CULTURAL EVOLUTION IN THE WILD

TRACKING THE LANDSCAPE OF LEARNING IN BIRD SONG

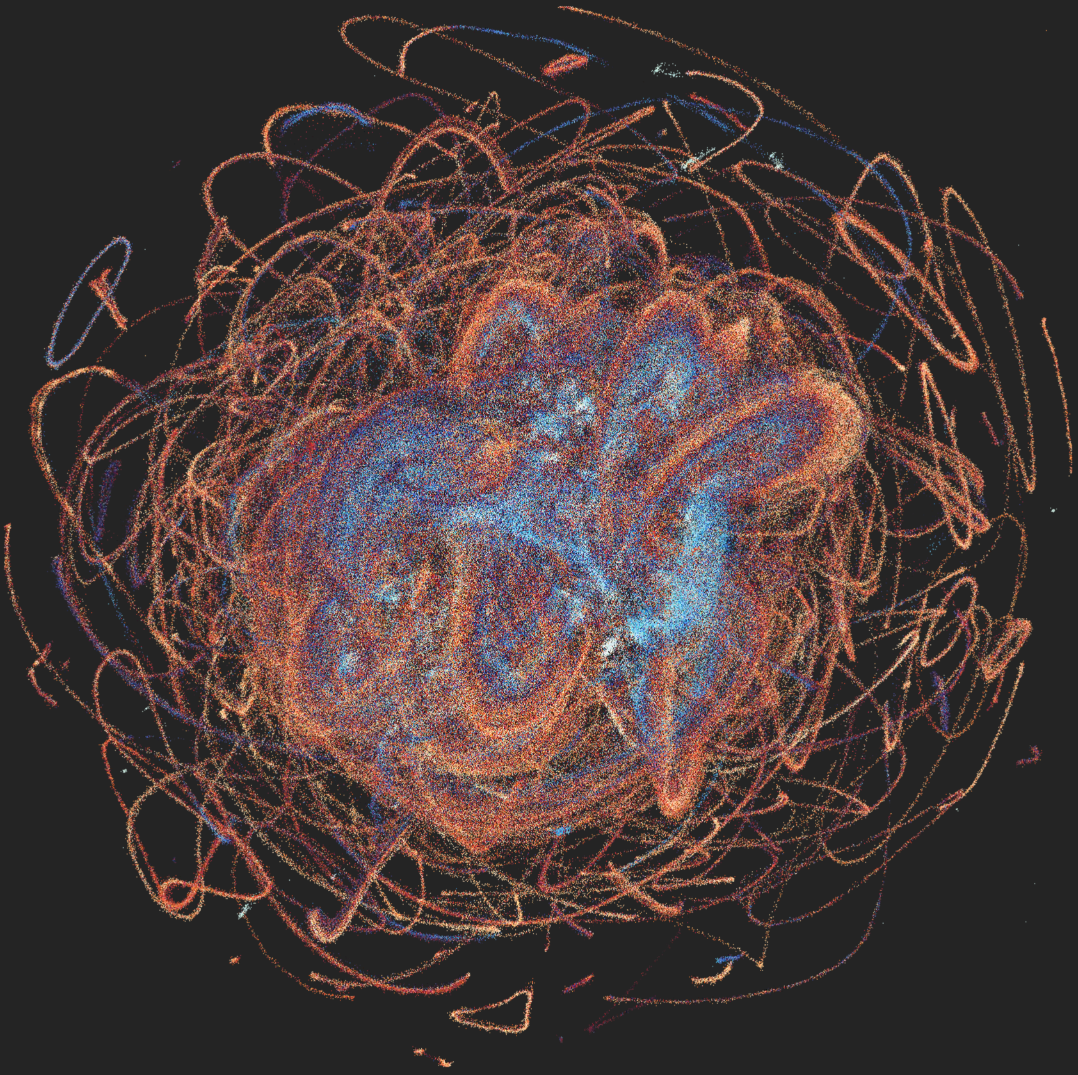
Nilo Merino Recalde

St Hilda's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Michaelmas 2023

DECLARATION

I declare that this thesis is entirely my own work, and except where otherwise stated, describes my own research. This work has not been submitted for any degree or professional qualification except as specified.



2D UMAP projection of song trajectories across acoustic space.
Sung by great tits in Wytham Woods, 2020

ACKNOWLEDGEMENTS

Spending nearly four years writing a thesis on bird song is undoubtedly one of the weirdest, most privileged things a human can do. And yet, somehow, it's still not easy: I couldn't have done it without the help of many people, so some thanks are in order.

Around five years ago, as I was completing my degree in Anthropology, I cold-emailed Ben, who would later become my supervisor: 'Hey, I'm interested in cultural change in bird song, can I come and record your birds?' (I'm paraphrasing, of course, it was more like 'Dear Professor Sheldon ...'), To my surprise, he generously provided me with a scientific home and the necessary infrastructure for my project. I'd like to think that he'd have second thoughts about that decision now that he knows me better. As a consequence, I've now learnt more than I ever wanted to know about phenology and breeding ecology in a lab that turned out to be made up of a bunch of kind and supportive people. This may or may not be a coincidence, given that Ben is a kind and supportive individual himself. So, thank you, Ben, for being a good scientist, mentor, and human.

Ella Cole has been my co-supervisor, which I'm very thankful for. She has done, and still does, a very good job at providing scientific advice to students, and, perhaps most importantly, helping to create an atmosphere where people can discuss their woes and these are not just brushed off—because my oh my are we not all a bit broken. I already said that the people in the lab are nice, but seriously, they are. I am tremendously lucky to be able to call some of them my friends. To all of them, from the bottom of my heart: thanks for putting up with me these last few years. Even after only a few years as a lowly student, I have already had a glimpse of an academia that can be competitive, a tad nasty, and motivated by all the wrong reasons, but none of it I've seen within the current iteration of this group: everyone understands that the science is, as it should be, secondary.

Whatever I've learned about programming, statistics and bioacoustics - all new to me - I fully owe it to all the people who share their code, data, and expertise altruistically on the internet, so thank you, faceless and often thankless open source and open science community! I would have definitely been more productive if I hadn't spent so much time learning about packaging, reproducibility, and sustainability, but I choose to believe that time was not wasted.

And then there's Carys Jones, who went from another student that I would bump into in the woods to an indispensable part of life, and has made this time at Oxford—and, I hope, many times to come, here and elsewhere—much better, fun, and interesting. I am grateful beyond words for your friendship and patience, your help and support, all the good and bad times already spent together in the last few years. This thesis owes a lot to you, too.

I am extremely lucky to have the family I have. It has been a while since I left the nest myself (almost 12 years now!), but they are in everything I do—well, not in everything, just the good and reasonable bits. My parents, Olga and Nilo, taught me to love nature and nurtured my curiosity, allowed me to explore the world, and have always been exceptionally loving and non-judgemental. My brothers, David and Juan, are also wonderful individuals who have taught me a lot, and I'm immensely proud of them. I don't say this nearly often enough, and a PhD thesis is kind of a weird place to do so, but I love them all dearly. The rest of my family is pretty great too!

More than seven years ago, shortly after I quit my old job, and before I had gone to university, I found myself hitchhiking across the Balkans with Andrea. It was a lot of fun, but neither of us could have possibly imagined that we'd be finishing our PhDs together all these years later. (As a matter of fact she's helping me draft the conclusions as I write this, because, very psychosomatically in character for me, I got shingles on my face and a headache like a hammer.) Back then I was debating whether it was worth going to uni at all, concluded 'well I got some savings, what can I lose' [narrator: he lost all of his hard-earned savings] and moved to the perfidious Albion; had an absolutely miserable, pitiful first year; but then Andrea came to work and study here also, and pretty much everything good that has happened ever since I have her to thank for. There's no way on earth I could have done any of this without you, science and life. Hopefully if my examiners like this thesis and I get ~~ordained~~ a PhD in Biology you'll stop saying that I'm not a real biologist? If so, this thesis, which is already partly yours, is dedicated to you.

OUTPUTS

CHAPTER 2 is published as

Merino Recalde, N. (2023). pykanto: A python library to accelerate research on wild bird song. *Methods in Ecology and Evolution*, 14, 1994–2002. 10.1111/2041-210X.14155

Code repository | Documentation | PyPI

CHAPTER 3 is in press at *Animal Behaviour* as

Merino Recalde, N., Estandía, A., Pichot, L., Vansse, A., Cole, E. F., & Sheldon, B. C. (2024). A densely sampled and richly annotated acoustic dataset from a wild bird population. *Animal Behaviour*. 10.1016/j.anbehav.2024.02.008

Website | Code repository | Data repository

CHAPTER 4 is under review

The main repository with code and data to reproduce the analysis and figures in this chapter can be found at birdsong-demography. Dataset available from osf.io/n8ac9. The code used to perform the song type validation process, along with the figures generated during it, can be found in a narrative notebook and its dedicated repository.

CHAPTER 5 is in preparation

The code to replicate these analyses and figures can be found at github/nilomr/greti-song-intervals. All the data are available from osf.io/n8ac9, and documented here.

Author contributions are detailed within each chapter.

CONTENTS

LIST OF FIGURES

CULTURAL EVOLUTION IN THE WILD: TRACKING THE LANDSCAPE OF LEARNING IN BIRD SONG

DPhil thesis by Nilo Merino Recalde,
Submitted Michaelmas Term 2023
Thesis advisors: Prof Ben C. Sheldon and Dr Ella F. Cole

ABSTRACT

The songs of some birds are learnt, much like human music and languages, and they change over time—they evolve culturally. Learning is imperfect and more likely to occur from nearby individuals, and this simple fact opens the door for behaviours to evolve differently in distinct populations. When these differences accumulate and persist over time, cultural traditions emerge. The resulting cultures can be transient or long-lasting, disorderly diverse, or monolithically uniform. This thesis addresses some of the challenges associated with studying cultural diversity in natural populations, and provides empirical evidence of factors that drive change and stability within vocal cultures.

After a general introduction to bird song and cultural evolution in the first chapter, the second introduces Pykanto, an open-source Python library designed to process and help analyse large datasets of animal vocalizations recorded in the field. The third chapter presents a comprehensive and richly annotated open dataset of wild bird songs collected over three years from a single population of Great Tits (*Parus major*) in the UK, which provides the data used in the remaining chapters. Then, the fourth chapter explores how demographic factors influence the frequency and diversity of cultural traits on small spatiotemporal scales, revealing that dispersal, immigration, and age structure are associated with variation in cultural diversity and turnover, and that this process is driven by individual-level differences in song repertoires. The fifth and final chapter characterises variation in the melodic and rhythmic structure of songs, suggesting that it is both non-random and categorically organised, which might contribute to large-scale stability in bird song cultures.

In summary, this thesis hopes to contribute to animal culture research by providing a substantial acoustic dataset of wild bird songs, introducing new computational tools for data analysis, and examining how extrinsic and intrinsic factors influence the diversity of bird song, rhythm, and melody; all of which are important for understanding emergent cultural processes at larger scales.

Great tit male singing by a nest box at sunrise.
Vector Illustration, Nilo Merino Recalde, 2023

1. INTRODUCTION

> *[T]he birds are in misery. I don't think they—they sing, they just screech in pain.*
> – Werner Herzog, *Burden of Dreams (1982)*

ANIMAL CULTURE AND SOCIAL LEARNING

THE idea that culture demarcates humans from other animals was widely prevalent in Western academia. Over the past few decades this view was steadily challenged, and today it is common to find references to non-human animal cultures in scientific journals and the popular press alike (Whiten, 2019). To be sure, some energetically oppose the notion (Galef, 1992; Laland & Janik, 2006; Sahlins, 1976), and there is no shortage of disagreement over the definition of the term 'culture' (Heyes, 2020; Kroeber & Kluckhohn, 1952; Laland & Hoppitt, 2003). But intricate and distinctive as human culture might be, the now burgeoning field of animal cultural research is showing us that the difference is one of degree and not kind (Whiten et al., 2017).

So what do we mean by culture in this context? For our purposes, we can define it as any behavioural trait or information that is maintained in a population by virtue of being learnt from others—not genetically inherited, or independently acquired (see definitions in Laland and Hoppitt, 2003; Whiten et al., 2017). Human ritual funerary practices are cultural; so are religions, the game of croquet, and PhD degrees. Crucially, under this definition, so is tool use in capuchin monkeys, homing efficiency in pigeons, the songs of many birds, feeding behaviours in humpback whales, and even mate preferences in fruit flies (Allen et al., 2013; Danchin et al., 2018; Falótico et al., 2019; Sasaki & Biro, 2017; Slater, 2003).

Social learning, where animals learn by observing or interacting with others, is widespread and a prerequisite for culture. While it may not always be advantageous (Giraldeau et al., 2002; Henrich & Boyd, 1998; Whitehead & Richerson, 2009), there is ample evidence that many of the skills that animals need to survive and reproduce can only be acquired by observing or interacting with others (Galef & Laland, 2005). Learning is more likely to occur from animals in close proximity or within the same social group, and this simple fact opens the door for behaviours to evolve differently in distinct populations, which can happen due to variation in learning abilities, ecological differences, or stochastic and neutral processes (Aplin, 2016; Araya-Salas et al., 2019; Mesoudi et al., 2016). When these differences, advantageous or not, accumulate and persist over time, cultural traditions emerge (Nunn et al., 2009; Tchernichovski et al., 2017). The resulting cultures can be transient or long-lasting, disorderly diverse, or monolithically uniform: in two primate examples, chimpanzees may have used stone tools in a similar way for thousands of years (Carvalho et al., 2008; Mercader et al., 2007), while white-faced capuchin monkeys frequently invent and abandon quirky social conventions such as eyeball-poking, hand-sniffing, and tail-sucking (Perry et al., 2003).

CULTURAL BIRDS—AND THEIR STUDY

The fact that animal lives have a cultural dimension was perhaps recognized earliest in birds. In 1920s South East England, some birds in the tit family started perforating the wax board or metal foil that sealed milk bottles to guzzle the cream accumulated at the top. This behaviour increased in frequency and geographic spread in the following decades, in what became a famous case of likely cultural transmission (Fisher & Hinde, 1949). Many years later, Aplin et al. (2015b) carried out experiments in a wild population of great tits *Parus major* which demonstrated that new foraging behaviours can indeed spread socially, and even persist over more than one generation. Similarly, information acquired by individuals and groups of birds when flying along a route can accumulate in populations and, over time and even generations, lead to distinct migratory cultures (Berdahl et al., 2018; Byholm et al., 2022; Jesmer et al., 2018; Sasaki & Biro, 2017).

While numerous examples of social learning and cultural phenomena exist in birds, it is their songs—thanks to their music-like qualities, and the remarkable ability of some species to imitate a wide range of sounds—that have garnered the most interest. Humanity's captivation with bird songs is also far from a recent development. As far back as 350 BCE., in his work 'Historia Animalium'. Aristotle noted

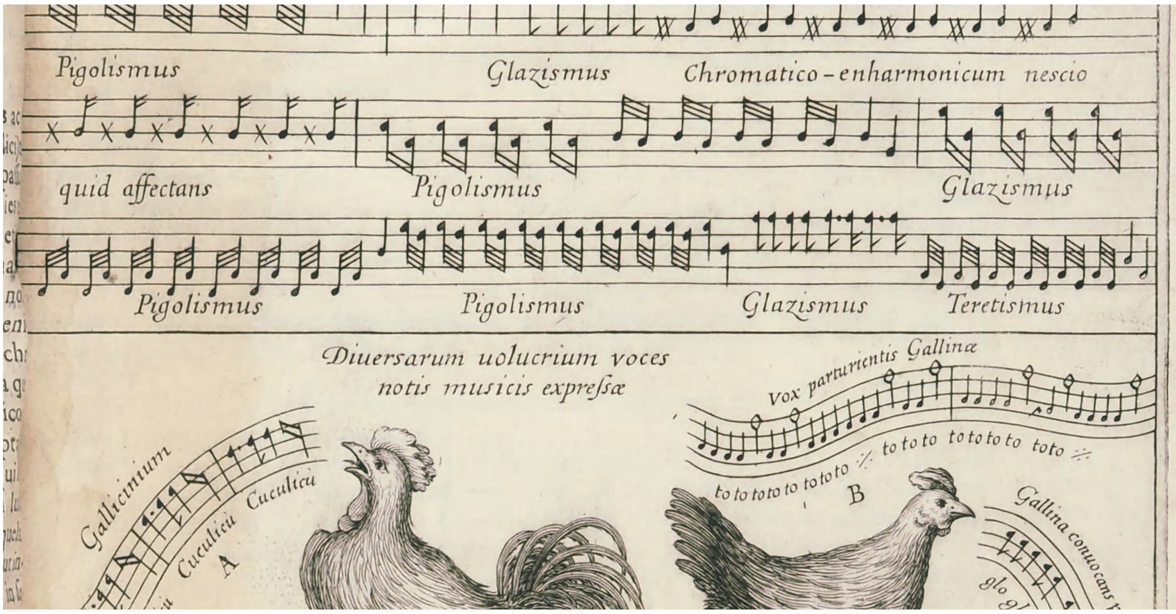



Figure 1.1. ***Musurgia Universalis, sive Ars Magna Consoni et Dissoni.***
'The Universal Musical Art, of the Great Art of Consonance and Dissonance', is a work by the Jesuit scholar Athanasius Kircher, published in 1650. Book one includes a rich discussion of bird song, including this illustration (Iconismus III), where the top part contains a study of *Glottismi modulationum sibilo exprimendi in Luscinia observati*, or 'Modulations of the glottis to express whistling observed in Luscinia', which is contrasted with the much simpler vocalizations of cocks and hens. Reproduction in the public domain.

that birds, especially 'broad-tongued' ones, were capable of learning their songs; and that, sometimes, their voices changed with the 'diversity of locality' (Book 4, Chapter 9). Aristotle's is one of the earliest recorded examples in a long tradition of analogies drawn between bird songs and human language and music across times and cultures (Kleczkowska, 2015; Zirin, 1980).

Many centuries later, in 1650, the German Jesuit Athanasius Kircher would use musical notation to transcribe and analyse bird songs in his early musicology treatise, 'Musurgia Universalis', at a time when incorporating bird song-inspired phrases into instrumental compositions had become rather popular (See Figure 1.1). However, it would not be until after the Industrial Revolution that technological advancements enabled the first recordings of singing birds, a crucial step in the study of changes and variations in their songs. (One earliest, if not the earliest, dates back to 1889 and was made by the pioneering sound recordist and broadcaster, Ludwig Koch, at the age of eight; British Library, 2023.) Fast-forward to the 1940s, and the invention of the sound spectrograph at the Bell Telephone Laboratories paved the way for a generation of researchers interested in bird song who were, for the first time, able to measure songs in unprecedented detail (Baker, 2001; Koenig et al., 1946; see Figure 1.2). An exhaustive study of the development of the song of the chaffinch *Fringilla coelebs* by Thorpe (1958) was followed by an explosion of interest in the matter, both in the field (Marler & Tamura, 1962; Marler & Tamura, 1964) and the laboratory (see, for example, Nottebohm et al., 1976 on the neurobiology of song production), which continues unabated to this day.

As researchers' fascination with animal cultures and the population consequences of song learning grew, they began describing cultural variation and longitudinal change in bird songs in great detail (Baptista, 1977; Lachlan & Slater, 2003; Nelson, 2017; Zann, 1993; Zimmerman et al., 2016). This is well exemplified by the extensive research, over four decades, on a population of Savannah sparrows *Passerculus sandwichensis* in Kent Island, New Brunswick, Canada (Dixon, 1978; Hensel et al., 2022; Mennill et al., 2018; Williams et al., 2013, 2019). Our growing understanding of bird song cultures, which forms the backdrop for this thesis, provides the most comprehensive evidence to date for the presence of culture and its evolution in non-human animals (Aplin, 2019; Slater, 2003). However, as with any behaviour, song learning is a complex process governed by numerous factors of diverse nature. To understand why change happens, we first need to understand why birds learn songs in the first place and the mechanisms behind change and variation.

WHY DO SOME BIRDS LEARN THEIR SONGS?

Most birds vocalize in one way or another, but the ability for vocal learning is a specialized trait found in only three orders: Psittaciformes, Apodiformes, and Passeriformes (Kroodsma, 2004; Päckert, 2018). Among passerines, commonly known as the perching birds, oscines (suborder Passeri) learn their songs, while suboscines (Tyranni) do not—but see Searcy et al., 2021; ten Cate, 2021. Song learning requires a significant investment of time, is metabolically costly, and requires a specialized syrinx and brain system, so the question arises: Why did it evolve, and why did it persist? Many hypotheses attempt to explain the potential advantages of vocal learning, which include a better ability to match the acoustic requirements of different habitats (Hansen, 1979; Ríos-Chelén et al., 2012), the potential benefits of sharing songs with neighbours (Payne, 1982), and the ability of learned songs to serve as an 'honest signal' of developmental stability, nutritional status, and other indicators of fitness (Nowicki et al., 2002; Ritchie et al., 2008). Existing evidence suggests two primary functions: attracting and securing mates and deterring rivals (Collins, 2004). Nevertheless, there is considerable variation among species, and the current fitness consequences of song learning, on which these ideas are based, do not necessarily bear on its evolutionary origins.

Not all birds that learn their songs do so at the same point in their lives or for the same length of time. Some can only learn during the first months after they are born: Zebra finch *Taeniopygia guttata* juveniles, for example, can only learn and develop their song roughly between 20 and 80 days after hatching (Liu et al., 2004). Other birds, such as the European starling *Sturnus vulgaris*, can continue to learn, mimic, and perhaps invent new sounds throughout their lives. Furthermore, many birds that disperse early in their lives have a 'sensitive period' for learning that extends to the time when they establish their first territory (Beecher & Brenowitz, 2005; Liu et al., 2004). While this variation is often categorized as either 'open-ended' learning, where adults can keep learning, or 'closed-ended' learning, where songs solidify before the first year, it may be better viewed as a spectrum (Brenowitz & Beecher, 2005). Lastly, there is remarkable variability in the size and complexity of the song repertoires among vocal learners. Some birds learn and endlessly sing a simple, single-syllable song, while others, like the brown thrasher *Toxostoma rufum*, never seem to stop incorporating new sounds into their repertoires (Boughey & Thompson, 1981).

(SOME OF) THE FORCES THAT SHAPE BIRD SONG

Various factors shape the evolution of bird songs on a broad scale. These include physiological constraints on song production, the influence of past evolutionary trajectories (phylogenetic inertia), the acoustic environment, and sexual selection. Others, such as the dynamics of cultural transmission, learning biases, and the social, spatial, and demographic structure of populations, lead to changes in the relative frequencies of different song variations—that is, cultural evolution (Whiten, 2019). Although these processes and their often complex interactions all influence bird song, their relative importance is not well understood. We will briefly touch on some of them and then turn to those that have been least studied

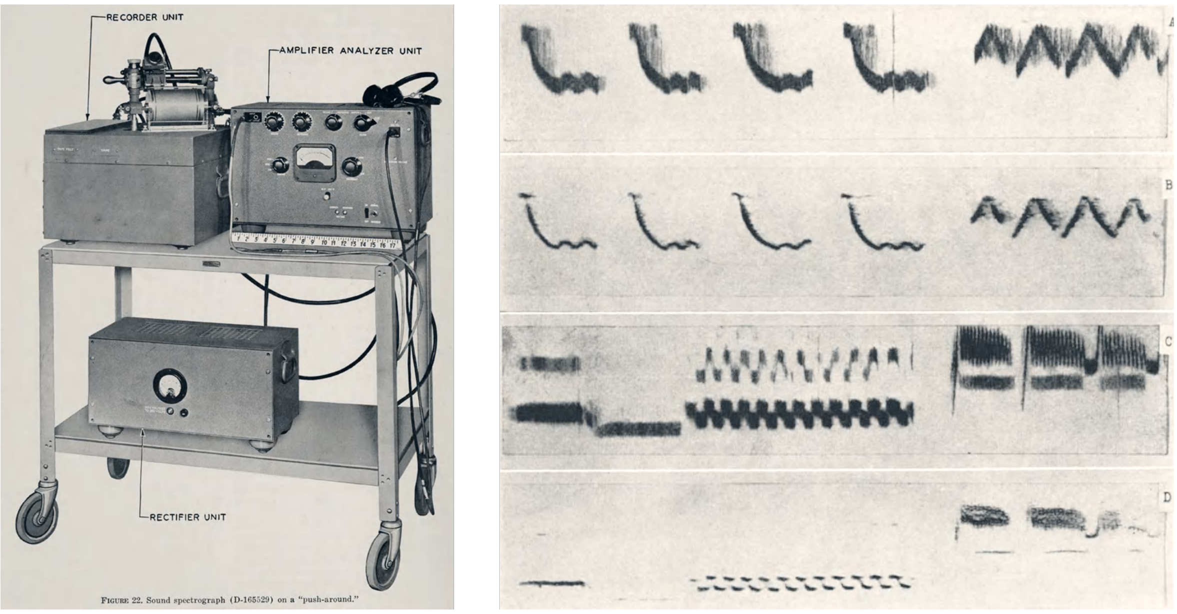


Figure 1.2. Early spectrograph device and spectrograms of bird songs.
Left: Early magnetic tape spectrograph, built by Bell Telephone Laboratories circa 1946 around a modified commercial turntable. Right: Spectrograms of the songs of a nightingale *Luscinia megarhynchos* (A, B) and a wood thrush *Hylocichla mustelina* (C, D), generated from uncredited phonograph records. Both reproduced from Koenig et al. (1946).

and are part of the focus of this thesis: demography's influence on cultural change and diversification in bird song, and structural biases that underpin cultural stability.

Physiological constraints, phylogenetic inertia

Birds produce their songs using an organ called the syrinx, which is located at the base of the trachea in an air sac close to the lungs (Larsen & Goller, 2002). Its structure and precise location vary between species; those with more complex songs tend to have more syringeal muscles, which allows for precise control of sound (Suthers, 2004). The sounds produced by the syrinx are then filtered and modulated by the trachea and the beak, which together constitute the vocal tract (Podos et al., 2004). The body mass of a species is correlated with the size of its syrinx, which in turn influences the fundamental or lowest resonant frequency of the sounds that it can produce (Martin et al., 2011; Ryan & Brenowitz, 1985a). This means that vocal evolution will be constrained by any selective pressures impinging on body size, as well as the particular evolutionary trajectory of the species. The same is true for the morphology of the beak, which influences the pace at which a bird can produce notes and the range of their frequencies (Derryberry et al., 2012, 2018; Podos, 2001; Seddon, 2005). These and other constraints, not least those derived from neural development and energy expenditure, determine the degrees of freedom available to other forces driving song evolution.

Ecological factors

Sound propagates differently in different habitats. For example, vegetation attenuates sound amplitude, filters some frequencies more than others, and causes reverberation. Different physical environments have different levels of background noise; there may be sounds of abiotic origin, such as wind, streams, and rain, and a cacophony of other species striving to be heard. Songs, like any other signal, need to be detectable to be effective, and this process of acoustic adaptation is thought to be important in driving the evolution of song and other social signals (Endler, 1992; Grant and Grant, 2010; Tobias et al., 2010; but see Mikula et al., 2020). On a shorter timescale, birds can respond to increased noise levels by adjusting the properties of their songs to reduce overlap with background noise, which they can do, for example, by increasing their minimum frequency (Brumm, 2004; Derryberry et al., 2020; Halfwerk et al., 2011; Liu et al., 2020; Nemeth et al., 2013; Roca et al., 2016; Winandy et al., 2021). However, it's not entirely clear whether this response is always driven by individual plastic responses or something that can also evolve across generations.

There are other ecological factors that can lead to song divergence: body size and beak morphology often change when a population of birds adapts to a different foraging niche, which, as mentioned earlier, can influence some of the acoustic and temporal features of songs as a by-product (Mayr, 1963; Podos, 2001). Additionally, songs can undergo changes when different bird species co-exist in the same habitat (sympatry). In such cases, selective processes may favour songs that can more easily distinguish birds from different species, which can help avoid the fitness costs of hybridization or aggressive interactions; a mechanism known as character displacement (Seddon, 2005). Interestingly, the opposite—convergent character displacement in sympatry—might also happen if better recognition of competitors is advantageous (Grant, 1972; Tobias & Seddon, 2009; Tobias et al., 2014).

Sexual selection

Bird song has traditionally been considered a male ornament that evolved through female preference. However, recent phylogenetic analyses suggest that, at least in oscine passerines, dimorphism in this trait is a consequence of multiple losses of elaborate female songs rather than gains in males (Odom & Benedict, 2018; Odom et al., 2014), and that the paucity of research on female song is significantly influenced by sociological factors (Haines et al., 2020). Indeed, most of the work done in the field of behavioural ecology over the past decades has focused on how females respond to songs and how males use them in territorial interactions. And, as a result, there is a vast literature detailing how songs can encode multiple messages and serve different, sexually selected functions, with great variation between species and sometimes inconsistent results (see Catchpole and Slater, 2008 for reviews; also Sierro et al., 2023 for recent work on this). To mention just a few themes: song frequency can be a reliable indicator of body size (Ryan & Brenowitz, 1985b), temporal and spectral consistency can evidence the nutritional stress suffered in early life (MacDonald et al., 2006), and sharing songs with other males can influence territory acquisition and the outcome of agonistic encounters (Demko et al., 2016; Krebs et al., 1978).

Learning biases and cultural dynamics

For birds that learn their songs, the process itself introduces a host of new variables that can influence which songs a bird ultimately sings, their structure, and how they evolve over time. Initially, a bird needs to learn songs from another bird. Some birds may prefer specific 'tutors' over others, such as those that are older, louder, or sing more frequently (Greig et al., 2012). This preference is variously referred to as a tutor, model, or demonstrator bias in the field of cultural evolution (Kendal et al., 2015; Van De Waal et al., 2010). Cultural transmission of songs can occur in different ways. When songs are passed from parents to offspring, this is referred to as ’vertical transmission’. If songs are shared among birds in the same generation, it is known as ’horizontal transmission’, and ’oblique transmission’ when songs are learned from unrelated individuals in the previous generation (Cavalli-Sforza et al., 1982; Ram et al., 2018). Sometimes, tutor ’choice’ can have extreme consequences: if a female pairs up with a male of a different species because he sings the same song as her father, and her father learnt the ’wrong’ song, this can lead to the production of hybrid offspring (Grant & Grant, 1997a; Grant & Grant, 1997b).

Then, some songs may be preferred independently of who sings them—a ’content bias’ (Richerson & Boyd, 2005). Song production is costly, and birds could favour learning easy-to-sing variants, or the opposite if singing difficult songs is rewarded with better reproductive resources. More broadly, however, there is a well-documented tendency for birds to preferentially learn the songs sung by members of their own species (Slabbekoorn & Smith, 2002), or even subspecies (Nelson, 2000). These preferences can arise through the interaction of innate biases for some regions of the syntactic or acoustic trait space, exposure to sounds during early life, and feedback loops during the learning process (Fehér et al., 2009, 2017; Verzijden et al., 2012). This interaction can be highly complex and shows great variability between species and even individuals (James et al., 2020; Mets & Brainard, 2017, 2019; ten Cate & Rowe, 2007).

Song acquisition in birds can also be influenced by how frequently they have heard a particular variant, or how many other birds sing it (Aplin et al., 2015a; van Leeuwen et al., 2015). The relationship between the frequency of a trait in a population and the probability that it is learned can be linear: if this is the case, transmission is unbiased. However, if more popular traits are more likely to be learned than would be expected given their frequency, learning is said to be conformist, or positively frequency-dependent. When the opposite is true the pattern is one of anti-conformism, or negative frequency dependence. Conformist learning seems to be a common strategy in nature, perhaps because it can help individuals leverage collective information to make decisions about locally adaptive behaviour (Danchin et al., 2018; Pike & Laland, 2010; Whiten, 2019); all else being equal, cultural traits will have a slower rate of change if their learning is conformist. Conversely, anti-conformist biases will cause faster turnover (Acerbi & Alexander Bentley, 2014) and might arise if individuals have a preference for novelty (Smaldino & Epstein, 2015).

Another determinant of the nature and pace of change in a population’s songs is the precision with which birds learn them. Some authors have tried to calculate ’cultural mutation rates’ for different species to reflect how much songs change from one generation to the next: Chaffinches, for example, appear to learn their songs very accurately (Lachlan & Slater, 2003; Slater, 1986), which slows change. However, determining this quantity is complicated by the fact that many other factors also influence the pace of cultural change. One is the learning strategy used by individuals, as already mentioned. Another is the fact that songs can be modified during the learning process in non-random directions, in a process sometimes called convergent transformation or cultural attraction (Claidière et al., 2018; Gray et al., 2007; Heyes, 1993; Morin, 2016). This can be illustrated by an elegant experiment carried out by Fehér et al. (2009): song tutoring lineages were started with birds that were raised in isolation, and each new bird learned from the previous—mimicking natural generations of tutors and pupils. Many of the characteristic song features of the species emerged over the course of a few generations, presumably as a consequence of directional modifications introduced by the learners.

Consequences of learning

The available evidence suggests that the changes brought about by cultural dynamics are often neutral and can even be negative with respect to individual fitness (Langin et al., 2017; Slater & Lachlan, 2003), akin to genetic drift (Grant & Grant, 2010). However, in some cases, cultural change might help birds adapt to new acoustic landscapes (Ríos-Chelén et al., 2012; Slater & Lachlan, 2003), or fine-tune the song’s characteristics to better match the shifting perceptual preferences of the receivers (Renoult & Mendelson, 2019). At deeper timescales, the accumulation of neutral or directional cultural changes in songs might promote reproductive isolation between populations. However, although this possibility has attracted much interest, strong evidence for it is lacking (Freeman et al., 2022; Lachlan & Servedio, 2004; Verzijden et al., 2012; Yeh & Servedio, 2015). More generally, and further complicating things, songs are not necessarily fixed behavioural packages, and different elements (notes, phrases, spectral properties, syntax) might change at different rates and be affected by different cultural, sexual, or ecological pressures (see, for example, Williams et al., 2013).

Spatial, social, and demographic factors

The most fundamental determinant of the interactions that lead to vocal learning in a population is the habitat that it occupies. Its physical features, vegetation structure, and distribution of resources, which are all tightly linked, bound the use of space made by individuals (Albery et al., 2021; Firth & Sheldon, 2016b; Spiegel et al., 2016). This influences their social interactions and emergent social organization (see He et al., 2019 for a review), which, in conjunction with dispersal and learning norms, ultimately dictates who learns from whom. For instance, spatial proximity will typically be correlated with vocal similarity if birds in a population learn their songs with some accuracy and their dispersal is limited—although the precise nature of this pattern depends on whether birds disperse before or after they learn, how many songs they learn, and from how many birds (Ellers & Slabbekoorn, 2003; Williams & Slater, 1990), as well as the many sources of learning bias discussed above. When birds remain close to the location where they learned, or if they selectively settle in places where they hear familiar songs, dialects or local song 'neighbourhoods' can emerge (Podos & Warren, 2007). It then follows that increased dispersal between any two areas within a population will promote vocal sharing and that the influx of immigrants from other populations can introduce new variation.

