both outcomes. For each case, I predict the optimal amount of convergence or divergence effort an individual should exert given possible values of the parameters listed in Table 2.

Table 2. Model parameters and variables.

Parameter or variable	Definition	Explanation and examples
ω	Focal individual's fitness	The fitness of the focal individual, which can increase when the individual exerts effort for convergence or divergence under the appropriate conditions. Examples of fitness benefits due to signal convergence or divergence include increased access to mating opportunities and acceptance into a social group that enables increased access to resources.
d	Acoustic distance	The initial distance between the signal of the focal individual and the signals of all neighbors. The manner in which distance is calculated depends on the vocal characteristics of the system in question. For example, among birds that have repertoires of multiple song types, d might represent the amount of song sharing between individuals.
x	Change in acoustic distance	The amount by which d increases or decreases due to effort for convergence or divergence made by the focal individual. For example, x could reflect the reduction in acoustic distance resulting from an immigrant bird acquiring a local song type.

Table 2 (Continued)

Parameter or variable	Definition	Explanation and examples
<i>a</i>	Cost of convergence or divergence effort	This parameter controls the degree to which a given convergence or divergence effort is costly. This includes energetic costs (e.g., the physiological costs of signal production), and learning costs (e.g., time required to acquire new songs).
<i>cc , cd</i>	Sensitivity of social benefits to convergence or divergence effort	This term defines how quickly social benefits increase as the focal individual's signal becomes increasingly similar to or dissimilar from neighbors' signals. A social benefit can result from a reduction in some kind of social cost, e.g., increased viability as the chance of misdirected aggression received from a neighbor decreases.

The convergence model (1) describes the relationship between the focal individual's fitness, ω , and the reduction in distance between their signal and neighbors' signals due to positive convergence effort, x , when there is selection for signal convergence.

$$\omega = e^{-cc*(d-x)^2} * e^{-ax} \quad (1)$$

The first exponential term on the right side of (1) describes the social benefit of convergence, e.g., the 1-probability of the focal individual receiving injurious aggression from its neighbor as a function of the focal individual's effort in reducing their acoustic distance. The initial acoustic distance between the focal individual's signal and collective neighbors' signals is represented by d , and the term $(d - x)^2$ is this distance after some conformity effort, x , is made by the focal individual. Thus, the form of (1) encodes the assumption that the social benefit of convergence is maximized when $x = d$, i.e. the acoustic distance is reduced to zero. The second exponential term represents the focal individual's loss in survival or fecundity from expending effort x (time or energy) in reducing the acoustic difference. Thus, the fitness-maximizing value of the effort x must be one that optimally balances the social benefit of convergence and the direct cost of convergence effort.

As the cost of conformity effort, a , increases, the maximum possible fitness decreases, as is expected in cases where the acquisition of new signals is costly, e.g., when reducing acoustic distance requires precisely adjusting fine structure of signals, and therefore a large time investment in vocal learning. Additionally, as the sensitivity to mismatch, cc , increases, the amount of signal convergence necessary to achieve the maximum possible fitness increases, as would be expected when receivers are more discerning of differences among signals. The peak fitness of the focal individual occurs when $\frac{d\omega}{dx} = 0$ and ω is a local

maximum. We solve for x in $\frac{d\omega}{dx}$ (2) to find the optimal reduction in acoustic distance, x^* , shown in (3).

$$\frac{d\omega}{dx} = -e^{-cc*(d-x)^2-ax} * (a + 2cc(-d + x)) \quad (2)$$

$$x^* = \frac{-a}{2cc} + d \quad (3)$$

We can also use the solution for x in (3) to verify that the second derivative is negative, confirming that the solution above is the fitness maximum. Thus, the optimal convergence effort increases as the initial acoustic distance d increases, the effort cost rate, a , decreases, and as the sensitivity of social benefits to acoustic distance, cc , increases. An immediate consequence is that the focal individual should try harder to converge the greater the initial acoustic distance and the more potent the social benefit of convergence (Fig. 1a,b).

The divergence model (4) describes the change in a focal individual's fitness, ω , as a function of divergence effort, x .

$$\omega = e^{-cd/(d+x)^2} * e^{-ax} \quad (4)$$

The first exponential term in (4) describes the social benefit of divergence, e.g., increased mating opportunities or reduced aggression intended for others as a focal individual becomes increasingly distinct from neighbors. The term $(d + x)^2$ is the signal distance from collective neighbors' signals after some divergence effort, x , to increase acoustic distance beyond the starting value of d . The value of the first exponential term increases as x increases, which reflects the model assumption that social benefits increase as the focal individual increases

divergence effort. Consequently, the model predicts that fitness increases as divergence effort increases, i.e., when acoustic distance between the focal individual and neighbors grows larger.

As in the previous model, the second exponential term in (4) represents the costs of divergence effort, such as time or energy invested in acquiring or producing divergent signals. The focal individual's fitness will be maximized when the tradeoffs between costs of divergence effort and social benefits of divergence are optimized. To find this fitness maximum, we calculate $\frac{d\omega}{dx}$ in (5), and solve for x^* in (6).

$$\frac{d\omega}{dx} = -e^{\frac{-cd}{(d+x)^2} - ax} * \left(-a + \frac{2cd}{(d+x)^3}\right) \quad (5)$$

$$x^* = \frac{2^{1/3} cd^{1/3}}{a^{1/3}} - d \quad (6)$$

The solution for the fitness maximum in (6) shows that the optimal divergence effort decreases as the initial signal distance, d , increases and as the cost of divergence, a , increases. Additionally, when the effort cost, a , is small relative to the sensitivity, cd , the focal individual will exert more divergence effort as the sensitivity increases, as would be expected when receivers are increasingly better at distinguishing between similar signals and the costs of divergence effort are negligible. Therefore, this model predicts that individuals will exert more divergence effort when initial acoustic distance is smaller and the ratio cd to a is higher (Fig. 1c,d).

The combined model (7) describes individual fitness when there is opposing selection for both convergence and divergence. We assume that there are two multiplicative social benefits, one due to convergence benefits and one due to divergence benefits, and ignore the direct cost term for simplicity.

$$\omega = e^{-cc*(d-x)^2} * e^{\frac{-cd}{(d-x)^2}} \quad (7)$$

Here, the term x is the change in signal distance resulting from the focal individual's effort, which may result in being either more similar to or more different from neighbors (i.e., x in this model is allowed to be either positive or negative, with a positive value indicating convergence and a negative value indicating divergence). This model describes a scenario in which an individual experiences two multiplicative social benefits, one maximized by signal convergence and the other by signal divergence. Thus, the relative values of cd and cc influence the focal individual's optimal strategy. For example, when cc is large and cd approaches zero, the second exponential term in (7) can be ignored and the focal individual's fitness is maximized when x approaches d , i.e., when there is effort for convergence.

We seek to find the conditions under which the product of the social benefits should lead to positive x (convergence) or negative x (divergence). We first find the derivative of (7) with respect to x , yielding

$$\frac{d\omega}{dx} = e^{-cc*(d-x)^2 - cd/(d-x)^2} * \left(\frac{-2cd}{(d-x)^3} + 2 * cc * (d - x) \right) \quad (8)$$

We then evaluate the sign of this derivative for $x = 0$, which tells us whether selection at an acoustic difference d should cause subsequent increases in x (convergence) or decreases in x (divergence). The result is that convergence should be favored over divergence as long as

$$cc > \frac{cd}{d^4} \quad (9)$$

Thus, convergence should occur when cc is large relative to cd , and the initial acoustic distance, d , is small.

Suppose that there is stronger selection for convergence than for divergence (i.e., $cc \gg cd$) so that the pure convergence model described above applies. Suppose further that an immigrant enters into and acquires a territory in a stable population. If $cc \gg cd$, it follows that all individuals in the local population, hereafter residents, have previously attained some optimal amount of convergence with one another. For the case where the residents have negligible signal distance with each other, let the immigrant differ from each of the residents by an initial distance, d . In this case, the immigrant's fitness as a function of its own convergence effort x is given by:

$$\omega_{imm} = (e^{-cc*(d-x)^2})^n * e^{-ax} \quad (10)$$

Here, n is the number of neighbors (i.e., residents) surrounding the immigrant and the fitness expressions that it now must survive costly interactions with each of the n residents. Using the approaches above, the immigrant's optimal convergence effort is equal to

$$x^* = d - \frac{a}{2ccn} \quad (11)$$

Therefore, an immigrant should exert more effort to converge the local signal type (i.e., x must more closely approach d) when there are more residents (Fig. 2a).

We can also describe the fitness of a resident in relation to their change in signal distance after a single immigrant arrives as a function of its own effort x to converge with the immigrant:

$$\omega_{res} = e^{-cc*(d-x)^2} * (e^{-cc*(x)^2})^{n-1} * e^{-ax} \quad (12)$$

In this case, a resident whose signal approaches an immigrant's signal by a distance x gains convergence benefits with the immigrant, but lowers its convergence benefits from the other $n-1$ residents because it has increased its signal distance by x with the signals of the other residents. Using the approaches above, the resident's optimal convergence effort is equal to:

$$x^* = \frac{d}{n} - \frac{a}{2ccn} \quad (13)$$

A comparison of the immigrants and resident's optimal convergence efforts in (11) and (13), respectively, reveals that the optimal convergence effort for the resident is lower than that of immigrant, particularly when n is high. In fact, for high enough n , the resident will not be favored to exert any convergence effort, as the arrival of a single immigrant into a population of many residents does not significantly alter the signal space in which residents exist. These differences in optimal convergence effort of immigrants and residents are illustrated in Fig. 2.

An assumption in all models is that selection acts upon the amount of signal similarity between the focal individual and neighbors. Therefore, it is expected that there will be variation in amount of vocal convergence focal individuals exhibit, but does not necessitate that individuals can actively adjust the amount of effort they invest in convergence or mismatching. For example, when considering vocal signals used by territorial songbirds, the mechanism underlying convergence effort may be song acquisition via post-dispersal social learning, intentional settlement near territory holders using similar songs, or it could be that individuals exhibit natal philopatry and therefore can maximize convergence by remaining near their natal nest. In all models, fitness costs result from a combination of ecological and social factors, and thus representing these using a single term allows our models to be broadly applicable.

I created plots of model predictions using MATLAB 2015a (The Mathworks Inc., Natick, MA, USA).

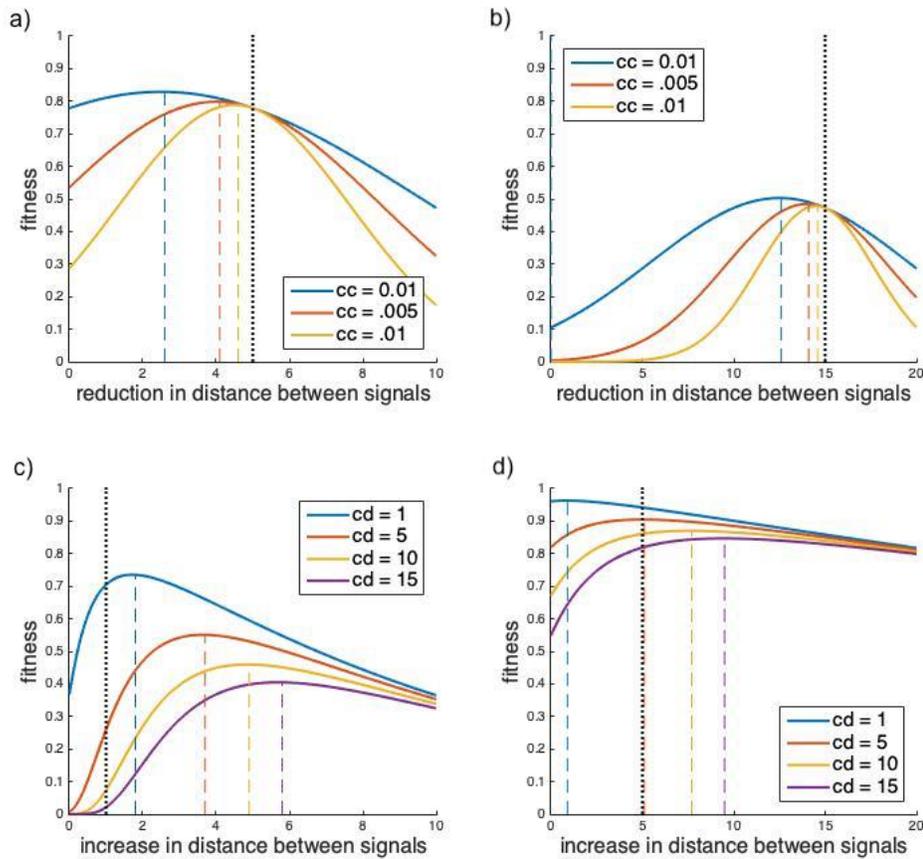


Figure 1. Convergence and divergence model predictions. The fitness of focal individuals (y-axis) corresponds to the change in signal distance from neighbors due to effort from the focal individual, x (x-axis). Under selection for convergence, as the sensitivity to mismatch, cc , increases, higher levels of convergence effort, x , are required to achieve maximum fitness, and higher initial signal distance, d , reduces fitness maxima, as shown in a) $d = 5$, and b) $d = 15$. Under selection for divergence, higher sensitivity to mismatch, cd , requires higher levels of divergence effort, x , to reach maximum possible fitness, and higher initial signal distance, d , increases fitness maxima, as shown in c) $d = 1$, and d) $d = 5$. In all plots, the initial distance, d , between the focal individual's signal and neighbors' signals is shown by the dotted black line, and fitness maxima are indicated with dashed lines.

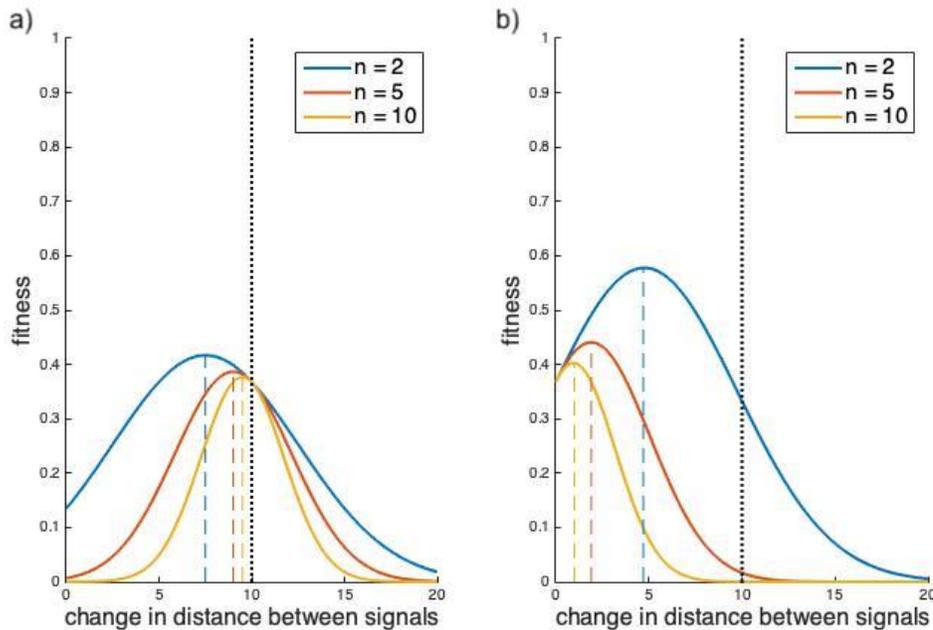


Figure 2. Convergence model predictions for immigrants and residents. As the number of neighbors increases, a) immigrants will exert more effort to converge with residents' signals (shown with $a = 0.1$), and b) residents will exert less effort to match immigrants (shown with $a = 0.01$). The initial distance, d , between the focal individual's signal and neighbors' signals is shown by the dotted black line, and fitness maxima are indicated with dashed lines.

Study system. I collected songs from a wild population of great tits in Wytham Woods, Oxfordshire, UK (51460 N, 01200 W). This population is part of a long-term breeding study that annually monitors great tits using nest boxes placed within the woods. Great tits are preferential cavity nesters, and nearly all individuals breeding in Wytham Woods use nest boxes that are monitored by field assistants during annual data collection. Birds in this study are fitted with standard British Trust for Ornithology (BTO) metal leg bands as well as Passive Integrated Transponders (PIT tags), which were used to identify individuals in this study. Great tits are year-round residents and begin claiming territories in February or March each year, approximately four weeks before the onset of breeding (Firth and Sheldon 2015). Juvenile male great tits acquire songs from nearby adults and are thought to be able to acquire

new songs after dispersal to breeding territories in their first year (Rivera-Gutierrez et al 2011). Males use repertoires that include between one to nine unique song types, which they use during dawn chorus displays during the breeding system (McGregor and Krebs 1982).

Using data from the long-term study, I identified all focal birds as residents, dispersers, or immigrants. Residents were defined as any birds that were born in Marley Plantation (the region of Wytham Woods where data were collected for this study) or any bird that had bred previously within Marley Plantation. Dispersers were defined as birds born within or having previously bred within Wytham Woods, but outside of Marley Plantation. Immigrants were defined as birds that were not born in and had not previously bred within Wytham Woods. Birds that could not be identified using BTO rings or PIT tags were classified as unknown. Using nest monitoring data collected by field assistants, I identified all nest boxes occupied by great tits in the study area, and mapped territories using Thiessen polygons, which has been shown to closely match the regions occupied by breeding males in this system (Wilkin et al. 2007). I defined all pairs of birds sharing a territory boundary as neighbors.

Data collection. I collected recordings at 54 great tit nests between 30 March and 15 May in 2017-2019 (N = 21, 16, 17 for each year) using Swift acoustic recorders (Cornell Center for Conservation Bioacoustics, Ithaca, NY). The recorders collected sounds from focal nests continuously from 0500-0900 daily and saved recordings in as WAV files with a 32kHz sampling rate and 16-bit precision. I analyzed the songs used by focal birds during dawn chorus displays on three consecutive mornings, as this has been shown to be sufficient to sample an individual's complete repertoire (Rivera-Gutierrez et al. 2011). I ensured that the

mornings sampled from each individual were collected either during the egg laying period of the mate of the focal male or within five days before the onset of egg laying, as this is the period of peak dawn chorus output (Mace 1987). I analyzed only songs used in males' dawn chorus display, which was defined as any songs produced within 90 min of civil twilight (*sensu* Mace 1987). This ensured the dataset did not include songs used during counter-singing. I used Raven Pro 1.5 (Cornell Bioacoustics Research Program, 2014) to generate spectrograms of recordings with a Hann window function, 1024-point Fourier transforms, and 50% window overlap. A research assistant trained to identify great tit songs created separate Raven selections for all dawn chorus songs. Using the technique described in Keen et al. (in prep), songs from an individual male were classified as distinct song types. I then selected the five samples of each song type produced by a single bird that had the highest signal-to-noise ratio, and used this representative subset of songs to obtain acoustic measurements for every bird.

Acoustic analysis. In order to calculate the amount of vocal similarity among birds, I collected several acoustic measurements of all songs in the dataset. These included 28 spectro-temporal measurements from every recorded song using the WarbleR R package (Araya-Salas and Smith-Vidaurre, 2017), 181 descriptive statistics of Mel Frequency cepstral coefficients (MFCCs; Lyon and Ordubadi 1982, *sensu* Salamon et al. 2014), as well measurements generated from similarity matrices produced by spectrogram cross correlation (Clark et al. 1987) and dynamic time warping (Wolberg 1990). The song measurement vectors were collated into a single matrix for each bird. Using the method described in Chapter 2, I assigned every recorded song into a class representing a distinct song type, and

selected the 15 songs with highest signal-to-noise ratio from each class. I then averaged all measurements from this subset of songs from a single bird, resulting in one measurement vector per individual. Finally, I calculated vocal differences between pairs of individuals as Euclidean distance between these vectors, and hereafter refer to this value as pairwise acoustic distance.