Space and movement play a significant role in how information flows and its impact on diversity. However, other factors, such as group size and turnover, also affect the speed of cultural change. First, smaller groups are more susceptible to stochastic factors, paralleling the effects of genetic drift on small populations (Kimura & Crow, 1964). As a consequence, smaller group sizes can increase the probability of extinction of any given song variant (Nunn et al., 2009), and lead to idiosyncratic changes in song structure (Lachlan et al., 2013). Second, higher background mortality and emigration will also lead to faster rates of change, independently of the learning strategy employed by individuals or the fidelity of their learning (Nunn et al., 2009; Slater, 1986).

Finally, a number of recent theoretical and experimental results suggest that the number, distribution, and connectedness of individuals in a population can influence the diversity and complexity of cultural traits in addition to their frequency distribution (see, for example, Creanza et al., 2017; Derex and Boyd, 2016; Derex et al., 2018; Kempe and Mesoudi, 2014). While this phenomenon might be of great importance in human culture, bird songs are heavily constrained by ecological, cognitive, and sexual selective factors, which limits the potential of extrinsic factors to lead to appreciable gains in diversity or complexity. There is, however, some evidence from laboratory experiments that suggests that the availability of tutors and the quality of their instruction can reduce the influence of genetic contributions to the song phenotype (Mets & Brainard, 2017, 2019), and this mechanism could help explain the loss of song cultural diversity observed in some fragmented populations (Hart et al., 2018; Paxton et al., 2019).

THIS THESIS

STRUCTURE

STudying the processes that influence temporal and spatial patterns of bird song diversity in natural populations presents significant challenges. Although a long and fruitful history of field studies has provided us with examples of some of the processes outlined in the introduction, they often suffer from limitations such as small sample sizes, a lack of data for individual birds, and idiosyncratic analytical choices.

To address these challenges and contribute to the understanding of animal cultural diversity, as part of this thesis I introduce computational tools designed to efficiently collect and analyse extensive amounts of bird songs from the field (Chapter 2). Then, I present and document the largest open-access dataset of songs from a single wild bird population to date, which includes comprehensive song and individual metadata, allowing researchers to rigorously test hypotheses that previously lacked robust support because of the limited data available for individual birds (Chapter 3). It is perhaps worth noting that this field, like so many others, is increasingly reliant on technology and computational tools, which may pose a steep learning curve. In my case, transitioning from a background in fine arts and anthropology meant dedicating a substantial portion of my PhD to acquiring skills in statistics, high-performance computing, and programming—although I was also lucky to spend almost 9 months quite literally running around the woods!

Having introduced the new software and data, which form the core of this thesis, I move on to its empirical section. As I briefly reviewed in the general introduction above, there has been extensive research on macroevolutionary patterns and social learning strategies or biases (Aplin et al., 2017; Kendal et al., 2015; Lachlan et al., 2018; Pike & Laland, 2010; Tchernichovski et al., 2021), but there remains a gap in our understanding of how demography contributes to the emergence and persistence of cultural traits in wild populations. Factors extrinsic to culture, like juvenile recruitment, emigration, mortality, and age structure, are thought to impact individuals' opportunities for learning and exposure to cultural variants (Barta et al., 2023; Deffner & McElreath, 2020, 2022; Deffner et al., 2022; Derex & Boyd, 2016; Fogarty et al., 2019; Kandler et al., 2023; Kirby & Tamariz, 2021; Nunn et al., 2009). But translating these expectations into empirical evidence remains a challenge; In this context, Chapter 4 of this thesis provides evidence that demographic factors are associated with cultural diversity and turnover at the individual and neighbourhood level, identifying the relevant spatial and temporal scale that drives emergent patterns within a wild bird population.

On the other hand, while we think that demographic factors can influence the pace of cultural change and promote or restrict diversity, they alone cannot explain why some song cultures are stably polymorphic, without collapsing into single variants or experiencing complete cultural replacement. It is generally assumed that innate mechanisms for accurate and conformist learning are the primary forces behind cultural stability in bird song (Lachlan & Slater, 2003; Lachlan et al., 2018). However, there is also growing support for the idea that perceptual, cognitive, and production biases, innate or culturally acquired, serve as attractors that can align the perception, memory, and reproduction of information among individuals and, as a consequence, enhance the fidelity of cultural transmission and limit the drift of cultural traits (Fehér et al., 2009; Saldana et al., 2019; Trehub, 2015). In Chapter 5 I present some early evidence that the melodic space explored by a population of wild birds is highly non-random, in a way that parallels human musical traditions.

STUDY SYSTEM

I conducted all the research presented in this thesis within the long-term population study of great tits *Parus major* in Wytham Woods, near Oxford, UK, which spans 76 uninterrupted years and has already recorded the life histories of nearly 120,000 individual birds (Sheldon et al., 2022). It began in 1947 with John Gibb and David Lack from the University of Oxford, initially with 100 nest boxes and later expanding to over 1000 around 1960. This study focuses primarily on great tits, a species that has become somewhat of a model species for various evolutionary, ecological, and cognitive studies in natural populations (Aplin et al., 2017; Boyce & Perrins, 1987; Charmantier et al., 2008; Cole et al., 2012; Firth & Sheldon, 2016a; Firth et al., 2018; Spurgin et al., 2019). Their well-documented life history and behaviour, coupled with the characteristics of their songs, also make them ideal subjects for research on song learning and cultural change (see, for example, Lambrechts and Dhondt, 1990; Lind et al., 1996; Ritschard et al., 2012; Rivera-Gutierrez, Matthysen, et al., 2010; Rivera-Gutierrez, Pinxten, and Eens,

Figure 1.3. The author (me!) in the field office, Wytham Woods, 2021.

Figure 1.4. Wytham Woods 2021 fieldwork season group photo.
From left to right: Sam Crofts, Nilo Merino Recalde, Kristina Beck, Charlotte Regan, Joe Woodman, Anett Kiss, Carys Jones, Loanne Pichot, Andrea Estandía, Ben Sheldon, Keith McMahon, and Julia Haynes.

2010; Rivera-Gutierrez et al., 2011; Slagsvold et al., 1994).

Male great tits sing a repertoire of one to over ten simple song variants, each characterized by a small number of repeated notes that are repeated in a stereotypical manner (Krebs et al., 1978; Rivera-Gutierrez, Matthysen, et al., 2010). They primarily learn these songs from conspecifics during their first year, up to the point where they establish their own territory, and, consequently, are more likely to share their songs with neighbouring males than with other birds (McGregor & Krebs, 1982a; McGregor & Krebs, 1989). Research on great tit song conducted within the Wytham Woods population, in particular, has given rise to influential ideas and insights into bird singing behaviour, encompassing themes such as neighbour interactions, song matching, the connection between song repertoires and reproductive success (McGregor & Krebs, 1989; McGregor et al., 1981; Mcgregor et al., 1983), dynamics of song learning (McGregor & Krebs, 1982b, 1989), the role of song repertoires in maintaining territories (Krebs, 1976; Krebs et al., 1978), the functions of dawn song (Kacelnik & Krebs, 1983; Mace, 1987), and the influence of spatial factors and movement on song culture (Fayet et al., 2014).

I am not one to particularly cherish tradition in other aspects of life, but I feel honoured to be able to contribute—if much more modestly—to the rich history of research on bird song, and especially to do so here at Wytham. If you are about to read this thesis: thank you, and I hope it is not too painful.

A FEW NOTES TO THE READER

Each chapter in this thesis is a standalone piece of work. This means that some information, like the introduction to the study system and certain methods, is repeated in each chapter. To save time and avoid redundancy, Chapter 3 offers the most detailed description of the study system, fieldwork, and data annotation process. I have invested significant time into ensuring that all the code (over 25,000 lines, many more than in this thesis!) and the data generated during my PhD are easily accessible and reasonably well-documented, so that they may be useful to other students and researchers. You will find links to these resources within the manuscripts and highlighted in the Outputs page at the beginning. Lastly, I am told that the hyperlinks will not work properly if, for some reason, you have printed this thesis on paper. Well, you shouldn't have, what can I say?

REFERENCES

Acerbi, A., & Alexander Bentley, R. (2014). Biases in Cultural Transmission Shape the Turnover of Popular Traits. *Evolution and Human Behavior*, *35*(3), 228–236. https://doi.org/10.1016/j.evolhumbehav.2014.02.003

Albery, G. F., Morris, A., Morris, S., Pemberton, J. M., Clutton-Brock, T. H., Nussey, D. H., & Firth, J. A. (2021). Multiple spatial behaviours govern social network positions in a wild ungulate. *Ecology Letters*, *24*(4), 676–686. https://doi.org/10.1111/ele.13684

Allen, J., Weinrich, M., Hoppitt, W., & Rendell, L. (2013). Network-Based Diffusion Analysis Reveals Cultural Transmission of Lobtail Feeding in Humpback Whales. *Science*, *340*(6131), 485–488. https://doi.org/10.1126/science.1231976

Aplin, L. (2016). Understanding the multiple factors governing social learning and the diffusion of innovations. *Current Opinion in Behavioral Sciences*, *12*, 59–65. https://doi.org/10.1016/j.cobeha.2016.09.003

Aplin, L. M. (2019). Culture and Cultural Evolution in Birds: A Review of the Evidence. *Animal Behaviour*, *147*, 179–187. https://doi.org/10.1016/j.anbehav.2018.05.001

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015a). Counting conformity: Evaluating the units of information in frequency-dependent social learning. *Animal Behaviour*, *110*, e5–e8. https://doi.org/10.1016/j.anbehav.2015.09.015

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015b). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*, *518*(7540), 538–541. https://doi.org/10.1038/nature13998

Aplin, L. M., Sheldon, B. C., & McElreath, R. (2017). Conformity does not perpetuate suboptimal traditions in a wild population of songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7830–7837. https://doi.org/10.1073/pnas.1621067114

Araya-Salas, M., Smith-Vidaurre, G., Mennill, D. J., González-Gómez, P. L., Cahill, J., & Wright, T. F. (2019). Social group signatures in hummingbird displays provide evidence of co-occurrence of vocal and visual learning. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1903). https://doi.org/10.1098/rspb.2019.0666

Baker, M. (2001). Bird song research: The past 100 years. *Bird Behavior*, *14*(1), 3–50.

Baptista, L. F. (1977). Geographic Variation in Song and Dialects of the Puget Sound White-Crowned Sparrow. *The Condor*, *79*(3), 356–370. https://doi.org/10.2307/1368014

Barta, K. A., Garamszegi, L. Z., Scheuring, I., & Zsebk, S. (2023). Effects of positive frequency-dependent learning, learning mistakes, and immigration on complex culturesValidation on the song of collared flycatcher (Ficedula albicollis) by individual-based modeling. *Frontiers in Ecology and Evolution*, *11*.

Beecher, M. D., & Brenowitz, E. A. (2005). Functional Aspects of Song Learning in Songbirds. *Trends in Ecology and Evolution*, *20*(3), 143–149. https://doi.org/10.1016/j.tree.2005.01.004

Berdahl, A. M., Kao, A. B., Flack, A., Westley, P. A., Codling, E. A., Couzin, I. D., Dell, A. I., & Biro, D. (2018). Collective animal navigation and migratory culture: From theoretical models to empirical evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1746), 20170009. https://doi.org/10.1098/rstb.2017.0009

Boughey, M. J., & Thompson, N. S. (1981). Song Variety in the Brown Thrasher (Toxostoma rufum). *Zeitschrift für Tierpsychologie*, *56*(1), 47–58. https://doi.org/10.1111/j.1439-0310.1981.tb01283.x

Boyce, M. S., & Perrins, C. M. (1987). Optimizing great tit clutch size in a fluctuating environment. *Ecology*, *68*(1), 142–153. https://doi.org/10.2307/1938814

Brenowitz, E. A., & Beecher, M. D. (2005). Song learning in birds: Diversity and plasticity, opportunities and challenges. *Trends in Neurosciences*, *28*(3), 127–132. https://doi.org/10.1016/j.tins.2005.01.004

British Library, T. (2023). The first wildlife recording.

Brumm, H. (2004). The impact of environmental noise on song amplitude in a territorial bird. *Journal of Animal Ecology*, *73*(3), 434–440. https://doi.org/10.1111/j.0021-8790.2004.00814.x

Byholm, P., Beal, M., Isaksson, N., Lötberg, U., & Åkesson, S. (2022). Paternal transmission of migration knowledge in a long-distance bird migrant. *Nat Commun*, *13*, 1566. https://doi.org/10.1038/s41467-022-29300-w

Carvalho, S., Cunha, E., Sousa, C., & Matsuzawa, T. (2008). Chaînes opératoires and resource-exploitation strategies in chimpanzee (Pan troglodytes) nut cracking. *Journal of Human Evolution*, *55*(1), 148–163. https://doi.org/10.1016/j.jhevol.2008.02.005

Catchpole, C. K., & Slater, P. J. B. (2008). *Bird Song*. Cambridge University Press. https://doi.org/10.1017/CBO9780511754791

Cavalli-Sforza, L. L., Feldman, M. W., Chen, K. H., & Dornbusch, S. M. (1982). Theory and observation in cultural transmission. *Science*, *218*(4567), 19–27. https://doi.org/10.1126/science.7123211

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E., & Sheldon, B. C. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, *320*(5877), 800–803. https://doi.org/10.1126/science.1157174

Claidière, N., kuma Amedon, G. K., André, J. B., Kirby, S., Smith, K., Sperber, D., & Fagot, J. (2018). Convergent transformation and selection in cultural evolution. *Evolution and Human Behavior*, *39*(2), 179–190. https://doi.org/10.1016/j.evolhumbehav.2017.12.007

Cole, E. F., Morand-Ferron, J., Hinks, A. E., & Quinn, J. L. (2012). Cognitive ability influences reproductive life history variation in the wild. *Current Biology*, *22*(19), 1808–1812. https://doi.org/10.1016/j.cub.2012.07.051

Collins, S. (2004). Vocal fighting and flirting: The functions of birdsong. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 39–79). Elsevier. https://doi.org/10.1016/B978-012473070-0/50005-0

Creanza, N., Kolodny, O., & Feldman, M. W. (2017). Greater than the sum of its parts? Modelling population contact and interaction of cultural repertoires. *Journal of the Royal Society Interface*, *14*(130). https://doi.org/10.1098/rsif.2017.0171

Danchin, E., Nöbel, S., Pocheville, A., Dagaeff, A.-C., Demay, L., Alphand, M., Ranty-Roby, S., van Renssen, L., Monier, M., Gazagne, E., Allain, M., & Isabel, G. (2018). Cultural flies: Conformist social learning in fruitflies predicts long-lasting mate-choice traditions. *Science (New York, N.Y.)*, *362*(6418), 1025–1030. https://doi.org/10.1126/science.aat1590

Deffner, D., Kandler, A., & Fogarty, L. (2022). Effective population size for culturally evolving traits. *PLOS Computational Biology*, *18*(4), e1009430. https://doi.org/10.1371/journal.pcbi.1009430

Deffner, D., & McElreath, R. (2020). The importance of life history and population regulation for the evolution of social learning. *Philosophical Transactions of the Royal So-*

ciety B: Biological Sciences, *375*(1803), 20190492. https://doi.org/10.1098/rstb.2019.0492

Deffner, D., & McElreath, R. (2022). When does selection favor learning from the old? Social learning in age-structured populations. *PLOS ONE*, *17*(4), e0267204. https://doi.org/10.1371/journal.pone.0267204

Demko, A. D., Reitsma, L. R., & Staicer, C. A. (2016). Repertoire structure, song sharing, reproductive success, and territory tenure in a population of Canada Warblers (Cardellina canadensis) in central New Hampshire. *Canadian Journal of Zoology*, *94*(4), 283–290. https://doi.org/10.1139/cjz-2015-0213

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987. https://doi.org/10.1073/pnas.1518798113

Derex, M., Perreault, C., & Boyd, R. (2018). Divide and conquer: Intermediate levels of population fragmentation maximize cultural accumulation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170062. https://doi.org/10.1098/rstb.2017.0062

Derryberry, E. P., Phillips, J. N., Derryberry, G. E., Blum, M. J., & Luther, D. (2020). Singing in a silent spring: Birds respond to a half-century soundscape reversion during the COVID-19 shutdown. *Science*. https://doi.org/10.1126/science.abd5777

Derryberry, E. P., Seddon, N., Claramunt, S., Tobias, J. A., Baker, A., Aleixo, A., & Brumfield, R. T. (2012). Correlated evolution of beak morphology and song in the neotropical woodcreeper radiation. *Evolution*, *66*(9), 2784–2797. https://doi.org/10.1111/j.1558-5646.2012.01642.x

Derryberry, E. P., Seddon, N., Derryberry, G. E., Claramunt, S., Seeholzer, G. F., Brumfield, R. T., & Tobias, J. A. (2018). Ecological drivers of song evolution in birds: Disentangling the effects of habitat and morphology. *Ecology and Evolution*, *8*(3), 1890–1905. https://doi.org/10.1002/ece3.3760

Dixon, C. L. (1978). Breeding Biology of the Savannah Sparrow on Kent Island. *The Auk*, *95*(2), 235–246. https://doi.org/10.1093/auk/95.2.235

Ellers, J., & Slabbekoorn, H. (2003). Song divergence and male dispersal among bird populations: A spatially explicit model testing the role of vocal learning. *Animal Behaviour*, *65*(4), 671–681. https://doi.org/10.1006/anbe.2003.2081

Endler, J. A. (1992). Signals, Signal Conditions, and the Direction of Evolution. *The American Naturalist*, *139*, S125–S153. https://doi.org/10.1086/285308

Falótico, T., Proffitt, T., Ottoni, E. B., Staff, R. A., & Haslam, M. (2019). Three thousand years of wild capuchin stone tool use. *Nature Ecology and Evolution*. https://doi.org/10.1038/s41559-019-0904-4

Fayet, A. L., Tobias, J. A., Hintzen, R. E., & Seddon, N. (2014). Immigration and dispersal are key determinants of cultural diversity in a songbird population. *Behavioral Ecology*, *25*(4), 744–753. https://doi.org/10.1093/beheco/aru047

Fehér, O., Ljubii, I., Suzuki, K., Okanoya, K., & Tchernichovski, O. (2017). Statistical learning in songbirds: From self-tutoring to song culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160053. https://doi.org/10.1098/rstb.2016.0053

Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, *459*(7246), 564–568. https://doi.org/10.1038/nature07994

Firth, J. A., Cole, E. F., Ioannou, C. C., Quinn, J. L., Aplin, L. M., Culina, A., McMahon, K., & Sheldon, B. C. (2018). Personality shapes pair bonding in a wild bird social system. *Nature Ecology and Evolution*, *2*(11), 1696–1699. https://doi.org/10.1038/s41559-018-0670-8

Firth, J. A., & Sheldon, B. C. (2016a). Social carry-over effects underpin trans-seasonally linked structure in a wild bird population. *Ecology Letters*, *19*(11), 1324–1332. https://doi.org/10.1111/ele.12669

Firth, J. A., & Sheldon, B. C. (2016b). Social carry-over effects underpin trans-seasonally linked structure in a wild bird population. *Ecol Lett*, *19*(11), 1324–1332. https://doi.org/10.1111/ele.12669

Fisher, J., & Hinde, R. A. (1949). The opening of milk bottles by birds. *British Birds*, *42*(2), 347–357.

Fogarty, L., Creanza, N., & Feldman, M. W. (2019). The life history of learning: Demographic structure changes cultural outcomes. *PLOS Computational Biology*, *15*(4), e1006821. https://doi.org/10.1371/journal.pcbi.1006821

Freeman, B. G., Rolland, J., Montgomery, G. A., & Schluter, D. (2022). Faster evolution of a premating reproductive barrier is not associated with faster speciation rates in New World passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1966), 20211514. https://doi.org/10.1098/rspb.2021.1514

Galef, B. G. (1992). The question of animal culture. *Human Nature*, *3*(2), 157–178. https://doi.org/10.1007/BF02692251

Galef, B. G., & Laland, K. N. (2005). Social learning in animals: Empirical studies and theoretical models. *BioScience*, *55*(6), 489–499. https://doi.org/10.1641/0006-3568(2005)055[0489:SLIAES]2.0.CO;2

Giraldeau, L. A., Valone, T. J., & Templeton, J. J. (2002). Potential disadvantages of using socially acquired information (R. A. Johnstone & S. R. X. Dall, Eds.). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*(1427), 1559–1566. https://doi.org/10.1098/rstb.2002.1065

Grant, B. R., & Grant, P. R. (2010). Songs of Darwin's finches diverge when a new species enters the community. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(47), 20156–20163. https://doi.org/10.1073/pnas.1015115107

Grant, P. R. (1972). Convergent and divergent character displacement. *Biological Journal of the Linnean Society*, *4*(1), 39–68. https://doi.org/10.1111/j.1095-8312.1972.tb00690.x

Grant, P. R., & Grant, B. R. (1997a). Genetics and the origin of bird species. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(15), 7768–7775. https://doi.org/10.1073/pnas.94.15.7768

Grant, P. R., & Grant, B. R. (1997b). Hybridization, Sexual Imprinting, and Mate Choice. *The American Naturalist*, *149*(1), 1–28. https://doi.org/10.1086/285976

Gray, R. D., Greenhill, S. J., & Ross, R. M. (2007). The Pleasures and Perils of Darwinizing Culture (with Phylogenies). *Biological Theory*, *2*(4), 360–375. https://doi.org/10.1162/biot.2007.2.4.360

Greig, E. I., Taft, B. N., & Pruett-Jones, S. (2012). Sons learn songs from their social fathers in a cooperatively breeding bird. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1741), 3154–3160. https://doi.org/10.1098/rspb.2011.2582

Haines, C. D., Rose, E. M., Odom, K. J., & Omland, K. E. (2020). The role of diversity in science: A case study of women advancing female birdsong research. *Animal Behaviour*, *168*, 19–24. https://doi.org/10.1016/j.anbehav.2020.07.021

Halfwerk, W., Bot, S., Buikx, J., Van Der Velde, M., Komdeur, J., Ten Cate, C., & Slabbekoorn, H. (2011). Low-frequency songs lose their potency in noisy urban conditions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(35), 14549–14554. https://doi.org/10.1073/pnas.1109091108

Hansen, P. (1979). Vocal learning: Its role in adapting sound structures to long-distance propagation, and a hypothesis on its evolution. *Animal Behaviour*, *27*, 1270–1271. https://doi.org/10.1016/0003-3472(79)90073-3

Hart, P. J., Sebastián-González, E., Tanimoto, A., Thompson, A., Speetjens, T., Hopkins, M., & Atencio-Picado, M. (2018). Birdsong characteristics are related to fragment size in a neotropical forest. *Animal Behaviour*, *137*, 45–52. https://doi.org/10.1016/j.anbehav.2017.12.020

He, P., Maldonado-Chaparro, A. A., & Farine, D. R. (2019). The role of habitat configuration in shaping social structure: A gap in studies of animal social complexity. *Behavioral Ecology and Sociobiology*, *73*(1). https://doi.org/10.1007/s00265-018-2602-7

Henrich, J., & Boyd, R. (1998). The Evolution of Conformist Transmission and the Emergence of Between-Group Differences. *Evolution and Human Behavior*, *19*(4), 215–241. https://doi.org/10.1016/S1090-5138(98)00018-X

Hensel, A. L. J., Dobney, S. L., Doucet, S. M., Ryan Norris, D., Newman, A. E. M., Williams, H., & Mennill, D. J. (2022). Microgeographical variation in birdsong: Savannah sparrows exhibit microdialects in an island population. *Animal Behaviour*, *188*, 119–131. https://doi.org/10.1016/j.anbehav.2022.04.008

Heyes, C. (2020). Culture. *Current Biology*, *30*(20), R1246–R1250. https://doi.org/10.1016/j.cub.2020.08.086

Heyes, C. (1993). Imitation, culture and cognition. *Animal Behaviour*, *46*(5), 999–1010. https://doi.org/10.1006/anbe.1993.1281

James, L. S., Davies, R., Mori, C., Wada, K., & Sakata, J. T. (2020). Manipulations of sensory experiences during development reveal mechanisms underlying vocal learning biases in zebra finches. *Developmental Neurobiology*, *80*(3-4), 132–146. https://doi.org/10.1002/dneu.22754

Jesmer, B. R., Merkle, J. A., Goheen, J. R., Aikens, E. O., Beck, J. L., Courtemanch, A. B., Hurley, M. A., McWhirter, D. E., Miyasaki, H. M., Monteith, K. L., & Kauffman, M. J. (2018). Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. *Science*, *361*(6406), 1023–1025. https://doi.org/10.1126/science.aat0985

Kacelnik, A., & Krebs, J. R. (1983). The Dawn Chorus in the Great Tit (Parus Major): Proximate and Ultimate Causes. *Behaviour*, *83*(3-4), 287–308. https://doi.org/10.1163/156853983X00200

Kandler, A., Fogarty, L., & Karsdorp, F. (2023). The interplay between age structure and cultural transmission. *PLOS Computational Biology*, *19*(7), e1011297. https://doi.org/10.1371/journal.pcbi.1011297

Kempe, M., & Mesoudi, A. (2014). An experimental demonstration of the effect of group size on cultural accumulation. *Evolution and Human Behavior*, *35*(4), 285–290. https://doi.org/10.1016/j.evolhumbehav.2014.02.009

Kendal, R., Hopper, L. M., Whiten, A., Brosnan, S. F., Lambeth, S. P., Schapiro, S. J., & Hoppitt, W. (2015). Chimpanzees copy dominant and knowledgeable individuals: Implications for cultural diversity. *Evolution and Human Behavior*, *36*(1), 65–72. https://doi.org/10.1016/j.evolhumbehav.2014.09.002

Kimura, M., & Crow, J. F. (1964). The Number of Alleles That Can Be Maintained in a Finite Population. *Genetics*, *49*(4), 725–738.

Kirby, S., & Tamariz, M. (2021). Cumulative cultural evolution, population structure and the origin of combinatoriality in human language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1843), 20200319. https://doi.org/10.1098/rstb.2020.0319

Kleczkowska, K. (2015). Bird Communication in Ancient Greek and Roman Thought. *Maska*, *28*, 95–106.

Koenig, W., Dunn, H. K., & Lacy, L. Y. (1946). The Sound Spectrograph. *Journal of the Acoustical Society of America*, *18*(1), 19–49. https://doi.org/10.1121/1.1916342

Krebs, J. R. (1976). Habituation and song repertoires in the great tit. *Behav Ecol Sociobiol*, *1*(2), 215–227. https://doi.org/10.1007/BF00299198

Krebs, J., Ashcroft, R., & Webber, M. (1978). Song repertoires and territory defence in the great tit. *Nature*, *271*(5645), 539–542. https://doi.org/10.1038/271539a0

Kroeber, A. L., & Kluckhohn, C. (1952). Culture: A critical review of concepts and definitions. *Papers. Peabody Museum of Archaeology & Ethnology, Harvard University*, *47*(1), viii, 223–viii, 223.

Kroodsma, D. E. (2004). The diversity and plasticity of birdsong. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 108–131). Elsevier.

Lachlan, R. F., & Servedio, M. R. (2004). Song learning accelerates allopatric speciation. *Evolution*, *58*(9), 2049. https://doi.org/10.1554/03-708

Lachlan, R. F., & Slater, P. J. (2003). Song learning by chaffinches: How accurate, and from where? *Animal Behaviour*, *65*(5), 957–969. https://doi.org/10.1006/anbe.2003.2091

Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, *9*(1), 2417. https://doi.org/10.1038/s41467-018-04728-1

Lachlan, R. F., Verzijden, M. N., Bernard, C. S., Jonker, P. P., Koese, B., Jaarsma, S., Spoor, W., Slater, P. J., & Ten Cate, C. (2013). The progressive loss of syntactical structure in bird song along an Island colonization chain. *Current Biology*, *23*(19), 1896–1901. https://doi.org/10.1016/j.cub.2013.07.057

Laland, K. N., & Hoppitt, W. (2003). Do Animals Have Culture? *Evolutionary Anthropology*, *12*(3), 150–159. https://doi.org/10.1002/evan.10111

Laland, K. N., & Janik, V. M. (2006). The animal cultures debate. *Trends in Ecology & Evolution*, *21*(10), 542–547. https://doi.org/10.1016/j.tree.2006.06.005

Lambrechts, M. M., & Dhondt, A. A. (1990). A relationship between the composition and size of great tit song repertoires. *Animal Behaviour*, *39*(2), 213–218. https://doi.org/10.1016/S0003-3472(05)80865-6

Langin, K. M., Sillett, T. S., Morrison, S. A., & Ghalambor, C. K. (2017). Bill morphology and neutral genetic structure both predict variation in acoustic signals within a bird population. *Behavioral Ecology*, *28*(3), 866–873. https://doi.org/10.1093/beheco/arx051

Larsen, O. N., & Goller, F. (2002). Direct observation of syringeal muscle function in songbirds and a parrot. *Journal of Experimental Biology*, *205*(1), 25 LP –35.

Lind, H., Dabelsteen, T., & McGregor, P. K. (1996). Female great tits can identify mates by song. *Animal Behaviour*, *52*(4), 667–671. https://doi.org/10.1006/anbe.1996.0211

Liu, W. C., Gardner, T. J., & Nottebohm, F. (2004). Juvenile zebra finches can use multiple strategies to learn the same song. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(52), 18177–18182. https://doi.org/10.1073/pnas.0408065101

Liu, Y., Zollinger, S. A., & Brumm, H. (2020). Chronic exposure to urban noise during the vocal learning period does not lead to increased song frequencies in zebra finches. *Behav Ecol Sociobiol*, *75*(1), 3. https://doi.org/10.1007/s00265-020-02942-w

MacDonald, I. F., Kempster, B., Zanette, L., & MacDougall-Shackleton, S. A. (2006). Early nutritional stress impairs development of a song-control brain region in both male and female juvenile song sparrows (Melospiza melodia) at the onset of song learning. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1600), 2559–2564. https://doi.org/10.1098/rspb.2006.3547

Mace, R. (1987). The dawn chorus in the great tit Parus major is directly related to female fertility. *Nature*, *330*(6150), 745–746. https://doi.org/10.1038/330745a0

Marler, P., & Tamura, M. (1962). Song "Dialects" in Three Populations of White-Crowned Sparrows. *The Condor*, *64*(5), 368–377. https://doi.org/10.2307/1365545

Marler, P., & Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. *Science*, *146*(3650), 1483–1486. https://doi.org/10.1126/science.146.3650.1483

Martin, J. P., Doucet, S. M., Knox, R. C., & Mennill, D. J. (2011). Body size correlates negatively with the frequency of distress calls and songs of Neotropical birds. *Journal of Field Ornithology*, *82*(3), 259–268. https://doi.org/10.1111/j.1557-9263.2011.00329.x

Mayr, E. (1963, January). *Animal Species and Evolution*. Harvard University Press. https://doi.org/10.4159/harvard.9780674865327

McGregor, P. K., & Krebs, J. R. (1982a). Mating and song sharing in the great tit. *Nature*, *297*(5861), 60–61. https://doi.org/10.1038/297060a0

McGregor, P. K., & Krebs, J. R. (1982b). Song Types in a Population of Great Tits (Parus major): Their Distribution, Abundance and Acquisition by Individuals. *Behaviour*, *79*(2/4), 126–152.

McGregor, P. K., & Krebs, J. R. (1989). Song learning in adult great tits (Parus major): Effects of neighbours. *Behaviour*, *108*(1-2), 139–159. https://doi.org/10.1163/156853989X00105

McGregor, P. K., Krebs, J. R., & Perrins, C. M. (1981). Song Repertoires and Lifetime Reproductive Success in the Great Tit (Parus major). *The American Naturalist*, *118*(2), 149–159. https://doi.org/918

Mcgregor, P. K., Krebs, J. R., & Ratcliffe, L. M. (1983). The Reaction of Great Tits (Parus major) to Playback of Degraded and Undegraded Songs: The Effect of Familiarity with the Stimulus Song Type. *The Auk*, *100*(4), 898–906.

Mennill, D. J., Doucet, S. M., Newman, A. E., Williams, H., Moran, I. G., Thomas, I. P., Woodworth, B. K., & Norris, D. R. (2018). Wild Birds Learn Songs from Experimental Vocal Tutors. *Current Biology*, *28*, 3273–3278. https://doi.org/10.1016/j.cub.2018.08.011

Mercader, J., Barton, H., Gillespie, J., Harris, J., Kuhn, S., Tyler, R., & Boesch, C. (2007). 4,300-Year-old chimpanzee sites and the origins of percussive stone technology. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3043–3048. https://doi.org/10.1073/pnas.0607909104

Mesoudi, A., Chang, L., Dall, S. R., & Thornton, A. (2016). The Evolution of Individual and Cultural Variation in Social Learning. *Trends in Ecology and Evolution*, *31*(3), 215–225. https://doi.org/10.1016/j.tree.2015.12.012

Mets, D. G., & Brainard, M. S. (2017). Genetic variation interacts with experience to determine interindividual differences in learned song. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(2), 421–426. https://doi.org/10.1073/pnas.1713031115

Mets, D. G., & Brainard, M. S. (2019). Learning is enhanced by tailoring instruction to individual genetic differences. *eLife*, *8*, 1–15. https://doi.org/10.7554/eLife.47216

Mikula, P., Valcu, M., Brumm, H., Bulla, M., Forstmeier, W., Petrusková, T., Kempenaers, B., & Albrecht, T. (2020). A global analysis of song frequency in passerines provides no support for the acoustic adaptation hypothesis but suggests a role for sexual selection. *bioRxiv*, 2020.06.30.179812. https://doi.org/10.1101/2020.06.30.179812

Morin, O. (2016). Reasons to be fussy about cultural evolution. *Biology and Philosophy*, *31*(3), 447–458. https://doi.org/10.1007/s10539-016-9516-4

Nelson, D. A. (2000). A preference for own-subspecies' song guides vocal learning in a song bird. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(24), 13348–13353. https://doi.org/10.1073/pnas.240457797

Nelson, D. A. (2017). Geographical variation in song phrases differs with their function in white-crowned sparrow song. *Animal Behaviour*, *124*, 263–271. https://doi.org/10.1016/j.anbehav.2016.05.016

Nemeth, E., Pieretti, N., Zollinger, S. A., Geberzahn, N., Partecke, J., Miranda, A. C., & Brumm, H. (2013). Bird song and anthropogenic noise: Vocal constraints may explain why birds sing higher-frequency songs in cities. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1754), 20122798. https://doi.org/10.1098/rspb.2012.2798

Nottebohm, F., Stokes, T. M., & Leonard, C. M. (1976). Central control of song in the canary, Serinus canarius. *Journal of Comparative Neurology*, *165*(4), 457–486. https://doi.org/10.1002/cne.901650405

Nowicki, S., Searcy, W. A., & Peters, S. (2002). Quality of song learning affects female response to male bird song. *Proceedings of the Royal Society B: Biological Sciences*, *269*(1503), 1949–1954. https://doi.org/10.1098/rspb.2002.2124

Nunn, C. L., Thrall, P. H., Bartz, K., Dasgupta, T., & Boesch, C. (2009). Do transmission mechanisms or social systems drive cultural dynamics in socially structured populations? *Animal Behaviour*, *77*(6), 1515–1524. https://doi.org/10.1016/j.anbehav.2009.02.023

Odom, K. J., & Benedict, L. (2018). A call to document female bird songs: Applications for diverse fields. *The Auk*, *135*(2), 314–325. https://doi.org/10.1642/AUK-17-183.1

Odom, K. J., Hall, M. L., Riebel, K., Omland, K. E., & Langmore, N. E. (2014). Female song is widespread and ancestral in songbirds. *Nature Communications*, *5*, 1–6. https://doi.org/10.1038/ncomms4379

Päckert, M. (2018). Song: The Learned Language of Three Major Bird Clades, 75–94. https://doi.org/10.1007/978-3-319-91689-7_5

Paxton, K. L., Sebastián-González, E., Hite, J. M., Crampton, L. H., Kuhn, D., & Hart, P. J. (2019). Loss of cultural song diversity and the convergence of songs in a declining Hawaiian forest bird community. *Royal Society Open Science*, *6*(8). https://doi.org/10.1098/rsos.190719

Payne, R. B. (1982). Ecological Consequences of Song Matching: Breeding Success and Intraspecific Song Mimicry in Indigo Buntings. *Ecology*, *63*(2), 401–411. https://doi.org/10.2307/1938958

Perry, S., Baker, M., Fedigan, L., Gros-Louis, J., Jack, K., MacKinnon, K. C., Manson, J. H., Panger, M., Pyle, K., & Rose, L. (2003). Social conventions in wild white-faced capuchin monkeys: Evidence for traditions in a neotropical primate. *Current Anthropology*, *44*(2), 241–268. https://doi.org/10.1086/345825

Pike, T. W., & Laland, K. N. (2010). Conformist learning in nine-spined sticklebacks' foraging decisions. *Biology Letters*, *6*(4), 466–468. https://doi.org/10.1098/rsbl.2009.1014

Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, *409*(6817), 185–188. https://doi.org/10.1038/35051570

Podos, J., Southall, J. A., & Rossi-Santos, M. R. (2004). Vocal mechanics in Darwin's finches: Correlation of beak gape and song frequency. *Journal of Experimental Biology*, *207*(4), 607–619. https://doi.org/10.1242/jeb.00770

Podos, J., & Warren, P. S. (2007). The Evolution of Geographic Variation in Birdsong. *Advances in the Study of Behavior*, *37*, 403–458. https://doi.org/10.1016/S0065-3454(07)37009-5

Ram, Y., Liberman, U., & Feldman, M. W. (2018). Evolution of vertical and oblique transmission under fluctuating selection. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(6), E1174–E1183. https://doi.org/10.1073/pnas.1719171115

Renoult, J. P., & Mendelson, T. C. (2019). Processing bias: Extending sensory drive to include efficacy and efficiency in information processing. *Proceedings. Biological sciences*, *286*(1900), 20190165. https://doi.org/10.1098/rspb.2019.0165

Richerson, P. J., & Boyd, R. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press.

Ríos-Chelén, A. A., Salaberria, C., Barbosa, I., Macías Garcia, C., & Gil, D. (2012). The learning advantage: Bird species that learn their song show a tighter adjustment of song to noisy environments than those that do not learn. *Journal of Evolutionary Biology*, *25*(11), 2171–2180. https://doi.org/10.1111/j.1420-9101.2012.02597.x

Ritchie, G. R., Kirby, S., & Hawkey, D. J. (2008). Song learning as an indicator mechanism: Modelling the developmental stress hypothesis. *Journal of Theoretical Biology*, *251*(4), 570–583. https://doi.org/10.1016/j.jtbi.2007.12.013

Ritschard, M., van Oers, K., Naguib, M., & Brumm, H. (2012). Song Amplitude of Rival Males Modulates the Territorial Behaviour of Great Tits During the Fertile Period of Their Mates. *Ethology*, *118*(2), 197–202. https://doi.org/10.1111/j.1439-0310.2011.01999.x

Rivera-Gutierrez, H. F., Matthysen, E., Adriaensen, F., & Slabbekoorn, H. (2010). Repertoire sharing and song similarity between great tit males decline with distance between forest fragments. *Ethology*, *116*(10), 951–960. https://doi.org/10.1111/j.1439-0310.2010.01813.x

Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2010). Multiple signals for multiple messages: Great tit, Parus major, song signals age and survival. *Animal Behaviour*, *80*(3), 451–459. https://doi.org/10.1016/j.anbehav.2010.06.002

Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2011). Difficulties when Assessing Birdsong Learning Programmes under Field Conditions: A Re-Evaluation of Song Repertoire Flexibility in the Great Tit (M. Giurfa, Ed.). *PLoS ONE*, *6*(1), e16003. https://doi.org/10.1371/journal.pone.0016003

Roca, I. T., Desrochers, L., Giacomazzo, M., Bertolo, A., Bolduc, P., Deschesnes, R., Martin, C. A., Rainville, V., Rheault, G., & Proulx, R. (2016). Shifting song frequencies in response to anthropogenic noise: A meta-analysis on birds and anurans. *Behavioral Ecology*, *27*(5), 1269–1274. https://doi.org/10.1093/beheco/arw060

Ryan, M. J., & Brenowitz, E. A. (1985a). The Role of Body Size, Phylogeny, and Ambient Noise in the Evolution of Bird Song. *The American Naturalist*, *126*(1), 87–100. https://doi.org/10.1086/284398

Ryan, M. J., & Brenowitz, E. A. (1985b). The Role of Body Size, Phylogeny, and Ambient Noise in the Evolution of Bird Song. *The American Naturalist*, *126*(1), 87–100. https://doi.org/10.1086/284398

Sahlins, M. (1976). *The use and abuse of biology: An anthropological critique of sociobiology*. U Michigan Press.

Saldana, C., Fagot, J., Kirby, S., Smith, K., & Claidière, N. (2019). High-fidelity copying is not necessarily the key to cumulative cultural evolution: A study in monkeys and children. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2019.0729

Sasaki, T., & Biro, D. (2017). Cumulative culture can emerge from collective intelligence in animal groups. *Nature Communications*, *8*, 15049. https://doi.org/10.1038/ncomms15049

Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2021). Variation in vocal production learning across songbirds. *Phil. Trans. R. Soc. B*, *376*(1836), 20200257. https://doi.org/10.1098/rstb.2020.0257

Seddon, N. (2005). Ecological adaptation and species recognition drives vocal evolution in neotropical suboscine birds. *Evolution*, *59*(1), 200–215. https://doi.org/10.1111/j.0014-3820.2005.tb00906.x

Sheldon, B. C., Kruuk, L. E. B., & Alberts, S. C. (2022). The expanding value of long-term studies of individuals in the wild. *Nat Ecol Evol*, *6*(12), 1799–1801. https://doi.org/10.1038/s41559-022-01940-7

Sierro, J., de Kort, S. R., & Hartley, I. R. (2023). Sexual selection for both diversity and repetition in birdsong. *Nat Commun*, *14*(1), 3600. https://doi.org/10.1038/s41467-023-39308-5

Slabbekoorn, H., & Smith, T. B. (2002). Bird song, ecology and speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*(1420), 493–503. https://doi.org/10.1098/rstb.2001.1056

Slagsvold, T., Sætre, G.-P., & Dale, S. (1994). Dawn Singing in the Great Tit (Parus Major): Mate Attraction, Mate Guarding, or Territorial Defence? *Behaviour*, *131*(1-2), 115–138. https://doi.org/10.1163/156853994X00244

Slater, P. J. (1986). The cultural transmission of bird song. *Trends in Ecology and Evolution*, *1*(4), 94–97. https://doi.org/10.1016/0169-5347(86)90032-7

Slater, P. J. (2003). Fifty years of bird song research: A case study in animal behaviour. *Animal Behaviour*, *65*(4), 633–639. https://doi.org/10.1006/anbe.2003.2051

Slater, P. J. B., & Lachlan, R. F. (2003). Is Innovation in Bird Song Adaptive? In S. M. Reader & K. N. Laland (Eds.), *Animal Innovation* (pp. 117–135). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198526223.003.0005

Smaldino, P. E., & Epstein, J. M. (2015). Social conformity despite individual preferences for distinctiveness. *Royal Society Open Science*, *2*(3). https://doi.org/10.1098/rsos.140437

Spiegel, O., Leu, S. T., Sih, A., & Bull, C. M. (2016). Socially interacting or indifferent neighbours? Randomization of movement paths to tease apart social preference and spatial constraints. *Methods in Ecology and Evolution*, *7*(8), 971–979. https://doi.org/10.1111/2041-210X.12553

Spurgin, L., Bosse, M., Adriaensen, F., Albayrak, T., Barboutis, C., Belda, E., Bushuev, A., Cecere, J., Charmantier, A., Cichon, M., Dingemanse, N., Doligez, B., Eeva, T., Erikstad, K. E., Fedorov, V., Griggio, M., Heylen, D., Hille, S., Hinde, C., … Slate, J. (2019). The great tit HapMap project: A continental-scale analysis of genomic variation in a songbird. *bioRxiv*, 561399. https://doi.org/10.1101/561399

Suthers, R. A. (2004). How birds sing and why it matters. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 272–295). https://doi.org/10.1016/B978-012473070-0/50012-8

Tchernichovski, O., Eisenberg-Edidin, S., & Jarvis, E. D. (2021). Balanced imitation sustains song culture in zebra finches. *Nat Commun*, *12*(1), 2562. https://doi.org/10.1038/s41467-021-22852-3

Tchernichovski, O., Feher, O., Fimiarz, D., & Conley, D. (2017). How social learning adds up to a culture: From birdsong to human public opinion. *Journal of Experimental Biology*, *220*(1), 124–132. https://doi.org/10.1242/jeb.142786

ten Cate, C. (2021). Re-evaluating vocal production learning in non-oscine birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1836), 20200249. https://doi.org/10.1098/rstb.2020.0249

ten Cate, C., & Rowe, C. (2007). Biases in signal evolution: Learning makes a difference. *Trends in Ecology and Evolution*, *22*(7), 380–387. https://doi.org/10.1016/j.tree.2007.03.006

Thorpe, W. H. (1958). The Learning of Song Patterns By Birds, With Especial Reference To the Song of the Chaffinch Fringilla Coelebs. *Ibis*, *100*(4), 535–570. https://doi.org/10.1111/j.1474-919X.1958.tb07960.x

Tobias, J. A., Aben, J., Brumfield, R. T., Derryberry, E. P., Halfwerk, W., Slabbekoorn, H., & Seddon, N. (2010). Song divergence by sensory drive in amazonian birds. *Evolution*, *64*(10), 2820–2839. https://doi.org/10.1111/j.1558-5646.2010.01067.x

Tobias, J. A., Cornwallis, C. K., Derryberry, E. P., Claramunt, S., Brumfield, R. T., & Seddon, N. (2014). Species coexistence and the dynamics of phenotypic evolution in adaptive radiation. *Nature*, *506*(7488), 359–363. https://doi.org/10.1038/nature12874

Tobias, J. A., & Seddon, N. (2009). Signal design and perception in hypocnemis antbirds: Evidence for convergent evolution via social selection. *Evolution*, *63*(12), 3168–3189. https://doi.org/10.1111/j.1558-5646.2009.00795.x

Trehub, S. E. (2015). Cross-cultural convergence of musical features. *Proceedings of the National Academy of Sciences*, *112*(29), 8809–8810. https://doi.org/10.1073/pnas.1510724112

Van De Waal, E., Renevey, N., Favre, C. M., & Bshary, R. (2010). Selective attention to philopatric models causes directed social learning in wild vervet monkeys. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1691), 2105–2111. https://doi.org/10.1098/rspb.2009.2260

van Leeuwen, E. J., Kendal, R. L., Tennie, C., & Haun, D. B. (2015). Conformity and its look-a-likes. *Animal Behaviour*, *110*, e1–e4. https://doi.org/10.1016/j.anbehav.2015.07.030

Verzijden, M. N., ten Cate, C., Servedio, M. R., Kozak, G. M., Boughman, J. W., & Svensson, E. I. (2012). The impact of learning on sexual selection and speciation. *Trends in Ecology and Evolution*, *27*(9), 511–519. https://doi.org/10.1016/j.tree.2012.05.007

Whitehead, H., & Richerson, P. J. (2009). The evolution of conformist social learning can cause population collapse in realistically variable environments. *Evolution and Human Behavior*, *30*(4), 261–273. https://doi.org/10.1016/j.evolhumbehav.2009.02.003

Whiten, A. (2019). Cultural Evolution in Animals. *Annual Review of Ecology, Evolution, and Systematics*, *50*(1), 1–22. https://doi.org/10.1146/annurev-ecolsys-110218-025040

Whiten, A., Ayala, F. J., Feldman, M. W., & Laland, K. N. (2017). The extension of biology through culture. *Pro-*

ceedings of the National Academy of Sciences, *114*(30), 7775–7781. https://doi.org/10.1073/pnas.1707630114

Williams, H., Levin, I. I., Norris, D. R., Newman, A. E. M., & Wheelwright, N. T. (2013). Three decades of cultural evolution in Savannah sparrow songs. *Animal Behaviour*, *85*(1), 213–223. https://doi.org/10.1016/j.anbehav.2012.10.028

Williams, H., Robins, C. W., Norris, D. R., Newman, A. E. M., Freeman-Gallant, C. R., Wheelwright, N. T., & Mennill, D. J. (2019). The buzz segment of Savannah sparrow song is a population marker. *J Ornithol*, *160*(1), 217–227. https://doi.org/10.1007/s10336-018-1611-7

Williams, J. M., & Slater, P. J. (1990). Modelling bird song dialects: The influence of repertoire size and numbers of neighbours. *Journal of Theoretical Biology*, *145*(4), 487–496. https://doi.org/10.1016/S0022-5193(05)80483-7

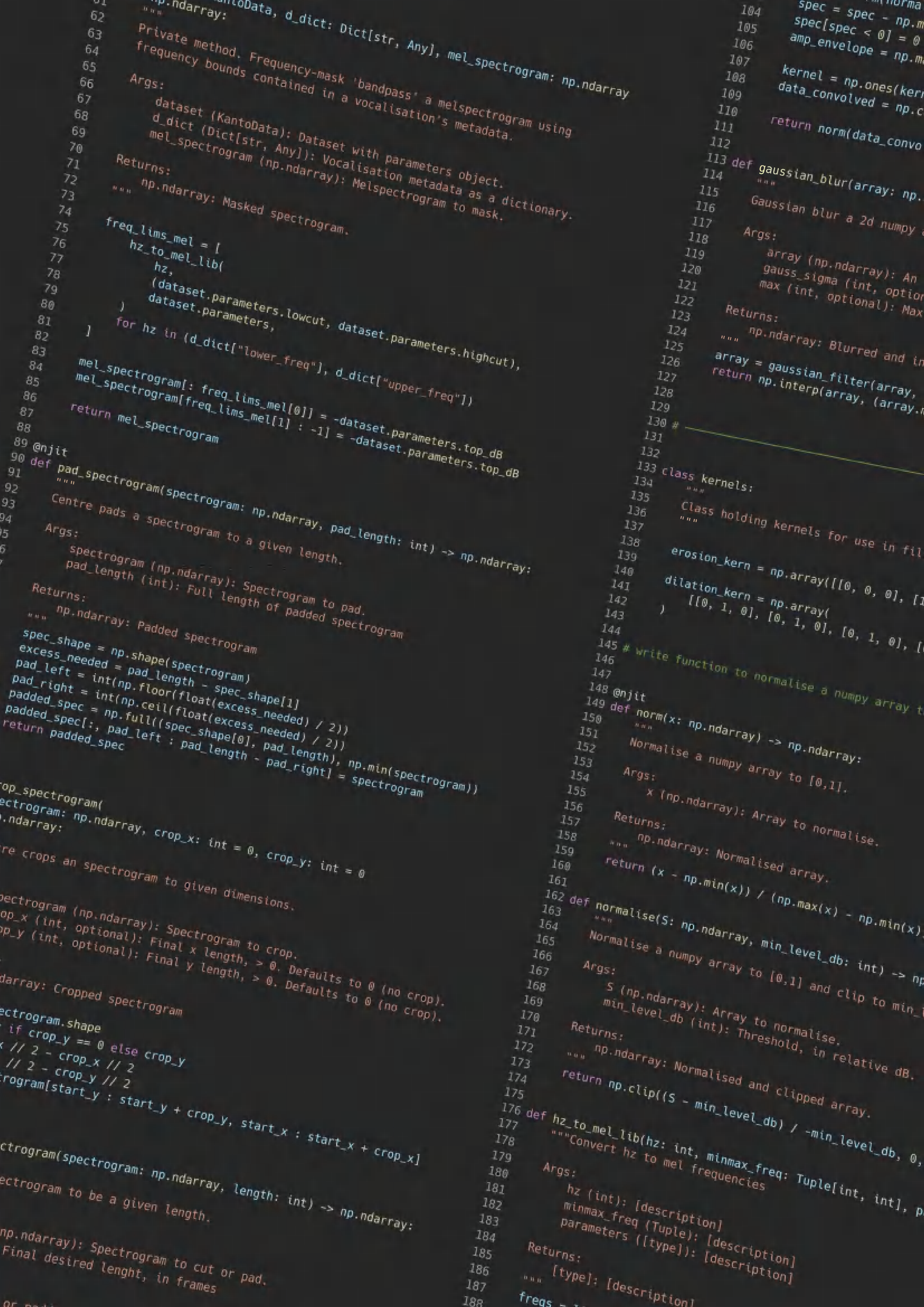
Winandy, G. S. M., Félix, R. P., Sacramento, R. A., Mascarenhas, R., Batalha-Filho, H., Japyassú, H. F., Izar, P., & Slabbekoorn, H. (2021). Urban Noise Restricts Song Frequency Bandwidth and Syllable Diversity in Bananaquits: Increasing Audibility at the Expense of Signal Quality. *Frontiers in Ecology and Evolution*, *9*.

Yeh, D. J., & Servedio, M. R. (2015). Reproductive isolation with a learned trait in a structured population. *Evolution*, *69*(7), 1938–1947. https://doi.org/10.1111/evo.12688

Zann, R. (1993). Structure, Sequence and Evolution of Song Elements in Wild Australian Zebra Finches. *The Auk*, *110*(4), 702–715. https://doi.org/10.2307/4088626

Zimmerman, H. D., Ramsay, S. M., Mesias, V., Mora, M., Murray, B. W., & Otter, K. A. (2016). Evolution of white-throated sparrow song: Regional variation through shift in terminal strophe type and length. *Behaviour*, *153*(15), 1839–1861. https://doi.org/10.1163/1568539X-00003394

Zirin, R. A. (1980). Aristotle's Biology of Language. *Transactions of the American Philological Association*, *110*, 325–347. https://doi.org/10.2307/284226



```
              oData, d_dict: Dict[str, Any], mel_spectrogram: np.ndarray
 p.ndarray:
    """
    Private method. Frequency-mask 'bandpass' a melspectrogram using
    frequency bounds contained in a vocalisation's metadata.
    Args:
        dataset (KantoData): Dataset with parameters object.
        d_dict (Dict[str, Any]): Vocalisation metadata as a dictionary.
        mel_spectrogram (np.ndarray): Melspectrogram to mask.
    Returns:
        np.ndarray: Masked spectrogram.
    """
    freq_lims_mel = [
        hz_to_mel_lib(
            hz,
            (dataset.parameters.lowcut, dataset.parameters.highcut),
            dataset.parameters,
        )
        for hz in (d_dict["lower_freq"], d_dict["upper_freq"])
    ]
    mel_spectrogram[: freq_lims_mel[0]] = -dataset.parameters.top_dB
    mel_spectrogram[freq_lims_mel[1] : -1] = -dataset.parameters.top_dB
    return mel_spectrogram

@njit
def pad_spectrogram(spectrogram: np.ndarray, pad_length: int) -> np.ndarray:
    """
    Centre pads a spectrogram to a given length.
    Args:
        spectrogram (np.ndarray): Spectrogram to pad.
        pad_length (int): Full length of padded spectrogram
    Returns:
        np.ndarray: Padded spectrogram
    """
    spec_shape = np.shape(spectrogram)
    excess_needed = pad_length - spec_shape[1]
    pad_left = int(np.floor(float(excess_needed) / 2))
    pad_right = int(np.ceil(float(excess_needed) / 2))
    padded_spec = np.full((spec_shape[0], pad_length), np.min(spectrogram))
    padded_spec[:, pad_left : pad_length - pad_right] = spectrogram
    return padded_spec


rop_spectrogram(
ectrogram: np.ndarray, crop_x: int = 0, crop_y: int = 0
.ndarray:

re crops an spectrogram to given dimensions.

pectrogram (np.ndarray): Spectrogram to crop.
op_x (int, optional): Final x length, > 0. Defaults to 0 (no crop).
op_y (int, optional): Final y length, > 0. Defaults to 0 (no crop).

darray: Cropped spectrogram

ectrogram.shape
 if crop_y == 0 else crop_y
 // 2 - crop_x // 2
 // 2 - crop_y // 2
rogram[start_y : start_y + crop_y, start_x : start_x + crop_x]


ctrogram(spectrogram: np.ndarray, length: int) -> np.ndarray:

ectrogram to be a given length.

np.ndarray): Spectrogram to cut or pad.
Final desired lenght, in frames

or padd
```

```
            (norma
        spec = spec - np.m
        spec[spec < 0] = 0
        amp_envelope = np.m

        kernel = np.ones(ker
        data_convolved = np.c

        return norm(data_convo

def gaussian_blur(array: np.
    """
    Gaussian blur a 2d numpy a
    Args:
        array (np.ndarray): An
        gauss_sigma (int, optio
        max (int, optional): Max
    Returns:
        np.ndarray: Blurred and in
    """
    array = gaussian_filter(array,
    return np.interp(array, (array.


#


class kernels:
    """
    Class holding kernels for use in fil
    """

    erosion_kern = np.array([[0, 0, 0], [1
    dilation_kern = np.array(
        [[0, 1, 0], [0, 1, 0], [0, 1, 0], [
    )

# write function to normalise a numpy array t


@njit
def norm(x: np.ndarray) -> np.ndarray:
    """
    Normalise a numpy array to [0,1].
    Args:
        x (np.ndarray): Array to normalise.
    Returns:
        np.ndarray: Normalised array.
    """
    return (x - np.min(x)) / (np.max(x) - np.min(x)

def normalise(S: np.ndarray, min_level_db: int) -> np
    """
    Normalise a numpy array to [0,1] and clip to min_
    Args:
        S (np.ndarray): Array to normalise.
        min_level_db (int): Threshold, in relative dB.
    Returns:
        np.ndarray: Normalised and clipped array.
    """
    return np.clip((S - min_level_db) / -min_level_db, 0,
def hz_to_mel_lib(hz: int, minmax_freq: Tuple[int, int], p
    """Convert hz to mel frequencies
    Args:
        hz (int): [description]
        minmax_freq (Tuple): [description]
        parameters ([type]): [description]
    Returns:
        [type]: [description]
    """
    freqs = l
```

2. PYKANTO: A PYTHON LIBRARY TO ACCELERATE RESEARCH ON WILD BIRD SONG

Nilo Merino Recalde [1,*]

[1] Edward Grey Institute, Department of Biology, University of Oxford, Oxford, UK
*Corresponding author: nilo.recalde@biology.ox.ac.uk

ABSTRACT

Studying the vocalizations of wild animals can be a challenge due to the limitations of traditional computational methods, which often are time-consuming and lack reproducibility. Here, I present pykanto, a new software package that provides a set of tools to build, manage, and explore large sound databases. It can automatically find discrete units in animal vocalizations, perform semi-supervised labelling of individual repertoires with a new interactive web app, and feed data to deep learning models to study things like individual signatures and acoustic similarity between individuals and populations. To demonstrate its capabilities, I put the library to the test on the vocalizations of male great tits in Wytham Woods, near Oxford, UK. The results show that the identities of individual birds can be accurately determined from their songs and that the use of pykanto improves the efficiency and reproducibility of the process.

Keywords: bioacoustics; animal vocalizations; python; bird song

INTRODUCTION

Collecting large amounts of acoustic data from wild bird populations has traditionally been very difficult. Due to technical limitations, studies have often been constrained to tens of individuals and tens, or at best hundreds of vocalizations. But this has changed rapidly within the last decade: compact and economic autonomous recording units, such as the AudioMoth (Hill et al., 2019), now make it possible to collect orders of magnitude more data from many more individuals at once—and to do so much more cheaply. As a direct consequence, many of the computational tools traditionally employed with bioacoustic data have quickly become obsolete: they require manual curation, segmentation, and labelling of data, which are extremely time-consuming and prone to errors.

To illustrate this point, as part of our research on a wild population of great tits (*Parus major*), we record around 50,000 songs every year, which translates to well over half a million discrete acoustic units. Any analysis that required finding, labelling, and characterizing them, if done manually—as it is still often the case in wild bird vocalization research (Beecher et al., 2020; Demko & Mennill, 2018; McLean & Roach, 2020; Pipek et al., 2018; Youngblood & Lahti, 2022)—would take a very long time to complete. This bottleneck, in turn, severely limits researchers' ability to ask questions that require large datasets to answer—such as those about social learning, vocal development, large-scale cultural diversity, and the syntactic organization of animal vocalizations (Aplin, 2019; Kollmorgen et al., 2020; Lachlan et al., 2018; Sainburg et al., 2019).

In addition to concerns over the scalability of existing data analysis pipelines, there is now a demand for tools that are freely accessible and promote transparent, reproducible research. Existing proprietary software, such as the widely used Raven Pro (up to $800; K. Lisa Yang Center for Conservation Bioacoustics, 2019) and Avisoft-SASLab Pro (up to $2,835; Specht, 2002) are difficult to reconcile with contemporary data science practices that rely on open-source programming languages such as R (R Core Team, 2021) and Python (van Rossum, 1995). There exist some excellent open-source options, such as Luscinia (Lachlan, 2016), Sound Analysis Pro (Tchernichovski et al., 2000) and the more recent Koe (Fukuzawa et al., 2020). However, these were generally not designed to cope with large volumes of data, and their reliance on point-and-click graphical user interfaces limits their flexibility and hinders reproducibility.

As a response to the need for scalable and open-source tools for vocalization data analysis and related issues, the field of bioacoustics has recently started to experiment with a new suite of methods based on deep-learning artificial neural network architectures, the same that excel at, for example, com-

You can use this app to explore and label the vocal repertoire of an individual animal or a population. Hover the mouse cursor over a point in the acoustic space to see its corresponding spectrogram.

Use the lasso tool to select multiple points at once. | ⇧+ **CLICK** to select multiple points one by one | ⇧+ to make multiple lasso selections.

// **0 out of 1 done** // Use the 'Label' button to change the label of the selected points. Label data are automatically saved when you click 'Next'.
Points labelled '-1' (red asterisks) are considered noise.

Figure 2.1. Interactive web app to review and correct cluster assignment.
Interface of the interactive web app in pykanto. This app can be used to explore datasets as well as to review and correct automatically assigned class labels in bulk.

puter vision and speech recognition tasks (Stowell, 2021). Segmentation and annotation pipelines based on deep neural networks have already been shown to work well in laboratory settings, where three conditions hold: i) acoustic data have a high signal-to-noise ratio, ii) there are orders of magnitude more examples per vocalization type than there are vocalization types, and iii) vocalizations are produced by relatively few individuals (fewer than ten to a few tens) that do so in a stereotyped manner (Coffey et al., 2019; Cohen et al., 2022; Steinfath et al., 2021). Unfortunately, none of these conditions tend to be the case in field studies, and this creates a barrier to the adoption of new methods by researchers working with natural populations.

This is the context in which I present `pykanto` (pronounced pɪɰ'kæntəʊ). This software library was born of three needs, which can be summarized as follows.

First, it needed to provide the infrastructure necessary to catalogue, explore and label large acoustic datasets collected in often suboptimal field conditions.

Second, it had to serve as a flexible starting point that would allow researchers to perform both traditional analyses (such as extracting hand-picked features from the vocalizations) and to use machine learning algorithms to learn low-dimensional representations of the data (Goffinet et al., 2021; Kollmorgen et al., 2020; Morfi et al., 2021; Sainburg et al., 2020), train classifiers, or detect vocalizations in unseen recordings (Cohen et al., 2022; Kahl et al., 2021; Stowell & Plumbley, 2014).

Third, I wanted to build a tool that was free, open source, followed sustainable software practises, and geared towards computational reproducibility and transparency.

PYKANTO: IMPLEMENTATION

`pykanto` is a software library designed to streamline the process of analysing animal vocalizations. It is programmed in Python and offers various modules to assist users in their work (see Figure 2.2). The central module is `pykanto.dataset`, which serves as a database for vocalizations and includes methods to visualize, segment, and label them. The `pykanto.signal` module provides tools for signal processing and creating spectrograms, while `pykanto.parameters` contains classes and functions for managing parameters. The web application `pykanto.app` allows users to explore and label large numbers of vocalizations (Figure 2.1) and `pykanto.plot` provides functions for plotting spectrograms. Finally, `pykanto.utils` includes parsers, I/O tools, custom typing, and general computing functions. The documentation for `pykanto` is available at nilomr.github.io/pykanto.

Dependencies

`pykanto` was written in Python 3.8 and tested in Python 3.8, 3.9 and 3.10. Its interactive web application also relies on JavaScript, HTML, and CSS. External dependencies are automatically downloaded during package installation (see the `pyproject.toml` file for a full list of dependencies).

API and documentation

`pykanto` is a well-documented code library, making it easier to use and contribute to its development. The methods and functions in `pykanto` have clear and concise documentation, including type annotations and descriptions of their intended use. Its API (Application Programming Interface) reference, along with tutorials and practical examples, can be found in the online documentation at nilomr.github.io/pykanto.

Reproducibility and open research

`pykanto` encourages the user to create reproducible data science projects. For example, one of its modules is dedicated to creating consistent project structures, inspired by popular utilities such as cookiecutter. Using the library requires writing simple scripts in Python, which allows every step of the research, from data ingestion to eventual model training and reporting, to be explicitly reproduced. The documentation includes a complete user guide with examples of best practices.

The input and output files use open data formats, and all code is available under the MIT licence (a simple and very permissive licence). Where applicable, we have followed the guidelines and recommendations of the Software Sustainability Institute, a UK-based facility dedicated to research software sustainability (software.ac.uk).

Many of the processes that `pykanto` carries out are computationally intensive, such as calculating spectrograms, performing operations on large arrays, and running dimensionality reduction and clustering algorithms. High-level, interpreted languages—like R or Python—are notoriously slow: where possible, we have optimized performance by both a) translating functions to optimized machine code at runtime using Numba (Lam et al., 2015) and b) parallelizing tasks using Ray, a state-of-the-art platform for distributed computing (Moritz et al., 2018). As an example, the `segment_into_units()` function can find and segment 20,000 discrete acoustic units in approximately 16s on a desktop, 8-core machine; a dataset with over half a million (556.472) units takes 132s on a standard 48-core compute node. If `pykanto` detects a suitable GPU unit and the optional dependencies are installed, algorithms such as UMAP (McInnes et al., 2018) switch to their GPU implementation, which provides a 15-100x speed-up (Nolet et al., 2021; Raschka et al., 2020). The library has a module dedicated to making it easy for users to run their scripts in a high-performance computing context (for example, a university compute cluster), and its documentation includes examples of configuration and submission scripts.

Limitations

This final section discusses some of the main limitations of `pykanto`. Although it will hopefully offer a flexible solution for researchers, it is also limited in important ways.

Limitation 1 Vocalization unit segmentation via the very simple amplitude thresholding algorithm will not work well with species whose vocalizations vary greatly in amplitude, or with very noisy datasets. In those cases, and depending on data volume, segmentation might better be performed either manually or in a semi-automated way. For example, one could use *chipper* (Searfoss et al., 2020) or train a neural net-

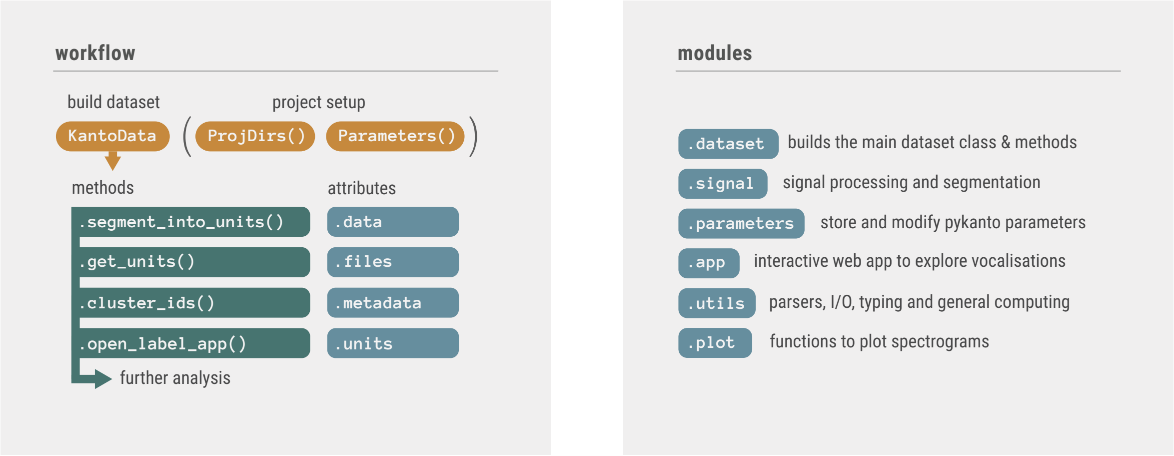


Figure 2.2. General structure of the library: main class and modules.
`pykanto` is written around a central dataset class, `KantoData`, which provides methods to segment, visualize and label vocalizations. The library contains six modules with functions and classes to carry out common tasks in animal vocalization analysis.

work like TweetyNet, (Cohen et al., 2022) on a manually annotated subset of the data.

Limitation 2 `pykanto` has been tested on species that produce vocalizations made up of a small or moderate number of different but distinct elements (variously referred to as notes or syllables). It will be useful for researchers working with any species, but the automatic part of the clustering process will work increasingly poorly with those that have many, very variable elements. This is true of any clustering method: they will fail or produce spurious results if variation in the data is continuous.