To determine whether geographic distance between birds is correlated with acoustic similarity, I used Mantel tests to compare pairwise acoustic similarity with pairwise distance between breeding territories. For resident birds born within Wytham Woods, I used a separate Mantel test to compare pairwise acoustic similarity with pairwise distance between natal nests. This allowed me to evaluate whether birds exhibited vocal convergence with nearby conspecifics after dispersing to breeding territories, which suggest that there is context-dependent selection pressure for vocal convergence.

To calculate levels of vocal convergence among neighbors, I first calculated the mean pairwise acoustic similarity between a focal individual and all of its neighbors, as well as the mean pairwise acoustic similarity between that individual and all non-neighbors. This was repeated for every bird in the dataset. Non-neighbors were defined as all birds recorded within the same season that did not share a territory boundary with the focal bird. To test whether vocal converge among neighbors was higher than among non-neighbors, I used a linear mixed model (LMM) with pairwise acoustic similarity as a response variable, pairwise relationship (neighbors or non-neighbors) as a fixed effect, and reference bird as a random effect. To evaluate whether the levels of acoustic similarity with neighbors and non-neighbors was different for residents, immigrants, and dispersers I compared acoustic similarity with

neighbors and acoustic similarity with non-neighbors using two one-way ANOVAs and Tukey post hoc tests to make comparisons among birds in from each class.

Novel song playbacks. In order to test the convergence model predictions for residents and immigrants, I used a playback experiment to simulate the arrival of an immigrant bird into an established population of residents. I conducted ten playbacks using unique recordings of non-local great tit songs at ten locations within the study system (Fig. 3). All song recordings used in playbacks were acquired from the online archive Xeno-Canto (xeno-canto.org). I selected songs that were recorded in mainland Europe and ensured that songs did not resemble those used by local birds. I created ten unique MP3 sound files using Audacity 2.3.1 (audacityteam.org). Playback sounds comprised 1 minute of a novel song type sung repeatedly followed by 10 seconds of silence, repeated 15 times. I programmed AGPTEK A02 MP3 players to automatically play MP3 files at 0900, 0930, 1000, 1030, and 1100 daily for five consecutive days. All MP3 players were connected to ANKER SoundCore 2 Bluetooth speakers which were placed on an empty neighboring territory and facing in the direction of the focal nest. For all replicates, speakers were approximately 100 m from the focal nest. I placed Swift acoustic recorders at focal nests approximately 5 days before playbacks began. The recorders collected data continuously from 0500-0900 as in the song analysis described above.



Figure 3. Playback locations. Black dots represent nest boxes in the study site and red stars indicate focal nests at which playbacks took place.

To determine whether focal birds adjusted their singing behavior after the onset of playbacks, I compared songs used during the three mornings preceding playbacks to the songs used on the third, fourth, and fifth mornings after playbacks began. Songs used by focal birds were identified within recordings and analyzed following the procedure described above. I tested for changes in songs used by focal males before and after playbacks using three approaches. First, I visually inspected spectrograms to determine whether focal birds added or removed songs from their repertoires after playbacks began. Second, I calculated the acoustic distance between the playback songs used at a focal nest and all songs used by the focal bird. I

found the means of these distances for songs used before and after the start of playbacks for all focal nests. To test whether there was an overall effect in singing behavior for all birds in the experiment, I used an LMM with acoustic distance as a response variable, order (before or after) as a fixed effect, and nest as a random effect. Lastly, I used separate LMMs for all focal birds to test the effects of playbacks on singing behavior, using acoustic distance as a response variable, order (before or after) as a fixed effect, and date of recording as a random effect. I conducted all analyses in R (R Core Team 2015) and used LmerTest package for model analysis (Kuznetsova et al. 2015).

RESULTS

I found that pairwise acoustic distance was correlated with distance between breeding nests (Mantel test: correlation = 0.28, $p = 0.004$ $N = 54$ pairs; Fig 4a), and did not find a significant correlation between pairwise acoustic similarity and distance between natal nests (Mantel test: correlation = 0.01, $p = 0.43$, $N = 21$ pairs; Fig 4b). Birds in the study exhibited significantly higher acoustic similarity with neighbors as compared to non-neighbors (LMM: $t = 3.02$, $df = 54.9$, $p = 0.003$; Fig 4c), although the observed levels of acoustic similarity did not suggest complete vocal convergence (acoustic similarity between focal bird and neighbors (mean \pm SE): 0.72 ± 0.02 , focal bird and non-neighbors: 0.64 ± 0.02). There was not a significant difference in the amount of acoustic similarity that residents, immigrants, and dispersers exhibited with neighbors (ANOVA: $F(2) = 0.55$, $p = 0.58$; Tukey posthoc test: immigrants vs. residents: $p = 0.56$, dispersers vs. residents: $p = 0.7$, dispersers vs. immigrants: $p = 0.99$) or non-neighbors (ANOVA: $F(2) = 1.01$, $p = 0.37$; Tukey posthoc test: immigrants vs. residents: $p = 0.91$, dispersers vs. residents: $p = 0.36$, dispersers vs. immigrants: $p = 0.45$; Fig. 4d).

I collected sufficient recordings for analysis at eight of the ten playback locations. Visual inspection of spectrograms showed that focal birds did not adopt novel songs used in playbacks. One individual added a song to their repertoire after playbacks began that had not been used previously, although this song did not resemble the playback song. This individual was also the only bird in our sample to show significant changes in singing after the start of playbacks (Fig. 5, Table 3). I observed that several focal birds adjusted the ratios with which they used different songs in their repertoires, but playback timing (before or after) had no detectable influence on these adjustments (Table 3).

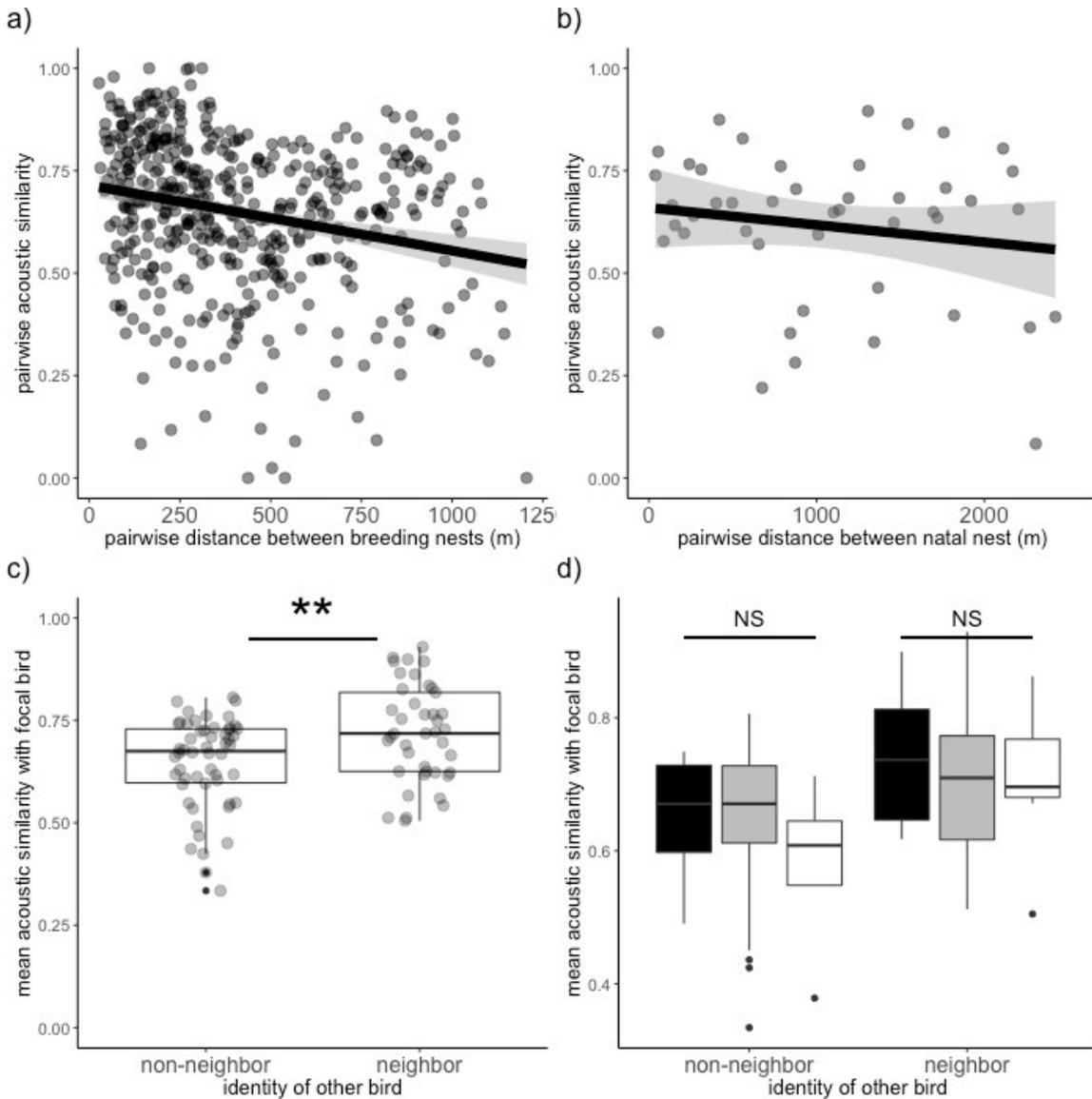


Figure 4. a) Pairwise acoustic similarity is significantly correlated with distance between breeding territories, b) Pairwise acoustic similarity is not significantly correlated with distance between natal nests, (c) focal birds have significantly higher pairwise acoustic similarity with neighbors versus non-neighbors, d) There is not a significant difference in the amount of acoustic similarity residents, immigrants, and dispersers exhibit with either neighbors or non-neighbors.

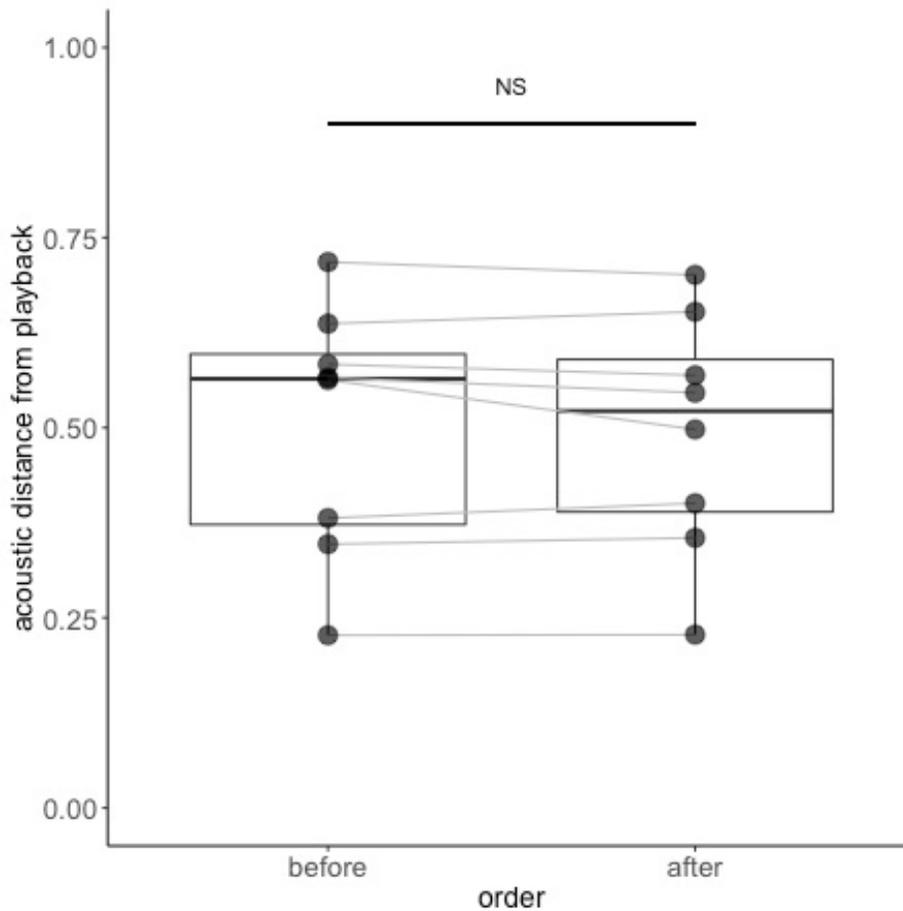


Figure 5. Birds did not adjust songs after playbacks simulating immigrant arrival. The mean acoustic distance of focal birds' repertoires from playback songs, shown on the y-axis, did not significantly change after playbacks began ($t = 1.24$, $p = 0.21$). Black dots represent acoustic means of focal birds ($N = 8$) and grey lines indicate measurements from the same bird.

Table 3. Individual results for playback replicates. I used order (before or after) as a fixed effect and recording day as random effect, and found that only the bird in replicate 2 significantly changed songs after the onset of playbacks.

Replicate	Estimate	Standard error	t	p
1	0.028	0.024	1.18	0.24
2	0.082	0.032	2.55	0.011
3	-0.018	0.044	-0.41	0.69
4	0.081	0.059	1.38	0.17
5	0.0	0.017	-0.03	0.98
6	0.0	0.015	0.05	0.96
7	0.052	0.042	1.22	0.22
8	0.012	0.011	1.09	0.28

DISCUSSION

Although many studies have investigated vocal convergence, the factors that determine when convergence is expected and to what extent individuals will converge are poorly understood. The models presented here make it possible to predict these outcomes for a given social and ecological context and to consider how context changes may lead to changes in vocal behavior. I find that both my observations of vocal similarity among resident birds and my results from playback experiments simulating immigrant arrival are most consistent with the convergence model predictions. Below, I discuss the implications of these results and suggest that this model could be generalized to explain signal convergence in many scenarios.

Empirical tests of model predictions. I analyzed songs used during the breeding season in a population of great tits, during which male territory defense occurs (Falls et al. 1982, McGregor et al. 1992). If a defending male confronts a newly settled territorial neighbor that has a novel song (high acoustic difference) compared to the other established territorial neighbors, the latter male might be seen as an enhanced threat for a territory take-over, so there might be an advantage for the male with the novel song to converge its song with those of the other established territorial neighbors. The converging male could thereby reduce the chance of receiving mistaken costly aggression. This sets up a benefit for social convergence, and this benefit will be strong given the presence of multiple territorial male neighbors.

However, past studies also suggest that females prefer males that use some unfamiliar song types (McGregor and Krebs 1982) as well as males with larger repertoires (Baker et al. 1986), which often correlates with using more unshared songs (Keen et al. in prep). Thus, to the extent that females are receivers, one might expect some sexually-selected divergence among male songs, opposing the selection pressures favoring convergence at least to some degree.

Thus, in this study population, one might expect that senders optimally should make some convergence effort but do not fully match neighbors. This is seen in my results: birds sharing territory boundaries had higher levels of acoustic similarity than non-neighbors, but did not approach full convergence. Additionally, I find a negative correlation between pairwise acoustic similarity and distance between current nests, but do not find this relationship when considering natal nests. This suggests that birds exert effort to converge with current neighbors, and thus adapt to changing social contexts. This effort might include

acquiring songs that are acoustically similar to neighbors' songs after dispersing, or searching for neighbors with similar songs before territory establishment.

I find further support for context-dependent convergence when making comparisons between residents and immigrants. Previous work shows that great tits from different populations often use acoustically dissimilar songs (Rivera-Gutierrez et al. 2010), suggesting that upon arrival into a new population, immigrants have higher initial acoustic distance from neighbors than residents. The model predicts that because immigrants and residents experience identical social benefits of convergence, they will have the same optimal level of convergence relative to neighbors' signals. Therefore, immigrants and residents are predicted to exhibit rapidly the same levels of vocal convergence with neighbors, although achieving this will be costlier for immigrants. This prediction would also apply to dispersers that arrive from other regions of the study system. In accordance with model predictions, I find that that residents, dispersers, and immigrants did not differ significantly in levels of acoustic similarity with neighbors. In other words, all individuals exhibited similarly high levels of vocal similarity with neighbors regardless of immigration status. This is in agreement with the model prediction that the optimal level of vocal convergence is a fixed distance from neighbors' signals, regardless of the initial acoustic distance between the focal individual and neighbors. Unfortunately, because few birds were resampled in subsequent years or sampled at both the beginning and end of the breeding season, it was not possible to evaluate the change in immigrant vocal behavior over time.

The observational analysis allowed me to evaluate the study population after convergence effort was made by focal individuals, which presumably happens before or during territory establishment. In contrast, the playback experiment made it possible to test

predictions about changes in convergence effort in response to changing contexts. For the specific scenario of an immigrant arriving into a population of established residents, the convergence model predicts that a resident will exert little or no effort to converge with an immigrant, because doing so would then maladaptively diverge with its other neighbors. This was supported by findings that resident focal birds did not adopt highly novel playback songs and exhibited little or no change in acoustic distance with the novel playback song.

Implications and broader relevance. Taken together, the empirical results suggest that breeding great tit males adjust their singing behavior in the presence of neighboring territorial males in accordance with the convergence model. This model also predicts that low initial signal distance and low costs to convergence can facilitate the emergence of dialects. Why, then, do great tits not exhibit dialects? One possible explanation is that the sensitivity to convergence is relatively low in this species, which may correspond to the tradeoff between territory defense and mate attraction mentioned above. This presents an interesting implication of our model: when using a single mode of communication, a signaler's peak fitness may be lower than the peak fitness that could be attained by using multimodal communication, some components of which exhibit high convergence and others of which exhibit high divergence. This therefore suggests that multimodal signaling may be favored in contexts where senders must communicate with multiple receivers.

We can also consider possible outcomes in species that experience different magnitudes of social benefits and effort costs. For example, when signaling social group membership strongly influences survival because it enables group recognition, our model predicts high levels of convergence because social benefits and sensitivity are high. This

aligns with evidence of vocal convergence in lekking birds (hermit hummingbirds, Kapoor 2016), group-living species (e.g., budgerigars, Farabaugh et al. 1994), and cooperative breeders (e.g., wood hoopoes, Radford 2005; superb starlings, Keen et al. 2013; western bluebirds, Akcay et al. 2014). Similarly, vocal convergence with mates may augment social benefits in pair bonding species, as has been observed budgerigars (Hile et al. 2000), crossbills (Sewall 2009), and ravens (Luef et al. 2017).

This framework may also be applied to non-vocal signals. For example, social insects are known to identify nest mates using cuticular hydrocarbons and are highly sensitive to differences in hydrocarbon profiles (Hölldobler and Wilson 2009). Given the high sensitivity to convergence and high social benefits of signaling group membership, our model predicts that fitness is maximized when signal distance from neighbors approaches zero. This prediction is supported by evidence that nest mates exhibit colony-specific hydrocarbon profiles that result from contact between individuals and nest materials, and are highly similar among nest mates (Lenoir et al. 1999). A different example from social insects is the use of facial patterning in social wasps, which has been shown to signal individual identity in at least one species (Tibbetts 2002). Within wasps that form dominance hierarchies, individuals that are more distinct from others receive less misdirected aggression (Sheehan and Tibbetts 2009), which aligns with the predictions of the divergence model (Dale et al. 2001). Another example in which divergence is favored is found in sciurid rodents, which have been shown to use more individually distinct vocal signals when living in larger groups because receivers must discriminate among higher numbers of senders (Pollard and Blumstein 2011). We can also consider the case of egg mimicry in brood parasites. In cases where host defenses enable receivers to be very discerning, brood parasites show high levels of convergence in egg

appearance with hosts (Spottiswoode and Stevens 2011), as predicted by high sensitivity in the convergence model. We could use the same approach to describe the occurrence of mimicry to avoid aggression, or, more broadly, the maintenance of cultural norms.