Limitation 3 The library does not include methods to train models intended to find analysable vocalizations in long recordings of entire soundscapes. This is a particularly challenging problem (Priyadarshani et al., 2018) without a universal solution. However, `pykanto` can be used to generate and organize the training data required by these models (Kahl et al., 2021; Stowell & Plumbley, 2014; Stowell et al., 2019), and to work with their output annotations.

Limitation 4 `pykanto` is intended as a flexible solution for managing and preparing animal vocalization data for further analysis. It provides tools that can save researchers a great deal of time while making analysis pipelines more reproducible. However, it does not implement any specific analysis or feature extraction methods, since these will vary greatly by use case. This means that researchers using the library as part of their work will need to either have or develop familiarity with bioacoustic analysis and scripting in Python.

USING PYKANTO: CAN INDIVIDUAL BIRDS BE IDENTIFIED FROM THEIR SONGS?

I now provide a worked example of how `pykanto` can be used to help answer real research questions about vocalizations—bird song in this case:

Introduction

Great tits are small, short-lived birds (average lifespan: 1.9 years) that sing acoustically simple yet highly diverse songs. In Wytham Woods, Oxfordshire (UK), a population of these birds has been the focus of a long-term study that is now in its 75th year. For the past three years, I have recorded the song repertoires of hundreds of individual males when they sing close to their nest before their partner begins laying. With the help of these data, we are trying to answer questions about song learning and cultural change in natural populations.

To do this we first need to know which individuals are present in the breeding population for the first time, and which were already around in previous years. However, individual survival over the winter months is low and detection by traditional means—such as mist-netting or identification in the nest—is imperfect. So we would first like to test whether individual birds can be identified based on their songs alone, and then quantify how much variation in song types occurs within and between years.

Our example dataset consists of 5293 songs from 12 males that were known (from physical recaptures) to be present in the breeding population in two different years, 2020 and 2021. Although this is a small subset of our data, it is large enough that it would still take weeks to process and analyse using traditional methods. We demonstrate the use of `pykanto` to a) organize, segment and label the dataset, and b) prepare it so that we can train a deep neural network to recognize the bird's song types. The entire process, which takes under an hour to complete, can be computationally reproduced using its dedicated repository. The repository includes raw data, auxiliary scripts and detailed instructions. Below is a short narrative description of the process.

Data collection

Most great tits in our population nest in nest boxes with known locations. Every year, fieldworkers record the identities of breeding males and females, clutch initiation and egg-hatching dates, clutch size, and fledgling success using standardized protocols. A significant number of birds in the population are fitted with a unique British Trust for Ornithology (BTO) metal leg ring as nestlings or adults. During the breeding season (March to June), great tit pairs are socially monogamous and protect territories around their nest boxes (Hinde, 1952).

We collected data during the breeding seasons of 2020 and 2021, from early April to late May, using a dense sampling design with multiple recorders placed in nest boxes throughout the study site. Fieldworkers checked every nest box in the study site at least once a week before and during egg laying, which can last from one to 14 days (Perrins, 1965). Once a nest box was believed to be in use by a great tit, we placed an autonomous sound recorder in its vicinity, either in the same tree or in a suitable neighbouring tree. We left each recorder in the same location for at least three consecutive days before moving it to a different nest box. Throughout the recording period, we relocated 20 recorders every day.

We used 60 AudioMoth recorders (Hill et al., 2019) in 2021 and 30 in 2020, which were housed in waterproof custom-built enclosures. Recording began about an hour before sunrise (from 05:36 to 04:00 UTC during the recording period) and consisted of seven 60-minute recordings with a sample rate of 48 kHz. Since the recording process was automated, there is a possibility that some of the songs recorded in the immediate vicinity of a given nest box do not belong to the focal bird. To reduce the risk of false positives, we discarded recordings with more than one vocalizing bird, unless one was distinctly louder than the others. We also discarded all songs with a maximum amplitude below $-16dB$, calculated as $20\log_{10}\left(\frac{A}{A_0}\right)$, with $A = 5000$ and $A_0 = 32767$ (the maximum value for 16-bit digital audio). This threshold was determined from the observation that, in cases where we had simultaneous recordings of close neighbours from the centres of their respective territories, an amplitude cut-off greater than 4000 always separated a focal bird from its neighbours. It should be noted that these values are not calibrated and are relative to the recording equipment and settings used, as well as other factors such as sound directionality and vegetation cover.

Running the analysis

Installation

`pykanto` can be used outside a virtual environment, but this is not encouraged. Using clean environments for each project will allow you to avoid dependency issues. Once inside a new environment with Python 3.8 or above, you can install `pykanto` by simply running `pip install pykanto`, then install the package containing this example. See detailed installation and use instructions in the `.README`.

Creating a new project and dataset

Our first step will be to define a directory structure for our project and a `ProjDirs` object to hold everything together. Then, we can test and set adequate parameters for our dataset. These include things like low- and high-cut filters, spectrogram settings, amplitude thresholding, and whether the analysis will be carried out at the song or note level. The data folder in the project already contains `.wav` audio files and their corresponding `.json` with annotations, so we can create a `KantoData` instance: this will be our database.

Segmenting songs and using the interactive app

Then, using the `.segment_into_units()` method, we find segment onsets, offsets, unit and silence durations and add them to `KantoData.data`, the main data frame in our database object. At this point, we could already carry out most of the analyses common in the bird song literature, for example, by extracting some simple acoustic parameters from the segmented data. Instead, we want to preserve all the temporal and spectral information that is available in the spectrograms to train a more accurate classifier.

The next step is to compute and store spectrograms for each unit under examination, and then reduce their dimensionality and group them into clusters. This can be achieved by using the `.get_units()` and `.cluster_ids()` methods, which employ algorithms such as UMAP (McInnes et al., 2018) and HDBSCAN (McInnes et al., 2017). Afterwards, we can launch the interactive web app by calling `.open_label_app()`. Using this app, we can review the automatic labels for up to tens of thousands of vocalizations at once, splitting or combining clusters as needed. Once completed, we will have a fully annotated dataset, which can be divided into training and testing sets and exported as labelled spectrograms using `pykanto`.

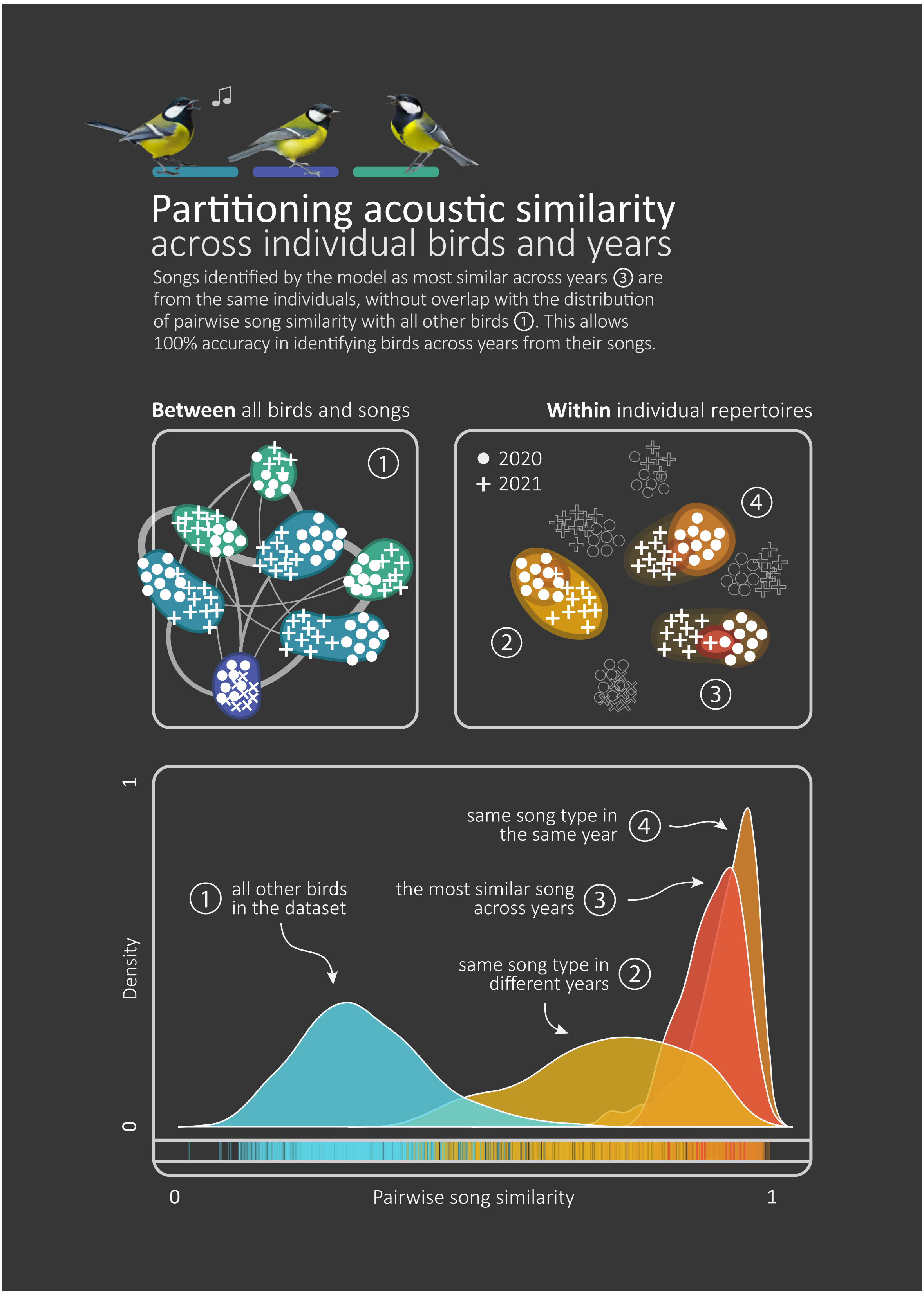
Training a convolutional neural network classifier

Our goal is to generate compressed representations of songs that can facilitate comparisons and identification of those sung by the same individual, even in the presence of variations in performance and noise. To do this, we can train a model to distinguish between 44 different song types, which are categorically distinct within individual song repertoires, so that we can try to assign song IDs back to individual birds. This is a small example; see Chapter 3 and specially Chapter 4 for a more precise bird re-identification method used on a much larger dataset.

In this example, we use weights from a pre-trained ResNet50 backbone (He et al., 2015) and gradually unfreeze the earlier layers of the network during training. By doing so, we can fine-tune the network to attain better performance in our task, while still benefiting from the weights learned on a much larger dataset (Zhuang et al., 2021).

The distribution of song sample sizes per individual approximately follows a power law, so there is a very large amount of data for a few birds and very little for most. This imbalance of data is problematic for learning algorithms, as they may develop a bias toward larger classes and perform poorly on rarer ones. There are different ways to deal with this (see, e.g., Krawczyk, 2016; Thabtah et al., 2020); however, to keep things simple, here we will just undersample majority classes so that all birds have the smallest common sample size for each song type.

Background noise can also bias our analysis. Each bird's acoustic environment is unique, and the network may learn to distinguish between songs based on this noise rather than the signal of interest. To address this issue, we remove most background noise by thresholding the spectrograms, taking advantage of the difference in amplitude between the focal bird singing near the recording and its acoustic background. Additionally, we apply a series of data augmentation techniques during the training process to prevent over-fitting and ensure that the algorithm does not memorize irrelevant or highly variable ele-



Partitioning acoustic similarity across individual birds and years

Songs identified by the model as most similar across years ③ are from the same individuals, without overlap with the distribution of pairwise song similarity with all other birds ①. This allows 100% accuracy in identifying birds across years from their songs.

Figure 2.3. Partitioning acoustic similarity between and within birds.
(From left to right and top to bottom.) First, we calculate the pairwise similarity between the songs of all birds, which provides a baseline distribution of similarity in the population (1). Then, we compare song types from the same bird in different years (2), find the pair of songs within song types that are most similar across years (3), and compare songs within the same year (4). The probability density estimates in the bottom panel show how pairwise similarities in (3) allow us to re-identify birds across as they do not overlap with any other birds (1).

ments. These techniques include semi-random cropping in the time domain, blurring and sharpening, random erasing of parts of the spectrogram, and contrast and brightness changes. After training the model, we can verify that background noise is not causing bias by checking whether songs recorded in the same location are classified as more similar than expected by chance. In addition, we can create class activation maps to identify the regions of the spectrograms used by the model to generate predictions.

Once the model is trained, we can assess its ability to classify unseen songs drawn from the held-out dataset, which can reach an accuracy of around 92%. As described below, most of the remaining 8% can be attributed to confusion between the same song types sung by the same birds in different years.

Finally, we can use the model to extract a compressed representation of each song in the entire dataset. This is achieved by passing each song through the trained network and obtaining the output from the last hidden layer. This output consists of a feature vector that captures the most relevant information to distinguish between different types of songs; these feature vectors can then be used to determine whether two songs are very similar. In our case, we do this by taking the complement of the cosine distance between each pair to build a similarity matrix.

The repository that supports this paper contains a streamlined way of doing this using PyTorch (2019) and PyTorch Lightning (2019) that can be easily adapted for use with other datasets.

Results & Discussion

After calculating the similarity scores between all pairs of songs, we grouped them according to song type, the bird that sang them, and the year they were sung. For each bird in the first year, we identified the individual who sang the song with the highest similarity to any within its repertoire during the following year. We found that we were able to successfully identify the correct bird in all cases, even though the baseline probability was only 2.27% (1 out of 44 song types). This suggests that we would have been able to re-identify the individuals even if they had not been observed or captured again.

As shown in Figure 2.3, the highest similarity values correspond to comparisons of the same song types within years and birds. The similarity between the same song types sung by the same bird across different years is consistently higher than that between different birds, even when some song types were shared by individuals: this means that individual vocal signatures are at least partly maintained across their lifespan.

The conclusions drawn from this analysis are limited by the small size of the dataset: including more birds would likely lead to noisier results, as it increases the chances of finding a second bird with even more similar songs. Nonetheless, in combination with other information (such as spatial location), they might allow high-confidence identification of individuals between years without physical capture.

This example illustrates how `pykanto` can be used to help address a specific research question. The model-based feature vectors used to describe each song can be imported back into the `KantoData` database as a new column, enabling a wide range of research possibilities while maintaining a clear project structure.

DATA AVAILABILITY

We distribute `pykanto` with three sample datasets that are used to run unit tests and as examples in the documentation (Merino Recalde, 2023b).

Great tit songs 20 songs recorded from male birds during the dawn chorus in a population in Oxford, UK. Recorded by the author and accessible at pykanto/data/great_tit.

European storm-petrel purr songs Two males singing from burrows in the Shetland and Faroe Islands. Source: XC46092 (Iˊ Dougie Preston), XC663885 (Iˊ Simon S. Christiansen). Under CC BY-NC-ND 2.5 licence.

Bengalese finch songs Recordings from 2 isolated Bengalese finches. Originally published in Tachibana, Koumura and Okanoya (Tachibana et al., 2015), data can be accessed at OSF.

They can be found under `pykanto/data` when you install the package, as well as in the GitHub repository.

Additionally, the worked example in this article uses 5293 songs from male great tit songs recorded by the author between 2020 and 2021 in Wytham Woods, Oxfordshire, UK. They are available from pykanto-example/data on GitHub, along with detailed metadata (Merino Recalde, 2023a).

CODE AVAILABILITY

The latest version of `pykanto` is available from PyPI (`pip install pykanto`) and its source repository (github.com/pykanto). See the repository for detailed installation instructions.

`pykanto` and the example in this article rely on the following open-source scientific libraries or tools: numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), pandas (The pandas development team, 2023), numba (Lam et al., 2015), pytorch (Paszke et al., 2019), torchvision (TorchVision maintainers and contributors, 2016), pytorch lightning (Falcon & The PyTorch Lightning team, 2019), tqdm (da Costa-Luis, 2019), ray (Moritz et al., 2018), soundfile (Bechtold & Geier, 2022), umap (McInnes et al., 2018), joblib (Joblib Development Team, 2020), hdbscan (McInnes et al., 2017), seaborn (Waskom, 2021), scikit-image (van der Walt et al., 2014), librosa (McFee et al., 2015), bokeh (Bokeh Development Team, 2018), ujson (van Kemenade et al., 2023), psutil (Rodola, 2023), attrs (Schlawack, 2019).

ACKNOWLEDGEMENTS

I thank Ben Sheldon and the Sheldon lab for their support and patience. Ben Sheldon, Carys Jones and Andrea Estandía provided useful comments on a draft of this manuscript. Carys Jones and Antoine Vansse tried early versions of the interactive app in `pykanto` and provided valuable feedback.

Some of the methods in `pykanto` are directly inspired by or adapted from Sainburg et al., 2020. I have indicated where this is the case in the relevant method's docstring. The dereverberation function is based on code by Robert Lachlan that is part of Luscinia (Lachlan, 2016), a software for bioacoustic archiving, measurement and analysis. Please consider citing these two publications if you use `pykanto` on your own projects.

I have learnt a great deal about packaging and developing in Python by browsing the structure of existing open source projects, for example some by David Nicholson (@NickleDave). I only became aware of VocalPy, a project that aims to 'develop an ecosystem of interoperable packages' for 'computational vocal communication and learning research' when I had already written most of `pykanto`, but eventually, I would like to make it compatible with it: standardization is direly needed in the field and I don't want to contribute to the chaos.

This work was supported by a Clarendon-Mary Frances Wagley Graduate Scholarship and an EGI scholarship to Nilo Merino Recalde, and made use of the University of Oxford Advanced Research Computing facility (Richards, 2015).

CONFLICT OF INTEREST

The author declares no conflict of interest.

AUTHOR CONTRIBUTIONS

Nilo Merino Recalde wrote the software library and its documentation, collected the data, conducted the analyses, and wrote the manuscript.

REFERENCES

Aplin, L. M. (2019). Culture and Cultural Evolution in Birds: A Review of the Evidence. *Animal Behaviour*, *147*, 179–187. https://doi.org/10.1016/j.anbehav.2018.05.001

Bechtold, B., & Geier, M. (2022). *Soundfile*. https://python-soundfile.readthedocs.io/en/0.11.0/

Beecher, M. D., Akçay, Ç., & Campbell, S. E. (2020). Birdsong learning is mutually beneficial for tutee and tutor in song sparrows. *Animal Behaviour*, *166*, 281–288. https://doi.org/10.1016/j.anbehav.2020.05.015

Bokeh Development Team. (2018). *Bokeh: Python library for interactive visualization*. Manual.

Coffey, K. R., Marx, R. G., & Neumaier, J. F. (2019). DeepSqueak: A deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacol.*, *44*(5), 859–868. https://doi.org/10.1038/s41386-018-0303-6

Cohen, Y., Nicholson, D. A., Sanchioni, A., Mallaber, E. K., Skidanova, V., & Gardner, T. J. (2022). Automated annotation of birdsong with a neural network that segments spectrograms (J. H. Goldberg, R. L. Calabrese, J. H. Goldberg, & M. Brainard, Eds.). *eLife*, *11*, e63853. https://doi.org/10.7554/eLife.63853

da Costa-Luis, C. O. (2019). Tqdm: A fast, extensible progress meter for python and CLI. *Journal of Open Source Software*, *4*(37), 1277. https://doi.org/10.21105/joss.01277

Demko, A. D., & Mennill, D. J. (2018). Rufous-capped Warblers Basileuterus rufifrons show seasonal, temporal and annual variation in song use. *Ibis*. https://doi.org/10.1111/ibi.12666

Falcon, W., & The PyTorch Lightning team. (2019, March). PyTorch lightning. https://doi.org/10.5281/zenodo.3828935

Fukuzawa, Y., Webb, W. H., Pawley, M. D., Roper, M. M., Marsland, S., Brunton, D. H., & Gilman, A. (2020). Koe: Web-based software to classify acoustic units and analyse sequence structure in animal vocalizations. *Methods in Ecology and Evolution*, *11*(3), 431–441. https://doi.org/10.1111/2041-210X.13336

Goffinet, J., Brudner, S., Mooney, R., & Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires (J. H. Goldberg, T. E. Behrens, J. H. Goldberg, O. Tchernichovski, & S. W. Linderman, Eds.). *eLife*, *10*, e67855. https://doi.org/10.7554/eLife.67855

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

He, K., Zhang, X., Ren, S., & Sun, J. (2015, December). Deep Residual Learning for Image Recognition. https://doi.org/10.48550/arXiv.1512.03385

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, *6*, e00073. https://doi.org/10.1016/j.ohx.2019.e00073

Hinde, R. A. (1952). The Behaviour of the Great Tit (Parus Major) and Some Other Related Species. *Behaviour. Supplement*, (2), III–201.

Joblib Development Team. (2020). Joblib: Running Python functions as pipeline jobs.

K. Lisa Yang Center for Conservation Bioacoustics. (2019). Raven Pro: Interactive Sound Analysis Software.

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

Kollmorgen, S., Hahnloser, R. H. R., & Mante, V. (2020). Nearest neighbours reveal fast and slow components of motor learning. *Nature*, *577*(7791), 526–530. https://doi.org/10.1038/s41586-019-1892-x

Krawczyk, B. (2016). Learning from imbalanced data: Open challenges and future directions. *Prog Artif Intell*, *5*(4), 221–232. https://doi.org/10.1007/s13748-016-0094-0

Lachlan, R. F. (2016, October). Luscinia: A bioacoustics analysis computer program.

Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, *9*(1), 2417. https://doi.org/10.1038/s41467-018-04728-1

Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: A LLVM-based Python JIT compiler. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, 1–6. https://doi.org/10.1145/2833157.2833162

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and Music Signal Analysis in Python. *Python in Science Conference*, 18–24. https://doi.org/10.25080/Majora-7b98e3ed-003

McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *The Journal of Open Source Software*, *3*(29), 861.

McLean, L. C., & Roach, S. P. (2020). Markov dependencies in the song syntax of Hermit Thrush (Catharus guttatus). *Journal of Ornithology*. https://doi.org/10.1007/s10336-020-01840-2

Merino Recalde, N. (2023a, May). Nilomr/pykanto-example: Publication-ready release of pykanto (use example). https://doi.org/10.5281/zenodo.7941355

Merino Recalde, N. (2023b, May). Nilomr/pykanto: V0.1.5. https://doi.org/10.5281/zenodo.7888656

Morfi, V., Lachlan, R. F., & Stowell, D. (2021). Deep perceptual embeddings for unlabelled animal sound events. *The Journal of the Acoustical Society of America*, *150*(1), 2–11. https://doi.org/10.1121/10.0005475

Moritz, P., Nishihara, R., Wang, S., Tumanov, A., Liaw, R., Liang, E., Elibol, M., Yang, Z., Paul, W., Jordan, M. I., & Stoica, I. (2018). Ray: A distributed framework for emerging AI applications. *Proceedings of the 13th USENIX Conference on Operating Systems Design and Implementation*, 561–577.

Nolet, C. J., Lafargue, V., Raff, E., Nanditale, T., Oates, T., Zedlewski, J., & Patterson, J. (2021). Bringing UMAP Closer to the Speed of Light with GPU Acceleration. *arXiv:2008.00325 [cs, stat]*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., … Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc.

Perrins, C. M. (1965). Population Fluctuations and Clutch-Size in the Great Tit, Parus major L. *Journal of Animal Ecology*, *34*(3), 601–647.

Pipek, P., Petrusková, T., Petrusek, A., Diblíková, L., Eaton, M. A., & Pyek, P. (2018). Dialects of an invasive songbird are preserved in its invaded but not native source range. *Ecography*, *41*(2), 245–254. https://doi.org/10.1111/ecog.02779

Priyadarshani, N., Marsland, S., & Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: A review. *Journal of Avian Biology*, *49*(5), jav–01447. https://doi.org/10.1111/jav.01447

R Core Team. (2021). R: A Language and Environment for Statistical Computing.

Raschka, S., Patterson, J., & Nolet, C. (2020). Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *arXiv preprint arXiv:2002.04803*.

Richards, A. (2015, August). *University of oxford advanced research computing*. Manual. https://doi.org/10.5281/zenodo.22558

Rodola, G. (2023). *Psutil (python system and process utilities)*. https://psutil.readthedocs.io/en/latest/

Sainburg, T., Theilman, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11605-y

Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Computational Biology*, *16*(10), e1008228. https://doi.org/10.1371/journal.pcbi.1008228

Schlawack, H. (2019). *Attrs*. https://www.attrs.org/en/stable/

Searfoss, A. M., Pino, J. C., & Creanza, N. (2020). Chipper: Open-source software for semi-automated segmentation and analysis of birdsong and other natural sounds. *Methods in Ecology and Evolution*, *11*(4), 524–531. https://doi.org/10.1111/2041-210X.13368

Specht, R. (2002). Avisoft-SASLab Pro. *Avisoft Bioacoustics, Berlin*, *2002*, 1–723.

Steinfath, E., Palacios-Muñoz, A., Rottschäfer, J. R., Yuezak, D., & Clemens, J. (2021). Fast and accurate annotation of acoustic signals with deep neural networks (R. L. Calabrese, S. R. Egnor, & T. Troyer, Eds.). *eLife*, *10*, e68837. https://doi.org/10.7554/eLife.68837

Stowell, D. (2021). Computational bioacoustics with deep learning: A review and roadmap. *arXiv:2112.06725 [cs, eess, q-bio]*.

Stowell, D., & Plumbley, M. D. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ*, *2*, e488. https://doi.org/10.7717/peerj.488

Stowell, D., Wood, M. D., Pamua, H., Stylianou, Y., & Glotin, H. (2019). Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge. *Methods in Ecology and Evolution*, *10*(3), 368–380. https://doi.org/10.1111/2041-210X.13103

Tachibana, R. O., Koumura, T., & Okanoya, K. (2015). Variability in the temporal parameters in the song of the Bengalese finch (Lonchura striata var. domestica). *J Comp Physiol A*, *201*(12), 1157–1168. https://doi.org/10.1007/s00359-015-1046-z

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour*, *59*(6), 1167–1176. https://doi.org/10.1006/anbe.1999.1416

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, *513*, 429–441. https://doi.org/10.1016/j.ins.2019.11.004

The pandas development team. (2023, January). Pandas-dev/pandas: Pandas. https://doi.org/10.5281/zenodo.7549438

TorchVision maintainers and contributors. (2016). TorchVision: PyTorch's computer vision library.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., & scikit-image contributors, t. (2014). Scikit-image: Image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

van Kemenade, H., Woodsend, B., Hamrén, J., JustAnotherArchivist, Crall, J., O'Mahony, K., Lay, E. L.,), H. M. (, Sychev, M., Swenson, D. W., Dawborn, T., Garcia-Armas, M., Ceccon, R., Moiron, J., Górny, M., NaN-git, Naoki, I., Beasley, B., Zhou, L., . . . CozyDoomer. (2023, January). Ultrajson. https://doi.org/10.5281/zenodo.7510698

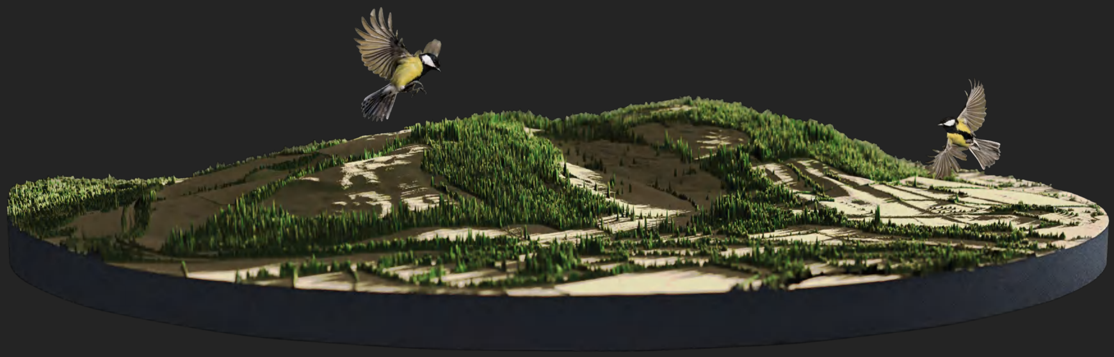
van Rossum, G. (1995, January). Python reference manual.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Youngblood, M., & Lahti, D. C. (2022). Content bias in the cultural evolution of house finch song. *Animal Behaviour*, *185*, 37–48. https://doi.org/10.1016/j.anbehav.2021.12.012

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, *109*(1), 43–76. https://doi.org/10.1109/JPROC.2020.3004555



3D render Wytham Woods based on first return LiDAR data.
Made with rayshader; elevation is exaggerated.

3. A DENSELY SAMPLED AND RICHLY ANNOTATED ACOUSTIC DATASET FROM A WILD BIRD POPULATION

Nilo Merino Recalde [1,*], Andrea Estandía [1], Loanne Pichot[2],
Antoine Vansse[2], Ella F. Cole [1], and Ben C. Sheldon [1]

[1]Edward Grey Institute, Department of Biology, University of Oxford, Oxford, UK
[2] École Normale Supérieure de Lyon, Lyon, France
*Corresponding author: nilo.recalde@biology.ox.ac.uk

ABSTRACT

We present a high-resolution, densely-sampled dataset of wild bird songs collected over multiple years from a single population of great tits (*Parus major*) in the UK. The dataset includes over 1,100,000 individual acoustic units from 109,963 richly annotated songs, sung by more than 400 individual birds, and provides unprecedented detail on the vocal behaviour of wild birds. Here, we describe the data collection and processing procedures and provide a summary of the data. We also discuss potential research questions that can be addressed using this dataset, including behavioural repeatability and stability, links between vocal performance and reproductive success, the timing of song production, syntactic organization of song production, and song learning in the wild. We have made the dataset and associated software tools publicly available with the aim that other researchers can benefit from this resource and use it to further our understanding of bird vocal behaviour in the wild.

Keywords: animal culture; bird song; demography, cultural evolution

INTRODUCTION & BACKGROUND

Despite a long history of scientific interest from disciplines as diverse as behavioural ecology, neurobiology and physiology, there is still much to learn regarding the evolution and function of animal vocalizations. Ongoing research covers a wide range of topics, including speech recognition and language evolution in humans, animal welfare, and even fish vocal communication. The study of animal vocalizations offers valuable insights into the intricacies of social interactions and reproductive strategies. They frequently convey crucial information about an individual's condition and identity (Lehmann & Seufert, 2017; Linhart et al., 2019), the cohesion of social groups, and the structure of social hierarchies (Bell et al., 2010; Engesser & Manser, 2022; Radford & Ridley, 2007). Additionally, animal vocalizations play a substantial role in the formation of social bonds, the selection of mates, and the provision of parental care (Behr & von Helversen, 2004; Gerhardt, 1991; Pitcher et al., 2010; Roulin, 2001).

For those interested in social learning and cultural evolution, animal vocalizations, particularly those of birds, have long been a focus of research. This interest dates back at least to the pioneering work of Marler and Thorpe with Chaffinches *Fringilla coelebs* and White-crowned Sparrows *Zonotrichia leucophrys* (Marler & Tamura, 1962; Marler, 1952; Marler & Tamura, 1964; Thorpe, 1958), which paved the way for what continues to be a thriving field today (see Mets and Brainard, 2019; Riebel et al., 2015; Williams and Lachlan, 2021; Youngblood and Lahti, 2022). In addition, and from a more mechanistic point of view, they offer a window into the physiological and neural mechanisms underlying vocal production and perception, as well as the consolidation of memories and motor coordination, to name but a few (Davenport & Jarvis, 2023).

Beyond their fundamental scientific importance, animal vocalizations have practical applications in various fields. For example, there is increasing recognition of their potential as a non-invasive tool for monitoring populations. By analysing entire soundscapes, researchers can gather crucial information about population dynamics, species distribution, and the presence of rare or elusive species (Kahl et al., 2021; Sethi et al., 2020; Sugai et al., 2019).

However, despite the growing interest in animal vocalizations and their potential applications, publicly available data from wild populations are still scarce—with the xeno-canto community science project as a prominent exception, focusing primarily on sparse recordings of most of the world's bird species rather than dense sampling of populations within the same species. This can severely limit researchers' ability to ask questions that require large datasets to answer, such as those about social learning, vocal development, large-scale cultural diversity, and the syntactic structure of animal vocalizations (Aplin, 2019; Kollmorgen et al., 2020; Lachlan et al., 2018; Sainburg et al., 2019). Indeed, while controlled laboratory settings allow researchers to track vocal development and production in minute detail, it is much harder to obtain finely-grained data from animals in their natural habitats. The process of collecting such data can be quite demanding and requires significant time, technical expertise, and resources: this includes both data collection itself and the subsequent processing of acoustic data files.

A second limitation arises after data have been collected, due to (i) researchers' understandable focus on specific, often narrowly defined questions, (ii) practical constraints, and (iii) scientific cultural norms that have not encouraged data-sharing. Combined, these factors often lead to a tendency of not publishing or only partially publishing the data collected during research. This lack of data sharing can hinder scientific progress and make it difficult to reproduce research findings (Jenkins et al., 2023; Powers & Hampton, 2019; Reichman et al., 2011; Wilkinson et al., 2016); hence, we argue that there is great intrinsic value in publishing fully curated acoustic datasets. If this practice becomes widespread, it would allow scientists to explore a broader range of research questions, improve reproducibility, and facilitate the validation of findings across different studies and populations (Hersh et al., 2023; Powers & Hampton, 2019).

In line with this perspective, we present a comprehensive dataset of wild bird songs recorded from a single population of great tits (*Parus major*) in Wytham Woods, Oxford, UK. We collected 21,283 hours of continuous recordings across 703 nesting sites over three spring seasons, which resulted in the annotation of over 1,100,000 notes or acoustic units from more than 100,000 songs (see below for definitions of these terms), sung by approximately 400 different male great tits. Among these birds, we have detailed information on the identity and life history of 242 individuals, including 50 that were recorded in multiple years. This information includes the time and location of breeding attempts, clutch size, number of fledglings, age of the bird, and basic morphological traits. For birds born in the population (106, or 43% of the total), we also include details such as birthplace, postnatal dispersal distance, mother, and social father.

To complement the song recordings, we have prepared extensive metadata for each of the more than 100,000 songs. This includes details such as the onset and offset times of each note within the song, a song type label, and the time of recording. We also provide the time of the first song during dawn. Finally, we augment the dataset by providing embeddings of each song, which are vector representations derived from a deep metric learning model specifically trained on this dataset. These can be used to identify individuals and in tasks that require similarity judgements.

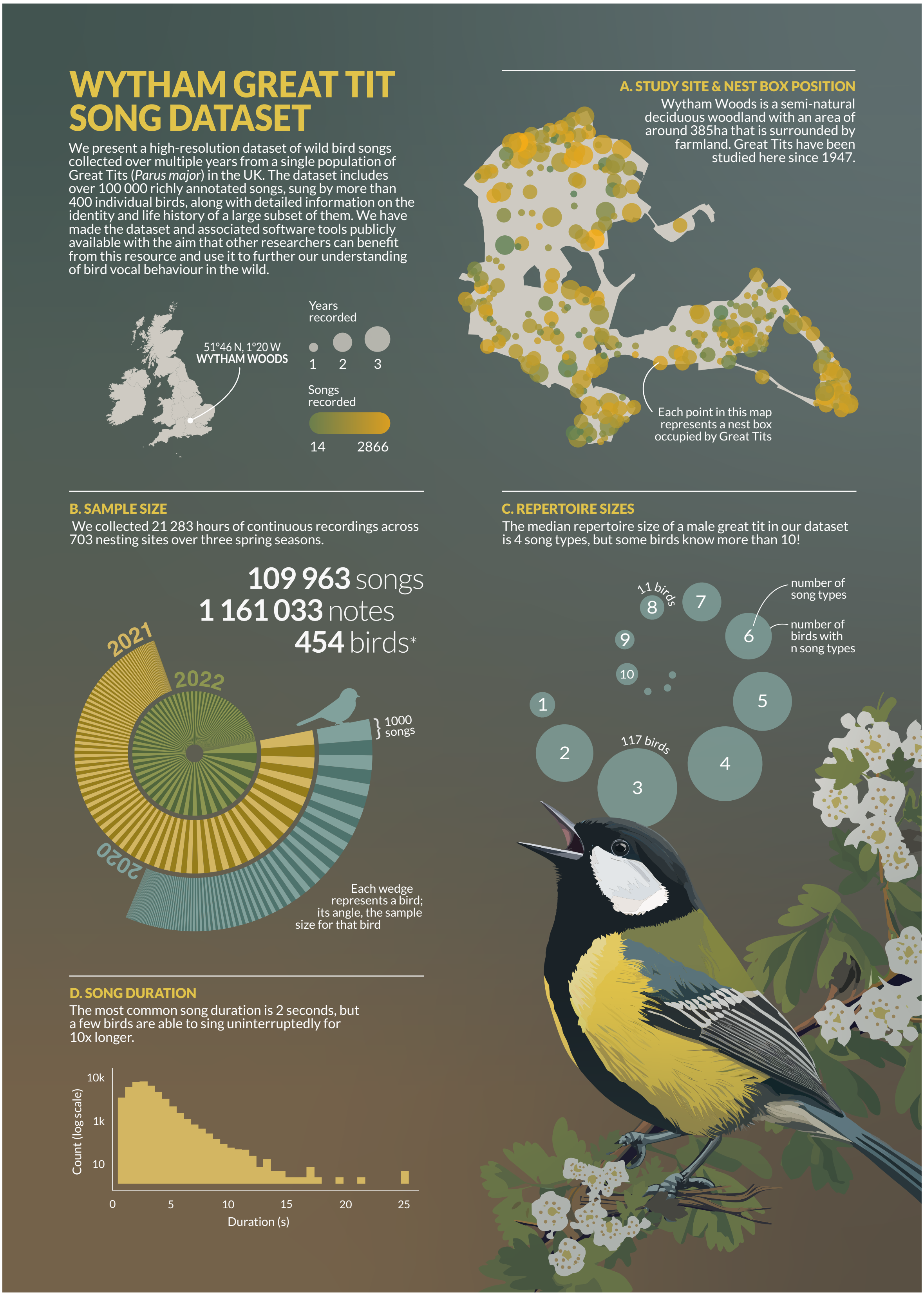
Great tit song has been the subject of extensive research activity (see, for example, Lambrechts and Dhondt, 1990; Lind et al., 1996; Ritschard et al., 2012; Rivera-Gutierrez, Matthysen, et al., 2010; Rivera-Gutierrez, Pinxten, and Eens, 2010; Rivera-Gutierrez et al., 2011; Slagsvold et al., 1994). Research conducted within the Wytham Woods population, in particular, has given rise to many influential ideas and insights into bird singing behaviour. These include investigations into neighbour interactions, song matching and the connection between song repertoires and reproductive success (McGregor & Krebs, 1989; McGregor et al., 1981; Mcgregor et al., 1983), the dynamics of song learning from neighbouring individuals and the acquisition of distinct song types (McGregor & Krebs, 1982, 1989), as well as the role of song repertoires in maintaining territories and reducing listener habituation (Krebs, 1976; Krebs et al., 1978), the functions of dawn song (Kacelnik & Krebs, 1983; Mace, 1987), and the influence of spatial factors and movement on song culture (Fayet et al., 2014). We hope that this dataset—which is, to the best of our knowledge, the largest publicly available collection of bird songs from a single wild population—will contribute to that effort by providing valuable insights into a range of scientific questions, including behavioural repeatability and stability, links between vocal performance and reproductive success, the timing of song production, the syntactic organization of song production, and song learning in the wild.

What follows is a detailed description of the data collection and curation process and the resulting dataset, together with some discussion around potential uses of data presented in this format.

DATA COLLECTION

Study system & fieldwork

Great tits are small, short-lived birds—average lifespan: 1.9 years—that sing acoustically simple yet



WYTHAM GREAT TIT SONG DATASET

We present a high-resolution dataset of wild bird songs collected over multiple years from a single population of Great Tits (*Parus major*) in the UK. The dataset includes over 100 000 richly annotated songs, sung by more than 400 individual birds, along with detailed information on the identity and life history of a large subset of them. We have made the dataset and associated software tools publicly available with the aim that other researchers can benefit from this resource and use it to further our understanding of bird vocal behaviour in the wild.

A. STUDY SITE & NEST BOX POSITION

Wytham Woods is a semi-natural deciduous woodland with an area of around 385ha that is surrounded by farmland. Great Tits have been studied here since 1947.

Each point in this map represents a nest box occupied by Great Tits

B. SAMPLE SIZE

We collected 21 283 hours of continuous recordings across 703 nesting sites over three spring seasons.

109 963 songs
1 161 033 notes
454 birds*

Each wedge represents a bird; its angle, the sample size for that bird

C. REPERTOIRE SIZES

The median repertoire size of a male great tit in our dataset is 4 song types, but some birds know more than 10!

D. SONG DURATION

The most common song duration is 2 seconds, but a few birds are able to sing uninterruptedly for 10x longer.

Figure 3.1. Description of the study site and dataset.
(A) Map of the study site and sample locations, (B) total sample sizes for each bird and year, (C) distribution of repertoire sizes, and (D) distribution of song lengths. *The exact number of individual birds is not known exactly.

highly diverse songs. During the breeding season, from March to June, great tit pairs are socially monogamous and defend territories around their nests (Hinde, 1952). In Wytham Woods, Oxfordshire, UK (51°46 N, 1°20 W), a population of these birds has been the focus of a long-term study since 1947 (Lack, 1964). Wytham Woods is a semi-natural predominantly deciduous woodland that spans an area of approximately 385 hectares and is surrounded by farmland. Most great tits in this population breed in nest boxes with known locations (see map in Figure 3.1), and the majority of individuals are marked with a unique British Trust for Ornithology (BTO) metal leg ring as nestlings or adults.

We collected data from late March to mid-May during the breeding seasons of 2020, 2021, and 2022. Every year, fieldworkers checked each of the 1018 nest boxes at least once a week before and during the egg-laying period, which typically lasts from one to 14 days (Perrins, 1965), and recorded the identities of breeding males and females, the dates of clutch initiation and egg hatching, clutch size, and fledgling number and condition under standardized protocols. We found the first egg date by assuming that one egg is laid every day and counting back from the day of observation. In cases where we did not observe the chicks on the day of hatching, the actual hatching date was determined by assessing the weight of the heaviest chicks and extrapolating their age from established growth curves.

To record the vocalizations of male great tits, we took advantage of their behaviour during the reproductive period, when they engage in continuous singing near their nests at dawn before and during egg laying (Mace, 1987). Collectively, this vocal display is referred to as the dawn chorus and has been demonstrated to yield a reliable estimation of the song repertoire of individuals when recorded in full (Rivera-Gutierrez et al., 2012; Van Duyse et al., 2005). As soon as we suspected that a pair of great tits were using a nest box based on nest lining materials, egg size if present, or other signs of activity, we deployed an autonomous sound recorder nearby. These recorders were placed on the trunk of the same tree or on a nearby tree, between 1 and 2 metres above the ground and no more than 5 metres away, depending on tree availability. We aimed to keep the recorder in a consistent position and orientation. The microphone pointed upwards and slightly away from the nest box, in the same direction as the entrance hole. The birds sang close to the recorder—we were not able to collect data on this, but the mean distance to the nest box was 10 m in a different population (Halfwerk et al., 2012), which matches our anecdotal observations— and moved around. Although changes in amplitude due to distance and directionality impacted song selection, we didn't observe any systematic bias.

Ethical note

All work involving birds was subject to review by the University of Oxford, Department of Zoology, Animal Welfare and Ethical Review Board (approval number: APA/1/5/ZOO/NASPA/ Sheldon/TitBreedingEcology). Data collection adhered to local guidelines for the use of animals in research and all birds were caught, tagged, and ringed by BTO licence holders.

Recording equipment and schedule

We used 60 (30 in 2020) AudioMoth recorders (Hill et al., 2019), which were housed in waterproof, custom-built enclosures. Recording began approximately one hour before sunrise (05:36 – 04:00 UTC during the recording period) and consisted of seven consecutive 60-minute-long recordings with a sample rate of 48 kHz, and a depth of 16-bit. To sample as many birds as possible, we left each recorder in the same location for at least three consecutive days before moving it to a different nest box. We relocated 20 recorders (10 in 2020) every day throughout the recording period.

A note on terminology

There is no consistent terminology used to describe the various hierarchical levels of a bird's vocal production. For clarity, we adopt the terminology outlined in Thompson et al., 1994; see also Figure 3.2B for a graphical explanation. The fundamental temporal unit is referred to as a **note**. Notes are represented by continuous traces on the sound spectrogram and are separated by silences. Moving up the hierarchy, **syllables** are sequences of one or more notes that are always repeated in the same order. Beyond syllables, we have **songs**, which consist of clusters of the same type of syllables punctuated by longer pauses, often in the order of seconds. Lastly, song **bouts** are uninterrupted performances of songs of the same type. great tits tend to sing the same song type repeatedly before transitioning to a different type. They continue this pattern until they stop singing altogether, often after having performed their entire song repertoire.

DATA PROCESSING AND ANNOTATION

We processed and annotated the recordings using custom software and scripts written in Python 3

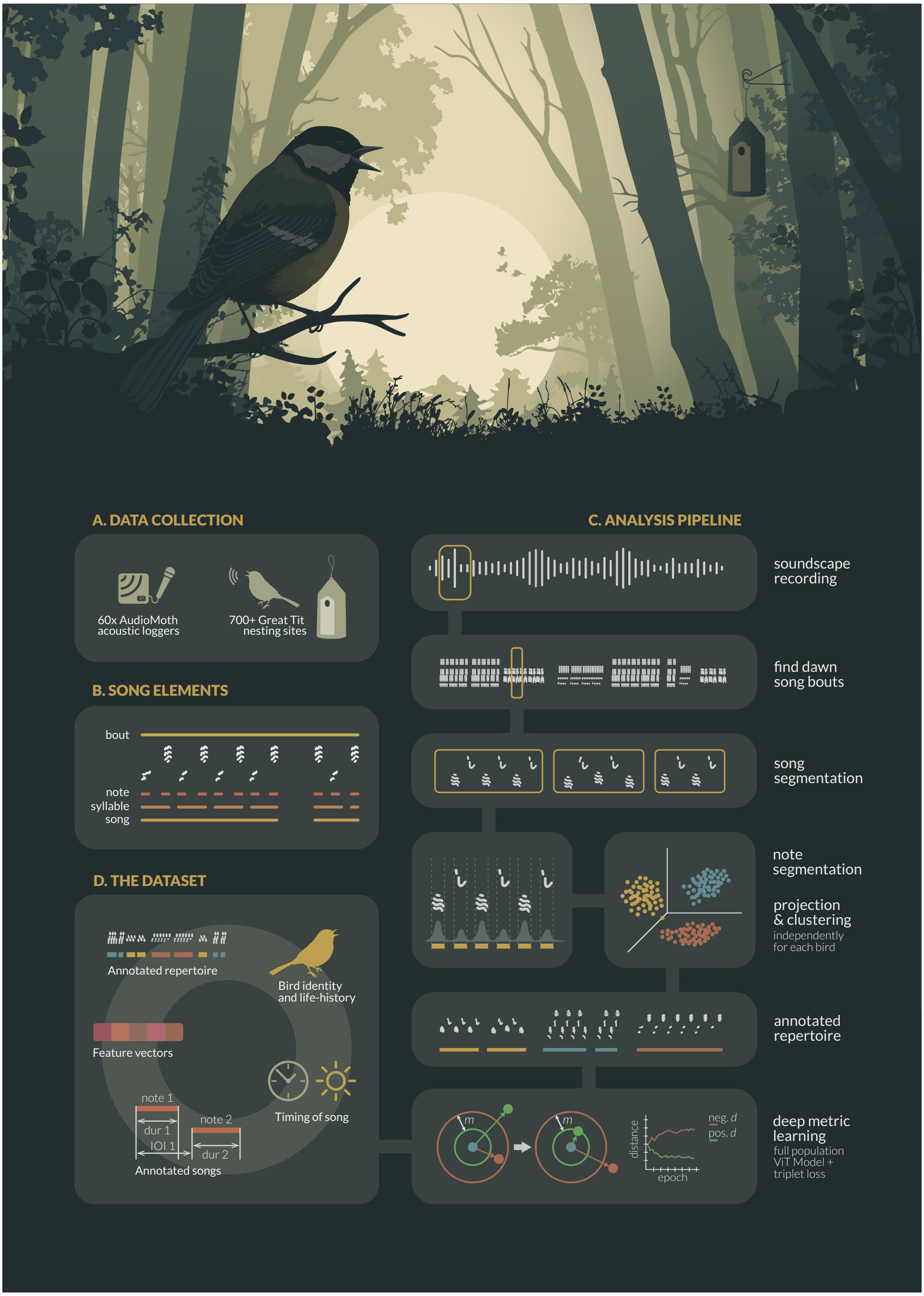


Figure 3.2. Data collection and analysis pipeline used to prepare the Wytham great tit Song Dataset. A brief visual summary of the data collection and analysis pipeline used to prepare the Wytham great tit Song Dataset. (A) Data collection in the field. (B) The terminology used to describe the various hierarchical levels at which we can describe great tit's singing. (C) Computational pipeline. (D) Main outputs included as part of the dataset.

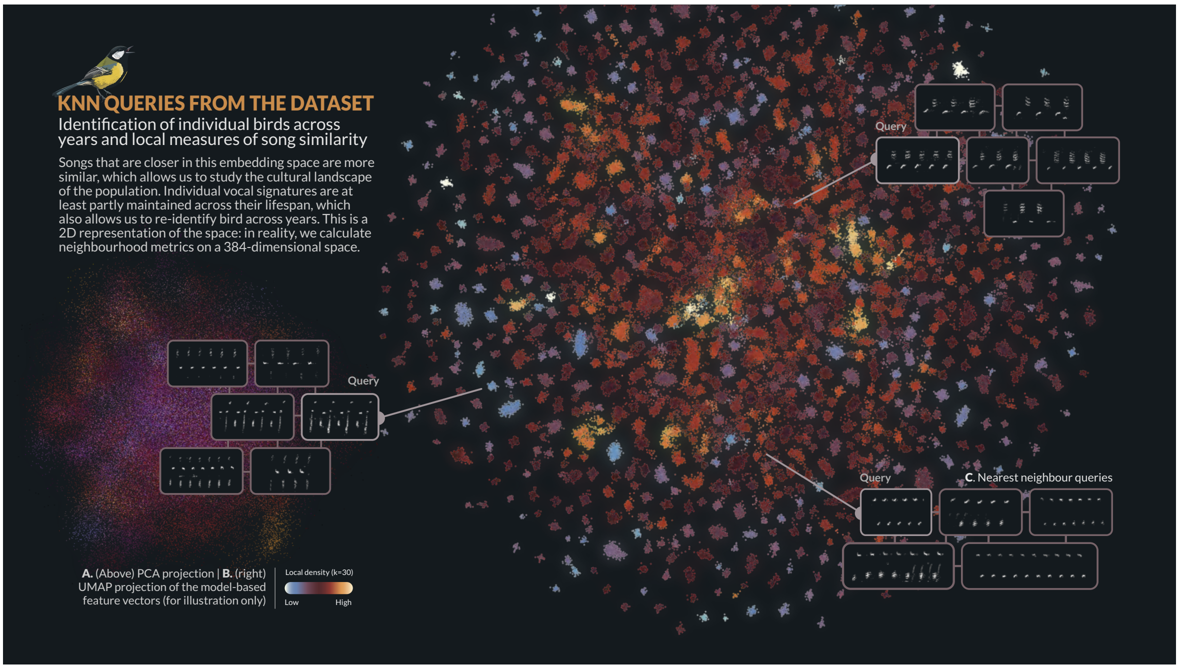


Figure 3.3. A visual representation of the song spectrogram embedding space.
Measuring similarity is a very hard problem, in large part because there is often no objective way to compare the performance of different methods. Here, we took a data-based approach by training a Vision Transformer (ViT) model as a feature extractor in a Euclidean metric learning task. The resulting embedding space allows us to judge if two songs are very similar, and to re-identify birds. (**A**) PCA projection of the feature vectors: two orthogonal linear components do not capture much of the high-level distinguishing features. (**B**) This figure shows a UMAP projection of the 384-dimensional vectors for each song in the dataset into 2D, which leads to a fairly arbitrary but useful visualization where tight clusters of points correspond to song types in the repertoire of individual birds. They are coloured by how densely occupied that region of space is in the high-dimensional space, based on k=30 neighbours from other song types. (**C**) A k-nearest neighbour search returns the closest matches for a query vector (highlighted).

(van Rossum, 1995), using the open-source package `pykanto` (Merino Recalde, 2023a). These are available from github.com/nilomr/great-tit-hits-setup (Merino Recalde, 2023b). Figure 3.2 shows a graphic illustration of the process. See also Box 1 for a note on the terminology used for different parts of the songs.

Song segmentation

We inspected spectrograms for each raw recording and selected songs based on a simple criterion: that its notes were clearly distinct from background noise and other bird vocalizations. We chose entire songs where it was possible; where it was not, we selected the longest contiguous segment possible. This process was carried out manually using the open-source software Sonic Visualizer (Cannam et al., 2010) by drawing boxes bounding songs in the time and frequency domains.

Assigning song bouts to individuals

Due to the automated recording process, there is a possibility that some of the recorded songs near a particular nest box may not originate from the focal bird. To minimize the chance of false positives, we discarded recordings with more than one vocalizing bird if one was not distinctly louder than the rest during the segmentation process. Additionally, we discarded all songs with a maximum amplitude below -16 dB, calculated as $20\log_{10}\left(\frac{A}{A_0}\right)$, with $A = 5000$ and $A_0 = 32767$ (the maximum value for 16-bit digital audio). This specific threshold was derived from observations indicating that when simultaneous recordings captured neighbouring birds, an amplitude cut-off greater than 4000 consistently differentiated the focal bird from its closest neighbours. It is important to note that these are not calibrated values and are, therefore, relative to the recording equipment and settings we used — as well as other factors like sound directionality and vegetation cover.

Spectrogramming

For most operations beyond this point, we used normalized, band-passed and log-scaled mel spectrogram representations of each of the songs (sampling rate = 22050, window length = 1024, hop length = 128, mel bins = 224; see the repository nilomr/great-tit-hits-setup for full details on the process).

Note segmentation

We segmented the resulting song selections into their constituent notes using a custom dynamic threshold algorithm implemented in pykanto (Merino Recalde, 2023a), based on the work of Sainburg et al. (2019). Briefly, the algorithm finds minima in the spectral envelope of a spectrogram, which are considered silences; if the length of the signal between these minima exceeds a maximum note duration, a new local minimum is defined that divides the signal into two shorter segments. This is repeated until multiple notes are defined or there are no local minima below a maximum amplitude threshold. Then, segments below a minimum note duration threshold are discarded. To make the algorithm more robust to noise, the spectrogram is subject to morphological transformations and de-echoing before amplitude information is extracted. The de-echoing algorithm implemented in pykanto is based on that in Luscinia (Lachlan, 2016), and works by subtracting a delayed version of the spectrogram from itself. We determined minimum and maximum note length ranges by manually segmenting a small, random subset of songs (n = 30).

It should be noted that the automated segmentation process is susceptible to various factors that can influence its accuracy. These include background noise, significant variation in amplitude between notes, attenuation caused by vegetation, changes in the direction of sound production, and even variations in performance where some notes may be much quieter. As a result, the algorithm may fail to detect or incorrectly delimit certain notes. Despite this, we estimate that approximately 96% of the notes are correctly segmented (0.037 error rate based on a random subset of n = 1048 notes that were checked manually). Still, depending on specific goals, we recommend manual verification of note segmentation if complete accuracy is crucial.

Song type annotation

We annotated each song type in the dataset using a semi-supervised approach implemented in `pykanto` . The process involved several steps to ensure accurate classification. First, we generated average unit spectrograms for each song by taking the mean of the centred and padded spectrograms of its units or notes, which provided a concise representation of the temporal and spectral characteristics of the syllable within it. Next, we performed non-linear dimensionality reduction using UMAP (McInnes et al., 2018) and a cluster search using HDBSCAN (McInnes et al., 2017) for each bird in the dataset. See (Sainburg & Hedley, 2020; Thomas et al., 2021) for similar approaches. This strategy, while useful, often leads to spurious outcomes. For instance, it may separate renditions of the same song type if variation in performance or background noise exists, or if certain song elements are sometimes attenuated. Such variation could be misinterpreted as distinct song types, leading to an overestimation of repertoire size. To address this, we used the interactive app in `pykanto` to review and split or combine clusters as necessary for each bird. It is worth mentioning that this process would be significantly more challenging in species with highly variable songs: our approach benefited from the great tits' relatively limited repertoires (1 to fewer than 15 song types in our population) and their tendency to produce stable and stereotyped songs.

Calculating song embeddings

Comparing animal vocalizations poses a significant challenge for researchers. Traditionally, two approaches have been used: visual comparisons of spectrograms and, more recently, measurement of hand-picked acoustic features (Goffinet et al., 2021). However, these methods have limitations when dealing with noise, variations in performance, and changes in syntax (where compositional syntax is not relevant). For instance, if a song with the sequence 'tea-cher, tea-cher' is recorded as 'cher-tea, cher', it might be wrongly perceived as highly dissimilar, despite being the same song (see Stowell, 2021; Zandberg et al., 2022 for a good overview of these issues). Additionally, these methods often fail to capture high-level features such as the syntactic relationships between notes and other complex spectrotemporal characteristics that cannot be easily characterized by an orthogonal combination of simple acoustic features.

Unfortunately, we cannot rely on the birds' perceptual judgements due to the lack of a hard to obtain experimental data (though recent studies, such as Morfi et al., 2021; Zandberg et al., 2022, have explored this avenue). This can be an issue where the focus of research is behavioural interactions or the social functions of song. At the same time, for monitoring or individual identification purposes, fully mimicking the bird's perceptual space may not be ideal: the performance of metric learning or classification algorithms trained for narrow purposes can surpass the organism's abilities, as exemplified by facial recognition in humans (Lu & Tang, 2014). Here, our goal was to define a similarity space based on the inherent variation in the data and the only categorical labels that we know are perceptually and behaviourally significant: song types sung by individual birds. Given that great tits can recognize each other based on their vocalizations (Lind et al., 1996), we aimed to define a similarity space that facilitates similarity-based research and captures some of the song characteristics that birds themselves might attend to when distinguishing individuals. To do this, we took advantage of recent advances in the fields of deep learning and computer vision and used a data-driven approach. Below is a simple narrative description of the process. For further details, see the dedicated repository nilomr/open-metric-learning and the OML li-

Table 3.1. Short description of the files included in the dataset.
Short description of the files included in the dataset. See the docs for detailed documentation.

109,963 raw song files with their corresponding metadata
Model-based feature vectors, with 109,963 samples and 384 dimensions
A main derived dataset including information on broods, adult bird life-history traits, nesting locations, and acoustic recordings
Morphological measurements for birds captured and re-trapped in the study area. Includes information on species, age, sex, weight, and various other traits
Information on the location and characteristics of nest boxes in Wytham Woods, along with a map of the site

brary (Shabanov, 2023).

Metric Learning and Vision Transformers

Rather than focusing on classification, we aimed to develop semantically meaningful embeddings. To achieve this, we used a Vision Transformer (ViT) model as a feature extractor in a (Euclidean) metric learning task. These models, inspired by the success of transformers in natural language processing applications, process images by splitting them into patches, treating them as tokens similar to words in a natural language (Dosovitskiy et al., 2021; Raghu et al., 2022). In this case, we used the ViT-S/16 architecture (21.7 M parameters), pre-trained on ImageNet using the DINO method (self-distillation with no labels; Caron et al., 2021), and fine-tuned of spectrogram representations of songs.

Model Training

During the training phase, we fine-tuned the ViT model using the great tit song dataset. To optimize the performance of the model, we used Triplet loss, a loss function that ensures that the projection of a positive sample, which belongs to the same class as the anchor point (i.e., song-type-within-individual), is closer to the anchor's projection than that of a negative sample, which belongs to a different class, by at least a specified margin (Hermans et al., 2017; Hoffer & Ailon, 2018). This loss function enables embedding points of the same class to form clusters without collapsing into a single point, which allows us to also explore differences within song types. While training the model we mined hard triplets—where the negative sample is closer to the anchor than the positive—and used the Adam optimizer with a fixed learning rate of 1×10^{-5}.

Handling Data Imbalance and Batch Generation

The distribution of song sample sizes per individual in the great tit dataset approximately follows a power law, resulting in a significant data imbalance. Although the use of triplet loss already addresses this issue to some extent (Thakur et al., 2019), we adopt a random subsampling strategy where classes with more than 100 samples are reduced to 100 for computational efficiency, classes with fewer than 15 samples are excluded to allow a large enough query/gallery split for validation, and we ensure fair representation during training using a balanced sampler (Hermans et al., 2017). Our batch generation strategy involves uniformly sampling P song types without replacement and sampling K spectrograms for each song type, with replication as necessary. This guarantees that all labels are selected at least once in each epoch.

Train-Time Data Augmentation

To enhance model robustness and prevent overfitting, we apply various train-time data augmentation techniques (Mumuni & Mumuni, 2022; Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019). These include random cropping in the time domain, dropping out parts of the spectrogram, adding Gaussian and multiplicative noise, equalization, sharpening, changes to brightness and contrast, blurring, and slight shifting in both time and frequency domains. The latter augmentations are applied within the typical variation in performance observed in the great tit vocalizations.

Results

Our trained model shows very good performance, achieving a mean Average Precision at 5 (mAP5) of 0.98 and a Cumulative Matching Characteristic at 1 (CMC1) of 0.98. This indicates that in approximately 98% of the queries made to the similarity space, the returned candidate song type by a bird is the correct one. Errors primarily stemmed from instances where songs of the same type sung by the same bird appeared more than once in the dataset, which happened if a bird survived to the next year. Given that the model was trained on almost 2000 classes, this means that there is enough individual information contained in each song type to distinguish between birds with high confidence, which has important implications for both the study of individuality and population monitoring. See Figure 3.3 for a visual representation of the embedding space and nearest-neighbour queries.

Table 3.2. Description of the dataset and sample sizes.
A brief description of the dataset and sample sizes for different subsets of the data.

Description	Value
Number of segmented notes	1,161,033
Number of songs	109,963
Mean repertoire size	4.24
SD repertoire size	1.98
Median repertoire size	4
Number of unique classes	1,930
Mean class size	56.98
SD class size	68.35
Median class size	31
Number of nest sites recorded	706
Number of nest sites with data	454
Number of unique birds with data that were ID'd	242
Number of times each bird was recorded	192 (1y), 42 (2y), 8 (3y)

DATA RECORDS AND DESCRIPTION

Table A3.1 contains a summary of the files included with the dataset. Detailed data documentation, including variable descriptions, can be found online at nilomr.github.io/great-tit-hits.

The dataset provides a comprehensive view of the populations' natural dawn singing behaviour over three spring seasons. It documents changes in individual performance, the appearance and disappearance of birds—and with them, their songs—and highlights just how much behavioural variation there is along every dimension of what could at first seem a relatively simple trait. Table 3.2 presents some simple summary statistics, and Figure 3.1 provides a visual overview of the dataset.

Even though most birds in the dataset are one or two-year-olds recorded within a single year (which can be attributed to high turnover rates in the population given low annual survival), the dataset includes valuable data on much older individuals, including a 7-year-old. Among the recorded birds, some display metronome-like regularity in their performance, while others have highly variable or unusual songs, due to learning from allospecific vocalizations, or even issues with their vocal apparatus. You can find some interactive examples at nilomr.github.io/great-tit-hits. The longest song recorded is approximately 20 times longer than the shortest song (and, coincidentally, was sung by one of the largest great tits ever recorded in the Wytham population). The median number of songs per song type and per bird in the dataset is 31, with a significant number of birds having a much larger count, reaching into the thousands; the median repertoire size per bird is four song types, although some birds performed as many as 13 distinct song types.

Known biases and problems

Working with third-party datasets can be challenging, perhaps particularly so in the study of behaviour in natural populations. The familiarity that fieldworkers inevitably develop with the study system and the data is difficult to replace, and, as a result, there is a risk of unintentionally overlooking important sources of bias and variability. We have compiled a list of some key considerations, which, while not exhaustive, can serve as a starting point for identifying and addressing biases when testing hypotheses, estimating parameters, or evaluating findings from the data. These issues can be broadly classified into two groups: those around bird and song type labelling, which can be partially addressed, and those that are inherent in the data or how it was collected.

Individual and song identification

One factor that can be partially addressed is that the birds recorded in our dataset are not a random subset of the population; they are those that establish territories and begin the breeding process. In turn, birds that are subsequently identified are more likely to be those whose chicks hatch and survive for at least six days, when the first identification attempt is made. This may skew the distribution of certain behaviours within the dataset or lead to endogenous selection bias (Elwert & Winship, 2014). One way to quantify the extent to which the subset of identified birds is representative of the entire breeding population would be to compare the distribution of the trait of interest in both groups. See, for example, Kidd et al., 2015, who found that females in nests that fail early in our population are more likely to be immigrant birds breeding in poor-quality areas.

Another issue to consider is that birds may attempt to breed again in the same nest box or elsewhere after a failed attempt. This, coupled with a failure to identify the male associated with those attempts, means that it is conceivable (although likely very rare) that songs from the same bird could appear in the same year twice, leading to pseudoreplication. Similarly, unidentified birds present in the dataset for multiple years could contribute to this problem. One potential way to address these issues is by using song embeddings for identification based on similarity and assigning dummy IDs to birds believed to be the same individual. At least, this should be modelled to assess the sensitivity of any results to varying degrees of pseudoreplication from this source.

Finally, a few songs might have been mislabelled before model training, as it is not feasible to manually check such a large dataset. However, the model-based embeddings can help identify any mislabelled songs: they will be clear outliers within their respective classes, thanks to the relatively discrete nature of great tit repertoires.

Unequal samples; songs and calls—and female song

As is common in many complex systems, the interaction of the many processes involved in both song production and sampling results in a heavy-tailed frequency distribution of sample sizes. This variation stems from various sources, including characteristics inherent to the study system, such as individual differences in singing activity and temporal fluctuations throughout spring. The sampling process introduces further variation, through factors like equipment malfunctions causing small gaps in the data, variation in recording dates relative to peak activity, and the impact of rain and hail on singing activity and recording quality. We cannot assume these processes to be completely independent of each other. Therefore, when analysing song output or repertoire size, it is important to explicitly specify the assumed causal relationship between factors such as individual characteristics, sampling probability, and the outcome measure.

Another important aspect to consider is that, while we have said that the dataset consists of songs, the demarcation between songs and calls is not entirely straightforward. Some vocalizations that would typically be classified as calls, due to their acoustically simpler, shorter, and possibly more stereotyped nature, are actually used as part of the dawn vocal behaviour. These vocalizations are repeated in a manner that creates an impression of functional equivalence to songs. While we have followed criteria similar to other studies (Baker et al., 1986; Fayet et al., 2014; Krebs et al., 1978; Rivera-Gutierrez, Matthysen, et al., 2010) to maintain consistency, we believe that this phenomenon warrants further attention. These calls were not segmented and thus are not included in the dataset, but we are happy to provide soundscape recordings to anyone interested in exploring this aspect further.

Finally, although female song in birds has received relatively little historical attention (see Langmore, 2020; Odom and Benedict, 2018; Riebel et al., 2005 for further discussion), female great tits also sing (see a brief treatment in Gompertz, 1961; Hinde, 1952). The vast majority of songs in the dataset belong to the dawn song, a behaviour exclusively performed by the male prior to the female leaving the nest (a pattern observed in blue tits as well, as documented by Sierro et al., 2022). Females, on the other hand, vocalize within the nest, but these vocalizations (Gorissen & Eens, 2004, 2005) differ from songs and were not typically detectable by our recording devices. Nevertheless, Hinde, 1952 suggests that in the absence of males, females may be more inclined to engage in territorial behaviour that involves singing rather than just producing calls. If that is the case, it is possible that our dataset contains some isolated instances of female song.