Contextual importance. My results suggesting that complete convergence is non-optimal could seem inconsistent with studies showing that song matching is prevalent among territorial songbirds. Although one function of convergence is likely song matching during counter-singing (Bradbury and Vehrencamp 2011), I suggest that this is not mutually exclusive from the model predictions. This can be explained by considering that counter singing occupies a particular context because it is a different type recognition and communication. Song matching might serve the purpose of allowing birds to assess fighting ability, e.g. by evaluating song consistency (Byers 2007, Botero et al. 2009, Rivera-Gutierrez et al. 2010). In song matching, it is also expected that neighbors have the option to not match opponents so that encounters can be de-escalated (Beecher and Campbell 2005). Thus, neighbors are not predicted to exhibit complete repertoire sharing, which is in line with model predictions. This clarification may help to explain observations of rufous-and-white wrens, a species in which both sexes sing, but primarily males engage in counter singing, meaning that sexes experience different contexts. Males in this species exhibit only partial vocal convergence with neighbors, but levels of convergence among males were significantly higher than in females (Graham et al. 2017). The model could represent this difference as higher sensitivity to convergence among males. A similar explanation of sensitivity changing with context could be applied to reports of individuals adjusting levels of vocal convergence

in different social environments, such as Diana monkeys showing higher convergence with groupmates when non-group members are nearby (Candiotti et al. 2012).

In addition to social context, we might also consider how ecological context could influence levels of vocal convergence. For example, certain individuals might be more likely to occupy territories with particular habitat characteristics, as in the case of high quality individuals holding territories in high quality habitat. In this scenario, individuals of similar quality might cluster together within a population. Therefore, when signals encode individual quality and there is variation in habitat quality within a system, neighbors may be more likely to use similar signals than non-neighbors. Although my study took place in a homogeneous habitat and thus territory quality was not included in my analysis, future work may benefit from considering this factor.

Future directions. The most appropriate tests of model predictions would be experimental manipulations of the signals to which a focal individual is exposed, e.g., a relocation to a distant population in which conspecific signals are very different. In this case, the convergence model would predict that the relocated focal individual (i.e., the immigrant) will adopt signals used in the local environment, rather than residents adjusting their signals. It would also be possible to test predictions with a “natural experiment” the follows focal birds from their natal sites to dispersal sites or with a meta-analysis of past observational studies.

Conclusions. Patterns of vocal convergence cannot be universally explained by any one set of social or ecological factors. The proposed framework may help to explain observed patterns of signal convergence versus divergence in many taxa by considering how these factors

determine each individual's costs and benefits for convergence versus divergence. My empirical results for great tits align with convergence model predictions, both in the case of population-wide levels of song similarity, and in local interactions between residents and immigrants.

ACKNOWLEDGEMENTS

I thank H. Kern Reeve for his essential guidance during the development and writing of this paper. Thank you to Ana Verahami, Dallas Jordan, Benjamin Walton, Keith McMahon, Sam Crofts for their enthusiasm and help during data collection and analysis. This project was made possible by hardware and computing support provided by Holger Klinck and the Cornell Center for Conservation Bioacoustics. This project was supported by the Cornell Lab of Ornithology Athena Fund, the Edward Gray Institute for Field Ornithology, and funding from the Cornell Department of Neurobiology and Behavior. I am also grateful for many helpful suggestions during Cornell's Animal Behavior Lunch Bunch.

WORKS CITED

- Akçay, Ç., Hambury, K. L., Arnold, J. A., Nevins, A. M., and Dickinson, J. L. (2014). Song sharing with neighbours and relatives in a cooperatively breeding songbird. *Animal Behaviour*, 92: 55-62.
- Araya-Salas, M. and Smith-Vidaurre, G. (2017). warbleR: An R package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution*, 8: 184-191.
- Baker, M. C. (1982). Genetic population structure and vocal dialects in Zonotrichia (Emberizidae). *Acoustic communication in birds*, 2: 209-235.
- Baker, M. C., and Cunningham, M. A. (1985). The biology of bird-song dialects. *Behavioral and Brain Sciences*, 8: 85-133.
- Baker, M. C., Bjerke, T. K., Lampe, H., and Espmark, Y. (1986). Sexual response of female great tits to variation in size of males' song repertoires. *The American Naturalist*, 128: 491-498.
- Botero, C. A., Rossman, R. J., Caro, L. M., Stenzler, L. M., Lovette, I. J., de Kort, S. R., and Vehrencamp, S. L. (2009). Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Animal Behaviour*, 77: 701-706.
- Bradbury, J. W., and Vehrencamp, S. L. (2011). Principles of animal communication. 2nd. *Sunderland, Massachusetts: Sinauer*.
- Byers, B.E. (2007). Extra-pair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behavioral Ecology*, 18: 130-136.
- Candiotti, A., Zuberbühler, K., and Lemasson, A. (2012). Convergence and divergence in Diana monkey vocalizations. *Biology Letters*, 8: 382-385.
- Clark, C. W., Marler, P., and Beeman, K. (1987). Quantitative analysis of animal vocal phonology: an application to swamp sparrow song. *Ethology*, 76: 101-115.
- Dale, J., Lank, D. B., and Reeve, H. K. (2001). Signaling individual identity versus quality: a model and case studies with ruffs, queleas, and house finches. *The American Naturalist*, 158: 75-86.
- Deecke, V. B., Ford, J. K., and Spong, P. (2000). Dialect change in resident killer whales: implications for vocal learning and cultural transmission. *Animal behaviour*, 60: 629-638.
- Farabaugh, S. M., Linzenbold, A., and Dooling, R. J. (1994). Vocal plasticity in Budgerigars (Melopsittacus undulatus): evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108: 81.

Feekes, F. (1977). Colony-specific song in *Cacicus cela* (Icteridae, Aves): The password hypothesis. *Ardea*, 65: 197–202.

Firth, J. A., and Sheldon, B. C. (2015). Experimental manipulation of avian social structure reveals segregation is carried over across contexts. *Proceedings of the Royal Society B: Biological Sciences*, 282: 20142350.

Garland, E. C., Goldizen, A. W., Rekdahl, M. L., Constantine, R., Garrigue, C., Hauser, N. D., and Noad, M. J. (2011). Dynamic horizontal cultural transmission of humpback whale song at the ocean basin scale. *Current biology*, 21: 687-691.

Gaunt, S. L., Baptista, L. F., Sanchez, J. E., and Hernandez, D. (1994). Song learning as evidenced from song sharing in two hummingbird species (*Colibri coruscans* and *C. thalassinus*). *The Auk*, 111: 87-103.

Graham, B. A., Heath, D. D., and Mennill, D. J. (2017). Dispersal influences genetic and acoustic spatial structure for both males and females in a tropical songbird. *Ecology and evolution*, 7: 10089-10102.

Hile, A. G., Plummer, T. K., and Striedter, G. F. (2000). Male vocal imitation produces call convergence during pair bonding in budgerigars, *Melopsittacus undulatus*. *Animal Behaviour*, 59: 1209-1218.

Hölldobler B. and Wilson, E. O. (2009). *The superorganism: the beauty, elegance, and strangeness of insect societies*. New York, NY: WW Norton and Company.

Janik, V. M., and Slater, P. J. (1998). Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. *Animal behaviour*, 56: 829-838.

Johnstone, R. A. (1997). Recognition and the evolution of distinctive signatures: when does it pay to reveal identity? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264: 1547-1553.

Kapoor, V. (2016). The Functional Significance of Microgeographic Dialects in a Hermit Hummingbird. PhD Dissertation, Cornell University.

Keen, S. C., Meliza, C. D., and Rubenstein, D. R. (2013). Flight calls signal group and individual identity but not kinship in a cooperatively breeding bird. *Behavioral Ecology*, 24: 1279-1285.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). Package ‘lmerTest’. *R package version 2.0*.

Lachlan, R. F., Anderson, R. C., Peters, S., Searcy, W. A., and Nowicki, S. (2014). Typical versions of learned swamp sparrow song types are more effective signals than are less typical versions. *Proceedings of the Royal Society B: Biological Sciences*, 281: 20140252.

- Lenoir, A., Fresneau, D., Errard, C., and Hefetz, A. (1999). "The individuality and the colonial identity in ants: the emergence of the social representation concept," in *Information Processing in Social Insects*, eds C. Detrain, J. L. Deneubourg, and J. Pasteels. Basel: Birkhauser, 219–237.
- Luef, E. M., Ter Maat, A., and Pika, S. (2017). Vocal similarity in long-distance and short-distance vocalizations in raven pairs (*Corvus corax*) in captivity. *Behavioural processes*, 142: 1-7.
- Luther, D. A., and Derryberry, E. P. (2012). Birdsongs keep pace with city life: changes in song over time in an urban songbird affects communication. *Animal Behaviour*, 83: 1059-1066.
- Lyon, R. H., and Ordubadi, A. (1982). Use of cepstra in acoustical signal analysis. *Journal of Mechanical Design*, 104: 303-306.
- Macdougall-Shackleton, S. A. (1997). Sexual selection and the evolution of song repertoires. In *Current ornithology* (pp. 81-124). Springer, Boston, MA.
- Marler, P., and Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. *Science*, 146: 1483-1486.
- McGregor, P. K., and Krebs, J. R. (1982). Mating and song sharing in the great tit. *Nature*. 297: 60-61.
- McGregor, P. K., Dabelsteen, T., Shepherd, M. and Pedersen, S. B. 1992. The signal value of matched singing in great tits: evidence from interactive playback experiments. *Animal Behaviour*, 43: 987–998
- Nelson, D. A. (2000). Song overproduction, selective attrition and song dialects in the white-crowned sparrow. *Animal Behaviour*, 60: 887-898.
- Payne, R. B. (1982). Ecological consequences of song matching: breeding success and intraspecific song mimicry in indigo buntings. *Ecology*, 63: 401-411.
- Podos, J., and Warren, P. S. (2007). The evolution of geographic variation in birdsong. *Advances in the Study of Behavior*, 37: 403-458.
- Pollard, K. A., and Blumstein, D. T. (2011). Social group size predicts the evolution of individuality. *Current Biology*, 21: 413-417.
- Radford, A. N. (2005). Group-specific vocal signatures and neighbour–stranger discrimination in the cooperatively breeding green woodhoopoe. *Animal Behaviour*, 70: 1227-1234.

- Reeve, H.K. (1989). The evolution of conspecific acceptance thresholds. *The American Naturalist*, 133: 407–435.
- Rivera-Gutierrez, H. F., Matthysen, E., Adriaensen, F., and Slabbekoorn, H. (2010). Repertoire sharing and song similarity between great tit males decline with distance between forest fragments. *Ethology*, 116: 951–960.
- Rivera-Gutierrez, H. F., Pinxten, R., and Eens, M. (2011). Difficulties when assessing birdsong learning programmes under field conditions: a re-evaluation of song repertoire flexibility in the great tit. *PLoS One*, 6.
- Salamon, J., Jacoby, C., and Bello, J. P. (2014). A dataset and taxonomy for urban sound research. In *Proceedings of the 22nd ACM international conference on Multimedia*. 1041-1044.
- Searcy, W. A., Nowicki, S., Hughes, M., and Peters, S. (2002). Geographic song discrimination in relation to dispersal distances in song sparrows. *The American Naturalist*, 159: 221–230.
- Sewall, K. B. (2009). Limited adult vocal learning maintains call dialects but permits pair-distinctive calls in red crossbills. *Animal Behaviour*, 77: 1303-1311.
- Sheehan, M. J., and Tibbetts, E. A. (2009). Evolution of identity signals: frequency-dependent benefits of distinctive phenotypes used for individual recognition. *Evolution: International Journal of Organic Evolution*, 63: 3106-3113.
- Sheehan, M.J. and H.K. Reeve. (in press). Evolutionary stable investments in recognition systems explain patterns of discrimination failure and success. *Philosophical Transactions B: Biological Sciences*.
- Sherman PW, Reeve HK, Pfennig DW, Krebs JR, Davies NB. (1997) Recognition Systems. In *Behavioural ecology: an evolutionary approach*, Oxford: Blackell Science Ltd.
- Snell-Rood, E. C., and Badyaev, A. V. (2008). Ecological gradient of sexual selection: elevation and song elaboration in finches. *Oecologia*, 157: 545-551.
- Spottiswoode, C. N., and Stevens, M. (2011). How to evade a coevolving brood parasite: egg discrimination versus egg variability as host defences. *Proceedings of the Royal Society B: Biological Sciences*, 278: 3566-3573.
- Thom, M.D.F., Dytham C. Female choosiness leads to the evolution of individually distinctive males. *Evolution: International Journal of Organic Evolution*, 66: 3736-3742.

- Tibbetts, E. A. (2002). Visual signals of individual identity in the wasp *Polistes fuscatus*. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269: 1423-1428.
- Vehrencamp, S. L. (2001). Is song-type matching a conventional signal of aggressive intentions? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268: 1637-1642.
- Vehrencamp, S.L., Ritter, A.F., Keever, M. and Bradbury, J.W. (2003). Responses to playback of local vs. distant contact calls in the orange-fronted conure, (*Aratinga canicularis*). *Ethology*, 109: 37-54
- Wiens, J. A. (1982). Song pattern variation in the sage sparrow (*Amphispiza belli*): Dialects or epiphenomena?. *The Auk*, 99: 208-229.
- Wilkin, T. A., Perrins, C. M., and Sheldon, B. C. (2007). The use of GIS in estimating spatial variation in habitat quality: a case study of lay-date in the Great Tit *Parus major*. *Ibis*, 149: 110-1188.
- Wilkinson, G. S, Boughman, J. W. (1998). Social calls coordinate foraging in greater spear-nosed bats. *Animal Behaviour*, 55:337–350.
- Wolberg, G. (1990). *Digital image warping*. Vol. 10662. Los Alamitos, CA: IEEE computer society press.
- Wright, T. F. (1996). Regional dialects in the contact call of a parrot. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263: 867-872.
- Wright, T. F., and Dahlin, C. R. (2018). Vocal dialects in parrots: patterns and processes of cultural evolution. *Emu-Austral Ornithology*, 118: 50-66.

CHAPTER 4

SPATIAL AND TEMPORAL VARIATION IN SONGS IN A WILD POPULATION OF GREAT TITS, *PARUS MAJOR*

Sara Keen^{1,2}

¹ Department of Neurobiology and Behavior, Cornell University, Ithaca, NY

² Cornell Lab of Ornithology, 159 Sapsucker Woods Rd, Ithaca, NY

ABSTRACT

Vocal communication plays a crucial role in mediating social interactions in many taxa. Oftentimes among species that acquire vocalizations via social learning, vocal signals exhibit dynamic spatial variation thus signal structure may fluctuate on relatively short time scales. The songs of great tits, *Parus major*, have a simple, stereotyped structure, often comprising a single repeated two-note phrase. Across this species' continent-wide distribution, songs exhibit a similar structure and there is no evidence for distinct geographic dialects, though the particular song types found within a population may vary over time and space. Here, I analyze songs collected from a wild population of great tits in three consecutive years in order to investigate factors that might drive spatial and temporal variation in song on a microgeographic scale. I observed that the relative abundance of song types used in the population changed between years, and found no evidence that certain song types are used more often than expected by random chance. These results are consistent with a previous study conducted in this population 40 years prior, and upon comparing my dataset to earlier

records, I found that only a small proportion of previously recorded song types persist today. Although the likelihood of a song type being present in the population was positively correlated with the number of birds using the song in the preceding year, the appearance or disappearance of song types between years was nearly always explained by the appearance or disappearance of individual birds. Notably, all individuals in the study used at least one form of the common two-note song, and immigrant birds were more likely than residents to have more complex song types in their repertoires. I found that birds occupying breeding territories on forest edges used larger repertoires and shared fewer song types with the population than birds breeding in more central territories. Additionally, birds that shared few songs with the local population began breeding earlier and had larger clutches, which may be a consequence of breeding on higher quality territories. Together, these results suggest that the spatial distribution of song types may be influenced by a suite of factors, including competition for breeding territories and mates, immigration, individual survival, and nest-site characteristics. Lastly, I suggest that although there are high levels of spatial and temporal variation in song on a microgeographic scale, the ubiquity of common two-note song types may facilitate high levels of song similarity on a macrogeographic level.

INTRODUCTION

In many species, vocal communication plays an essential role in helping individuals navigate complex social environments. Bird song is a well-studied signal that is known to be a key determinant of survival and reproductive success (Catchpole and Slater 2008). In addition to being widely used for mate attraction and territory defense, songs can simultaneously signal a number of singer characteristics, including species, group and/or individual identity, and

individual quality (Searcy and Andersson 1986, Kroodsma and Byers 1991, Bradbury and Vehrenamp 2011). Owing to the fact that the songs are used to compete with or display to nearby conspecifics, and that signal transmission is mediated by habitat characteristics, it is critical that birds use songs which are best fit for both their social and ecological environment (Morton 1975, Hunter and Krebs 1979, Payne 1981, Nelson and Marler 1994, Vehrencamp 2001). However, this may be a moving target. Frequently, the songs which an individual should use to maximize their fitness are dependent on the precise location and timing of song production (Nelson 1992, Slabbekoorn and Smith 2002, Nordby et al. 2007). Several factors have been shown to influence temporal and spatial variation in songs, including sexual and/or natural selection, mating system, movement between populations, learning strategy, and repertoire size (Ellers and Slabbekoorn 2003, Kroodsma 2004, Derryberry 2009, Fayet 2014). Consequently, a continuum of vocal convergence is observed across species, ranging from relatively stable songs shared over large geographic scales, to songs that change from year to year and are shared by few individuals (Podos and Warren 2007, Wright and Dahlin 2018).

Within passerine birds, nearby conspecifics often converge upon similarly structured songs, which can lead to intra-specific geographic variation and the emergence of dialects (Baker and Cunningham 1985). Dialects often arise in species that have short range dispersal and in which juveniles learn songs by imitating nearby adults (Marler and Tamura 1962, Slabbekoorn and Smith 2002, Slater 1989). Territoriality has also been identified as a key ecological correlate among passerine species that exhibit geographic variation in songs, likely because the maintenance of territory boundaries often leads to song-sharing between adjacent males (McGregor and Krebs 1989, Kroodsma 2004, Beecher and Brenowitz 2005). A number of studies have proposed possible functions of vocal convergence, including enabling

recognition of group members and excluding non-members (Feekes 1977), facilitating assortative mating with individuals that are most fit for the local habitat (Marler and Tamura 1962), or encouraging adaptation to local environments (Slabbekoorn 2004). In all cases, the benefits an individual receives from resembling conspecifics and the amount of vocal convergence required to obtain benefits will be determined by unique ecological and social characteristics of the population (see Chapter 3).

Among avian species that use repertoires comprising several song types, geographic variation in songs may take the form of song sharing as well as high levels of similarity in acoustic structure of songs (McGregor and Krebs 1989, Nelson 1992, Beecher and Brenowitz 2005, Rivera-Gutierrez et al. 2010a). Because songbirds typically acquire songs via copying rather than innovation, they are usually limited to using songs of conspecific singers which were overheard during the sensitive period for vocal learning, although some variation may be introduced through copying errors (Slater 1989, Eilers and Slabbekoorn 2003, Slater and Lachlan 2003, Beecher and Brenowitz 2005). Unlike repertoire composition, repertoire size is thought to be a sexually selected trait in many species (Searcy 1992, MacDougall-Shackleton 1997, Catchpole and Slater 2008). An individual's ability to learn and produce multiple song types may correlate with age, experience, or developmental health, and therefore may be an honest indicator of quality in several species (Searcy and Nowicki 2005, Catchpole and Slater 2008, but see Byers and Kroodsma 2009). However, in species with small repertoires, the range of songs that an individual can produce is ultimately limited by learning opportunities and innate constraints (Gil and Gahr 2002), and oftentimes repertoire composition is further refined through selective attrition to best suit an individual's current social milieu (Marler and Peters 1982, Lachlan et al. 2018).

Movement and connectivity between populations can profoundly influence geographic variation in songs (Podos and Warren 2007). High levels of connectivity between populations, in the form of immigration, dispersal, or other manners of gene flow, have been shown to correlate with increased vocal similarity (MacDougall-Shackleton and MacDougall-Shackleton 2001). Inversely, lack of movement between populations is linked to increased vocal divergence, and previous studies have demonstrated that this may act as a barrier to reproduction and may promote speciation (Slabbekoorn and Smith 2002, Price 2008, Freeman and Montgomery 2017). However, among species that acquire vocalizations through social learning, geographic variation in song can often be maintained despite gene flow between populations (Ellers and Slabbekoorn 2003, Wright et al. 2005). Therefore, spatial patterns of variation do not always accurately reflect connectivity between populations, particularly among species that are capable of acquiring songs both before and after dispersal (Podos and Warren 2007). The relationship between individual movement within and between populations and microgeographic variation in song is not well understood, and few studies have explored how these processes might in turn influence spatial and temporal variation in songs on a larger scale.

To address these questions, I investigated variation in songs collected from a wild population of great tits (*Parus major*) in three consecutive years. Males of this species have repertoires comprising two to nine unique song types, with each song being a series of identical, repeated phrases composed of a unique combination of notes (Krebs 1976, Krebs 1977; Figure 1). Throughout their extensive range, which spans Europe and extends into the Middle East and parts of North Africa (BirdLife International 2020), great tits are known to use stereotyped songs composed of a repeated two-note phrase, often referred to with the

mnemonic “teacher” (Alexander 1935, Zollinger et al. 2017). Although many variants of this song exist, the distinctive pattern of two alternating high and low frequency notes is commonly used for species identification (Thomas 2019). Great tits also use song variants composed of phrases with three to five notes, but these account for a smaller proportion of songs observed in this species (McGregor and Krebs 1989). Past studies have suggested that habitat can influence the acoustic structure of great tit songs, with forest-dwelling birds having lower frequency songs than woodland birds (Hunter and Krebs 1979), and birds in urban areas singing at higher frequencies and using shorter notes than birds in rural areas (Slabbekoorn and Peet 2003, Slabbekoorn and den Boer-Visser 2006). Multiple studies have also demonstrated that repertoire size is positively correlated with survival and lifetime reproductive success (McGregor et al. 1981, Lambrechts and Dhont 1986, Rivera-Gutierrez et al. 2010b), and that song sharing between males decreases with distance between territories (McGregor and Krebs 1982, Rivera-Gutierrez et al. 2010a).

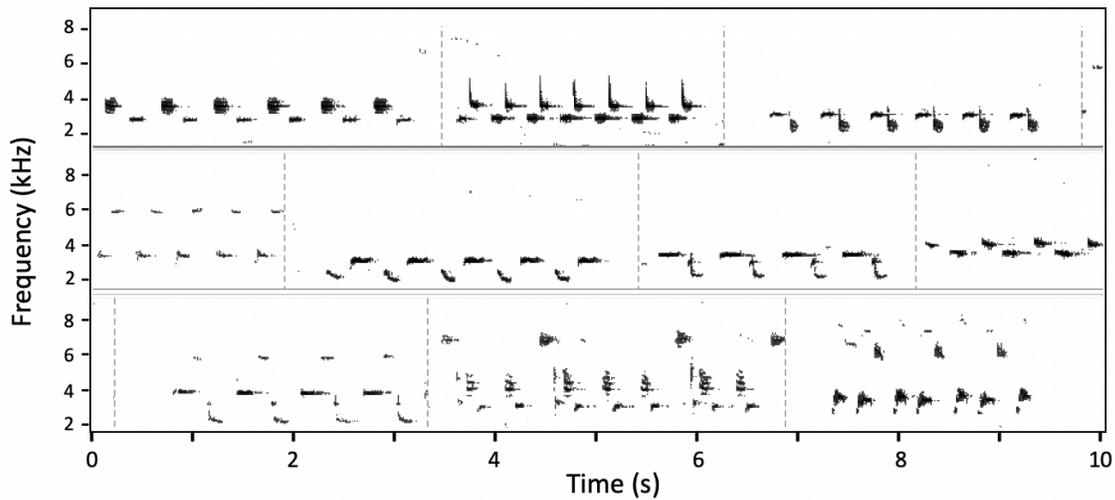


Figure 1. Spectrograms of great tit songs collected from the study population. Vertical dashed lines represent separations between sound files. Each sound file contains a different song type. Songs in the top two rows are composed of two-note phrases. Songs in the final row are composed of phrases with more than two notes and are termed complex songs in my analysis. Complex songs are infrequently shared with other birds in the study population. Spectrograms were generated using Raven Pro 1.5 using the parameters described below.

I analyzed songs used by tagged great tit males in consecutive years in order to determine the relative abundance of song types within the population and the extent to which these distributions varied over time and space. Using these data, I explore the how individual song characteristics vary with immigration status and reproductive success and investigate the role of individual movement and survival on microgeographic variation in songs. I also compare these findings with a previous study of songs in this population conducted by McGregor and Krebs (1982). Finally, I suggest possible explanations for the observed patterns in my data and make recommendations for future studies.

METHODS

Study population and data collection. I collected recordings of songs from male great tits in a wild population in Wytham Woods, Oxfordshire, UK (51460 N, 01200 W) between 1 -30 April during 2017-2019. Great tits in this population are part of a long-term breeding study begun in 1947 for which 1,018 woodcrete nest boxes were placed within the boundaries of the study site (Figure 2). Great tits in this population have been shown to breed almost exclusively in nest boxes (Hinde 1952, Firth and Sheldon 2015). More than 80% of the great tits in this population have been fitted with metal leg rings from the British Trust for Ornithology and an additional leg band with an identifiable passive integrated tag (PIT) produced by IB Technology, Aylesbury, U.K. Within Wytham Woods, great tits spend winters in mixed-species foraging flocks and often begin visiting and claiming breeding territories 4-6 weeks prior to the onset of egg laying every spring (Firth and Sheldon 2015). Individual lifespan ranges from 1 to 9 years, with the majority of birds surviving for one or two breeding seasons (Wilkin 2006). Great tits from nearby woodlands often immigrate into the study area and typically account for approximately one third of the breeding population in a given year (Wilkin 2006, Fayet et al. 2014). This population also experiences high levels of nest predation and offspring mortality; on average, approximately one offspring per breeding individual survives to breed in the following year (McCleery et al. 2004).



Figure 2. Map of Wytham Woods. Map areas with grey background represent land within study system boundaries. Axes represent Ordnance Survey National Grid coordinates and axis ticks indicate distance in meters. Points show locations of nest boxes; black points are nest boxes that were included in the study and gray points are nest boxes that were not included.

Great tit males typically acquire songs from nearby adults, including but not limited to their social fathers (McGregor and Krebs 1982). The sensitive period for song learning extends into the first breeding season and possibly beyond this point, meaning that males' repertoires can include songs acquired both before and after dispersal (Franco and Slabbekoorn 2009). During the breeding season, males sing a dawn chorus near their nest box before their female mate emerges each morning. This display is closely synchronized with

female fertility and male dawn chorus output reaches its maximum near the onset of female egg laying (Mace 1987). In addition to their ubiquitous two-note songs, which can be categorized into many different songs types, great tits can also use song types that consist of repeated sequences of more than two notes, hereafter referred to as complex songs (Krebs 1976, 1977a; Figure 1). Males frequently produce the same song type multiple times in succession, often cycling through all songs in their repertoires during their dawn chorus display (Lambrechts and Dhont 1986, Naguib et al. 2019, but see Rivera-Gutierrez et al. 2011).

I monitored great tits using nest boxes in Marley Plantation, the southeastern most corner of the study site (Figure 2). Field assistants visited nest boxes multiple times per week and identified all nests occupied by great tit pairs by catching birds in order to obtain identifying information from PIT tags. Using data collected from the long-term study, it was possible to reference records of natal nests of all birds in the population to classify males as residents, immigrants, dispersers, or of unknown origin. I defined residents as birds that were born in Wytham Woods or were previously recorded breeding within the system, immigrants as birds that were not born in Wytham and had not bred previously in the woods, and dispersers as birds that were born in Wytham Woods outside of Marley Plantation (the location of this study), and had not previously bred in this region. Birds that were unable to be caught or identified, including untagged birds, were classified as unknown. The sample included a total of 54 birds (birds sampled per year: 2017: N = 21, 2018: N = 16, 2019: N = 17; Table 1). Five birds from 2017 were resampled in 2018, and two birds sampled in 2018 were resampled in 2019. No birds were sampled in all three years.

Acoustic data collection and analysis. Acoustic recordings were collected by placing Swift recording units (Center for Conservation Bioacoustics, Cornell Lab of Ornithology, Ithaca, NY) directly below nest boxes occupied by breeding pairs of great tits. Recorders were programmed to record continuously from 0500-1000 daily, and recordings were saved as hour-long WAV files using a 32KHz sampling rate and 16-bit depth. I identified songs from focal males by using only the songs that were recorded at an amplitude threshold of 75 dB (as calculated by the peak power measurement for individual Raven selections) to ensure that songs were produced within close proximity to the nest. Throughout data collection, dawn chorus displays were regularly visually observed by field assistants and recorded using a Sennheiser shotgun microphone (Sennheiser Electronic, Old Lyme, CT) and a Marantz PMD661 digital recorder (Marantz, Mahwah, NJ) to confirm that songs captured by Swift recording units were indeed produced by focal males. Additionally, I compared recordings of songs produced at the same nest on different days to confirm that songs were produced by the same bird, as great tits often exhibit consistent repertoire composition between days (Naguib et al. 2019).

For each male included in my sample, I selected three consecutive mornings of recordings to use for acoustic analysis. All samples were collected either during the egg-laying period of a focal male's mate, or in the three days preceding egg laying. Spectrograms of recordings were created with Raven Pro 1.5 (Bioacoustic Research Program 2015) using a Hann window function with a 512-sample window and 50% overlap. Trained research assistants manually reviewed spectrograms and created Raven selections around all songs in an individual's dawn chorus display. The dawn chorus was defined as any song produced within 90 mins of civil twilight, as in (Mace 1987). I identified the song types in males' dawn

chorus repertoire using the method described in Chapter 2, and used Raven Pro 1.5 to measure signal-to-noise ratio to determine the highest recording quality for all songs in a focal male's repertoire, and selected the five best exemplars of each song type from an individual. This resulted in an 865-song dataset. I then used the method described in Chapter 2 to construct an acoustic feature space within which songs were distributed. I also conducted a visual analysis of spectrograms to determine whether song types were composed of two-note phrases or were composed of phrases with more than two notes, i.e. complex songs (Figure 1).

Statistical analyses. Several analyses were used to evaluate spatial and temporal variation in songs. First, I calculated the number of unique song types that individuals used, i.e., repertoire size. For every year in the sample, I created distributions of the number of birds using each song type and calculated song type densities (i.e., relative abundance) as the number of birds using a song type divided by the sum of unique song types in that year each multiplied by the number of birds using a song type. I used a Chi-squared test to determine whether the distribution of song type densities differed from a broken stick distribution (MacArthur 1957), which is considered a null distribution and allows for comparison to a previous study of this population (McGregor and Krebs 1982).

To test whether the number of birds using a song type in the given year influence the number of birds using that song type the following year, I used a linear mixed model (LMM) with song type presence in the current year as a dependent variable, the number of birds using the song type in the next year as a fixed effect, and current year (2017 or 2018) and song type as random effects. Presence was defined as at least one bird using a song type. I also calculated the Pearson correlation between the number of birds using a song type in the

preceding year and the number of birds using that song type in the current year. I did not remove birds that were sampled in consecutive years for either analysis.

To evaluate whether neighbors share more songs than expected by chance, I first used Thiessen polygons (Aurenhammer 1991) to calculate territories around all great tit nests in the study area each year, regardless of whether they were included in my sample. This technique has previously been shown to correlate closely with territory boundaries (Wilkin et al. 2007a, Firth and Sheldon 2016). I then classified pairs of birds within the same year as neighbors if they shared a territory boundary, and as non-neighbors otherwise. I used the Dice similarity index (Dice 1945) to calculate pairwise repertoire similarity between all birds within the same year. For every song in a bird's repertoire, I determined whether other birds used the song type (assigning 0 or 1 for song absence or presence, respectively), and calculated the probability of neighbors and non-neighbors sharing the song type by finding the mean within each group. Comparisons were made only between birds sampled in the same year. I then used two LMMs with pairwise repertoire similarity and probability of sharing a song as dependent variables, identity of the other bird (either neighbor or non-neighbor) as a fixed effect, and reference bird identity as a random effect.

To test whether repertoire composition was correlated with nest location, I first developed a metric to evaluate individual acoustic dissimilarity relative to the rest of the population. To calculate this, I found the median distance of all songs in a bird's repertoire from the centroid of the acoustic feature space. Because an inherent property of the acoustic feature space is that more common songs are in the center of the space and less common songs are farther from the center, this is therefore a proxy for how different a bird sounds from the local population. I used a Pearson correlation to test whether nest distance from the

edge of the woods was correlated with acoustic dissimilarity. To further explore the relationship between song repertoire and nest location, I used an analysis of variance (ANOVA) to test whether the number of multi-note songs (i.e., songs with phrases composed of more than two notes) in a bird's repertoire predicted nest distance from the forest edge, as well as a Pearson correlation to compare repertoire and nest location. Lastly, to test whether repertoire size was correlated with the use of complex songs, I used an LMM with repertoire size as the response variable, presence of complex songs as a fixed effect (binary), and bird identity as a random effect.

To determine whether immigration status influenced singing behavior, I used four separate ANOVAs to test whether residents, immigrants, and dispersers differed in acoustic dissimilarity from the local population, nest distance from forest edge, repertoire size, and number of complex songs used. Birds with unknown immigration status were excluded from this analysis. Lastly, I investigated the relationship between repertoire composition and breeding success by using two separate Pearson correlation tests to test whether acoustic dissimilarity was correlated with clutch size or the onset of egg laying. Clutch size was defined as the number of eggs that hatched in a bird's nest box, and onset of egg laying was calculated as days from April 1 on which the first egg hatched. To avoid sampling birds multiple times, I used records from only the first year in which a bird was included in the sample.

All statistical analyses were carried out in R (R Core Team, 2018) using the packages *PCDimension* (Coombes et al. 2019) and *lmerTest* (Kuznetsova et al. 2015).

RESULTS

I found that I found that there were 37 distinct song types used in the study population during 2017-2019 (mean \pm SE song per year: 30.66 ± 4.48 ; see Appendix C for further details). Birds in the sample had repertoire sizes of 3.63 ± 0.32 (mean \pm SE). In all years, a small number of song types were used by several birds and many song types were used by only one or two individuals (Figure 3a). The relative abundance of particular song types changed between years, meaning that songs used by many birds in a given year were not necessarily used by many birds in the subsequent year. However, the distribution of relative abundance of all song types used in the population did not change significantly between years and did not differ from a null distribution (Chi-squared test: 2017: $\chi^2(32) = 0.011$, $p = 1$, 2018: $\chi^2(26) = 0.08$, $p = 1$, 2019: $\chi^2(21) = 0.056$, $p = 1$; Figure 3b), and therefore certain song types were not more or less abundant than would be expected by chance. Nine of the 37 songs observed in the population were complex songs that were composed of phrases with more than two notes.

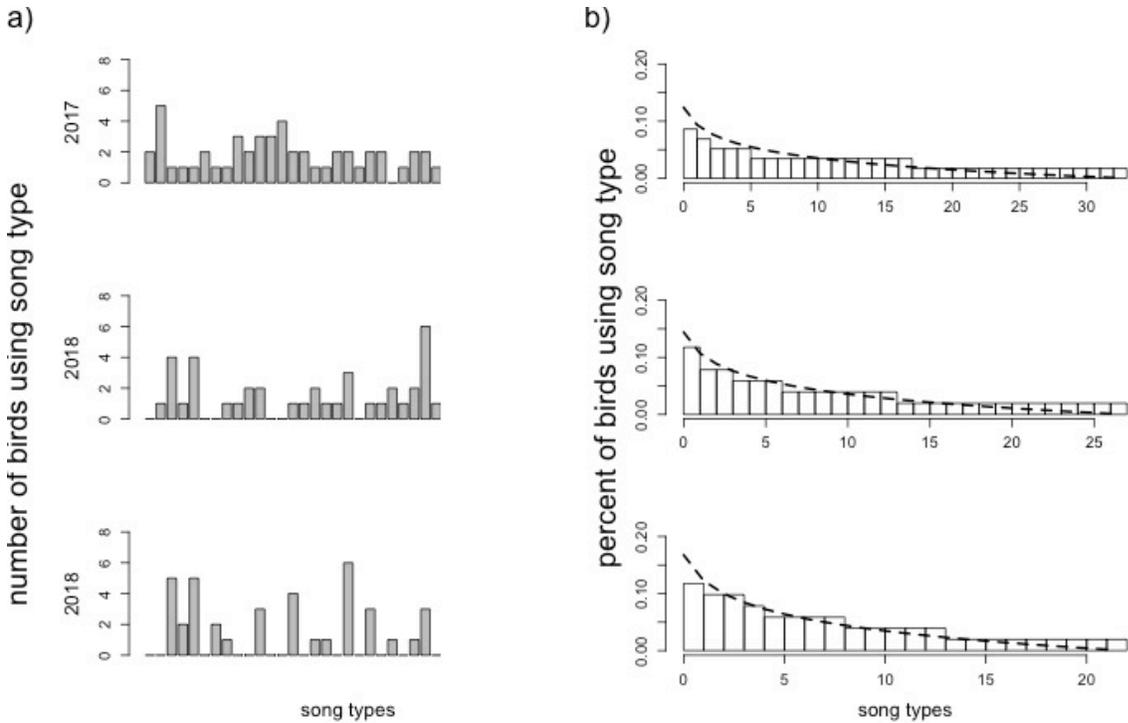


Figure 3. Song types used in study system during 2017-2019. The analysis found $N=37$ unique song types. Plots show a) the number of individuals using each song type per year and b) the proportion of birds using each song type, with song types ordered from most to least abundant within each year. The dashed line in (b) indicates the expected abundance of song types using a null distribution as predicted by a broken stick model.

In both 2018 and 2019 novel song types appeared and previously used song types disappeared (Table 1). In total, there were 14 song types that “disappeared” between years and, in all cases, these were previously used by birds that were not resampled in the subsequent year. In 2018 and 2019, there were three song types that appeared in the sample that were not recorded in the previous year, and in two cases these songs were used by previously unsampled birds. Thus, in almost all cases, song turnover between years (i.e., appearance or disappearance of song types) was linked to bird turnover. Although nine of the 48 song types recorded in this population 40 years previously by McGregor and Krebs (1982) persist today, the majority of previously recorded songs (39 song types) were not observed in

this study. Among the nine song types that were recorded previously, seven of these were songs composed of two-note phrases and two were complex songs. None of the nine songs that were found in both my study and the earlier study were previously classified as “rare” songs (McGregor and Krebs 1982).

Table 1. Summary of birds and song types included in sample each year. The letters R, I, D, and U indicate residents, immigrants, dispersers, and unknown individuals, respectively.

Year	Unique song types	Total birds	R	I	D	U	Song types appeared in current year	Song types disappeared from previous year
2017	33	21	13	4	2	2	-	-
2018	27	16	9	1	0	6	1	7
2019	22	17	10	6	1	0	2	7

I did not find that the number of birds using a song type in the previous year was a significant predictor of song presence in the following year, though the trend was in the predicted direction (LMM: $t(69.13) = 1.72$, $N = 72$, $p = 0.09$; Figure 4a). However, the number of birds using a song type in the previous year was positively correlated with the number of birds using that song type in the current year (Pearson correlation: $r = 2.44$, $N = 74$, $p = 0.017$; Figure 4b). Neighbors were significantly more likely to have higher levels of repertoire similarity and share song types than non-neighbors (LMM: repertoire similarity: $t =$

3.1, $N = 54$, $df = 54.91$, $p = 0.003$, song sharing: $t = 3.15$, $N = 320$, $df = 258.76$, $p = 0.002$;
 Figure 5).