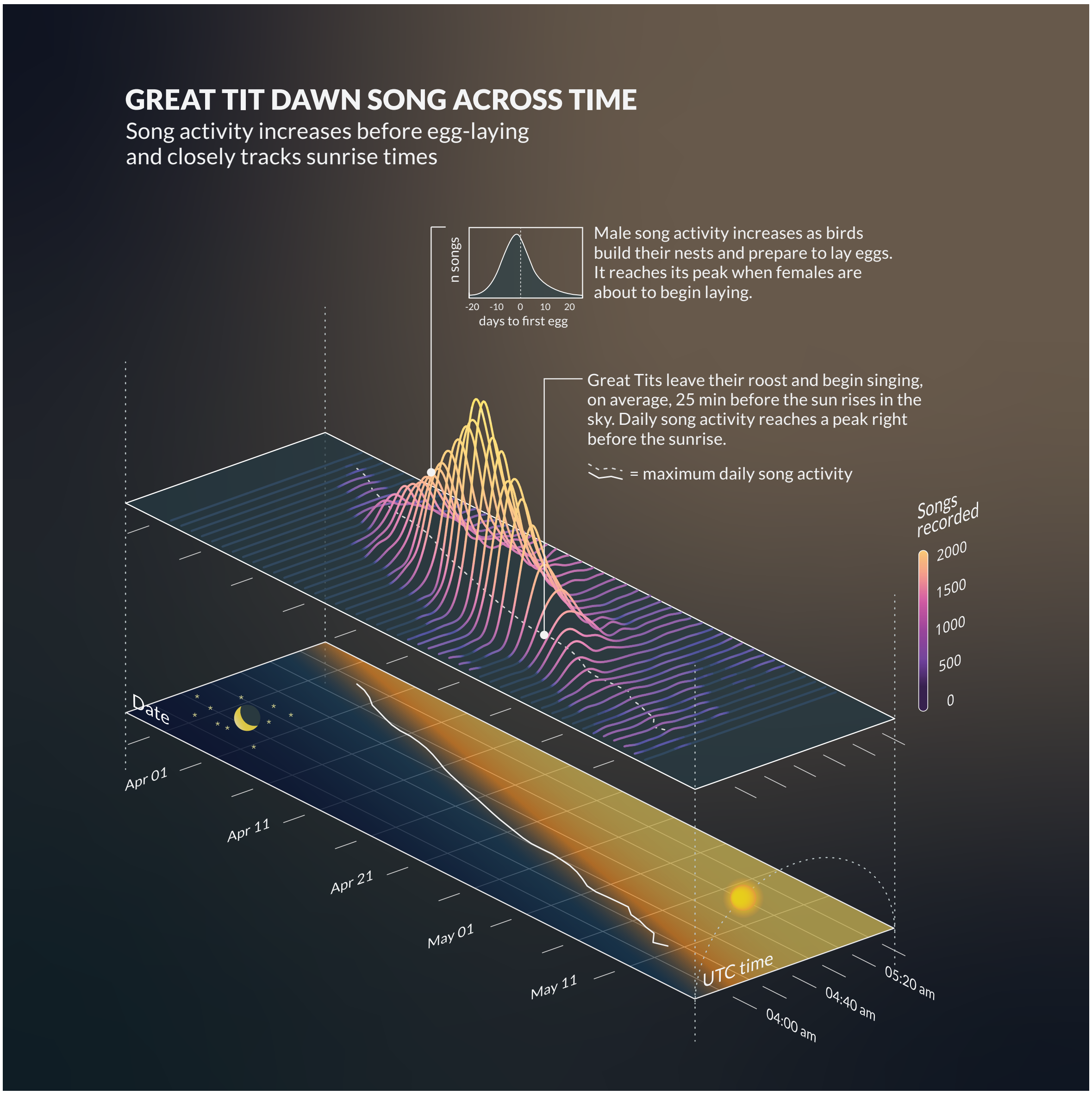
USES AND SUGGESTIONS

The dataset we are presenting contains detailed information about the vocal behaviour and life of wild birds, providing valuable opportunities for investigating a wide range of research questions. In this section, we suggest several research areas that can be explored using this dataset and provide references to relevant studies in the literature.

Behavioural repeatability and stability across multiple scales: Researchers can use the dataset to examine the repeatability and stability of song production and song characteristics across different temporal and spatial scales. This includes studying consistency in vocal behaviour within individuals over time and across different contexts, and its links to age (Rivera-Gutierrez et al., 2012; Zipple et al., 2019) and reproductive fitness (Sierro et al., 2023).

Links between vocal performance or diversity and reproductive success: Our data can be used to explore the relationships between vocal performance metrics, such as song complexity or vocal diversity, and individual breeding success on a dataset that is much larger than what is typical in the field (Beecher et al., 2020; Crates et al., 2021; Hiebert et al., 1989; Hutfluss et al., 2022; McGregor et al., 1981).

Spatial and temporal properties of acoustic communities: The dataset enables investigations into the spatial properties of acoustic communities, including the distribution of singing individuals within a given habitat and across time. This can provide valuable insights into the spatial dynamics of communication



GREAT TIT DAWN SONG ACROSS TIME

Song activity increases before egg-laying and closely tracks sunrise times

Male song activity increases as birds build their nests and prepare to lay eggs. It reaches its peak when females are about to begin laying.

Great Tits leave their roost and begin singing, on average, 25 min before the sun rises in the sky. Daily song activity reaches a peak right before the sunrise.

= maximum daily song activity

Figure 3.4. Great tit song activity closely tracks advancing sunrise times and female fertility. Days get longer as the spring progresses and male great tits track the advancing sunrise times with great precision, so that they always begin singing, on average, 25 minutes before the morning breaks. This figure also shows (z-axis) how song activity peaks alongside egg laying: males sing the most in the morning right before their partner lays the first egg.

networks and acoustic interaction among neighbour birds.

Timing and volume of song production: Researchers can use the dataset to analyse the temporal patterns and timing of song production in great tits. This might involve studying diurnal variation, seasonal trends, and the influence of environmental factors on the timing and abundance of vocal behaviour. As an example, Figure 3.4 provides an overview of key temporal shifts in dawn singing behaviour: male birds sing more during the fertile period of the female, and their activity closely tracks advancing sunrise times.

The syntactic organization of song production: The dataset captures song activity over entire dawn song periods, across days, and even years for many individuals. This would allow researchers to investigate the set of rules that govern the arrangement of song elements and transitions within the vocal repertoire of wild great tits, in terms of short and long-distance dependencies and other properties of their sequential dynamics (Hedley et al., 2018; Lachlan et al., 2010; Sainburg et al., 2019; Searcy et al., 2022).

Song learning in the wild: While this dataset does not directly provide evidence of song learning, researchers can use song similarity and proximity in time and space to infer cultural transmission processes. This allows for the exploration of the influence of spatial and social factors on song learning (James et al., 2020; Lachlan & Feldman, 2003; Nelson

& Poesel, 2014; Peters & Nowicki, 2017; Wheelwright et al., 2008).

CONCLUSION

With over 1,100,000 annotated notes and acoustic units from more than 100,000 songs, collected over three spring seasons, we hope that this dataset will offer valuable insights into bird vocal behaviour and song culture. The dataset is enriched with detailed metadata such as note onset and offset times, song type labels and embeddings derived from a deep metric learning model, as well as identity and life-history information for the birds, which makes it useful for a wide range of research purposes. By sharing this comprehensive dataset, we also aim to help promote data-sharing and scientific collaboration.

AUTHOR CONTRIBUTIONS

LP, AV, AE and NMR collected the data. NMR created the software and pipeline to plan fieldwork and analyse the data, annotated the dataset with help from AE, built the website, documentation, and visualizations, and wrote the original draft. BCS and EFC provided feedback and supervision throughout the research, and all authors contributed critically to drafts.

ACKNOWLEDGEMENTS

We thank all those who have contributed to the long-term nest box study in Wytham Woods and the collection of associated data. This work was supported by a Clarendon-Mary Frances Wagley Graduate Scholarship and an Edward Grey Institute scholarship to Nilo Merino Recalde and made use of the University of Oxford Advanced Research Computing facility (Richards, 2015).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AND CODE: AVAILABILITY AND USE

The complete Wytham great tit Song Dataset and its metadata are available at 10.17605/OSF.IO/N8AC9 (Merino Recalde, 2023c). The code to train the deep metric learning model can be found at nilomr/open-metric-learning. See the dataset website for documentation and more information.

All input and output data files use open data formats and are under a CC-BY-4.0 licence. The scripts and software used to create this dataset are available under the MIT licence from GitHub nilomr/great-tit-hits-setup and archived at Zenodo (Merino Recalde, 2023b).

REFERENCES

Aplin, L. M. (2019). Culture and Cultural Evolution in Birds: A Review of the Evidence. *Animal Behaviour*, *147*, 179–187. https://doi.org/10.1016/j.anbehav.2018.05.001

Baker, M. C., Bjerke, T. K., Lampe, H., & Espmark, Y. (1986). Sexual Response of Female Great Tits to Variation in Size of Males' Song Repertoires. *The American Naturalist*, *128*(4), 491–498. https://doi.org/10.1086/284582

Beecher, M. D., Akçay, Ç., & Campbell, S. E. (2020). Birdsong learning is mutually beneficial for tutee and tutor in song sparrows. *Animal Behaviour*, *166*, 281–288. https://doi.org/10.1016/j.anbehav.2020.05.015

Behr, O., & von Helversen, O. (2004). Bat serenadescomplex courtship songs of the sac-winged bat (Saccopteryx bilineata). *Behav Ecol Sociobiol*, *56*(2), 106–115. https://doi.org/10.1007/s00265-004-0768-7

Bell, M. B. V., Radford, A. N., Smith, R. A., Thompson, A. M., & Ridley, A. R. (2010). Bargaining babblers: Vocal negotiation of cooperative behaviour in a social bird. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1698), 3223–3228. https://doi.org/10.1098/rspb.2010.0643

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. *Proceedings of the International Conference on Multimedia - MM '10*, 1467. https://doi.org/10.1145/1873951.1874248

Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A. (2021, May). Emerging Properties in Self-Supervised Vision Transformers. https://doi.org/10.48550/arXiv.2104.14294

Crates, R., Langmore, N., Ranjard, L., Stojanovic, D., Rayner, L., Ingwersen, D., & Heinsohn, R. (2021). Loss of vocal culture and fitness costs in a critically endangered songbird. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1947), 20210225. https://doi.org/10.1098/rspb.2021.0225

Davenport, M. H., & Jarvis, E. D. (2023). Birdsong neuroscience and the evolutionary substrates of learned vocalization. *Trends in Neurosciences*, *46*(2), 97–99. https://doi.org/10.1016/j.tins.2022.11.005

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021, June). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. https://doi.org/10.48550/arXiv.2010.11929

Elwert, F., & Winship, C. (2014). Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. *Annu Rev Sociol*, *40*, 31–53. https://doi.org/10.1146/annurev-soc-071913-043455

Engesser, S., & Manser, M. B. (2022). Collective close calling mediates group cohesion in foraging meerkats via spatially determined differences in call rates. *Animal Behaviour*, *185*, 73–82. https://doi.org/10.1016/j.anbehav.2021.12.014

Fayet, A. L., Tobias, J. A., Hintzen, R. E., & Seddon, N. (2014). Immigration and dispersal are key determinants of cultural diversity in a songbird population. *Behavioral Ecology*, *25*(4), 744–753. https://doi.org/10.1093/beheco/aru047

Gerhardt, H. C. (1991). Female mate choice in treefrogs: Static and dynamic acoustic criteria. *Animal Behaviour*, *42*(4), 615–635. https://doi.org/10.1016/S0003-3472(05)80245-3

Goffinet, J., Brudner, S., Mooney, R., & Pearson, J. (2021). Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires (J. H. Goldberg, T. E. Behrens, J. H. Goldberg, O. Tchernichovski, & S. W. Linderman, Eds.). *eLife*, *10*, e67855. https://doi.org/10.7554/eLife.67855

Gompertz, T. (1961). The vocabulary of the great tit. *British Birds*, *54*(10-11), 369–394.

Gorissen, L., & Eens, M. (2004). Interactive communication between male and female great tits (Parus major) during the dawn chorus. *The Auk*, *121*(1), 184–191.

Gorissen, L., & Eens, M. (2005). Complex Female Vocal Behaviour of Great and Blue Tits inside the Nesting Cavity. *Behaviour*, *142*(4), 489–506.

Halfwerk, W., Bot, S., & Slabbekoorn, H. (2012). Male great tit song perch selection in response to noise-dependent female feedback. *Functional Ecology*, *26*(6), 1339–1347. https://doi.org/10.1111/j.1365-2435.2012.02018.x

Hedley, R. W., Logue, D. M., Benedict, L., & Mennill, D. J. (2018). Assessing the similarity of song-type transitions among birds: Evidence for interspecies variation. *Animal Behaviour*, *140*, 161–170. https://doi.org/10.1016/j.anbehav.2018.04.008

Hermans, A., Beyer, L., & Leibe, B. (2017, November). In Defense of the Triplet Loss for Person Re-Identification. https://doi.org/10.48550/arXiv.1703.07737

Hersh, T. A., Ravignani, A., & Burchardt, L. S. (2023). Robust rhythm reporting will advance ecological and evolutionary research. *Methods in Ecology and Evolution*, *n/a*(n/a). https://doi.org/10.1111/2041-210X.14118

Hiebert, S. M., Stoddard, P. K., & Arcese, P. (1989). Repertoire size, territory acquisition and reproductive success in the song sparrow. *Animal Behaviour*, *37*, 266–273. https://doi.org/10.1016/0003-3472(89)90115-2

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, *6*, e00073. https://doi.org/10.1016/j.ohx.2019.e00073

Hinde, R. A. (1952). The Behaviour of the Great Tit (Parus Major) and Some Other Related Species. *Behaviour. Supplement*, (2), III–201.

Hoffer, E., & Ailon, N. (2018, December). Deep metric learning using Triplet network. https://doi.org/10.48550/arXiv.1412.6622

Hutfluss, A., Bermúdez-Cuamatzin, E., Mouchet, A., Briffa, M., Slabbekoorn, H., & Dingemanse, N. J. (2022). Male song stability shows cross-year repeatability but does not affect reproductive success in a wild passerine bird. *Journal of Animal Ecology*, *91*(7), 1507–1520. https://doi.org/10.1111/1365-2656.13736

James, L. S., Davies, R., Mori, C., Wada, K., & Sakata, J. T. (2020). Manipulations of sensory experiences during development reveal mechanisms underlying vocal learning biases in zebra finches. *Developmental Neurobiology*, *80*(3-4), 132–146. https://doi.org/10.1002/dneu.22754

Jenkins, G. B., Beckerman, A. P., Bellard, C., Benítez-López, A., Ellison, A. M., Foote, C. G., Hufton, A. L., Lashley, M. A., Lortie, C. J., Ma, Z., Moore, A. J., Narum, S. R., Nilsson, J., O'Boyle, B., Provete, D. B., Razgour, O., Rieseberg, L., Riginos, C., Santini, L., … Peres-Neto, P. R. (2023). Reproducibility in ecology and evolution: Minimum standards for data and code. *Ecology and Evolution*, *13*(5), e9961. https://doi.org/10.1002/ece3.9961

Kacelnik, A., & Krebs, J. R. (1983). The Dawn Chorus in the Great Tit (Parus Major): Proximate and Ultimate Causes. *Behaviour*, *83*(3-4), 287–308. https://doi.org/10.1163/156853983X00200

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

Kidd, L. R., Sheldon, B. C., Simmonds, E. G., & Cole, E. F. (2015). Who escapes detection? Quantifying the causes and consequences of sampling biases in a long-term field study. *Journal of Animal Ecology*, *84*(6), 1520–1529. https://doi.org/10.1111/1365-2656.12411

Kollmorgen, S., Hahnloser, R. H. R., & Mante, V. (2020). Nearest neighbours reveal fast and slow components of motor learning. *Nature*, *577*(7791), 526–530. https://doi.org/10.1038/s41586-019-1892-x

Krebs, J. R. (1976). Habituation and song repertoires in the great tit. *Behav Ecol Sociobiol*, *1*(2), 215–227. https://doi.org/10.1007/BF00299198

Krebs, J., Ashcroft, R., & Webber, M. (1978). Song repertoires and territory defence in the great tit. *Nature*, *271*(5645), 539–542. https://doi.org/10.1038/271539a0

Lachlan, R. F., & Feldman, M. W. (2003). Evolution of cultural communication systems: The coevolution of cultural signals and genes encoding learning preferences. *Journal of Evolutionary Biology*, *16*(6), 1084–1095. https://doi.org/10.1046/j.1420-9101.2003.00624.x

Lachlan, R. F., Verhagen, L., Peters, S., & ten Cate, C. (2010). Are There Species-Universal Categories in Bird Song Phonology and Syntax? A Comparative Study of Chaffinches (Fringilla coelebs), Zebra Finches (Taenopygia guttata), and Swamp Sparrows (Melospiza georgiana). *Journal of Comparative Psychology*, *124*(1), 92–108. https://doi.org/10.1037/a0016996

Lachlan, R. F. (2016, October). Luscinia: A bioacoustics analysis computer program.

Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, *9*(1), 2417. https://doi.org/10.1038/s41467-018-04728-1

Lack, D. (1964). A Long-Term Study of the Great Tit (Parus major). *Journal of Animal Ecology*, *33*, 159–173. https://doi.org/10.2307/2437

Lambrechts, M. M., & Dhondt, A. A. (1990). A relationship between the composition and size of great tit song repertoires. *Animal Behaviour*, *39*(2), 213–218. https://doi.org/10.1016/S0003-3472(05)80865-6

Langmore, N. E. (2020). Female birdsong. *Current Biology*, *30*(14), R789–R790. https://doi.org/10.1016/j.cub.2020.05.042

Lehmann, J. A. M., & Seufert, T. (2017). The Influence of Background Music on Learning in the Light of Different Theoretical Perspectives and the Role of Working Memory Capacity. *Frontiers in Psychology*, *8*(October), 1–11. https://doi.org/10.3389/fpsyg.2017.01902

Lind, H., Dabelsteen, T., & McGregor, P. K. (1996). Female great tits can identify mates by song. *Animal Behaviour*, *52*(4), 667–671. https://doi.org/10.1006/anbe.1996.0211

Linhart, P., Osiejuk, T. S., Budka, M., álek, M., pinka, M., Policht, R., Syrová, M., & Blumstein, D. T. (2019). Measuring individual identity information in animal signals: Overview and performance of available identity metrics. *Methods in Ecology and Evolution*, *10*(9), 1558–1570. https://doi.org/10.1111/2041-210X.13238

Lu, C., & Tang, X. (2014). Surpassing human-level face verification performance on LFW with GaussianFace. *arXiv*. https://doi.org/10.48550/arXiv.1404.3840

Mace, R. (1987). The dawn chorus in the great tit Parus major is directly related to female fertility. *Nature*, *330*(6150), 745–746. https://doi.org/10.1038/330745a0

Marler, P., & Tamura, M. (1962). Song "Dialects" in Three Populations of White-Crowned Sparrows. *The Condor*, *64*(5), 368–377. https://doi.org/10.2307/1365545

Marler, P. (1952). Variation in the Song of the Chaffinch Fringilla Coelebs. *Ibis*, *94*(3), 458–472. https://doi.org/10.1111/j.1474-919X.1952.tb01845.x

Marler, P., & Tamura, M. (1964). Culturally transmitted patterns of vocal behavior in sparrows. *Science*, *146*(3650), 1483–1486. https://doi.org/10.1126/science.146.3650.1483

McGregor, P. K., & Krebs, J. R. (1982). Song Types in a Population of Great Tits (Parus major): Their Distribution, Abundance and Acquisition by Individuals. *Behaviour*, *79*(2/4), 126–152.

McGregor, P. K., & Krebs, J. R. (1989). Song learning in adult great tits (Parus major): Effects of neighbours. *Behaviour*, *108*(1-2), 139–159. https://doi.org/10.1163/156853989X00105

McGregor, P. K., Krebs, J. R., & Perrins, C. M. (1981). Song Repertoires and Lifetime Reproductive Success in the Great Tit (Parus major). *The American Naturalist*, *118*(2), 149–159. https://doi.org/918

Mcgregor, P. K., Krebs, J. R., & Ratcliffe, L. M. (1983). The Reaction of Great Tits (Parus major) to Playback of Degraded and Undegraded Songs: The Effect of Familiarity with the Stimulus Song Type. *The Auk*, *100*(4), 898–906.

McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *The Journal of Open Source Software*, *3*(29), 861.

Merino Recalde, N. (2023a). Pykanto: A python library to accelerate research on wild bird song. *Methods in Ecology and Evolution*, *00*, 1–9. https://doi.org/10.1111/2041-210X.14155

Merino Recalde, N. (2023b, June). Nilomr/great-tit-hits-setup: V0.1.0. https://doi.org/10.5281/zenodo.8092397

Merino Recalde, N. (2023c, June). Wytham Great Tit Song Dataset. https://doi.org/10.17605/OSF.IO/N8AC9

Mets, D. G., & Brainard, M. S. (2019). Learning is enhanced by tailoring instruction to individual genetic differences. *eLife*, *8*, 1–15. https://doi.org/10.7554/eLife.47216

Morfi, V., Lachlan, R. F., & Stowell, D. (2021). Deep perceptual embeddings for unlabelled animal sound events. *The Journal of the Acoustical Society of America*, *150*(1), 2–11. https://doi.org/10.1121/10.0005475

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258

Nelson, D. A., & Poesel, A. (2014). Tutor choice and imitation accuracy during song learning in a wild population of the Puget Sound white-crowned sparrow. *Behavioral Ecology and Sociobiology*, *68*(11), 1741–1752. https://doi.org/10.1007/s00265-014-1782-z

Odom, K. J., & Benedict, L. (2018). A call to document female bird songs: Applications for diverse fields. *The Auk*, *135*(2), 314–325. https://doi.org/10.1642/AUK-17-183.1

Perez, L., & Wang, J. (2017, December). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. https://doi.org/10.48550/arXiv.1712.04621

Perrins, C. M. (1965). Population Fluctuations and Clutch-Size in the Great Tit, Parus major L. *Journal of Animal Ecology*, *34*(3), 601–647.

Peters, S., & Nowicki, S. (2017). Overproduction and attrition: The fates of songs memorized during song learning in songbirds. *Animal Behaviour*, *124*, 255–261. https://doi.org/10.1016/j.anbehav.2016.09.019

Pitcher, B. J., Harcourt, R. G., & Charrier, I. (2010). Rapid Onset of Maternal Vocal Recognition in a Colonially Breeding Mammal, the Australian Sea Lion. *PLOS ONE*, *5*(8), e12195. https://doi.org/10.1371/journal.pone.0012195

Powers, S. M., & Hampton, S. E. (2019). Open science, reproducibility, and transparency in ecology. *Ecological Applications*, *29*(1), e01822. https://doi.org/10.1002/eap.1822

Radford, A. N., & Ridley, A. R. (2007). Individuals in foraging groups may use vocal cues when assessing their need for anti-predator vigilance. *Biology Letters*, *3*(3), 249–252. https://doi.org/10.1098/rsbl.2007.0110

Raghu, M., Unterthiner, T., Kornblith, S., Zhang, C., & Dosovitskiy, A. (2022, March). Do Vision Transformers See Like Convolutional Neural Networks? https://doi.org/10.48550/arXiv.2108.08810

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science*, *331*(6018), 703–705. https://doi.org/10.1126/science.1197962

Richards, A. (2015, August). *University of oxford advanced research computing*. Manual. https://doi.org/10.5281/zenodo.22558

Riebel, K., Hall, M., & Langmore, N. (2005). Female songbirds still struggling to be heard. *Trends in ecology & evolution*, *20*, 419–20. https://doi.org/10.1016/j.tree.2005.04.024

Riebel, K., Lachlan, R. F., & Slater, P. J. (2015). Learning and cultural transmission in chaffinch song. *Advances in the Study of Behavior*, *47*, 181–227. https://doi.org/10.1016/bs.asb.2015.01.001

Ritschard, M., van Oers, K., Naguib, M., & Brumm, H. (2012). Song Amplitude of Rival Males Modulates the Territorial Behaviour of Great Tits During the Fertile Period of Their Mates. *Ethology*, *118*(2), 197–202. https://doi.org/10.1111/j.1439-0310.2011.01999.x

Rivera-Gutierrez, H. F., Matthysen, E., Adriaensen, F., & Slabbekoorn, H. (2010). Repertoire sharing and song similarity between great tit males decline with distance between forest fragments. *Ethology*, *116*(10), 951–960. https://doi.org/10.1111/j.1439-0310.2010.01813.x

Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2010). Multiple signals for multiple messages: Great tit, Parus major, song signals age and survival. *Animal Behaviour*, *80*(3), 451–459. https://doi.org/10.1016/j.anbehav.2010.06.002

Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2011). Difficulties when Assessing Birdsong Learning Programmes under Field Conditions: A Re-Evaluation of Song Repertoire Flexibility in the Great Tit (M. Giurfa, Ed.). *PLoS ONE*, *6*(1), e16003. https://doi.org/10.1371/journal.pone.0016003

Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2012). Tuning and fading voices in songbirds: Age-dependent changes in two acoustic traits across the life span. *Animal Behaviour*, *83*(5), 1279–1283. https://doi.org/10.1016/j.anbehav.2012.03.001

Roulin, A. (2001). Food supply differentially affects sibling negotiation and competition in the barn owl (Tyto alba). *Behav Ecol Sociobiol*, *49*(6), 514–519. https://doi.org/10.1007/s002650100322

Sainburg, T., & Hedley, R. (2020). Finding , visualizing , and quantifying latent structure across diverse animal vocal repertoires.

Sainburg, T., Theilman, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11605-y

Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2022). Long-distance dependencies in birdsong syntax. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1967), 20212473. https://doi.org/10.1098/rspb.2021.2473

Sethi, S. S., Jones, N. S., Fulcher, B. D., Picinali, L., Clink, D. J., Klinck, H., Orme, C. D. L., Wrege, P. H., & Ewers, R. M. (2020). Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 17049–17055. https://doi.org/10.1073/pnas.2004702117

Shabanov, A. (2023, June). OML-Team/open-metric-learning.

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Sierro, J., de Kort, S. R., & Hartley, I. R. (2023). Sexual selection for both diversity and repetition in birdsong. *Nat Commun*, *14*(1), 3600. https://doi.org/10.1038/s41467-023-39308-5

Sierro, J., de Kort, S. R., Riebel, K., & Hartley, I. R. (2022). Female blue tits sing frequently: A sex comparison of occurrence, context, and structure of song. *Behavioral Ecology*, *33*(5), 912–925. https://doi.org/10.1093/beheco/arac044

Slagsvold, T., Sætre, G.-P., & Dale, S. (1994). Dawn Singing in the Great Tit (Parus Major): Mate Attraction, Mate Guarding, or Territorial Defence? *Behaviour*, *131*(1-2), 115–138. https://doi.org/10.1163/156853994X00244

Stowell, D. (2021). Computational bioacoustics with deep learning: A review and roadmap. *arXiv:2112.06725 [cs, eess, q-bio]*.

Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., Jr, & Llusia, D. (2019). Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience*, *69*(1), 15–25. https://doi.org/10.1093/biosci/biy147

Thakur, A., Thapar, D., Rajan, P., & Nigam, A. (2019). Deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss. *The Journal of the Acoustical Society of America*, *146*(1), 534–547. https://doi.org/10.1121/1.5118245

Thomas, M., Jensen, F. H., Averly, B., Demartsev, V., Manser, M. B., Sainburg, T., Roch, M. A., & Strandburg-Peshkin, A. (2021, December). A practical guide for generating unsupervised, spectrogram-based latent space representations of animal vocalizations. https://doi.org/10.1101/2021.12.16.472881

Thompson, N. S., LeDoux, K., & Moody, K. (1994). A System for Describing Bird Song Units. *Bioacoustics*, *5*(4), 267–279. https://doi.org/10.1080/09524622.1994.9753257

Thorpe, W. H. (1958). The Learning of Song Patterns By Birds, With Especial Reference To the Song of the Chaffinch Fringilla Coelebs. *Ibis*, *100*(4), 535–570. https://doi.org/10.1111/j.1474-919X.1958.tb07960.x

Van Duyse, E., Pinxten, R., Snoeijs, T., & Eens, M. (2005). Simultaneous treatment with an aromatase inhibitor and an anti-androgen decreases the likelihood of dawn song in free-living male great tits, Parus major. *Hormones and Behavior*, *48*(2), 243–251. https://doi.org/10.1016/j.yhbeh.2005.02.013

van Rossum, G. (1995, January). Python reference manual.

Wheelwright, N. T., Swett, M. B., Levin, I. I., Kroodsma, D. E., Freeman-Gallant, C. R., & Williams, H. (2008). The influence of different tutor types on song learning in a natural bird population. *Animal Behaviour*, *75*(4), 1479–1493. https://doi.org/10.1016/j.anbehav.2007.08.030

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Williams, H., & Lachlan, R. F. (2021). Evidence for cumulative cultural evolution in bird song. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1843), 20200322. https://doi.org/10.1098/rstb.2020.0322

Youngblood, M., & Lahti, D. C. (2022). Content bias in the cultural evolution of house finch song. *Animal Behaviour*, *185*, 37–48. https://doi.org/10.1016/j.anbehav.2021.12.012

Zandberg, L., Morfi, V., George, J., Clayton, D. F., Stowell, D., & Lachlan, R. F. (2022, December). Bird song comparison using deep learning trained from avian perceptual judgments. https://doi.org/10.1101/2022.12.23.521425

Zipple, M. N., Nowicki, S., Searcy, W. A., & Peters, S. (2019). Full life course analysis of birdsong reveals maturation and senescence of highly repeatable song characteristics (M. Naguib, Ed.). *Behavioral Ecology*, *30*(6), 1761–1768. https://doi.org/10.1093/beheco/arz146

A singing male great tit. Digital Illustration, Tawatha T. Elguero, 2022
In the background: 2D UMAP sliding window trajectories of the repertoire of a single bird.

4. THE DEMOGRAPHIC DRIVERS OF CULTURAL EVOLUTION IN BIRD SONG: A MULTILEVEL STUDY

Nilo Merino Recalde [1,*], Andrea Estandía [1], Sara C. Keen [1,2], Ella F. Cole [1], and Ben C. Sheldon [1]

[1] Edward Grey Institute, Department of Biology, University of Oxford, Oxford, UK
[2] Earth Species Project, 1536 Oxford St. Berkeley CA 94709, US
*Corresponding author: nilo.recalde@biology.ox.ac.uk

SUMMARY

Social learning within communities can lead to shared behavioural patterns that persist over time, which we know as culture. Examples of culture include bird and whale songs, cetacean feeding techniques, and avian and mammalian migratory routes. Shaped by neutral and selective forces, animal cultures evolve dynamically and form cultural traditions that vary greatly in their diversity and stability. These cultural traits can influence individual and group survival, population structure, and even inform conservation efforts, underscoring the importance of understanding how population processes interact with social learning to shape culture. Although the impact of social learning mechanisms and biases has been extensively explored, the role of demographic factors—such as population turnover, immigration, and age structure—on cultural evolution has received theoretical attention but rarely been subject to empirical investigation in natural populations. Doing so requires very complete trait sampling and detailed individual life history data, which are hard to acquire in combination. To this end, we compiled a multi-generational dataset containing over 109,000 songs from >400 repertoires in a population of great tits (*Parus major*), trained a deep metric learning model to re-identify individuals and quantify song similarity, and fit spatially explicit regression models of cultural diversity and turnover at the individual and neighbourhood scales. We show that demographic variation within the small spatial scales at which learning takes place has the potential to impact the diversity and pace of change of animal vocal cultures. In particular, longer natal dispersal distances within the population reduce cultural diversity and uniqueness, while immigrant birds seem to adopt existing song types rather than introducing new ones, but tend to have larger repertoires and so increase the absolute diversity of neighbourhoods. Birds of the same age tend to have more similar repertoires, and neighbourhoods comprising a mix of ages tend to have more cultural diversity. The main driver of cultural turnover is individual turnover; at the same time, dispersal, a higher proportion of immigrant birds, and an older population slow change down. Our analyses support theoretical expectations regarding a key role of demographic processes in cultural evolution, while highlighting their interaction with species-specific factors such as the timing of song acquisition.

Keywords: animal culture; bird song; demography, cultural evolution

RESULTS AND DISCUSSION

Some behavioural traits are shared and persist within communities due to social learning (Viciana, 2021). We refer to these behaviours as ’animal culture’, exemplified by tool use in capuchin monkeys (Falótico et al., 2019), the learned songs of oscine birds, migration routes (Berdahl et al., 2018; Byholm et al., 2022; Jesmer et al., 2018), and the feeding techniques of some cetaceans (Allen et al., 2013; Rendell & Whitehead, 2001). Animal cultures are not static: neutral and selective mechanisms influence the frequency of cultural traits (Potvin & Clegg, 2015; Williams & Lachlan, 2021), leading to a process of cultural evolution. The resulting cultural traditions vastly differ in their diversity and stability (Tchernichovski et al., 2017), determined by both learning biases and mechanisms and the demographic structure of populations (Deffner et al., 2022; Kandler et al., 2017).

While the role of social learning strategies and biases—frequency dependence, tutor biases, etc.—has been extensively studied (Aplin et al., 2017; Kendal et al., 2015; Lachlan et al., 2018; Pike & Laland, 2010; Tchernichovski et al., 2021), there exists a substantial gap in our understanding of how de-

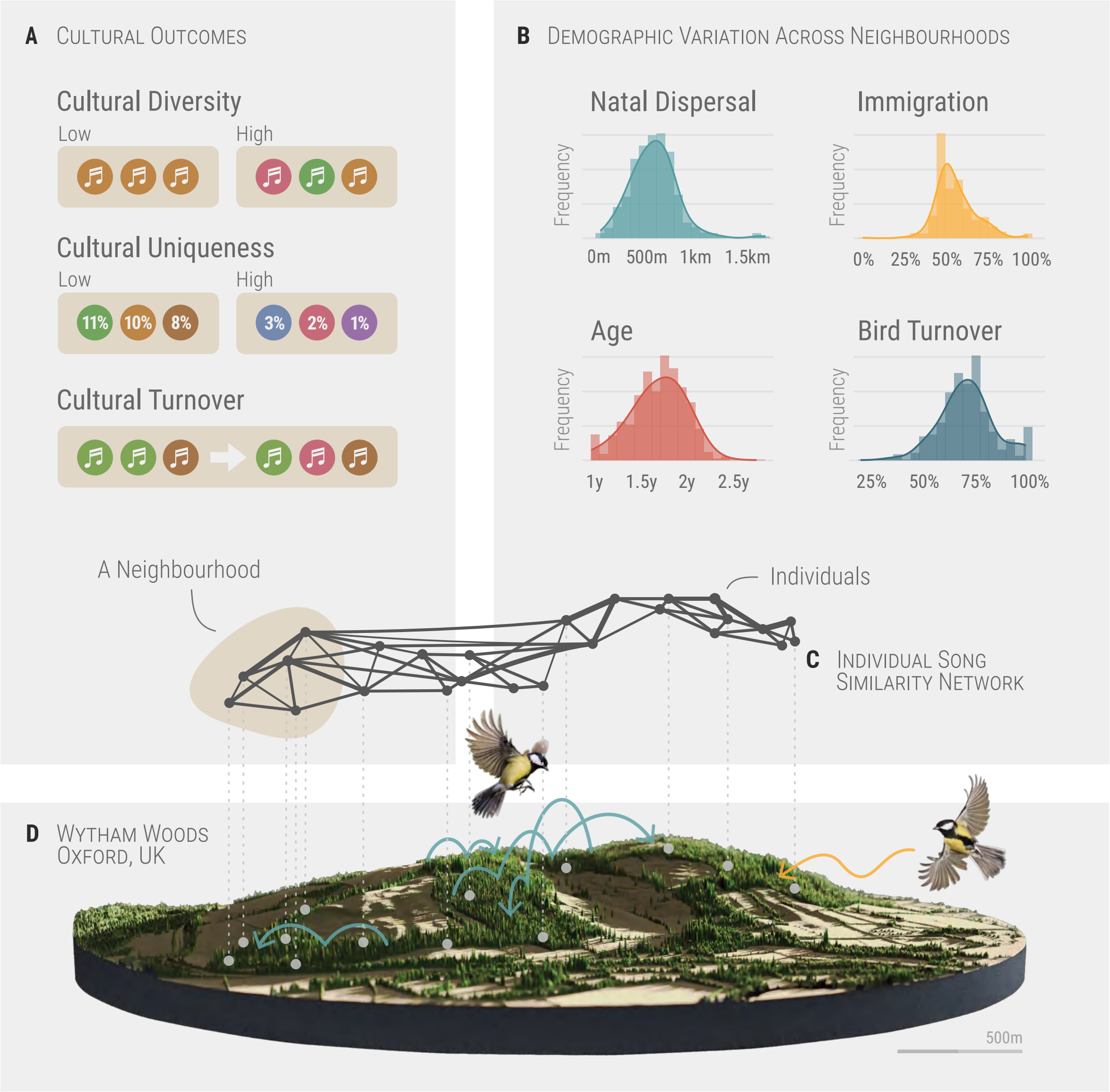


Figure 4.1. Study system and main variables in our analysis.
(A) Cultural variables measured at the neighbourhood level: Diversity is the relative number of unique song types in a neighbourhood; uniqueness is the complement of the average frequency of those song types in the entire population, and turnover is the proportion of unique song types in a given year not present in the same neighbourhood the preceding year. See methods for full definitions. (B) Variation in the demographic properties and composition of neighbourhoods across the population. See demographic variables for full definitions. (C) Cartoon representation of the pairwise continuous repertoire similarity network used in our individual level analyses. Each node represents an individual bird, and the edges represent the similarity between their song repertoires. The network is based on the similarity of the songs produced by each bird during the dawn chorus, and is used to estimate the cultural similarity between individuals. (D) 3D render of our study site, Wytham Woods, seen from the East. Image based on first return LiDAR data (Department for Environment Food and Rural Affairs, 2020) and made with rayshader (Morgan-Wall, 2023). Elevation is exaggerated. Aquamarine (darker) arrows represent natal dispersal within the population, and the yellow (lighter) arrow represents immigration into the population, two of the variables used in this work. Age and bird (individual) turnover are not depicted.

mography contributes to the emergence and persistence of distinct cultural traits within wild populations. Processes such as the recruitment of juveniles and immigration, emigration, mortality, and variation in age structure are likely to strongly affect an individual's opportunities for learning and exposure to different cultural variants, which has been amply emphasized by theoretical work (Barta et al., 2023; Chimento & Aplin, 2024; Deffner & McElreath, 2020, 2022; Deffner et al., 2022; Derex & Boyd, 2016; Dyble & Micheletti, 2024; Fogarty et al., 2019; Kandler et al., 2023; Kirby & Tamariz, 2021; Nunn et al., 2009). However, translating theoretical expectations into empirical evidence remains a challenge (see Chimento et al., 2021; Fayet et al., 2014; Payne and Payne, 1993 for exceptions).