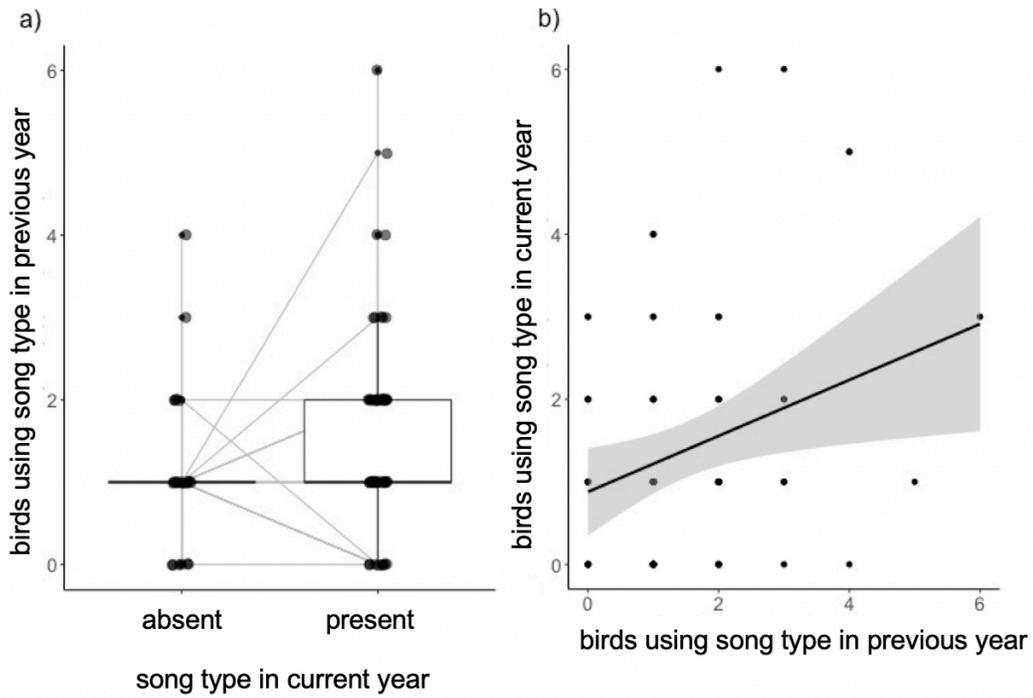


Figure 4. Song carryover between years. a) Song types present in the current year were used by more birds in the previous year, but this difference was not significant ($t = 1.72$, $p = 0.09$), b) there is a significant correlation between the number of birds using a song type from one year to the next ($r = 2.44$, $p = 0.017$).

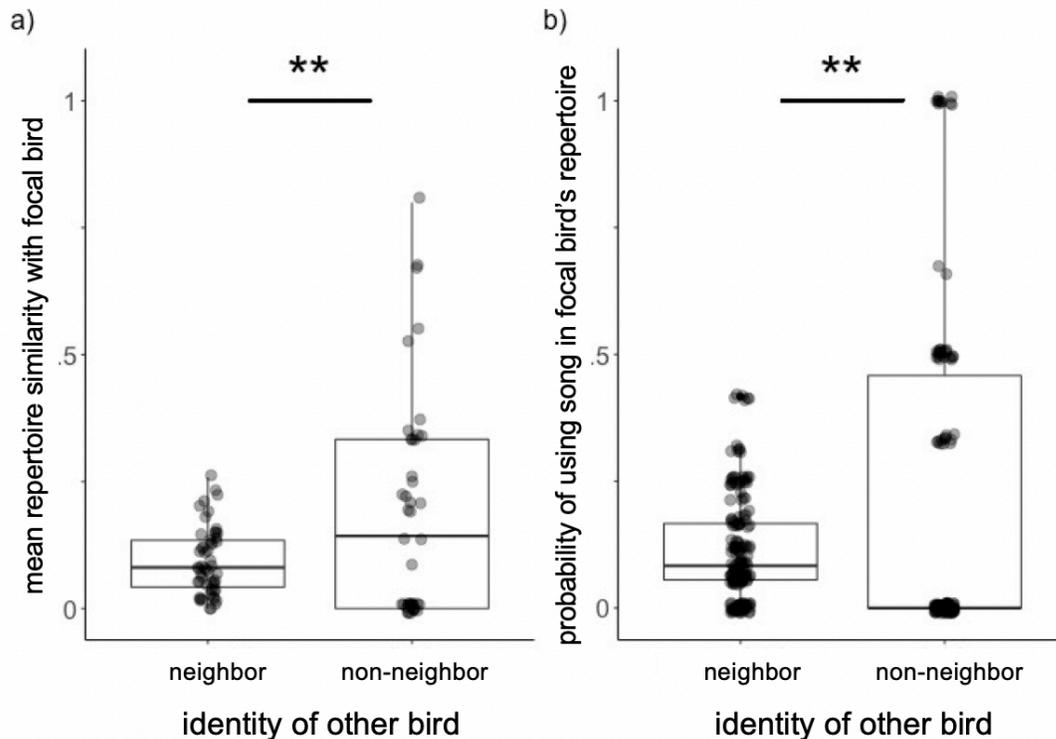


Figure 5. a) Vocal similarity between neighbors and non-neighbors. Neighbors were significantly more likely to have a) higher pairwise repertoire similarity ($t = 3.1$, $p = 0.003$), and b) share song types ($t = 3.15$, $p = 0.002$) than non-neighbors.

Birds that occupied breeding territories closer to forest edges used songs that were more dissimilar to the rest of the study population, as shown by several metrics. First, birds with nests closer to forest edges were closer to the edge of the acoustic feature space (Pearson correlation: $r = -0.31$, $N = 54$, $p = 0.025$; Figure 6a). Additionally, birds closer to forest edges had more complex song types in their repertoires (ANOVA: $F(1) = 4.87$, $p = 0.033$; Figure 6b) and had larger repertoires (Pearson correlation: $r = -2.48$, $N = 54$, $p = 0.016$; Figure 7a). Birds with larger repertoires were also more likely to use complex song type (t-test: $t = 2.55$, $df = 49.78$, $p = 0.014$; Figure 7b). Only one bird in the sample used two complex song types,

and this individual had a repertoire size of six songs types. Plots showing differences in repertoire size among territories are shown in Figure C1.

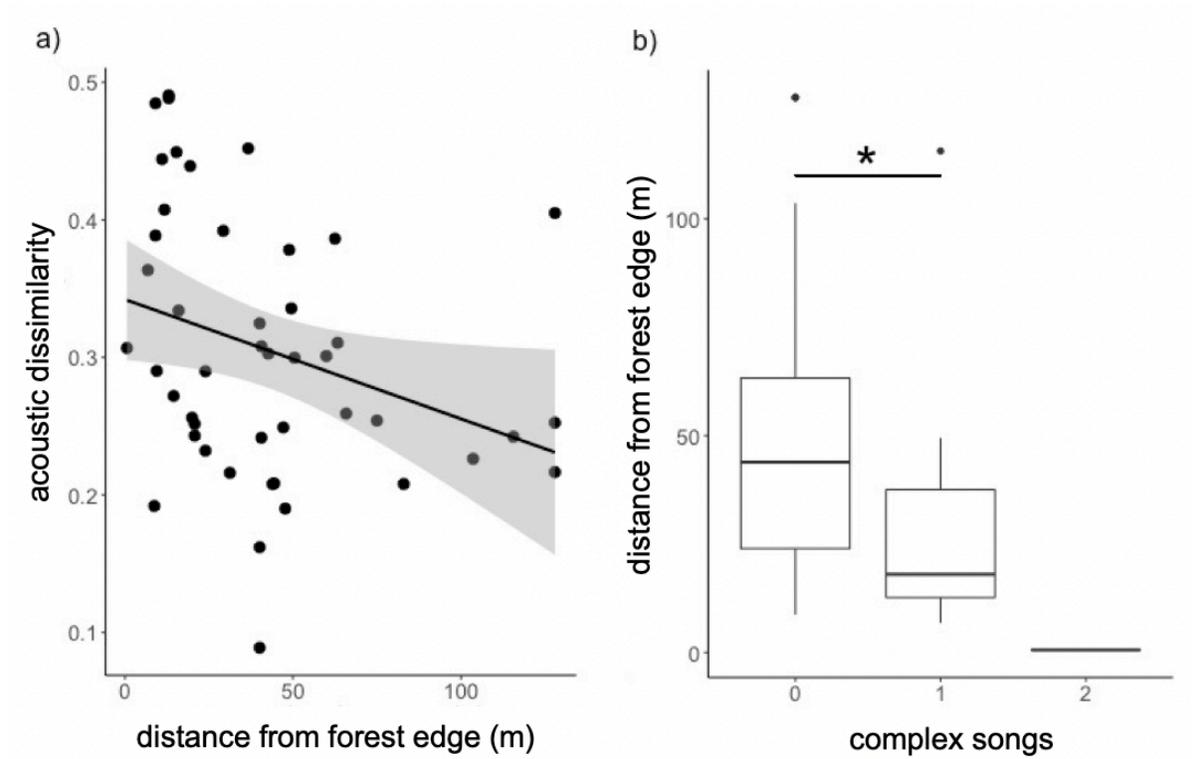


Figure 6. Birds on forest edges sound more dissimilar to other birds in the population. Birds nearer to edges a) had higher acoustic dissimilarity ($r = -0.31$, $p = 0.025$), and b) used more complex song types ($F = 4.87$, $p = 0.033$).

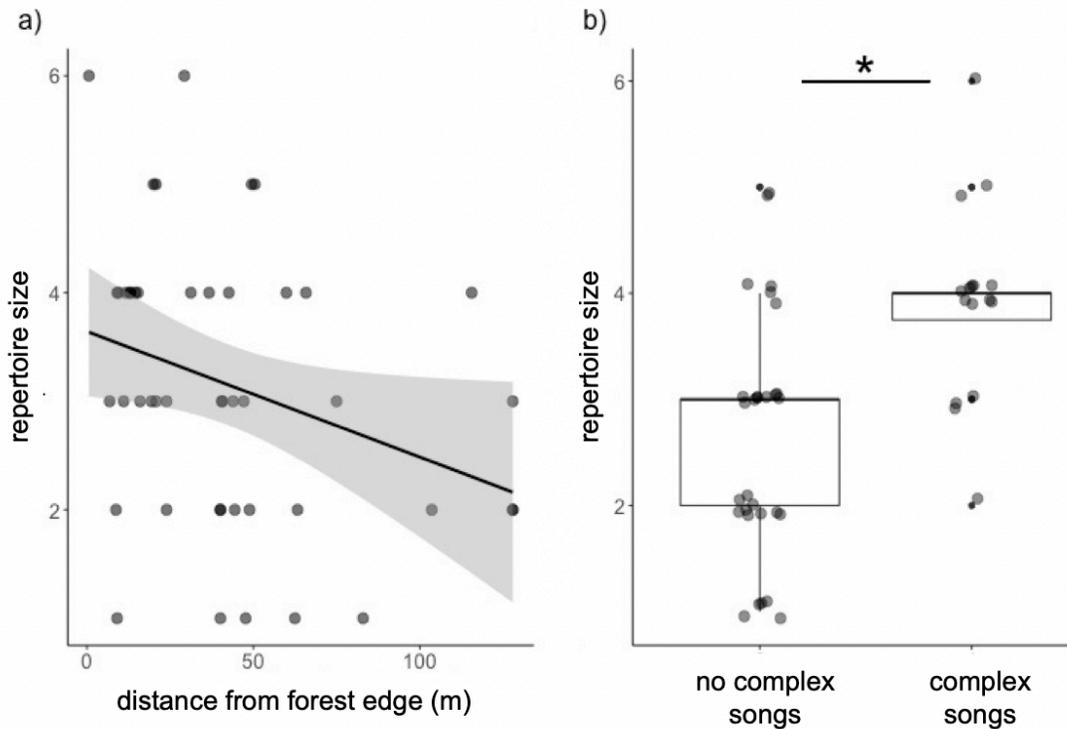


Figure 7. Repertoire size is correlated with nest location and repertoire composition. Birds with larger repertoires a) had nests closer to forest edges ($r = -2.48$, $p = 0.016$), and b) were more likely to use a complex song type ($t = 2.55$, $p = 0.014$).

Acoustic dissimilarity was not higher among either residents, immigrants, or dispersers (ANOVA: $F(2,43) = 0.81$, $p = 0.45$; Figure 8a), nor was any class more likely to breed closer to the forest edge (ANOVA: $F(2,43) = 2.05$, $p = 0.15$; Figure 8b). Although immigrants and dispersers tended to have larger repertoires than resident birds, this difference was not significant (ANOVA: $F(2,43) = 2.03$, $p = 0.15$; Tukey posthoc tests: residents vs. immigrants: $p = 0.15$; residents vs. dispersers: $p = 0.64$; immigrants vs. disperser: $p = 0.98$; Figure 8c). However, I did find that immigrants were more likely than residents to use complex song types (ANOVA: $F(2,43) = 3.49$, $p = 0.039$; Tukey posthoc tests: residents vs. immigrants: $p = 0.044$; residents vs. dispersers: $p = 0.44$; immigrants vs. disperser: $p = 0.98$;

Figure 8d). Plots of immigrant, resident, and disperser occupancy of territories in all years are shown in Figure C2.

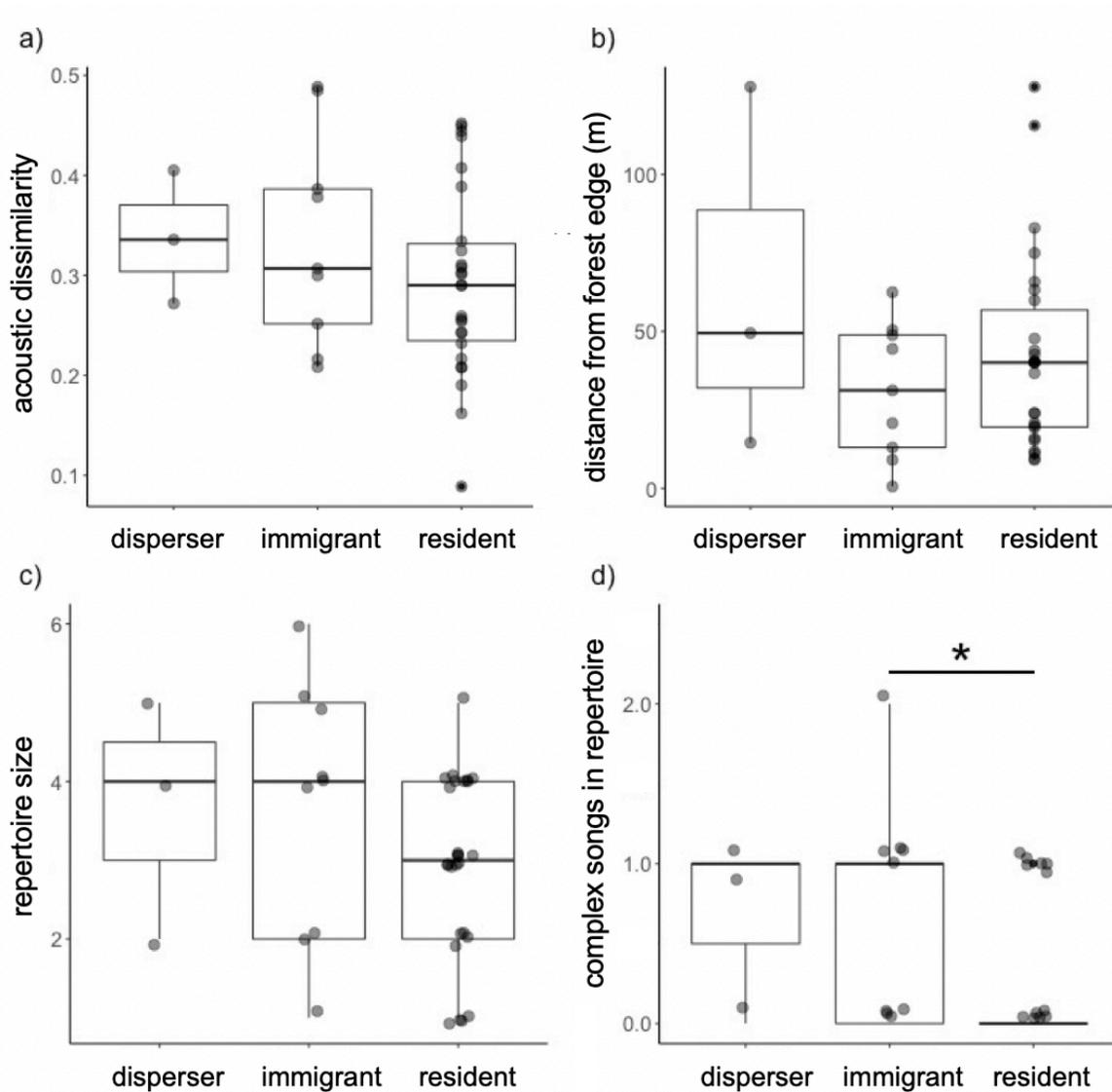


Figure 8. Comparisons between residents, immigrants, and dispersers. Immigration status is not a significant predictor of a) acoustic dissimilarity ($F = 0.81$, $p = 0.45$), b) nest distance from forest edge ($F = 2.05$, $p = 0.15$), or c) repertoire size ($F = 2.03$, $p = 0.15$), but d) immigrants are more likely than residents to use complex song types ($p = 0.044$).

Lastly, I observed that males with higher levels of acoustic dissimilarity had larger clutches (Pearson correlation: $r = 2.34$, $N = 45$, $p = 0.025$; Figure 9a) and mates that began egg laying earlier (Pearson correlation: $r = -2.0$, $N = 46$, $p = 0.05$; Figure 9b). However, when analyzing acoustic similarity versus lay date separately for each year, this was significant in only 2017 (2017: $r = -2.36$, $N = 21$, $p = 0.029$, 2018: $r = -2.84$, $N = 10$, $p = 0.43$; 2019: $r = 0.34$, $N = 15$, $p = 0.74$).

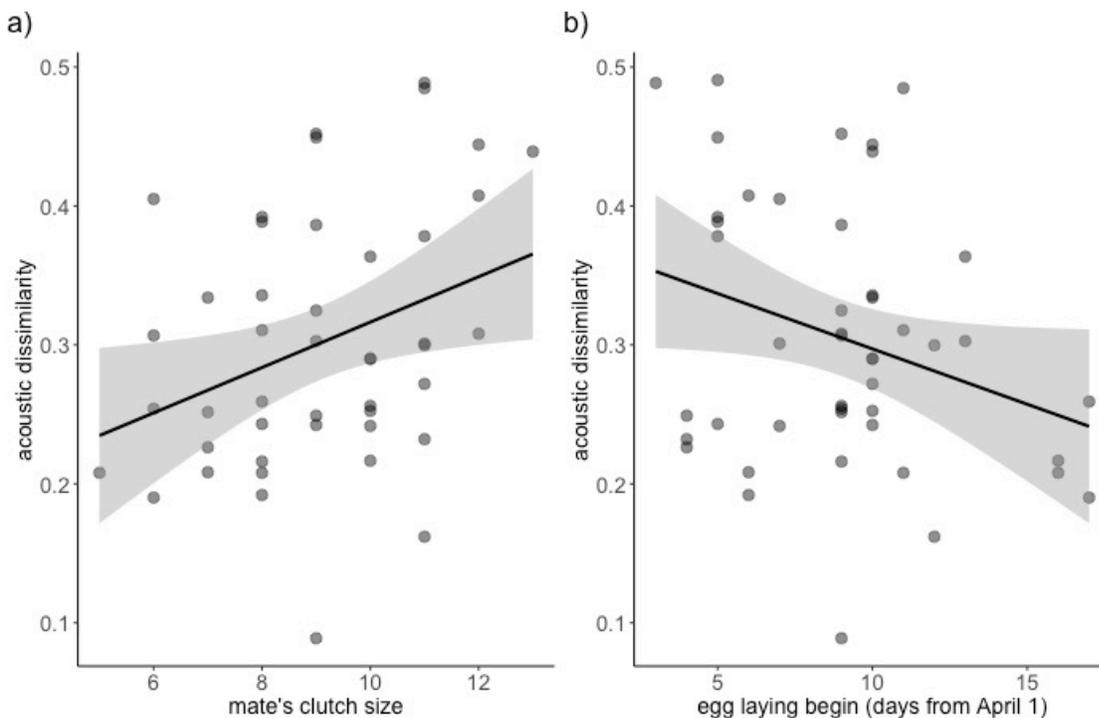


Figure 9. Acoustic dissimilarity may correlate with breeding success. Males that were more acoustically dissimilar from the local population (y-axis) had mates that a) had larger clutches Pearson correlation: $r = 2.34$, $p = 0.025$), and b) begin egg laying earlier ($r = -2.0$, $p = 0.05$), though when analyzed separately this relationship was only significant in one year of the study (2017: $p = 0.03$, 2018: $p = 0.16$, 2019: $p = 0.8$).

DISCUSSION

This study set out to evaluate how the songs used within a population vary over time and space and to identify the factors that might underlie this variation. I found that both the presence and abundance of unique song types changed every year, and that this appears to be linked to arrivals and deaths of birds in the study population. In accordance with previous studies of great tits, I found high levels of acoustic similarity between neighboring birds, but also noted that birds breeding near forest edges were more dissimilar to the population than centrally breeding birds. Additionally, although immigrants more often used complex song types, this alone did not explain the tendency for complex songs to be found near forest edges. Lastly, I observed that more acoustically dissimilar birds had larger clutches, but caution against over interpretation of this finding. I consider potential explanations for these results and discuss the implications of these findings below.

Relative abundance of song types. I find no evidence of particular song types being more or less common than expected by chance in my study population. This pattern was also reported by a similarly designed study of chaffinches (Slater et al. 1980) as well as an earlier study of great tits conducted in a different region in Wytham Woods (McGregor and Krebs 1982). The observation that song type abundance fits a null distribution suggests that birds do not obtain a selective advantage by using particular songs. However, certain conditions increase the likelihood of song types appearing in a given year, namely the number of birds using that song in the previous year, although even more widely used songs were found to disappear between years (Figure 4).

Several factors may help to explain the observed levels of temporal variability in songs. One likely driver of this variability is the demography of the study population. Great tit nests in Wytham Woods experience high levels of predation and offspring mortality, and an average of only one offspring per breeding pair goes on to breed in the population in the subsequent year (McCleery et al. 2004). Additionally, immigrants typically account for approximately one third of the population in the study area (Fayet et al. 2014), and 23% of birds in my sample were immigrants or dispersers. Although great tits can acquire songs after dispersing to breeding territories (Franco and Slabbekoorn 2009), previous work suggests that immigrants more often use rare or more dissimilar songs from local birds (McGregor and Krebs 1982, Fayet et al. 2014). Given the high levels of immigration and individual turnover and the low levels of offspring survival in this system, the likelihood of a meme surviving from one breeding season to the next may be relatively low.

It may also be possible that the process of song learning in great tits contributes to the observed levels of temporal variability. Previous work has suggested that songs may be acquired through a process of overproduction and selective attrition, meaning that song types might be present in a population but not observed, as they are not part of a bird's current functional repertoire (Franco and Slabbekoorn 2009). In this case, the interaction between birds' repertoires and their current social environment may prevent us from knowing which latent songs are present in the population. However, because songs nearly always appeared and disappeared from my sample with the appearance and disappearance individuals, my data do not support this hypothesis.

Although approximately 25% of the song types recorded in this study were also found in this system 40 years previously, the lack of continuous data collection makes it impossible

to conclude that these song types have persisted continuously over time. Regardless of whether this fraction of songs has remained stable in the population, the majority of songs used in this system previously have disappeared. This appears to be in contrast to vocal learning species that exhibit stable dialects, such as swamp sparrows (Lachlan et al. 2018) and white-crowned sparrows (Nelson et al. 2004), although methodological differences (e.g., measuring syllables rather than song types) might account for these differences. The songs of several other species have been shown to change over time, particularly when evaluated over several years or decades (e.g., chaffinches, Ince et al. 1980, budgerigars, Farabaugh et al. 1994, sparrows, Kopuchian et al. 2004, and chickadees, Baker and Gammon 2006).

Interestingly, all bird in this study used at least one variant of a two-note song, which may be linked to the ubiquity of these song types across the species distribution. It is also possible that there is some positive feedback in the high abundance of these songs, and their use for territory defense (Krebs et al. 1981) that might contribute to their persistence. Additionally, because great tits have been shown to favor songs which best transmit through their environments (Hunter and Krebs 1979), it is possible that songs composed of two-note phrases transmit more effectively in Wytham Woods, and are therefore more widely used.

Spatial distribution of song types and the effect of immigrant status. Birds that occupied territories closer to forest edges also had larger repertoires and exhibited higher levels of acoustic dissimilarity. In other words, edge birds used more song types, and often their songs were dissimilar to those of other birds in the population. Previous studies have found that immigrant birds more often nest near forest edges in this system (Wilkin et al. 2007b), and that immigrants often share fewer songs with the local population (McGregor and Krebs

1982). Therefore, one might expect that the finding of dissimilar songs being used near forest edges could be explained by immigrants more often breeding on edge territories. However, although immigrant birds more often used complex song types than residents, birds that bred closer to edges had higher acoustic dissimilarity regardless of immigration status (Figure 8). A possible explanation for this is that birds occupying edge territories have fewer neighbors, and thus birds with dissimilar songs may fare better on edges as they maintain territory boundaries with, and therefore may counter sing with, fewer individuals.

Why were immigrant birds more likely than residents to use complex song types? A likely explanation is that immigrants acquired these songs in their natal territories, and that birds born in Wytham were not exposed to complex songs during the sensitive period of song learning. It might also be possible that immigrants dispersed from habitats with different acoustic properties in which complex songs transmit more efficiently. Regardless of the drivers of differences in immigrant songs, my results support findings from previous studies suggesting that movement between populations is a driver of local song diversity (Fayet et al. 2014).

Past studies have shown that repertoire similarity decreases with distance in great tits (Rivera-Gutierrez et al. 2010a), and my findings of high levels of song sharing between neighbors further corroborate this work. In many species, nearby conspecifics converge upon similar vocalizations, and this may be influenced by several factors including dispersal tendencies, vocal learning process, and territoriality (Slabbekoorn and Smith 2002, Podos and Warren 2007). These drivers of vocal convergence may help to explain why high levels of acoustically similarity persist in this population despite yearly turnover in songs.

Repertoire size, composition, and reproductive success. Intriguingly, I found that males with higher levels of acoustic dissimilarity had female mates that began egg laying earlier and produced larger clutches. The correlation between onset of egg laying and clutch size has been shown previously in this population (Perrins and McCleery 1989), though it is unclear whether using dissimilar songs enables males to attract mates and claim territories earlier in the season. If using dissimilar songs indeed enabled such an advantage, acoustically dissimilar males might be expected to begin breeding sooner. However, these results must be interpreted with caution for several reasons. First, female immigration status has been shown to correlate with earlier egg laying (Wilkin et al. 2007b). Additionally, given the role of song sharing in male-male competition, further studies are needed to better understand the fitness consequences of acoustically dissimilarity. For example, in song sparrows, song sharing with neighbors has been shown to correlate with the amount of time a male holds a territory, and song sharing is known to play an important role in intra-specific competition in great tits (McGregor et al. 1992, Falls et al. 1982, Peake et al. 2005). Lastly, it is not possible to predict the direction of causality, i.e., whether using dissimilar songs enables birds to claim territories sooner or more easily attract mates, or whether birds of higher quality can learn more different songs because they are healthier or live longer.

Conclusions. I found annual changes in the presence and abundance of songs used in the study population, and suggest that this temporal variation can be only partially explained by song usage in the preceding year. I also found high levels of vocal similarity between neighbors and observed that both birds near forest edges and immigrants had higher levels of acoustically dissimilarity. Together, these findings support previous work showing that

songbirds do not exhibit a static tendency to use particular song types, and suggest that territoriality, habitat characteristics, and immigration may contribute to spatial and temporal variation in songs.

WORKS CITED

- Alexander, H. G. (1935). A chart of bird song. *British Birds*, 29, 190-198.
- Aurenhammer, F. (1991). Voronoi diagrams: a survey of a fundamental geometric data structure. *Computing Surveys*, 23: 345-405.
- Baker, M. C. and Cunningham, M. A. (1985). The biology of bird-song dialects. *Behavioral and Brain Sciences*, 8: 85-100.
- Baker, M. C., and Gammon, D. E. (2006). Persistence and change of vocal signals in natural populations of chickadees: annual sampling of the gargle call over eight seasons. *Behaviour*, 1473-1509.
- Beecher, M. D., Campbell, S. E., and Nordby, J. C. (2000). Territory tenure in song sparrows is related to song sharing with neighbours, but not to repertoire size. *Animal behaviour*, 59: 29-37.
- Beecher, M. D., and Brenowitz, E. A. (2005). Functional aspects of song learning in songbirds. *Trends in Ecology and Evolution*, 20: 143-149.
- Bioacoustics Research Program. (2011). Raven Pro: interactive sound analysis software. Version 1.5. The Cornell Lab of Ornithology. Ithaca, NY.
- Bradbury J. W., and Vehrencamp S. L. (2011). *Principles of animal communication*. Sunderland, MA: Sinauer.
- Byers, B. E., and Kroodsma, D. E. (2009). Female mate choice and songbird song repertoires. *Animal Behaviour*, 77: 13-22.
- Catchpole, C. K., and Slater, P. J. (2008). *Bird song: biological themes and variations*. Cambridge University Press.
- Coombes, K. R., Wang, M., and Coombes, M. K. R. (2019). Package PC Dimension. *R package version 1.0*.
- 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.
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26: 297-302.
- Ellers, J., and Slabbekoorn, H. (2003). Song divergence and male dispersal among bird populations: a spatially explicit model testing the role of vocal learning. *Animal Behaviour*, 65: 671-681.

Falls, J. B., Krebs, J. R., and McGregor, P. K. (1982). Song matching in the great tit (*Parus major*): the effect of similarity and familiarity. *Animal Behaviour*, 30: 997-1009.

Farabaugh, S. M., Linzenbold, A., and Dooling, R. J. (1994). Vocal plasticity in Budgerigars (*Melopsittacus undulatus*): evidence for social factors in the learning of contact calls. *Journal of Comparative Psychology*, 108: 81.

Fayet, A. L., Tobias, J. A., Hintzen, R. E., and Seddon, N. (2014). Immigration and dispersal are key determinants of cultural diversity in a songbird population. *Behavioral ecology*, 25: 744-753.

Feeke, F. (1977). Colony-specific song in *Cacicus cela* (Icteridae, Aves): The password hypothesis. *Ardea* 65: 197–202.

Firth, J. A., and Sheldon, B. C. 2015. Experimental manipulation of avian social structure reveals segregation is carried over across contexts. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 282: 20142350.

Firth, J. A., and Sheldon, B. C. (2016). Social carry-over effects underpin trans-seasonally linked structure in a wild bird population. *Ecology letters*, 19: 1324-1332.

Franco, P., and Slabbekoorn, H. (2009). Repertoire size and composition in great tits: a flexibility test using playbacks. *Animal Behaviour*, 77: 261-269.

Freeman, B. G., and Montgomery, G. A. (2017). Using song playback experiments to measure species recognition between geographically isolated populations: A comparison with acoustic trait analyses. *The Auk: Ornithological Advances*, 134: 857-870.

Gil, D., and Gahr, M. (2002). The honesty of bird song: multiple constraints for multiple traits. *Trends in Ecology and Evolution*, 17: 133-141.

Hunter, M. L., and Krebs, J. R. (1979). Geographical variation in the song of the great tit (*Parus major*) in relation to ecological factors. *The Journal of Animal Ecology*, 759-785.

Ince, S. A., Slater, P. J. B., and Weismann, C. (1980). Changes with time in the songs of a population of chaffinches. *The Condor*, 82: 285-290.

Kopuchian, C., Lijtmaer, D. A., Tubaro, P. L., and Handford, P. (2004). Temporal stability and change in a microgeographical pattern of song variation in the rufous-collared sparrow. *Animal Behaviour*, 68: 551-559.

Krebs, J. R. (1976). Habituation and song repertoires in the great tit. *Behavioral Ecology and Sociobiology*, 1: 215-227.

Krebs, J. R. (1977). Song and territory in the great tit *Parus major*. In *Evolutionary ecology*, pp. 47-62. Macmillan Education UK.

Krebs, J. R., and Kroodsma, D. E. (1980). Repertoires and geographical variation in bird song. *Advances in the Study of Behavior*, 11: 143-177.

Krebs, J. R., Ashcroft, R., and Van Orsdol, K. (1981). Song matching in the Great Tit *Parus major* L. *Animal Behaviour*, 29: 918-923.

Kroodsma, D. E. (1977). Correlates of song organization among North American wrens. *The American Naturalist*, 995-1008.

Kroodsma, D. E., and Byers, B. E. (1991). The function (s) of bird song. *American Zoologist*, 31: 318-328.

Kroodsma, D. E. (2004). The diversity and plasticity of birdsong. *Nature's music: the science of birdsong*, pp. 108-131. Elsevier Academic Press: Amsterdam.

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version 2.0*.

Lachlan, R. F., and Slater, P. J. B. (2003). Song learning by chaffinches: how accurate, and from where?. *Animal Behaviour*, 65: 957-969.

Lachlan, R. F., and Servedio, M. R. (2004). Song learning accelerates allopatric speciation. *Evolution*, 58: 2049-2063.

Lachlan, R. F., Ratmann, O., and Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature communications*, 9: 1-9.

Lambrechts, M., and Dhondt, A. A. (1986). Male quality, reproduction, and survival in the great tit (*Parus major*). *Behavioral Ecology and Sociobiology*, 19: 57-63.

MacArthur, R. H. (1957). On the relative abundance of bird species. *Proceedings of the National Academy of Sciences of the United States of America*, 43: 293.

Macdougall-Shackleton, S. A. (1997). Sexual selection and the evolution of song repertoires. In *Current ornithology*, pp. 81-124. Springer, Boston, MA.

MacDougall-Shackleton, E. A., and MacDougall-Shackleton, S. A. (2001). Cultural and genetic evolution in mountain white-crowned sparrows: song dialects are associated with population structure. *Evolution*, 55: 2568-2575.

Mace, R. (1987). The dawn chorus in the great tit *Parus major* is directly related to female fertility. *Nature*, 330: 745-746.

Marler, P., and Tamura, M. (1962). Song “dialects” in three populations of white-crowned sparrows. *Condor*, 64: 368–377.

Marler, P., and Peters, S. (1982). Developmental overproduction and selective attrition: new processes in the epigenesis of birdsong. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, 15: 369-378.

McCleery, R. H., Pettifor, R. A., Armbruster, P., Meyer, K., Sheldon, B. C., and Perrins, C. M. (2004). Components of variance underlying fitness in a natural population of the great tit *Parus major*. *The American Naturalist*, 164: E62-E72.

McGregor, P. K., Krebs, J. R., and Perrins, C. M. (1981). Song repertoires and lifetime reproductive success in the great tit (*Parus major*). *The American Naturalist*, 149-159.

McGregor PK, Krebs JR. 1982. Song types in a population of great tits (*Parus major*): their distribution, abundance and acquisition by individuals. *Behaviour*, 79:126–152.

McGregor, P. K., and Krebs, J. R. (1989). Song learning in adult great tits (*Parus major*): effects of neighbours. *Behaviour*, 108: 139-159.

McGregor, P. K., Dabelsteen, T., Shepherd, M., and Pedersen, S. B. (1992). The signal value of matched singing in great tits: evidence from interactive playback experiments. *Animal Behaviour*, 43: 987-998.

Morton, E. S. (1975). Ecological sources of selection on avian sounds. *The American Naturalist*, 109: 17-34.

Naguib, M., Diehl, J., Van Oers, K., and Snijders, L. (2019). Repeatability of signalling traits in the avian dawn chorus. *Frontiers in zoology*, 16: 27.

Nelson, D. A. (1992). Song overproduction and selective attrition lead to song sharing in the field sparrow (*Spizella pusilla*). *Behavioral Ecology and Sociobiology*, 30: 415-424.

Nelson, D. A., and Marler, P. 1994. Selection-based learning in bird song development. *Proceedings of the National Academy of Sciences U.S.A*, 91: 10498–10501.

Nelson, D. A., Hallberg, K. I., and Soha, J. A. (2004). Cultural evolution of Puget sound white-crowned sparrow song dialects. *Ethology*, 110: 879-908.

Nordby, J. C., Campbell, S. E., and Beecher, M. D. (2007). Selective attrition and individual song repertoire development in song sparrows. *Animal Behaviour*, 74: 1413-1418.

Nowicki, S., Peters, S., and Podos, J. (1998). Song learning, early nutrition and sexual selection in songbirds. *American Zoologist*, 38: 179-190.

- Payne, R. B. (1981). Song learning and social interaction in indigo buntings. *Animal Behaviour*, 29: 688-697.
- Peake, T. M., Matessi, G., McGregor, P. K., and Dabelsteen, T. (2005). Song type matching, song type switching and eavesdropping in male great tits. *Animal Behaviour*, 69: 1063-1068.
- Perrins, C. M., and McCleery, R. H. (1989). Laying dates and clutch size in the great tit. *The Wilson Bulletin*, 236-253.
- Podos, J., and Warren, P. S. (2007). The evolution of geographic variation in birdsong. *Advances in the Study of Behavior*, 37: 403-458.
- Price, T. (2008). *Speciation in Birds*. Roberts and Company Publishers, Greenwood Village, Colorado.
- R Core Team (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria.
- Rivera-Gutierrez, H. F., Matthysen, E., Adriaensen, F., and Slabbekoorn, H. (2010a). Repertoire sharing and song similarity between great tit males decline with distance between forest fragments. *Ethology*, 116: 951-960.
- Rivera-Gutierrez, H. F., Pinxten, R., and Eens, M. (2010b). Multiple signals for multiple messages: great tit, *Parus major*, song signals age and survival. *Animal Behaviour*, 80: 451-459.
- Rivera-Gutierrez, H. F., Pinxten, R., and Eens, M. (2011). Difficulties when assessing birdsong learning programmes under field conditions: a re-evaluation of song repertoire flexibility in the great tit. *PloS one*, 6: e16003.
- Searcy, W. A. (1992). Song repertoire and mate choice in birds. *American Zoologist*, 32: 71-80.
- Searcy, W. A., and Andersson, M. (1986). Sexual selection and the evolution of song. *Annual Review of Ecology and Systematics*, 17: 507-533.
- Searcy, W. A., and Nowicki, S. (2005). *The evolution of animal communication: reliability and deception in signaling systems*. Princeton University Press.
- Slabbekoorn, H., and Smith, T. B. (2002). Bird song, ecology and speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357: 493-503.
- Slabbekoorn, H., and Peet, M. (2003). Ecology: Birds sing at a higher pitch in urban noise. *Nature*, 424: 267-267.

Slabbekoorn, H. (2004). *Singing in the wild: The ecology of birdsong*. In 'Nature's Music'. Eds P. Marler and H. Slabbekoorn. Elsevier Academic Press: Amsterdam.

Slabbekoorn, H., and den Boer-Visser, A. (2006). Cities change the songs of birds. *Current Biology*, 16: 2326-2331.

Slater, P. J. B. (1989). Bird song learning: Causes and consequences. *Ethology, Ecology and Evolution*. 1: 19-46.

Slater, P. J. B., Ince, S. A., and Colgan, P. W. (1980). Chaffinch song types: their frequencies in the population and distribution between repertoires of different individuals. *Behaviour*, 207-218.