Culture is increasingly recognized as both a fundamental aspect of many animals' lives and a valuable tool in monitoring and conservation efforts (Brakes et al., 2019, 2021). Cultural traits play a role in the survival and reproduction of individuals and social groups; they reflect or even shape the structure of the population (Brakes et al., 2019), and can be lost when habitat fragmentation and population decline lead to reduced learning opportunities (Crates et al., 2021; Paxton et al., 2019). A comprehensive understanding of cultural change and loss, then, requires that we have the ability to detect and study not just intrinsic factors—social, cultural, cognitive—but also extrinsic, ecological and demographic processes. This entails identifying the relevant spatial and temporal scales at which these processes manifest within natural populations, as well as their relative importance.

To contribute to this goal, we built a comprehensive dataset that spans three years and documents the dawn songs produced by male great tit birds during 454 breeding attempts in a single population located in Wytham Woods, UK. The population's marked variation in individual turnover, postnatal dispersal distances, age structure, and immigration across space—known through ongoing long-term monitoring (Lack, 1964)—allowed us to estimate their effects on song cultural repertoires at both individual and group levels. First, we assign more than 109,000 songs in 330 song repertoires to 242 individual birds through a combination of direct physical capture, radio frequency identification microchips, and a novel song-based reidentification method using a deep metric learning model. Then, we quantified individual and group-level traits and analysed variation in song cultural similarity, diversity, and turnover (see definitions) using network and spatially explicit Bayesian multilevel regression models. See Figure 4.1 for a visual abstract of the study.

Our results reveal an interplay of demographics and song cultural dynamics that, albeit complex, largely matches theoretical expectations, as discussed below. This work also demonstrates that bird song, which already provides what is perhaps the largest body of evidence for cultural change in animals (Laland & Janik, 2006), also has the potential to help us shed light on the impact of other population processes on animal cultures, owing to the fact that we can sample song cultural repertoires with relative ease.

Reduced dispersal, higher immigration and age associated with higher cultural diversity

Population genetics provides robust evidence supporting the notion that high dispersal rates facilitate gene flow, which, in turn, reduces the efficacy of selection and diversification. Conversely, low dispersal facilitates genetic differentiation through mechanisms such as mutation and drift, leading to allopatric population divergence (Claramunt et al., 2011; Papadopoulou et al., 2009; Suárez et al., 2022). Were we to adhere to an analogy from genetics to culture, we would anticipate that reduced dispersal rates will decelerate the diffusion of cultural traits (Nunn et al., 2009). This, in turn, should result in the maintenance of distinct behavioural patterns within populations if learning is somewhat accurate, leading to a greater number of cultural variants unique to a neighbourhood or region (Planqué et al., 2014; Whitehead & Lusseau, 2012). Our analysis indeed indicates that neighbourhoods where more birds have remained in proximity to their natal areas harbour greater and more unique cultural diversity (*diversity:* $P(\beta_{disp(\overline{m})} < 0|D) = 1$, $mem = -0.018$, $CI_{95\%}$ $[-0.023, -0.012]$; *uniqueness:* $P(\beta_{disp(\overline{m})} < 0|D) = 0.96$, $mem = -0.005$, $CI_{95\%}$ $[-0.01, 0]$; Figure 4.2A&B, Table S4.3), in line with prior research at a much coarser grain (Fayet et al., 2014).

The analogy breaks down as soon as we con-

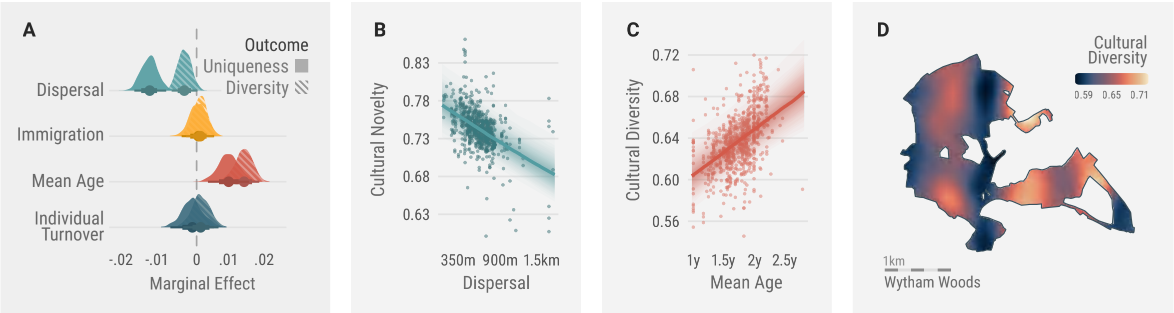


Figure 4.2. Influence of demographic variables on cultural diversity and uniqueness within neighbourhoods.
(A) Marginal effects at the mean of neighbourhood characteristics including mean dispersal distance, proportion of immigrant birds, average age, and individual turnover. (B) Adjusted predictions and partial residuals of the effect of mean neighbourhood dispersal distance on cultural uniqueness. Low-dispersal neighbourhoods are those in which birds were born in the same area. (C) Adjusted predictions and partial residuals of the effect of the mean age of the neighbourhood on cultural diversity. A neighbourhood with a mean age of 1 would be one where all birds are breeding for the first time. (D) The average distribution of cultural diversity in the population across space during the study period (2020-2022). This map captures the residual variation in cultural diversity after taking into account the demographic variables in (A). The map is based on a Gaussian process model with an exponentiated-quadratic kernel covariance function, which allows us to interpolate between the locations where we have data.

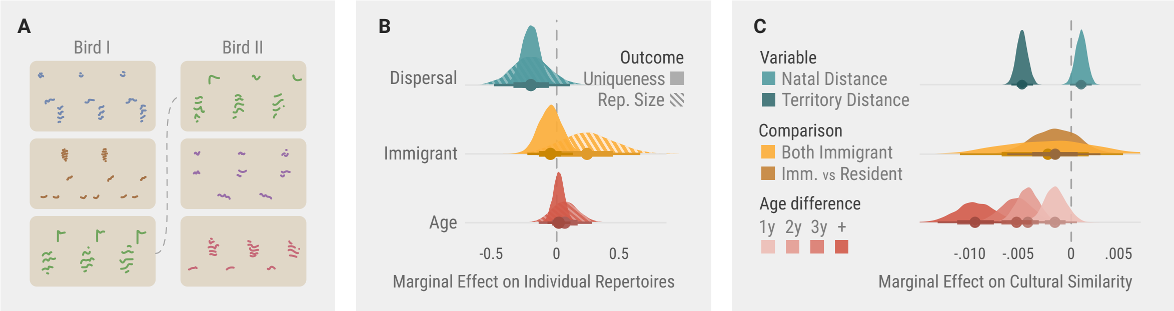


Figure 4.3. Individual and dyadic analysis of cultural diversity and similarity.
(A) Illustrative example showing the repertoires of two different birds in the population, with three songs in each, one of which is shared. Each sub-panel shows a cartoon spectrogram, with time in the horizontal axis and frequency in the vertical. Units not shown. (B) Marginal effects of dispersal, immigration status, and age on individual song repertoires, described in therms of their absolute size (the total number of distinct song types sang by that bird) and their relative uniqueness (how rare, on average, those songs are in the entire population within that year). (C) Marginal effects of dispersal, immigration status, and age on song repertoire similarity between individuals. Dispersal: Birds that are close neighbours are more culturally similar, regardless of where they were born, whereas natal distance may have a weak positive effect on cultural similarity. Immigration: There is no strong evidence that birds born outside the population are dissimilar from resident birds. Age difference: Birds are less culturally similar the greater the difference in their birth years.

sider the underlying individual-level mechanisms, however, due to complex interactions between the timing of dispersal and learning mechanisms that are specific to the cultural domain. Some species only learn songs early in life from their parents, in a manner reminiscent of genetic inheritance, while others learn continuously from their neighbours, or only after dispersal (see Searcy et al., 2021 for an overview of different strategies). In the case of our study species, the great tit, these learning mechanisms are thought to involve selective retention or modification of songs encountered early in life, while they disperse, and up until before they begin breeding for the first time—a process that results in crystallized repertoires of 1 to $\simeq$ 10 different song types (McGregor & Krebs, 1982b; Merino Recalde et al., 2023; Rivera-Gutierrez et al., 2011). In our individual-level analysis, we see that birds that dispersed over longer distances tend to have learned repertoires composed of songs that are more common within the population (*uniqueness:* $P(\beta_{disp(m)} < 0|D) = 1$, $mem = -0.2$, $CI_{95\%}$ $[-0.3, -0.09]$; Figure 4.3B, Table S4.3), and possibly smaller repertoires as well (*rep.size:* $P(\beta_{disp(m)} < 0|D) = 0.91$, $mem = -0.2$, $CI_{95\%}$ $[-0.44, 0.05]$; Figure 4.3B; Table S4.3). We hypothesize that birds with more extensive movements are more likely to sample and acquire a larger proportion of common cultural variants, simply because they are exposed to more songs across their learning period. This finds support in a spatially explicit simulation of song learning with dispersal, which shows this pattern would emerge under positively frequency-dependent learning or a process leading to similar acquisition curves (see Figure S4.10; note that we do not currently know which learning strategies are employed by great tits).

Building on our understanding of cultural dynamics in relation to dispersal we expect that, when song learning is relatively precise and dispersal is limited, cultural differences between populations will accumulate, and immigration will introduce cultural novelty to the recipient population. However, the extent to which immigration introduces new cultural variants also hinges on an interplay between the species' learning programme, the timing of dispersal, and the spatial movements of individuals. Animals that learn first and then disperse, for example, may bring cultural novelty with them. But this is not the case for great tits, whose young disperse in late summer and autumn, shortly after achieving independence, learn their songs until the end of their first winter (Rivera-Gutierrez et al., 2011), and become chiefly sedentary as adults (Dhondt, 1979; Dingemanse et al., 2003; Greenwood et al., 1979). In this species, then, we anticipate that immigrant birds will learn or retain songs they encounter upon arrival, either before or during the establishment of their territories (Graham et al., 2018; Keen, 2020).

Indeed, in our population, we find no evidence that the repertoires of birds originating from outside the population significantly differ acoustically from those of resident birds ($mem = -0.002$, $CI_{95\%}$ $[-0.006, 0.002]$; Figure 4.3C). This, in conjunction with the observation that song repertoire similarity between individuals is predicted by the distance between breeding territories ($mem = -0.005$, $CI_{95\%}$ $[-0.006, -0.004]$; Figure 4.3C; Table S4.3), supports the hypothesis that great tits are predominantly closed-end learners that learn primarily from territorial neighbours after dispersal (McGregor & Krebs, 1982b; Rivera-Gutierrez et al., 2011).

This leads to a somewhat contradictory scenario,

Figure 4.4. Influence of demographic variables on cultural turnover within neighbourhoods. (A) Marginal effect at the means for mean dispersal distance, proportion of immigrant birds, average age, and individual turnover on the rate of cultural turnover within neighbourhoods (B and C) Adjusted predictions and partial residuals of the effects of the proportion of individual turnover on cultural turnover (B) and the effect of mean neighbourhood age on cultural turnover within neighbourhoods (C). (D) The population's average distribution of cultural turnover across space during the study period (2020-2022). (E) Posterior counterfactual predictions for two scenarios: all variables at their maximum (Max.) and minimum (Min.) observed values in the population, holding turnover constant at its mean value. Cultural turnover is expected to be over two times higher if neighbourhood dispersal, immigration and age are low, as they slow down cultural change.

however: immigrant birds, while not acoustically distinct, tend to have larger repertoires compared to their resident counterparts (Figure 4.3B; $P(\beta_{\overline{imm.}} > 0|D) = 0.87$, $mem = 0.24$, $CI_{95\%}$ $[-0.098, 0.593]$; Table S4.3). At the group level, this small and uncertain effect amplifies, such that neighbourhoods with a higher proportion of immigrant birds do not have increased cultural diversity relative to the total number of songs ($mem = 0.002$ $CI_{95\%}$ $[-0.004, 0.007]$; Figure 4.2A); but they do have a higher absolute cultural diversity—above what would be expected based solely on the number of birds ($P(\beta_{\overline{imm.}} > 0|D) = 0.98$, $mem = 0.47$, $CI_{95\%}$ $[0.1, 0.84]$; Figure S4.9, Table S4.3).

Previous research (Verhulst et al., 1997) has revealed that most birds arriving from outside the population disperse over two kilometres, significantly farther than the typical distances observed within the population (median for males = 558 metres; Greenwood et al., 1979). This extended dispersal may have qualitative consequences for cultural diversity, through a combination of factors: first, an initial exposure to songs from the source population; then, a heightened pressure to adopt vocalizations similar to those of territorial neighbours to avoid any social or reproductive costs associated with non-local signals, as seen in other species (Baker et al., 1981; Beecher, 2008; Lachlan et al., 2014; Mortega et al., 2014; Payne, 1983).

Finally, we find that individual turnover does not significantly affect cultural diversity or uniqueness, and we uncover an association between age structure and cultural diversity and uniqueness. Individuals of the same generation share the most similar song repertoires, and while age itself does not directly relate to changes in the repertoires of individual birds (Figure 4.3B), the acoustic similarity between pairs of individuals decreases as the age gap between them widens (Figure 4.3C; Table S4.3). This is expected in birds that cease to learn new songs as they age, and has detectable consequences for neighbourhoods: those with a higher proportion of older individuals have heightened levels of cultural diversity and uniqueness (Figure 4.2A&C, Figure S4.9). Conversely, in areas where the majority of the population comprises younger birds surrounded by similar-aged peers, birds tend to produce fewer unique songs that are also more common within the population (*diversity:* $P(\beta_{\overline{age}} < 0|D) = 1$, $mem = 0.021$, $CI_{95\%}$ $[0.014, 0.027]$; *uniqueness:* $P(\beta_{\overline{age}} < 0|D) = 0.99$, $mem = 0.012$, $CI_{95\%}$ $[0.005, 0.019]$; Figure 4.2A&C, Table S4.3).

Demographic processes moderate the rate of cultural change at small spatio-temporal scales

We now shift our focus from static measures of cultural diversity to cultural turnover, examining how quickly song types disappear from neighbourhoods and the consequences this has for their cultural makeup. The primary driver of cultural turnover within neighbourhoods is individual turnover (total effect $mem = 0.072$ $CI_{95\%}$ $[0.051, 0.093]$): as birds leave or die, many song types disappear with them, and the young birds that replace them might speed up the adoption of new song types (Dyble & Micheletti, 2024). Across the three-year study period, now considering the entire population, cultural turnover between consecutive years averages 0.45 (0.47 and 0.44; note that specific values are contingent on the granularity of song type definitions, see manual categorization). If all variants faced an equal chance of disappearing, this would quickly lead to complete cultural replacement. However, after a two-year gap, turnover only slightly increases to 0.59%. We anticipate this rate to taper further over longer periods, as rare variants encounter greater stochasticity while common songs endure (Figure S4.8A). Indeed, some common song types documented over four decades ago persist within the population (Keen, 2020; McGregor & Krebs, 1982b). This persistence might be due to different factors, like accurate learning based on song frequency, or strong tendencies to converge on certain song types (Claidière & Sperber, 2007; James & Sakata, 2017; Lachlan et al., 2018; Tchernichovski et al., 2021).

After factoring in individual turnover, we examine how mean natal dispersal distance, the proportion of immigrant birds, and mean neighbourhood age affect cultural change within a neighbourhood. We find that higher levels of these factors correlate with slower cultural change (Figure 4.4A; Table S4.3). Specifically, when individuals have dispersed over greater distances to get to their breeding neighbourhood, there is a high influx of immigrants, and the distribution of age is skewed towards older individuals, the model predicts slower cultural change, at less than half the rate compared to the converse scenario ($[0.66, 0.77]0.39$ $CI_{95\%}$ $[0.24, 0.51]$ vs. 0.72 $CI_{95\%}$, as illustrated in Figure 4.4E). This twofold difference in cultural turnover, based on realistic parameter values, is likely to significantly influence the cultural composition of neighbourhoods over time if it persists, highlighting the importance of the demographic structure of populations in also moderating the pace of cultural change.

The observed slowdown in cultural change due to dispersal and immigration aligns with our findings that dispersal homogenizes song repertoires and that immigrants tend to adopt the existing population's variants rather than introducing new ones (see Figure 4.2A and Figure 4.3A). At the same time, our estimate for the effect of neighbourhood age ($P(\beta_{\overline{age}} < 0|D) = 1$, $mem = -0.044$, $CI_{95\%}$ $[-0.063, -0.026]$; Figure 4.4C) aligns with modelling work suggesting that learning from older individuals should slow down cultural change (Kirby & Tamariz, 2021). Indeed, age may serve as a brake on change, as older birds continue to sing song types that are becoming less frequent in the population, an idea supported by the observation that individual birds' repertoires are least similar when there is a large age difference (Figure 4.3C). The differences between the older and younger birds' repertoires also increases cultural diversity and uniqueness within neighbourhoods that include many older birds, as discussed above, suggesting an important role of age structure in shaping both cultural diversity and turnover.

Consequences for cultural structure, stability and diversity

Cultural traits, learnt bird song in this case, are shaped by many factors: some external, such as those discussed here, others intrinsic to learning and culture, and yet others that arise from selective processes driven by preference and function. Even within the confines of a relatively small population—Wytham Woods spans a mere four kilometres—we have been able to recover associations between heterogeneity in the demographic composition of neighbourhoods and cultural outcomes using a large dataset of song repertoires, and show that these are most likely underlied by differences in individual learning and exposure to cultural variants. In particular, dispersal within the population reduces cultural diversity and uniqueness. Birds that were born outside the population seem to adopt existing song types rather than introducing new ones, but tend to have larger repertoires and so increase the absolute diversity of neighbourhoods. Birds of the same age tend to share similar song types, while neighborhoods comprising both older and younger birds are more likely to exhibit a broader array of song types. Additionally, such neighborhoods are more likely to host a greater number of birds singing rare song types, perhaps because, as we also find, aged neighbourhoods have slower cultural turnover. The main driver of cultural turnover is individual turnover, and, at the same time, longer postnatal dispersal distances, a higher proportion of immigrant birds, and an older population slow it down. This emphasizes the need for both empirical studies and modelling efforts on cultural change to account for the population's demographic characteristics and their inherent heterogeneity across time and space, as these shape individuals' exposure to cultural variants and opportunities for learning and, therefore, emergent group-level cultural dynamics.

METHODS

Resource availability

The complete Wytham great tit song dataset is available in osf.io/n8ac9, and documented here. The main repository with code and data to reproduce the analysis and figures in this article can be found at github.com/nilomr/birdsong-demography.

Data collection

Study system and fieldwork

Great tits are small, short-lived birds—average lifespan: 1.9 years—that sing acoustically simple yet highly diverse songs. Each male great tit has a repertoire of one to over 10 song types, also referred to as song types, which are repeated multiple times in short bursts separated by longer periods of silence. Although detailed studies on how individual great tits learn their songs are limited, existing evidence suggests several key points. First, it appears that great tits do not learn their song repertoires from their fathers (McGregor & Krebs, 1982b). Instead, their song development is influenced by the songs they encounter during their early life until they establish a territory and breed for the first time. We do not currently know how precisely great tits learn songs, or how social interactions affect the process. This period of vocal learning results in a final crystallized repertoire that remains relatively stable afterward—a process known as close-ended learning (Rivera-Gutierrez et al., 2011). Additionally, while there is evidence that birds can continue to learn to recognize new songs later in life (that is, learning for discrimination, as opposed to learning for production), this ability seems to be limited compared to their early learning experiences (McGregor & Avery, 1986). Furthermore, females are able to individually recognize males based on their songs (Lind et al., 1996), and even across a large population, individual song renditions can accurately indicate the identity of the bird producing them (Merino Recalde et al., 2023).

During the breeding season, from March to June, great tit pairs are socially monogamous and defend territories around their nests (Hinde, 1952). In Wytham Woods, Oxfordshire, UK (51°46 N, 1°20 W), a population of these birds has been the focus of a long-term study since 1947 (Lack, 1964). Wytham Woods is a semi-natural, predominantly deciduous and partly discontinuous woodland that spans an area of approximately 385 hectares and is surrounded by farmland. Most great tits in this population breed in nest boxes with known locations, and the majority of individuals are marked with a unique British Trust for Ornithology (BTO) metal leg ring as either nestlings or adults.

We collected data from late March to mid-May during the 2020, 2021, and 2022 breeding seasons. Every year, fieldworkers checked each of the 1018 nest boxes at least once a week before and during the egg-laying period, which typically lasts from one to 14 days (Perrins, 1965), and recorded the identities of breeding males and females, the dates of clutch initiation and egg hatching, clutch size, and fledgling number and condition under standardized protocols. We found the first egg date by assuming that one egg is laid every day and counting back from the day of observation. In cases where we did not observe the chicks on their day of hatching, the actual hatching date was determined by assessing the weight of the heaviest chicks and extrapolating their age from established growth curves (Cresswell & Mccleery, 2003; Gibb, 1950).

Nest box occupancy and breeding density vary across the study area, with some areas having a higher density of nest boxes and a higher proportion of occupied boxes, which we account for in some of our analyses as described in the sections below. In the years of our study, 261, 289 and 278 nest boxes were occupied by pairs of great tits, with 173, 184 and 184 that lead to successful breeding attempts where at least one chick fledged. See Figure S4.1 for a map of the study site and sampling locations.

To record the vocalizations of male great tits we took advantage of their behaviour during the reproductive period, when they engage in continuous singing near their nests at dawn before and during egg laying (Mace, 1987). Collectively, this vocal display is referred to as the dawn chorus, and has been demonstrated to yield a reliable estimation of the song repertoire of individuals when recorded in full (Rivera-Gutierrez et al., 2012; Van Duyse et al., 2005). As soon as we suspected that a pair of great tits were using a nest box—based on nest lining materials, egg size if present, or other signs of activity—we deployed an autonomous sound recorder nearby. The microphone faced upwards and slightly away from the nest box, aligning with the nest box entrance hole's direction. The birds sang near the recorder: although we did not gather data on this aspect, our anecdotal observations were in line with a different population where the average distance to the nest box while singing was 10 metres (Halfwerk et al., 2012). The birds also changed perches and moved around during our recording. Although variation in sound amplitude due to changes in distance and direction could affect song selection, we did not observe systematic bias, ruling out potential issues like consistently low signal-to-noise ratios causing exclusion of entire song types.

For this study, we calculated postnatal dispersal distance was as the straight line distance from the natal site to the breeding site. The dispersal distances of birds classed as immigrants (not ringed as chicks in the population) are not known, but most are thought to come from other populations at least 1 km, and likely more than 2.5 km, away (Quinn et al., 2011; Verhulst et al., 1997). We determined age based on the year of hatching for birds born in the population and using plumage characteristics for immigrants, which are most often caught as yearlings (76%)—allowing us to age them accurately (Woodman et al., 2023).

All work involving birds was subject to review by the University of Oxford, Department of Zoology, Animal Welfare and Ethical Review Board (approval number: APA/1/5/ZOO/NASPA/ Sheldon/TitBreedingEcology). Data collection adhered to local guidelines for the use of animals in research and all birds were caught, tagged, and ringed by BTO licence holders (NMR's licence: C/6904).

Recording equipment and schedule

We used 60 (30 in 2020) AudioMoth recorders (Hill et al., 2019), which were housed in custom-built waterproof enclosures. Recording began approximately one hour before sunrise (05:36 – 04:00 UTC during the recording period) and consisted of seven consecutive 60-minute-long recordings with a sample rate of 48 kHz and a depth of 16-bit. To sample as many birds as possible, we left each recorder at the same location for at least three consecutive days before moving it to a different nest box. We relocated 20 recorders (10 in 2020) every day throughout the recording period.

Data processing and annotation

We processed and annotated the song recordings, 109,963 in total, using custom software and scripts written in Python 3 (van Rossum, 1995) and the open source package `pykanto` (Merino Recalde, 2023a). These are available from github.com/nilomr/great-tit-hits-setup (Merino Recalde, 2023b). Our annotated dataset and a detailed description of the process can be found in Merino Recalde et al(Merino Recalde et al., 2023).

Re-identifying individuals from their songs

In some cases, we failed to capture or identify birds using PIT (Passive Integrated Transponder) tags, as tags sometimes malfunction, birds can be difficult to catch, or disappear before their chicks are old enough to attempt this. We were able to match an additional 31 song repertoires belonging to this subset of birds with existing individual IDs by training a deep metric learning model to recognize individual songs. This increased the number of identified repertoires from 299 to 330, belonging to 242 unique birds. We summarise the process here, but see (Merino Recalde et al., 2023) for details on model design and training.

Briefly, we calculated pairwise song distances using the feature vectors obtained from a vision transformer on a metric learning task. This model achieved a mean average precision at 5 (mAP5) of 0.98 and a cumulative matching characteristic at 1 (CMC1) of 0.98, which roughly means that in 98% of the queries made to the similarity space on a held out test dataset the returned candidate song type by a bird is the correct one. Given that the model was trained on almost 2000 classes, this means that there is enough individual information contained in each song type to distinguish between birds with high confidence, even if we dont know just how precisely they are able to learn songs from a tutor.

We then use these distances between song embeddings to infer when two repertoires belong to the same bird, even if had not been physically identified more than once. To do this, we set very conservative criteria: that a minimum of two songs are more similar than the 0.025 quantile of the distribution of distances for all song types sung by birds already known to be different, or 0.9, and that the bird is no more than 100 metres apart from the reference bird. If we query the metric learning model across the songs of the entire population, which increases the chances of a false positive, we find that the percentage of wrongly assigned birds in our ground truth data (birds that we have physically re-identified across years) is 0.3%, or 34 out of 11,359 unique comparisons if we only use acoustic distance—further reduced to 0.04%, or 4 out of 11359, if we also add the spatial constraint (see Figure S4.7 for a graphic explanation).

Characterising repertoire similarity

Our analyses require i) a measure of the acoustic similarity between any two birds for the individual-level analyses, and ii) a way to identify categorical song types for the neighbourhood analyses. The underlying assumption is that song repertoires will be more similar if one bird has learned it at least in part from a second, or if they have both learnt from other individuals who are themselves similar due to intergenerational cultural descent. There is no single optimal solution for this problem, both due to technical challenges and because we do not know enough about song perception and learning mechanisms in this species. There are three main possible approaches, each with its own advantages and disadvantages, which we describe below.

Continuous similarity (individual-level analyses)

Traditional methods used to compare bird vocalizations include visual inspection of spectrograms and measurement of hand-picked acoustic features. However, these approaches have limitations in dealing with noise and variations in performance and can be extremely time-consuming. So, instead, we adopted a data-driven approach by training a Vision Transformer (ViT) model for feature extraction in a metric learning task on spectrogram representations of great tit songs. Our goal was to create a similarity space based on inherent variation in the data, using categorical labels of song types sung by individual birds, which we know to be perceptually and behaviourally significant (Lind et al., 1996). Further details, data, and code to reproduce this process are available at (Merino Recalde, 2023a) and (Merino Recalde et al., 2023); we include a summary here.

We optimized model performance using Triplet loss, a loss function that ensures the projection of a positive sample (belonging to the same song-type-within-individual class as the anchor point) is closer to the anchor's projection than that of a negative sample (belonging to a different class) by a specified margin (Hermans et al., 2017; Hoffer & Ailon, 2018). During training, we mined hard triplets—where the negative sample is closer to the anchor than the positive—and used the Adam optimizer with a fixed learning rate of 1×10^{-5}.

To prevent the model from overfitting to irrelevant features within localized portions of the training data or background noise, we applied various train-time data augmentation techniques (Mumuni & Mumuni, 2022; Perez & Wang, 2017; Shorten & Khoshgoftaar, 2019). These included random cropping in the time domain, dropping out parts of the spectrogram, adding Gaussian and multiplicative noise, equalization, sharpening, changes to brightness and contrast, blurring, and slight shifting in both time and frequency domains. This leads to a more robust model that can generalize better.

Finally, we used the resulting model to calculate feature vectors for each song in the dataset (109,963 samples x 384 dimensions), which serve as compressed representations that can be used to compare them. Great tits have variable repertoire sizes and there is no evidence that they ever learn them en bloc (McGregor & Krebs, 1982b; Rivera-Gutierrez et al., 2010). Therefore, the simplest continuous measure (an average pairwise Euclidean distance between all songs) would mask any signatures of learning if the average repertoire similarity is similar across the population, and does not take into account the asymmetry in total repertoire size. To improve on this, we define repertoire similarity as the average minimum Euclidean distance (AMED), given by

$$\text{AMED} = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} |a - b|_2 \qquad (4.1)$$

where we compare each song feature vector a in set A with all song feature vectors b in set B and compute their Euclidean distance $|a - b|_2$. We then retain the minimum distance for each element in set A and obtain the AMED by averaging these minimum distances over all elements in set A. The main advantage of this approach is that it allows us to avoid imposing discrete population-wide song categories. On the other hand, if song learning is categorical and not very precise in terms of fine song structure, this method could underestimate it or fail to detect it. We used this approach for all individual-level analyses in this paper.

Automated clustering (first step for neighbourhood analyses)

Instead of calculating a continuous measure of repertoire similarity, we can build a pairwise distance matrix for all songs, assign them to discrete clusters using a clustering algorithm, and then calculate the intersection between repertoires by using the Jaccard coefficient or modelling it as a binomial process, with n = the combined repertoire size and s = the number of songs in the same cluster. Here we used UPGMA hierarchical clustering and dynamic tree-cut techniques to classify the songs into distinct types (since songs consist of repeated syllables, syllable and song types are synonymous in this species). We allowed a minimum cluster size of 1 to ensure the representation of rare song types. The usefulness of this method relies on the global properties of the embedding space derived from the continuous similarity process. In a low-dimensional space where linear distances effectively capture meaningful variation, creating clusters by cutting the hierarchical tree at different heights yields varying cluster counts while maintaining meaningful groupings. However, in a high-dimensional space where global distances are not meaningful, only relatively small clusters of nearby points remain interpretable. This is the case with our dataset and embedding space: we find that the method reliably groups song renditions by the same bird across different years, alone or together with other birds with highly similar songs, yet consistently splits songs that are similar by human (and perhaps great tit, Falls et al., 1982) standards, ultimately leading to a very large number of clusters (the most stable clustering solutions were close to the total number of different individual song types, >1000). Due to these issues, we did not use song types defined in this way, and performed an additional manual categorization step.

Manual categorization (second step for neighbourhood analyses)

All great tit song research has, to date, visually classified songs into population-level types (Baker et al., 1987; Falls et al., 1982; Fayet et al., 2014; Hutfluss et al., 2022; McGregor & Krebs, 1982a; McGregor & Krebs, 1982b; McGregor et al., 1981). This process is both inevitable and very subjective. However, despite its clear problems, human perceptual judgments might be our best available substitute for those of the birds for some tasks (but see recent work by Morfi et al., 2021; Zandberg et al., 2022). Indeed, across fields, advanced classification algorithms are routinely evaluated against ground truth created by humans, and this is also the case in bird song research.

Our neighbourhood-level analyses require that we define discrete cultural units, so, given the difficulties with the alternatives described above, we adopted a variant of this approach and used the criteria followed by McGregor and Krebs McGregor and Krebs (1982b) and most subsequent work. With over 100,000 songs, our dataset is much larger than is common in the field and would have been impossible to label entirely manually. Instead, we used the output of the automated clustering process described above, consisting of labelled song repertoires (birdID x song type). This made the problem 57 times smaller: 1920 song types that were already assigned to small clusters of highly similar songs, which we reviewed manually.

Following common practice in the field, we validated our resulting 335 manually assigned labels statistically (see Figure S4.8), although we note that i) the ability of a statistical method to differentiate between manually defined clusters does not mean that these are perceptually meaningful, only that they can be distinguished in a manner that aligns with human classification, and ii) a large range of clustering solutions will be compatible with the data. To do this, we retrained the ResNet50-based classifier described in Merino Recalde Merino Recalde (2023a) to predict manual cluster ID using a random subset of the data and obtained an accuracy of 0.87 on the validation set (see other metrics in the repository). With the caveats already mentioned, this means that our manual classification following McGregor and Krebs McGregor and Krebs (1982b) is successful at finding a stable solution that reduces intraclass variation. A comparable process by Fayet et al. Fayet et al. (2014) was able to reach 0.71 accuracy for 374 songs. We further explored the result by building a dendrogram based on the confusion matrix during test time and reviewing the classes that were not well supported, which led us to combine seven classes into two. There is an inverse relationship between how densely occupied a region of the song space is and the ease with which we can find categorical divisions: the more examples the more graded the variation and, in consequence, what may have seemed like clear-cut categories if we had fewer data blend into one another without an obvious transition.