Slater, P. J., and Lachlan, R. F. (2003). *Is innovation in bird song adaptive?* In: *Animal innovation*, ed. S. M. Reader & K. N. Laland, pp. 117-36. Oxford University Press.

Thomas, A. (2019). *RSPB Guide to Birdsong*. Bloomsbury Publishing.

Vehrencamp, S. L. (2001). Is song-type matching a conventional signal of aggressive intentions? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268: 1637-1642.

Wilkin, T. (2006). *Environmental effects on great tit life-histories*. Doctoral dissertation, University of Oxford.

Wilkin, T. A., Perrins, C. M., and Sheldon, B. C. (2007a). The use of GIS in estimating spatial variation in habitat quality: a case study of lay-date in the Great Tit (*Parus major*). *Ibis*, 149: 110-118.

Wilkin, T. A., Garant, D., Gosler, A. G., and Sheldon, B. C. (2007b). Edge effects in the great tit: analyses of long-term data with GIS techniques. *Conservation Biology*, 21: 1207-1217.

Wright, T. F., Rodriguez, A. M., and Fleischer, R. C. (2005). Vocal dialects, sex-biased dispersal, and microsatellite population structure in the parrot *Amazona auropalliata*. *Molecular Ecology* 14: 1197-1205.

Wright, T. F., and Dahlin, C. R. (2018). Vocal dialects in parrots: patterns and processes of cultural evolution. *Emu-Austral Ornithology*, 118: 50-66.

Zollinger, S. A., Slater, P. J., Nemeth, E., and Brumm, H. (2017). Higher songs of city birds may not be an individual response to noise. *Proceedings of the Royal Society B: Biological Sciences*, 284: 20170602.

APPENDIX A: SUPPLEMENTARY MATERIALS FOR CHAPTER 1

Social transmission of antipredator behavior. We also used the combined z-transformed values shown in Fig. A1 as the response variable in a linear mixed model that included playback (treatment vs. control) and species as fixed effects and individual identity as a random effect. We observed a significant effect of treatment and no significant difference between species (LMM: playback: $F = 6.23$, $df = 45.06$, $p = 0.016$; species: $F = 2.09$, $df = 15.34$, $p = 0.17$). These tests suggest that perhaps further experiments may show a significant effect social transmission within species, which was not detected in our experiment. Analyses were conducted using the lme4 package (Bates et al. 2015) in R (R Core Team 2018).

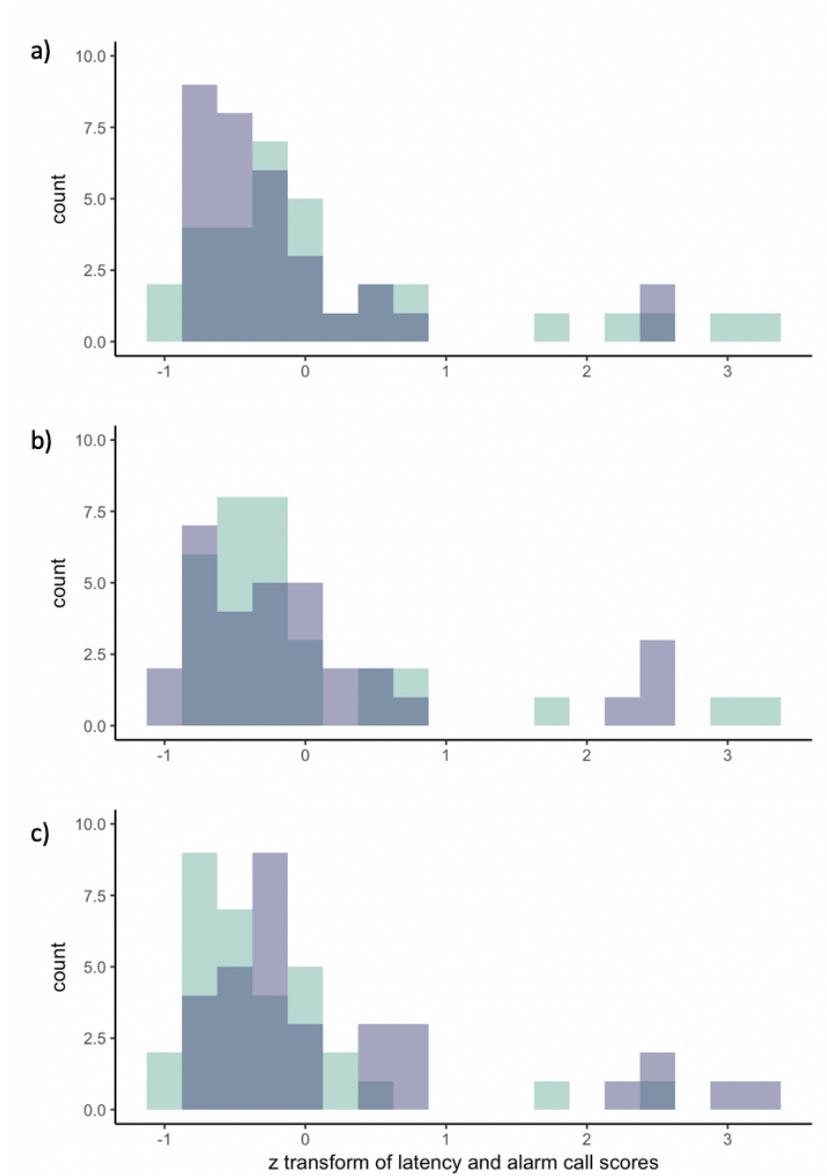


Figure A1. Histograms of separately z-transformed counts of number of alarm calls and latency to resume foraging for blue tit and great tit observers in five minutes following playbacks. Colours indicate distributions of counts for a) blue tit observers (dark blue) and great tit observers (light blue), b) alarm calls (light blue) and latency (dark blue), c) responses to control playbacks (light blue) and treatment playbacks (dark blue).

WORKS CITED

Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48

R Core Team. (2018). R: A language and environment for statistical computing. (R Foundation for Statistical Computing).

APPENDIX B: SUPPLEMENT FOR CHAPTER 2

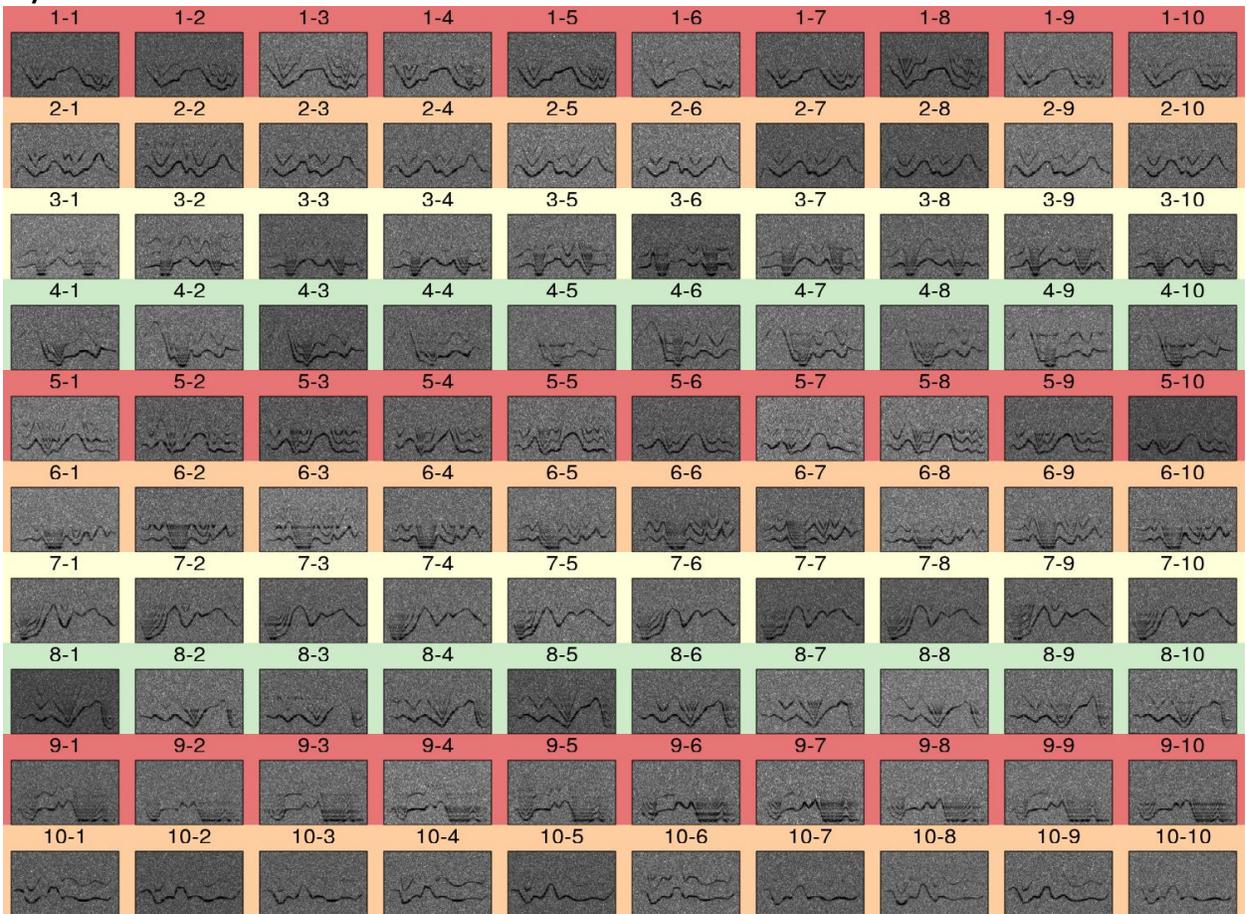
SUPPLEMENTARY METHODS

Data synthesis. Allowing for different levels of harmonic content made it possible to simulate recordings with low levels of signal attenuation, such as those collected at close range, as well as recordings with high levels of attenuation, which could be caused by environmental factors such as habitat type, as well as recording conditions. By using simulated data with known classes, we were able to make better predictions about which signal characteristics or recording conditions are likely to affect performance while also avoiding the time-consuming collection of data from live animals.

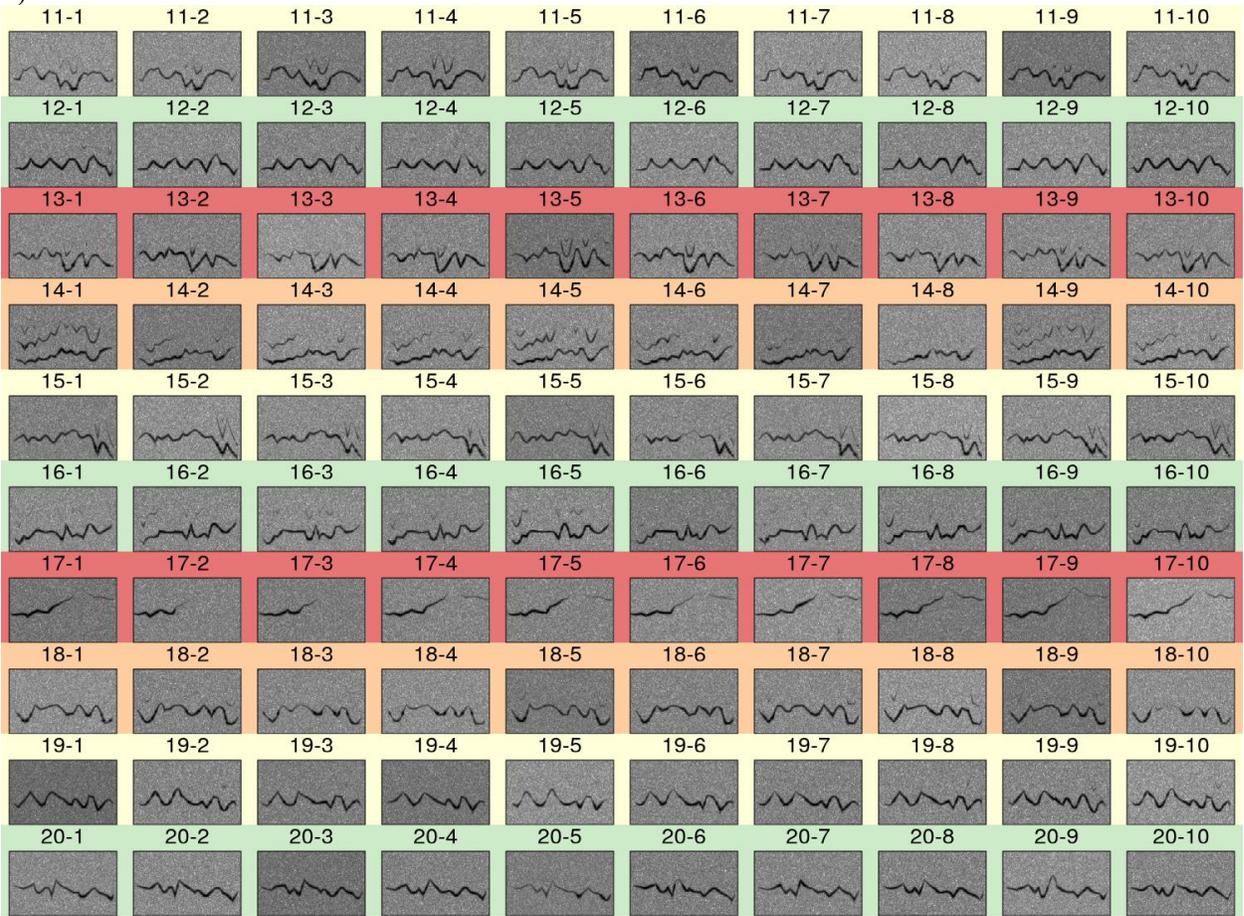
This approach of using synthetic data with known variation and class labels for every signal types is analogous to data augmentation in supervised machine learning. Data augmentation is a process in which labeled training data is slightly altered or modified in order to create additional annotated examples for training an algorithm, and is often employed when labeled data is scarce (Krizhevsky et al. 2012). Data augmentation has been shown to enhance performance of deep learning models in the classification of acoustic data (McFee et al. 2015, Salmon and Bello 2017). This approach may be particularly valuable when developing tools to help bioacoustics researchers in the analysis of field recordings because environmental conditions can alter acoustic structure in distinct ways through scattering, frequency-dependent attenuation and introduction of noise.

Figure B1. Spectrograms showing examples of signals in test datasets. a) synthetic budgerigar calls, b) synthetic long-billed hermit songs. Spectrograms in the same row show different synthetic signals that are considered to be the same element type.

a)



b)



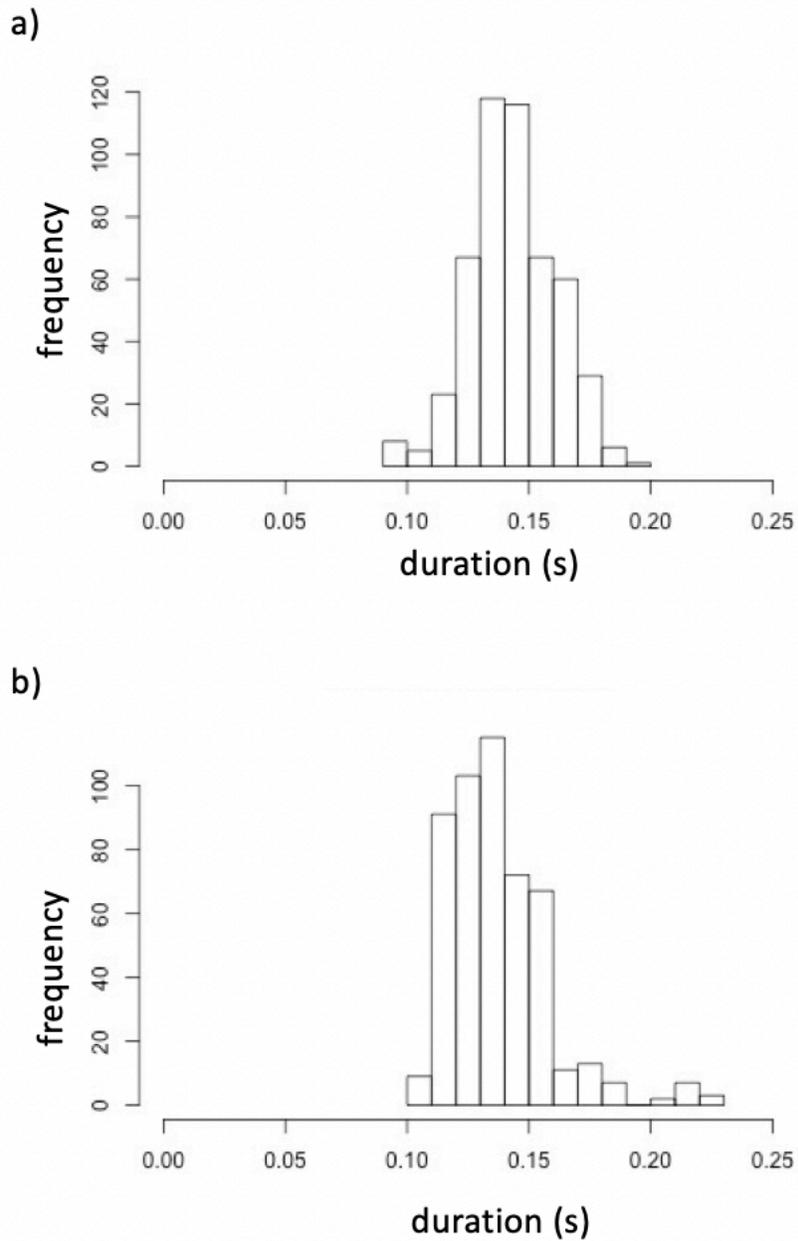


Figure B2. Histograms showing durations of a) field-recorded long-billed hermit songs, and b) lab-recorded budgerigar calls. Distributions of durations from live bird recordings were used to create synthetic datasets.

SUPPLEMENTARY RESULTS

To compare classification rates to those that would be expected by chance, we can calculate random chance of correct assignment as $1/c$, where c is the number of different classes. Note that to find statistical significance of observed correct classification rates versus those theoretically expected by chance one must adjust for a finite number of test datapoints (see Combrisson and Jerbi 2015). However, we use this value only as a point of reference for assessing supervised random forest performance. To evaluate the performance of our unsupervised method we use rigorous statistical testing.

Table B1. Variable importance rankings indicating which feature measurements were most useful in splitting data into distinct classes were different for each of the four dataset types used for testing. Variable rankings were produced by the separate unsupervised random forest analysis used for each data set. Rankings shown for synthetic data were randomly selected from a random forest model used for synthetic data sets with 100 unique elements. Variable names are listed as they are referred to by the R packages warbleR and seewave, and correspond to the feature measurements listed in the main text.

Variable ranking	Field-recorded long-billed hermit songs	Lab-recorded budgerigar calls	Synthetic long billed hermit songs	Synthetic budgerigar calls
1	var.cc23	xc.dim.1	min.cc12	max.cc13
2	var.cc16	freq.Q25	median.cc9	mean.cc24
3	var.cc24	median.cc2	kurt.cc21	kurt.cc25
4	var.cc15	freq.Q75	kurt.cc16	var.cc25
5	var.cc22	freq.median	var.cc4	var.cc8
6	median.cc4	dtw.dim.1	max.cc23	var.cc4
7	var.cc14	min.cc2	max.cc22	var.cc22
8	var.cc13	max.cc1	median.cc8	skew.cc2
9	var.cc11	median.cc6	mean.cc23	skew.cc22
10	sfm	mean.cc5	skew.cc8	kurt.cc20
11	entropy	median.cc3	kurt.cc1	kurt.cc23
12	median.cc3	var.cc5	kurt.cc19	kurt.cc21
13	median.cc5	var.cc6	var.cc22	skew.cc20
14	dtw.dim.1	median.cc7	skew.cc7	freq.IQR
15	var.cc9	min.cc6	max.cc21	time.Q25
16	kurt.cc15	max.cc5	min.cc9	maxdom
17	min.cc15	median.cc16	time.median	xc.dim.4
18	skew.cc4	xc.dim.2	dtw.dim.3	var.cc6
19	var.cc10	median.cc17	max.cc19	var.cc18
20	mean.cc6	max.cc2	kurt.cc11	var.cc19
21	kurt.cc14	time.ent	skew.cc22	var.cc5
22	mean.cc15	time.Q75	kurt.cc2	mean.cc22
23	freq.IQR	var.cc9	var.cc3	median.cc15
24	max.cc14	median.cc18	median.cc13	skew.cc19
25	skew.cc15	time.median	var.cc8	skew.cc25
26	var.cc25	median.cc15	kurt.cc20	var.cc15
27	max.cc13	kurt.cc6	mean.cc18	max.cc15
28	mean.cc14	median.cc5	var.cc25	max.cc20
29	max.cc16	median.cc8	max.cc24	median.cc13
30	skew.cc14	sfm	max.cc8	var.cc23
31	min.cc14	median.cc11	max.cc14	var.cc17
32	max.cc15	median.cc4	kurt.cc17	kurt.cc8
33	var.cc21	xc.dim.4	skew.cc6	var.cc20
34	min.cc10	max.cc3	skew.cc1	var.cc14
35	max.cc11	entropy	skew.cc10	mean.cc11

36	modindx	sd	var.cc16	median.cc7
37	skew.cc18	max.cc9	median.cc24	max.cc25
38	min.cc3	time.IQR	var.cc23	max.cc11
39	min.cc6	var.cc25	var.cc13	max.cc3
40	median.cc7	skew.cc1	max.cc20	min.cc11
41	skew.cc10	median.cc14	max.cc9	min.cc24
42	skew.cc16	min.cc16	max.cc15	min.cc15
43	kurt.cc16	var.cc1	xc.dim.4	max.cc4
44	max.cc3	median.cc19	min.cc10	max.cc5
45	max.cc5	var.cc24	max.cc3	min.cc19
46	time.ent	median.cc10	min.cc15	min.cc25
47	var.cc19	var.cc10	xc.dim.2	min.cc22
48	skew.cc13	var.cc8	sfm	min.cc18
49	skew.cc9	skew.cc6	dtw.dim.1	min.cc1
50	time.Q75	var.cc2	min.cc19	xc.dim.2
51	var.cc17	xc.dim.3	sp.ent	xc.dim.3
52	time.median	min.cc3	mindom	min.cc13
53	max.cc19	var.cc12	min.cc8	var.cc10
54	min.cc16	kurt.cc4	max.cc10	mean.cc23
55	var.cc8	var.cc4	median.cc2	kurt.cc19
56	mean.cc10	meanpeakf	mean.cc6	skew.cc13
57	max.cc10	min.cc1	var.cc24	var.cc13
58	var.cc18	var.cc3	median.cc22	median.cc19
59	mean.cc9	max.cc15	skew.cc20	median.cc9
60	kurt.cc11	skew.cc3	skew.cc18	median.cc10
61	max.cc23	skew.cc4	var.cc9	max.cc23
62	var.cc6	skew.cc5	skew.cc11	median.cc12
63	max.cc18	var.cc13	skew.cc3	median.cc8
64	kurt.cc13	var.cc7	max.cc25	max.cc16
65	kurt.cc7	skew.cc7	max.cc16	max.cc17
66	kurt.cc4	median.cc13	median.cc5	min.cc12
67	median.cc15	time.Q25	min.cc25	min.cc9
68	min.cc9	max.cc7	min.cc23	min.cc7
69	var.cc12	mean.cc12	min.cc21	min.cc21
70	skew.cc11	var.cc16	min.cc11	min.cc23
71	median.cc14	var.cc15	min.cc4	max.cc1
72	skew.cc3	kurt.cc7	dfrange	max.cc2
73	median.cc11	min.cc10	modindx	min.cc8
74	var.cc7	xc.dim.5	xc.dim.5	min.cc2

75	median.cc13	kurt.cc1	min.cc6	dtw.dim.1
76	min.cc4	dtw.dim.3	dtw.dim.5	dtw.dim.3
77	min.cc24	max.cc17	min.cc2	median.cc21
78	median.cc2	min.cc7	max.cc7	skew.cc10
79	min.cc5	median.cc20	var.cc14	kurt.cc17
80	median.cc18	min.cc11	kurt.cc14	kurt.cc16
81	skew.cc19	min.cc18	mean.d2.cc	skew.cc5
82	max.cc8	max.cc8	kurt.cc25	var.cc16
83	skew.cc17	max.cc12	kurt.cc3	mean.cc16
84	xc.dim.2	min.cc5	kurt.cc4	median.cc18
85	mean.cc8	dtw.dim.4	skew.cc17	median.cc6
86	var.cc20	min.cc4	var.cc17	max.cc12
87	skew.cc5	min.cc8	var.cc6	var.cc12
88	min.cc22	median.cc21	mean.cc11	median.cc17
89	skew.cc7	var.cc11	var.cc12	max.cc19
90	var.cc3	skew.cc8	skew.cc14	median.cc3
91	kurt.cc10	median.cc9	skew.cc5	skew.cc6
92	max.cc4	kurt.cc3	skew.cc24	var.cc2
93	min.cc18	skew.cc2	kurt.cc24	median.cc5
94	dtw.dim.2	kurt	kurt.cc23	skew.cc4
95	min.cc2	var.cc14	kurt.cc22	max.cc22
96	min.cc12	kurt.cc5	skew.cc23	max.cc18
97	median.cc21	modindx	skew.cc21	var.cc1
98	time.Q25	min.cc13	kurt.cc13	max.cc14
99	meanpeakf	freq.IQR	skew.cc25	var.cc7
100	mindom	max.cc6	kurt.cc18	skew.cc3
101	kurt.cc3	dtw.dim.2	kurt.cc9	kurt.cc13
102	median.cc23	dtw.dim.5	kurt.cc15	kurt.cc24
103	min.cc23	var.cc17	kurt.cc7	kurt.cc22
104	var.cc1	max.cc10	kurt.cc5	skew.cc15
105	startdom	var.cc23	kurt.cc12	var.cc21
106	kurt.cc23	kurt.cc25	skew.cc15	skew.cc7
107	min.cc7	max.cc11	kurt.cc10	skew.cc12
108	kurt.cc8	kurt.cc8	kurt.cc8	skew.cc11
109	max.cc24	var.cc22	skew.cc2	skew.cc23
110	kurt.cc22	median.cc22	skew.cc12	kurt.cc6
111	kurt.cc9	max.cc19	var.cc11	kurt.cc2
112	max.cc17	max.cc18	mean.cc10	skew.cc21
113	median.cc16	max.cc13	var.cc5	kurt.cc1

114	max.cc9	min.cc9	mean.cc4	skew.cc16
115	max.cc21	kurt.cc9	median.cc16	kurt.cc10
116	min.cc13	max.cc16	median.cc12	kurt.cc4
117	var.cc5	max.cc14	median.cc17	kurt.cc15
118	min.cc20	max.cc4	var.cc21	kurt.cc14
119	skew.cc6	min.cc14	var.cc15	kurt.cc5
120	median.cc12	kurt.cc24	var.cc18	skew.cc14
121	var.cc4	var.cc20	var.cc7	skew.cc18
122	dtw.dim.5	min.cc12	mean.cc7	skew.cc24
123	max.cc22	min.cc20	median.cc25	kurt.cc3
124	max.cc1	min.cc15	var.cc2	kurt.cc12
125	max.cc2	max.cc20	mean.cc2	kurt.cc9
126	skew.cc2	var.cc21	mean.cc15	kurt.cc7
127	median.cc24	median.cc23	median.cc21	kurt.cc18
128	skew.cc22	max.cc22	var.cc20	skew.cc17
129	min.cc21	dfslope	skew.cc4	skew.cc1
130	skew.cc21	min.cc21	skew.cc9	var.cc9
131	mean.cc25	min.cc19	skew.cc16	mean.cc15
132	var.cc2	var.cc19	kurt.cc6	max.cc24
133	maxdom	skew.cc10	skew.cc19	max.cc9
134	xc.dim.4	skew.cc11	skew.cc13	min.cc17
135	kurt	var.cc18	var.cc10	min.cc6
136	min.cc17	max.cc21	mean.cc19	meanpeakf
137	max.cc7	kurt.cc12	median.cc14	startdom
138	kurt.cc21	min.cc23	max.cc17	meandom
139	max.cc6	min.cc22	max.cc12	sfm
140	dfrange	max.cc24	max.cc1	time.Q75
141	min.cc19	mean.cc24	min.cc14	mindom
142	min.cc11	skew.cc9	meanpeakf	time.ent
143	median.cc22	kurt.cc2	kurt	kurt
144	dtw.dim.3	max.cc23	freq.IQR	time.median
145	median.cc19	min.cc25	freq.Q75	freq.Q25
146	skew.cc20	skew.cc23	duration	freq.median
147	median.cc20	min.cc17	sd	freq.Q75
148	dfslope	kurt.cc13	time.ent	xc.dim.1
149	kurt.cc6	maxdom	time.Q25	min.cc20
150	skew.cc1	median.cc25	min.cc5	median.cc20
151	xc.dim.1	kurt.cc11	min.cc7	var.cc24
152	max.cc25	skew.cc12	min.cc17	skew.cc8

153	kurt.cc24	min.cc24	max.cc5	var.cc11
154	skew.cc12	skew.cc15	min.cc22	median.cc14
155	min.cc8	skew.cc24	max.cc11	median.cc4
156	dtw.dim.4	max.cc25	max.cc4	max.cc21
157	skew.cc8	skew.cc19	max.cc6	max.cc10
158	skew.cc23	kurt.cc23	max.cc13	max.cc8
159	max.cc20	skew.cc16	min.cc20	max.cc7
160	kurt.cc12	skew.cc20	min.cc13	max.cc6
161	kurt.cc18	skew.cc14	min.cc18	min.cc10
162	median.cc17	skew.cc18	startdom	min.cc4
163	skew.cc24	skew.cc25	xc.dim.3	xc.dim.5
164	xc.dim.3	enddom	dtw.dim.4	dtw.dim.4
165	kurt.cc1	skew.cc17	xc.dim.1	dtw.dim.5
166	kurt.cc17	kurt.cc14	time.Q75	min.cc5
167	kurt.cc5	skew.cc21	maxdom	min.cc3
168	kurt.cc19	dfrange	enddom	dtw.dim.2
169	max.cc12	kurt.cc22	time.IQR	dfslope
170	min.cc25	skew.cc13	dfslope	modindx
171	xc.dim.5	kurt.cc10	dtw.dim.2	time.IQR
172	kurt.cc2	skew.cc22	min.cc16	enddom
173	skew.cc25	kurt.cc20	max.cc18	min.cc14
174	enddom	kurt.cc18	min.cc24	median.cc25
175	kurt.cc20	kurt.cc16	var.cc19	min.cc16
176	kurt.cc25	kurt.cc15	median.cc20	skew.cc9
177		kurt.cc21		kurt.cc11
178		kurt.cc17		var.cc3
179		mindom		elm.type
180		kurt.cc19		sd
181		startdom		duration

WORKS CITED

Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of neuroscience methods*, 250: 126-136.

Dalleau, K., Couceiro, M., & Smaïl-Tabbone, M. (2018). Unsupervised extremely randomized trees. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Springer, Cham, 2018.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097-1105.

McFee, B., E. Humphrey, and J. Bello. (2015). A software framework for musical data augmentation. In *16th International Society for Music Information Retrieval Conference*, pp. 248–254.

Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24: 279-283.

APPENDIX C: SUPPLEMENTARY MATERIALS FOR CHAPTER 4

The unsupervised algorithm which was used to assign songs to classes (see Chapter 2) found that optimal clustering occurred when using either 24, 37, or 54 classes of songs. In other words, when clustering similar songs together using either 24, 37, or 54 clusters, the algorithm was better able to maximize distance between clusters and minimize distance within clusters than when using other values. Because my analysis found that clustering accuracy was marginally better when using 37 song type clusters, I report the results using this number of song classes. However, it is critical to note that because the analysis also found that either 24 or 54 song classes offered comparable clustering accuracy, it is possible that the “true” number of song types in the study population may not be precisely 37 songs. Rather, this is the best approximation of the number of song types present given the inherent constraints of assigning continuous signals to discrete classes. Ultimately, the correct song type classifications are those that match birds’ perceptions of songs; here, I attempt to estimate those classifications using objective acoustic measurements. Importantly, the results shown in Chapter 4 do not change when using 24, 37, or 54 classes of song types. Thus, the results presented here accurately describe the variation in acoustic signals in the study population, regardless of the reported number of song types.

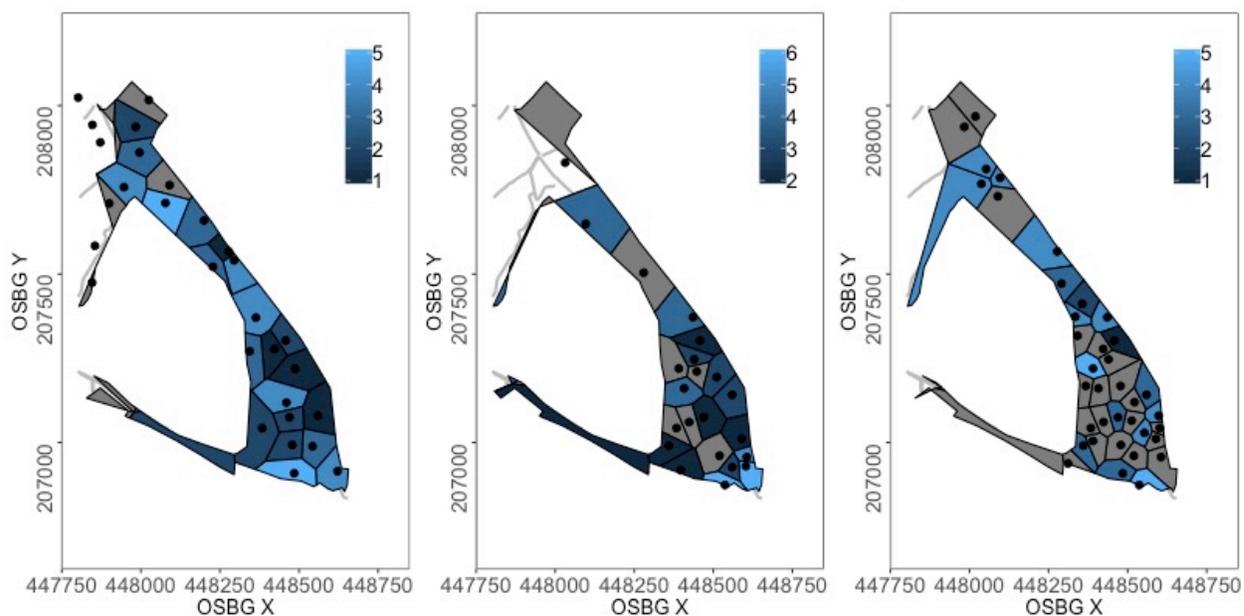
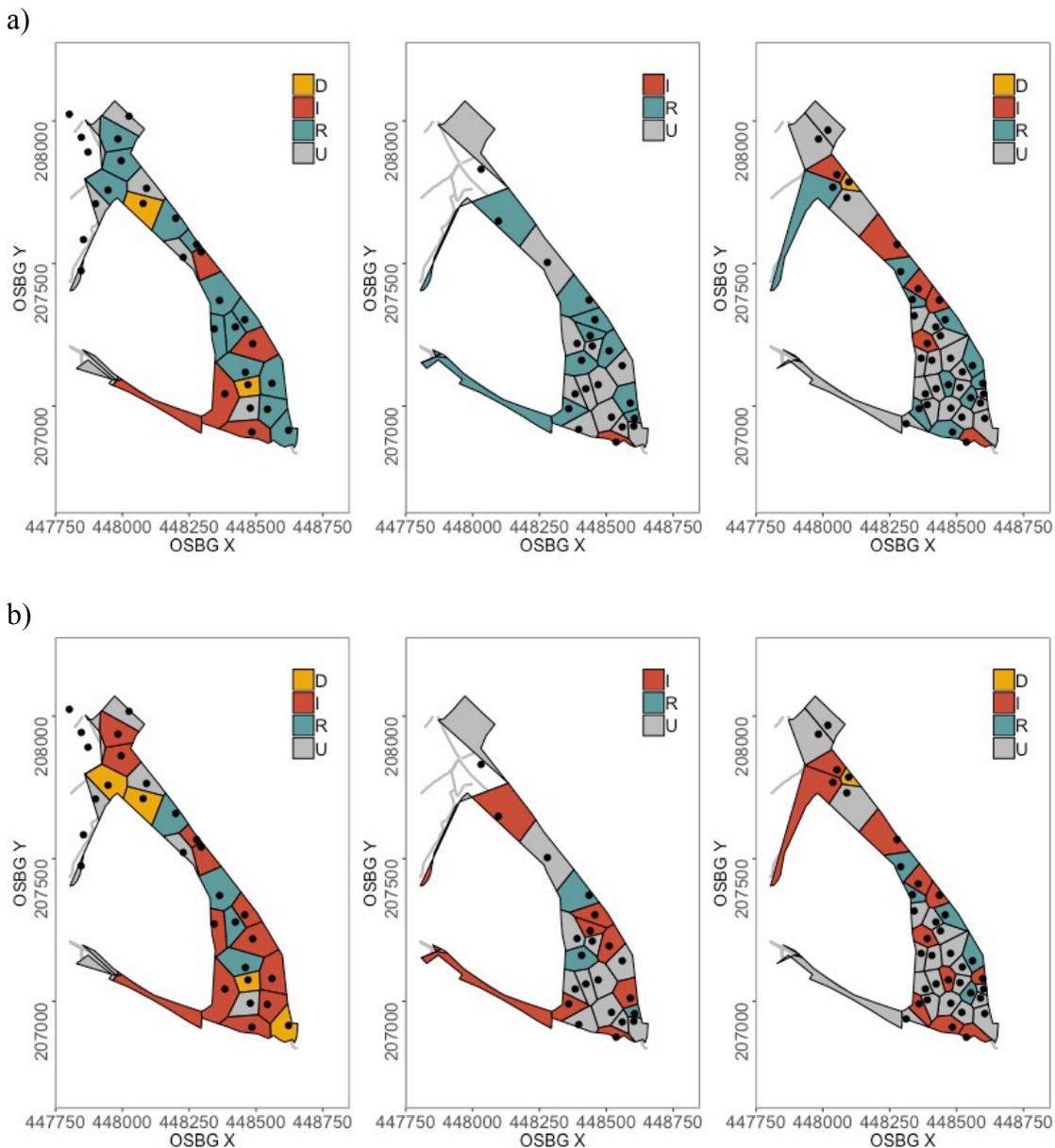


Figure C1. Repertoire size of birds occupying territories in the study area in 2017, 2018, 2019. Territories were calculated using Thiessen polygons. Grey polygons indicate territories of birds that were not sampled. Blue shading in polygons represents the repertoire size of the territory owner. White space surrounding territories is agricultural land that borders the study area. Axes indicate longitude and latitude coordinates used by the Ordnance Survey of Great Britain.



wFigure C2. a) Immigration status of birds occupying territories in the study area in 2017, 2018, 2019. b) Repertoire size of birds occupying territories in the study area in 2017, 2018, 2019. Territories were calculated using Thiessen polygons. Polygon shading represents immigration status, with residents, dispersers, immigrants, and unknown birds indicated by R, D, I, or U, respectively. In a), immigrants were defined as birds that were not born in Wytham and had not previously bred in Wytham (i.e., the definition used for the analysis in this study). In b), immigrants were defined as birds not born in Wytham but that may have previously bred within Wytham. White space surrounding territories is agricultural land that borders the study area. Axes indicate longitude and latitude coordinates used by the Ordnance Survey of Great Britain.