In practical terms, because most of the great tits in our population sing some variation of the well-known ’tea-cher, tea-cher’ song, these are much harder to categorize than the many rare songs with complex structures only sung by one or a few birds. This was our impression when manually labelling the songs, and it was also the case when applying the supervised classification algorithm. As mentioned in the main text, the direct consequence of this for our analysis is that the absolute estimates of cultural turnover depend on the granularity of this process: when we lump all similar ’tea-cher’ songs, as McGregor and Krebs McGregor and Krebs (1982b) do, the estimates of turnover are necessarily lower—but, crucially, any relative differences remain the same. Precisely quantifying turnover using categorical metrics might be altogether impossible in some speceis, because song categories are not likely to be natural kinds beyond the scale of interacting neighbours. See Figure S4.4 for examples of song types defined in this study.

The code used to perform this song type validation process, along with the figures generated during it, can be found in the main narrative notebook and a dedicated repository.

Quantification and statistical analysis

Modelling pairwise similarity and individual repertoires

It is common for analysis of song similarity to fit simple linear regression models using all pairwise comparisons in a population. However, this leads to very strong pseudo-replication and, therefore, an increased chance of Type I errors. To avoid this, we treat our song similarity data as a fully connected network and build Bayesian multilevel models with a multi-membership structure and the pairwise AMED described above as the response variable. The full model specifications can be found in the main repository for this project; also see a summary in Table S4.1 and variable descriptions in Table S4.2.

Individual repertoires We first modelled individual repertoire size using Poisson and negative binomial models, but this led to poor performance as assessed through posterior predictive checks (both underestimation of mean values and either under or overestimation of very low repertoire sizes). Instead, we built continuation ratio models, a type of sequential ordinal model where reaching a particular level (number of song types in the repertoire) requires first reaching all lower levels (Chambers & Drovandi, 2023; Warti et al., 2020). rep_{m_1}, $rep_{m_{1.1}}$, and $rep_{m_{1.2}}$ estimate the association between immigrant status, distance dispersed, age, and repertoire size. Three further log-normal models, $repnov_{m_1}$, $repnov_{m_{1.1}}$, $repnov_{m_{1.2}}$, do the same for the average cultural uniqueness of individual repertoires, defined as one minus the mean of the relative frequencies of each song type present in the repertoire.

Pairwise similarity Our first model ($disp_{m_1}$) explores the interaction between natal distance, that is, the distance between the nests where two resident birds were born, and the distance between the centre of their breeding territories, adjusting for year and absolute age difference. We do not have direct information on how long birds have spent around one another, so instead we estimate the effect of the interaction of the distance at which they were born and the distance at which they subsequently breed: If both are small, they will have had more opportunities for interaction and learning. We extract predictions for the interaction and calculate marginal effects at minimum distances, to answer the questions 'How does cultural similarity change with distance for birds that were born nearby' and 'Does how close a bird was born matter for birds that hold territories nearby'. We use a similar model structure (age_{m_1}) to estimate the marginal effects of the absolute age difference, this time adjusting for the natal and territorial distance between birds. Then, to study the effect of immigration, we fit a model (imm_{m_1}) with the possible combinations of immigration status (both immigrant, both residents, one of each) and adjust for age difference and territorial distance.

Modelling group-level outcomes

Defining neighbourhoods and their demographic properties Song turnover, diversity, and uniqueness are group-level properties. However, our study lacks naturally occurring distinct subpopulations that we can use as units for analysis. Rather than partitioning the population using a discrete polygonal grid or non-overlapping areas, we opted to model neighbourhoods continuously across space, with a radius of 200 m around each of 549 nest boxes occupied at least once during the study (Fayet et al., 2014), which we sampled across the duration of the study. This radius is necessarily arbitrary but strikes a good compromise between describing the relevant spatial scale at which vocal interactions occur, which extends up to around 180 metres (Bircher et al., 2021; Blumenrath & Dabelsteen, 2004), and maintaining an adequate sample size in areas of low density (neighbourhood sizes range from 7 to 60, mean 30.3 ± 14.9). Importanty, neighbourhoods defined in this way are highly non-independent, so we model both this methodological spatial dependence and other sources of complex spatial autocorrelation intrinsic to the study site by including a 2D Gaussian process (GP), which estimates a length-scale parameter defining a variance-covariance matrix for the spatial locations based on their distance (Dearmon & Smith, 2016; Gelfand & Schliep, 2016; Wright et al., 2021). We confirmed that this eliminated the residual spatial autocorrelation via Moran's I tests. Note that we fit a separate GP for each year, as treating the spatial dependence as fixed across the study duration, as is often done, risks further underestimating uncertainty.

We define our predictor variables in the following way:

- Individual turnover is the proportion of birds that were not already in a neighbourhood in the preceding year.
- Dispersal is the mean of the distances, in metres, that birds in the neighbourhood travelled to get from their natal territories to their current territories if they were hatched within the Wytham population.
- Immigration is the proportion of birds that were not ringed as nestlings in the population, and
- Neighbourhood age is the mean age of the birds within it.

Figure S4.2 illustrates that our sampling process did not introduce bias into any of these predictor variables: the birds from which we recorded song repertoires were, on average, representative of the true neighbourhood composition. Figure S4.3 shows the distribution of the 'raw' variables across the study site.

Operational definitions of cultural diversity, uniqueness, and turnover In this paper we use simple descriptions of cultural diversity, uniqueness, and turnover, that intuitively capture relevant aspects of the collective song repertoires of neighbourhoods in relation to the rest of the population. We define our outcome variables in the following way:

- We calculated a simple relative diversity index by dividing the number of unique song types by the total number of songs in a neighbourhood.
- To calculate the uniqueness index, we computed the relative frequency of each song type in the current year in the entire population. We then took the complement (1 -) of the logarithm of the mean of the relative frequencies of each song type present in the neighbourhood.
- We calculated the rate of song cultural turnover as the proportion of unique song types in a given year that were not already present in the same neighbourhood the preceding year.

In this way, 'diversity' describes the relative proportion of unique songs in a neighbourhood, and 'uniqueness' refers to how uncommon, on average, the songs of the birds in a neighbourhood are. These two ways of characterizing cultural diversity are (as expected) anti-correlated in our study site due to the effect of sampling: more frequent songs are sampled more readily, causing larger sample sizes—neighbourhoods with more density and therefore songs—to yield lower average estimates of diversity and higher average estimates of uniqueness, in a non-linear manner. Once this is adjusted for, which we do by including a GAM term capturing the nonlinear effect of the sample size, diversity and uniqueness are positively correlated, as expected (see Figure S4.5; models nov_{m_2} and $nov_{m_{2.1}}$; spatial distribution of these variables in Figure S4.6). All the relevant models adjust for these effects of neighbourhood density and sampling.

Models To study the effect of dispersal and immigration on local cultural diversity and uniqueness, we built log-normal models (div_{m_1}, nov_{m_1}) and estimated the marginal effects of the proportion of immigrants, mean dispersal distance, and mean neighbourhood age, while also adjusting for individual turnover, year, and spatial dependence. Lastly, to examine whether the effects of immigration and dispersal on cultural diversity were related to individual differences in repertoire size and uniqueness, we fit two further models predicting the absolute number of unique songs in a neighbourhood while also adjusting for the number of birds (div_{m_2}) and the number of songs ($div_{m_{2.1}}$).

The rate of song cultural turnover was the response variable in two models: one ($turn_{m_1}$) trying to estimate the total effect of turnover and a second ($turn_{m_2}$) estimating the marginal effects of the proportion of immigrants, mean dispersal distance, and mean age while also adjusting for individual turnover, year, and spatial dependence. In both cases, we modelled the response distribution as a truncated log-normal with a hurdle (logistic) part to account for the zeroes.

Model estimates and reporting

We built the models and approximated the posterior distributions of the parameters of interest using brms (Bürkner, 2017), an interface to the Hamiltonian Monte Carlo engine Stan (Stan Development Team, 2023). We then processed the posterior distributions with the help of the marginal effects package. We checked model convergence via the effective number of samples, visual inspection of the chain trace plots, and the Gelman-Rubin diagnostic. Estimation in a Bayesian framework returns a posterior distribution of possible values instead of point estimates. By convention, we report posterior central estimates (means or medians) and their 95% credible intervals, but also include plots with full posteriors. Note that categorical predictors are dummy-coded and continuous predictions z-score transformed.

For each parameter of interest, we calculated predictions or marginal effects at the means or other relevant values. Regression plots show predicted values of the mean and their credible intervals, as well as partial residuals adjusted to the means or other relevant values of the explanatory terms included in the model (Fox & Weisberg, 2018; Larsen & McCleary, 1972). We have tried to build reasonable models, but even then our estimates should not be interpreted causally. Figure S4.11 shows posterior predictive checks for all models in the study. See the software section at the end for a complete list of libraries used in the various analyses and the code repository for full model specifications.

Software

Song processing and analyses were carried out using pykanto (Merino Recalde, 2023a) and rely on the following open-source scientific libraries or tools: numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), pandas (The pandas development team, 2023), numba (Lam et al., 2015), pytorch (Paszke et al., 2019), torchvision (TorchVision maintainers and contributors, 2016), pytorch lightning (Falcon & The PyTorch Lightning team, 2019), tqdm (da Costa-Luis, 2019), ray (Moritz et al., 2018), soundfile (Bechtold & Geier, 2022), umap (McInnes et al., 2018), joblib (Joblib Development Team, 2020), hdbscan (McInnes et al., 2017), seaborn (Waskom, 2021), scikit-image (van der Walt et al., 2014), librosa (McFee et al., 2015), bokeh (Bokeh Development Team, 2018), ujson (van Kemenade et al., 2023), psutil (Rodola, 2023), attrs (Schlawack, 2019). We also used R version 4.2.1 (R Core Team, 2022) and the following R packages: bayesplot (Gabry & Mahr, 2022; Gabry et al., 2019), box (Rudolph, 2023), brms (Bürkner, 2017, 2021), ClustGeo (Chavent et al., 2021), colorspace (Zeileis et al., 2020), config (Allaire, 2020), devtools (Wickham et al., 2022), DHARMa (Hartig, 2022), fs (Hester et al., 2023), ggdist (Kay, 2023), ggraph (Pedersen, 2022a), httpgd (Rupprecht, 2023), igraph (Csardi & Nepusz, 2006; Csárdi et al., 2024), insight (Lüdecke et al., 2019), kableExtra (Zhu, 2021), languageserver (Lai, 2023), marginaleffects (Arel-Bundock, 2023), modelsummary (Arel-Bundock, 2022), patchwork (Pedersen, 2022b), progress (Csárdi & FitzJohn, 2019), renv (Ushey & Wickham, 2023), rprojroot (Müller, 2022), scales (Wickham & Seidel, 2022), scico (Pedersen & Crameri, 2023), sf (Pebesma, 2018; Pebesma & Bivand, 2023), svglite (Wickham et al., 2023), terra (Hijmans, 2023), tidygraph (Pedersen, 2023), tidyverse (Wickham et al., 2019), viridis (Garnier et al., 2023). See the code repositories for specific versions of these libraries.

ACKNOWLEDGEMENTS

We thank all those who have contributed to the long-term nest box study in Wytham Woods and the collection of associated data. We also thank three anonymous reviewers for their constructive feedback. This work was supported by a Clarendon-Mary Frances

Wagley Graduate Scholarship and an EGI scholarship to Nilo Merino Recalde, and made use of the University of Oxford Advanced Research Computing facility (Richards, 2015).

AUTHOR CONTRIBUTIONS

Nilo Merino Recalde: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Andrea Estandía**: Investigation, Data Curation, Writing - Review & Editing. **Sara C. Keen**: Writing – Review & Editing. **Ella F. Cole**: Supervision, Project Administration. **Ben C. Sheldon**: Supervision, Project Administration, Writing – Review & Editing, Funding Acquisition.

REFERENCES

Allaire, J. (2020). *config: Manage environment specific configuration values* [R package version 0.3.1]. https://CRAN.R-project.org/package=config

Allen, J., Weinrich, M., Hoppitt, W., & Rendell, L. (2013). Network-Based Diffusion Analysis Reveals Cultural Transmission of Lobtail Feeding in Humpback Whales. *Science*, *340*(6131), 485–488. https://doi.org/10.1126/science.1231976

Aplin, L. M., Sheldon, B. C., & McElreath, R. (2017). Conformity does not perpetuate suboptimal traditions in a wild population of songbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(30), 7830–7837. https://doi.org/10.1073/pnas.1621067114

Arel-Bundock, V. (2022). modelsummary: Data and model summaries in R. *Journal of Statistical Software*, *103*(1), 1–23. https://doi.org/10.18637/jss.v103.i01

Arel-Bundock, V. (2023). *marginaleffects: Predictions, comparisons, slopes, marginal means, and hypothesis tests* [R package version 0.13.0]. https://CRAN.R-project.org/package=marginaleffects

Baker, M. C., McGregor, P. K., & Krebs, J. R. (1987). Sexual Response of Female Great Tits to Local and Distant Songs. *Ornis Scandinavica*, *18*(3), 186. https://doi.org/10.2307/3676765

Baker, M. C., Thompson, D. B., & Sherman, G. L. (1981). Neighbor/Stranger Song Discrimination in White-Crowned Sparrows. *The Condor*, *83*(3), 265–267. https://doi.org/10.2307/1367322

Barta, K. A., Garamszegi, L. Z., Scheuring, I., & Zsebk, S. (2023). Effects of positive frequency-dependent learning, learning mistakes, and immigration on complex culturesValidation on the song of collared flycatcher (Ficedula albicollis) by individual-based modeling. *Frontiers in Ecology and Evolution*, *11*.

Bechtold, B., & Geier, M. (2022). *Soundfile*. https://python-soundfile.readthedocs.io/en/0.11.0/

Beecher, M. D. (2008, January). Chapter 4 Function and Mechanisms of Song Learning in Song Sparrows*. In *Advances in the Study of Behavior* (pp. 167–225, Vol. 38). Academic Press. https://doi.org/10.1016/S0065-3454(08)00004-1

Berdahl, A. M., Kao, A. B., Flack, A., Westley, P. A., Codling, E. A., Couzin, I. D., Dell, A. I., & Biro, D. (2018). Collective animal navigation and migratory culture: From theoretical models to empirical evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1746), 20170009. https://doi.org/10.1098/rstb.2017.0009

Bircher, N., van Oers, K., & Naguib, M. (2021). Anyone listening? No evidence for eavesdropping on male singing interactions in the great tit, Parus major. *Animal Behaviour*, *176*, 67–76. https://doi.org/10.1016/j.anbehav.2021.03.009

Blumenrath, S. H., & Dabelsteen, T. (2004). Degradation of Great Tit (Parus major) Song before and after Foliation: Implications for Vocal Communication in a Deciduous Forest. *Behaviour*, *141*(8), 935–958.

Bokeh Development Team. (2018). *Bokeh: Python library for interactive visualization*. Manual.

Brakes, P., Carroll, E. L., Dall, S. R. X., Keith, S. A., McGregor, P. K., Mesnick, S. L., Noad, M. J., Rendell, L., Robbins, M. M., Rutz, C., Thornton, A., Whiten, A., Whiting, M. J., Aplin, L. M., Bearhop, S., Ciucci, P., Fishlock, V., Ford, J. K. B., Notarbartolo di Sciara, G., … Garland, E. C. (2021). A deepening understanding of animal culture suggests lessons for conservation. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1949), 20202718. https://doi.org/10.1098/rspb.2020.2718

Brakes, P., Dall, S. R., Aplin, L. M., Bearhop, S., Carroll, E. L., Ciucci, P., Fishlock, V., Ford, J. K., Garland, E. C., Keith, S. A., McGregor, P. K., Mesnick, S. L., Noad, M. J., Notarbartolo di Sciara, G., Robbins, M. M., Simmonds, M. P., Spina, F., Thornton, A., Wade, P. R., … Rutz, C. (2019). Animal cultures matter for conservation. *Science*, *363*(6431), 1032–1034. https://doi.org/10.1126/science.aaw3557

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2021). Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software*, *100*(5), 1–54. https://doi.org/10.18637/jss.v100.i05

Byholm, P., Beal, M., Isaksson, N., Lötberg, U., & Åkesson, S. (2022). Paternal transmission of migration knowledge in a long-distance bird migrant. *Nat Commun*, *13*, 1566. https://doi.org/10.1038/s41467-022-29300-w

Chambers, M., & Drovandi, C. (2023, May). Many-levelled continuation ratio models for frequency of alcohol and drug use data.

Chavent, M., Kuentz, V., Labenne, A., & Saracco, J. (2021). *ClustGeo: Hierarchical clustering with spatial constraints* [R package version 2.1]. https://CRAN.R-project.org/package=ClustGeo

Chimento, M., Alarcón-Nieto, G., & Aplin, L. M. (2021). Population turnover facilitates cultural selection for efficiency in birds. *Current Biology*, *31*(11), 2477–2483.e3. https://doi.org/10.1016/j.cub.2021.03.057

Chimento, M., & Aplin, L. M. (2024). Understanding the Role of Naive Learners in Cultural Change. *The American Naturalist*. https://doi.org/10.1086/730110

Claidière, N., & Sperber, D. (2007). Commentary: The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1-2), 89–111. https://doi.org/10.1163/156853707X171829

Claramunt, S., Derryberry, E. P., Remsen, J. V., & Brumfield, R. T. (2011). High dispersal ability inhibits speciation in a continental radiation of passerine birds. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1733), 1567–1574. https://doi.org/10.1098/rspb.2011.1922

Crates, R., Langmore, N., Ranjard, L., Stojanovic, D., Rayner, L., Ingwersen, D., & Heinsohn, R. (2021). Loss of vocal culture and fitness costs in a critically endangered songbird. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1947), 20210225. https://doi.org/10.1098/rspb.2021.0225

Cresswell, W., & Mccleery, R. (2003). How great tits maintain synchronization of their hatch date with food supply in response to long-term variability in temperature. *Journal of Animal Ecology*, *72*(2), 356–366. https://doi.org/10.1046/j.1365-2656.2003.00701.x

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. https://igraph.org

Csárdi, G., & FitzJohn, R. (2019). *progress: Terminal progress bars* [R package version 1.2.2]. https://CRAN.R-project.org/package=progress

Csárdi, G., Nepusz, T., Traag, V., Horvát, S., Zanini, F., Noom, D., & Müller, K. (2024). *igraph: Network analysis and visualization in r* [R package version 1.5.0]. https://doi.org/10.5281/zenodo.7682609

da Costa-Luis, C. O. (2019). Tqdm: A fast, extensible progress meter for python and CLI. *Journal of Open Source Software*, *4*(37), 1277. https://doi.org/10.21105/joss.01277

Dearmon, J., & Smith, T. E. (2016). Gaussian Process Regression and Bayesian Model Averaging: An Alternative Approach to Modeling Spatial Phenomena: Gaussian Process Regression and BMA. *Geogr Anal*, *48*(1), 82–111. https://doi.org/10.1111/gean.12083

Deffner, D., Kandler, A., & Fogarty, L. (2022). Effective population size for culturally evolving traits. *PLOS Computational Biology*, *18*(4), e1009430. https://doi.org/10.1371/journal.pcbi.1009430

Deffner, D., & McElreath, R. (2020). The importance of life history and population regulation for the evolution of social learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1803), 20190492. https://doi.org/10.1098/rstb.2019.0492

Deffner, D., & McElreath, R. (2022). When does selection favor learning from the old? Social learning in age-structured populations. *PLOS ONE*, *17*(4), e0267204. https://doi.org/10.1371/journal.pone.0267204

Department for Environment Food and Rural Affairs. (2020). Defra Data Services Platform.

Derex, M., & Boyd, R. (2016). Partial connectivity increases cultural accumulation within groups. *Proceedings of the National Academy of Sciences*, *113*(11), 2982–2987. https://doi.org/10.1073/pnas.1518798113

Dhondt, A. A. (1979). Summer dispersal and survival of juvenile Great Tits in southern Sweden. *Oecologia*, *42*(2), 139–157. https://doi.org/10.1007/BF00344854

Dingemanse, N. J., Both, C., van Noordwijk, A. J., Rutten, A. L., & Drent, P. J. (2003). Natal dispersal and personalities in great tits (Parus major). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1516), 741–747. https://doi.org/10.1098/rspb.2002.2300

Dyble, M., & Micheletti, A. J. C. (2024). Population turnover, behavioral conservatism, and rates of cultural evolution. *Behavioral Ecology*, *35*(2), arae003. https://doi.org/10.1093/beheco/arae003

Falcon, W., & The PyTorch Lightning team. (2019, March). PyTorch lightning. https://doi.org/10.5281/zenodo.3828935

Falls, J. B., Krebs, J. R., & McGregor, P. K. (1982). Song matching in the great tit (Parus major): The effect of similarity and familiarity. *Animal Behaviour*, *30*(4), 997–1009. https://doi.org/10.1016/S0003-3472(82)80188-7

Falótico, T., Proffitt, T., Ottoni, E. B., Staff, R. A., & Haslam, M. (2019). Three thousand years of wild capuchin stone tool use. *Nature Ecology and Evolution*. https://doi.org/10.1038/s41559-019-0904-4

Fayet, A. L., Tobias, J. A., Hintzen, R. E., & Seddon, N. (2014). Immigration and dispersal are key determinants of cultural diversity in a songbird population. *Behavioral Ecology*, *25*(4), 744–753. https://doi.org/10.1093/beheco/aru047

Fogarty, L., Creanza, N., & Feldman, M. W. (2019). The life history of learning: Demographic structure changes cultural outcomes. *PLOS Computational Biology*, *15*(4), e1006821. https://doi.org/10.1371/journal.pcbi.1006821

Fox, J., & Weisberg, S. (2018). Visualizing Fit and Lack of Fit in Complex Regression Models with Predictor Effect Plots and Partial Residuals. *Journal of Statistical Software*, *87*, 1–27. https://doi.org/10.18637/jss.v087.i09

Gabry, J., & Mahr, T. (2022). bayesplot: Plotting for bayesian models [R package version 1.10.0]. https://mc-stan.org/bayesplot/

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., & Gelman, A. (2019). Visualization in bayesian workflow. *J. R. Stat. Soc. A*, *182*, 389–402. https://doi.org/10.1111/rssa.12378

Garnier, Simon, Ross, Noam, Rudis, Robert, Camargo, Pedro, A., Sciaini, Marco, Scherer, & Cédric. (2023). *viridis(Lite) - colorblind-friendly color maps for r* [viridis package version 0.6.3]. https://doi.org/10.5281/zenodo.4679424

Gelfand, A. E., & Schliep, E. M. (2016). Spatial statistics and Gaussian processes: A beautiful marriage. *Spatial Statistics*, *18*, 86–104. https://doi.org/10.1016/j.spasta.2016.03.006

Gibb, J. (1950). The Breeding Biology of the Great and Blue Titmice. *Ibis*, *92*(4), 507–539. https://doi.org/10.1111/j.1474-919X.1950.tb01759.x

Graham, B. A., Heath, D. D., Walter, R. P., & Mennill, D. J. (2018). Immigrant song: Males and females learn songs after dispersal in a tropical bird. *Behavioral Ecology*, *29*(3), 711–723. https://doi.org/10.1093/beheco/ary029

Greenwood, P. J., Harvey, P. H., & Perrins, C. M. (1979). The Role of Dispersal in the Great Tit (Parus major): The Causes, Consequences and Heritability of Natal Dispersal. *Journal of Animal Ecology*, *48*(1), 123–142. https://doi.org/10.2307/4105

Halfwerk, W., Bot, S., & Slabbekoorn, H. (2012). Male great tit song perch selection in response to noise-dependent female feedback. *Functional Ecology*, *26*(6), 1339–1347. https://doi.org/10.1111/j.1365-2435.2012.02018.x

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hartig, F. (2022). *DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models* [R package version 0.4.6]. https://CRAN.R-project.org/package=DHARMa

Hermans, A., Beyer, L., & Leibe, B. (2017, November). In Defense of the Triplet Loss for Person Re-Identification. https://doi.org/10.48550/arXiv.1703.07737

Hester, J., Wickham, H., & Csárdi, G. (2023). *fs: Cross-platform file system operations based on 'libuv'* [R package version 1.6.2]. https://CRAN.R-project.org/package=fs

Hijmans, R. J. (2023). *terra: Spatial data analysis* [R package version 1.7-39]. https://CRAN.R-project.org/package=terra

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, *6*, e00073. https://doi.org/10.1016/j.ohx.2019.e00073

Hinde, R. A. (1952). The Behaviour of the Great Tit (Parus Major) and Some Other Related Species. *Behaviour. Supplement*, (2), III–201.

Hoffer, E., & Ailon, N. (2018, December). Deep metric learning using Triplet network. https://doi.org/10.48550/arXiv.1412.6622

Hutfluss, A., Bermúdez-Cuamatzin, E., Mouchet, A., Briffa, M., Slabbekoorn, H., & Dingemanse, N. J. (2022). Male song stability shows cross-year repeatability but does not affect reproductive success in a wild passerine bird. *Journal of Animal Ecology*, *91*(7), 1507–1520. https://doi.org/10.1111/1365-2656.13736

James, L. S., & Sakata, J. T. (2017). Learning Biases Underlie "Universals" in Avian Vocal Sequencing. *Current Biology*, *27*(23), 3676–3682. https://doi.org/10.1016/j.cub.2017.10.019

Jesmer, B. R., Merkle, J. A., Goheen, J. R., Aikens, E. O., Beck, J. L., Courtemanch, A. B., Hurley, M. A., McWhirter, D. E., Miyasaki, H. M., Monteith, K. L., & Kauffman, M. J. (2018). Is ungulate migration culturally transmitted? Evidence of social learning from translocated animals. *Science*, *361*(6406), 1023–1025. https://doi.org/10.1126/science.aat0985

Joblib Development Team. (2020). Joblib: Running Python functions as pipeline jobs.

Kandler, A., Fogarty, L., & Karsdorp, F. (2023). The interplay between age structure and cultural transmission. *PLOS Computational Biology*, *19*(7), e1011297. https://doi.org/10.1371/journal.pcbi.1011297

Kandler, A., Wilder, B., & Fortunato, L. (2017). Inferring individual-level processes from population-level patterns in cultural evolution. *Royal Society Open Science*, *4*(9), 170949. https://doi.org/10.1098/rsos.170949

Kay, M. (2023). *ggdist: Visualizations of distributions and uncertainty* [R package version 3.3.0]. https://doi.org/10.5281/zenodo.3879620

Keen, S. C. (2020). The role of social environment in shaping vocal communication systems in wild songbirds. *Cornell University, PhD Thesis*. https://doi.org/10.7298/qz29-2v92

Kendal, R., Hopper, L. M., Whiten, A., Brosnan, S. F., Lambeth, S. P., Schapiro, S. J., & Hoppitt, W. (2015). Chimpanzees copy dominant and knowledgeable individuals: Implications for cultural diversity. *Evolution and Human Behavior*, *36*(1), 65–72. https://doi.org/10.1016/j.evolhumbehav.2014.09.002

Kirby, S., & Tamariz, M. (2021). Cumulative cultural evolution, population structure and the origin of combinatoriality in human language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1843), 20200319. https://doi.org/10.1098/rstb.2020.0319

Lachlan, R. F., Anderson, R. C., Peters, S., Searcy, W. A., & 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*(1785). https://doi.org/10.1098/rspb.2014.0252

Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, *9*(1), 2417. https://doi.org/10.1038/s41467-018-04728-1

Lack, D. (1964). A Long-Term Study of the Great Tit (Parus major). *Journal of Animal Ecology*, *33*, 159–173. https://doi.org/10.2307/2437

Lai, R. (2023). *languageserver: Language server protocol* [R package version 0.3.15]. https://CRAN.R-project.org/package=languageserver

Laland, K. N., & Janik, V. M. (2006). The animal cultures debate. *Trends in Ecology & Evolution*, *21*(10), 542–547. https://doi.org/10.1016/j.tree.2006.06.005

Lam, S. K., Pitrou, A., & Seibert, S. (2015). Numba: A LLVM-based Python JIT compiler. *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*, 1–6. https://doi.org/10.1145/2833157.2833162

Larsen, W. A., & McCleary, S. J. (1972). The Use of Partial Residual Plots in Regression Analysis. *Technometrics*, *14*(3), 781–790. https://doi.org/10.1080/00401706.1972.10488966

Lind, H., Dabelsteen, T., & McGregor, P. K. (1996). Female great tits can identify mates by song. *Animal Behaviour*, *52*(4), 667–671. https://doi.org/10.1006/anbe.1996.0211

Lüdecke, D., Waggoner, P., & Makowski, D. (2019). insight: A unified interface to access information from model objects in R. *Journal of Open Source Software*, *4*(38), 1412. https://doi.org/10.21105/joss.01412

Mace, R. (1987). The dawn chorus in the great tit Parus major is directly related to female fertility. *Nature*, *330*(6150), 745–746. https://doi.org/10.1038/330745a0

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2015). Librosa: Audio and Music Signal Analysis in Python. *Python in Science Conference*, 18–24. https://doi.org/10.25080/Majora-7b98e3ed-003

McGregor, P. K., & Avery, M. I. (1986). The Unsung Songs of Great Tits (Parus major): Learning Neighbours' Songs for Discrimination. *Behavioral Ecology and Sociobiology*, *18*(4), 311–316.

McGregor, P. K., & Krebs, J. R. (1982a). Mating and song sharing in the great tit. *Nature*, *297*(5861), 60–61. https://doi.org/10.1038/297060a0

McGregor, P. K., & Krebs, J. R. (1982b). Song Types in a Population of Great Tits (Parus major): Their Distribution, Abundance and Acquisition by Individuals. *Behaviour*, *79*(2/4), 126–152.

McGregor, P. K., Krebs, J. R., & Perrins, C. M. (1981). Song Repertoires and Lifetime Reproductive Success in the Great Tit (Parus major). *The American Naturalist*, *118*(2), 149–159. https://doi.org/918

McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *The Journal of Open Source Software*, *3*(29), 861.

Merino Recalde, N. (2023a). Pykanto: A python library to accelerate research on wild bird song. *Methods in Ecology and Evolution*, *00*, 1–9. https://doi.org/10.1111/2041-210X.14155

Merino Recalde, N. (2023b, June). Nilomr/great-tit-hits-setup: V0.1.0. https://doi.org/10.5281/zenodo.8092397

Merino Recalde, N., Estandía, A., Pichot, L., Vansse, A., Cole, E. F., & Sheldon, B. C. (2023). A densely sampled and richly annotated acoustic dataset from a wild bird population. *bioRxiv*. https://doi.org/10.1101/2023.07.03.547484

Morfi, V., Lachlan, R. F., & Stowell, D. (2021). Deep perceptual embeddings for unlabelled animal sound events. *The Journal of the Acoustical Society of America*, *150*(1), 2–11. https://doi.org/10.1121/10.0005475

Morgan-Wall, T. (2023). *Rayshader: Create maps and visualize data in 2d and 3d* [https://www.rayshader.com, https://github.com/tylermorganwall/rayshader].

Moritz, P., Nishihara, R., Wang, S., Tumanov, A., Liaw, R., Liang, E., Elibol, M., Yang, Z., Paul, W., Jordan, M. I., & Stoica, I. (2018). Ray: A distributed framework for emerging AI applications. *Proceedings of the 13th USENIX Conference on Operating Systems Design and Implementation*, 561–577.

Mortega, K. G., Flinks, H., & Helm, B. (2014). Behavioural response of a migratory songbird to geographic variation in song and morphology. *Frontiers in Zoology*, *11*(1), 85. https://doi.org/10.1186/s12983-014-0085-6

Müller, K. (2022). *rprojroot: Finding files in project subdirectories* [R package version 2.0.3]. https://CRAN.R-project.org/package=rprojroot

Mumuni, A., & Mumuni, F. (2022). Data augmentation: A comprehensive survey of modern approaches. *Array*, *16*, 100258. https://doi.org/10.1016/j.array.2022.100258

Nunn, C. L., Thrall, P. H., Bartz, K., Dasgupta, T., & Boesch, C. (2009). Do transmission mechanisms or social systems drive cultural dynamics in socially structured populations? *Animal Behaviour*, *77*(6), 1515–1524. https://doi.org/10.1016/j.anbehav.2009.02.023

Papadopoulou, A., Anastasiou, I., Keskin, B., & Vogler, A. P. (2009). Comparative phylogeography of tenebrionid beetles in the Aegean archipelago: The effect of dispersal ability and habitat preference. *Molecular Ecology*, *18*(11), 2503–2517. https://doi.org/10.1111/j.1365-294X.2009.04207.x

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. dAlché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc.

Paxton, K. L., Sebastián-González, E., Hite, J. M., Crampton, L. H., Kuhn, D., & Hart, P. J. (2019). Loss of cultural song diversity and the convergence of songs in a declining Hawaiian forest bird community. *Royal Society Open Science*, *6*(8). https://doi.org/10.1098/rsos.190719

Payne, R. B. (1983). The social context of song mimicry: Song-matching dialects in indigo buntings (Passerina cyanea). *Animal Behaviour*, *31*(3), 788–805. https://doi.org/10.1016/S0003-3472(83)80236-X

Payne, R. B., & Payne, L. L. (1993). Song copying and cultural transmission in indigo buntings. *Animal Behaviour*, *46*(6), 1045–1065. https://doi.org/10.1006/anbe.1993.1296

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, *10*(1), 439–446. https://doi.org/10.32614/RJ-2018-009

Pebesma, E., & Bivand, R. (2023). *Spatial Data Science: With applications in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429459016

Pedersen, T. L. (2022a). *ggraph: An implementation of grammar of graphics for graphs and networks* [R package version 2.1.0]. https://CRAN.R-project.org/package=ggraph

Pedersen, T. L. (2022b). *patchwork: The composer of plots* [R package version 1.1.2]. https://CRAN.R-project.org/package=patchwork

Pedersen, T. L. (2023). *tidygraph: A tidy api for graph manipulation* [R package version 1.2.3]. https://CRAN.R-project.org/package=tidygraph

Pedersen, T. L., & Crameri, F. (2023). *scico: Colour palettes based on the scientific colour-maps* [R package version 1.5.0]. https://CRAN.R-project.org/package=scico

Perez, L., & Wang, J. (2017, December). The Effectiveness of Data Augmentation in Image Classification using Deep Learning. https://doi.org/10.48550/arXiv.1712.04621

Perrins, C. M. (1965). Population Fluctuations and Clutch-Size in the Great Tit, Parus major L. *Journal of Animal Ecology*, *34*(3), 601–647.

Pike, T. W., & Laland, K. N. (2010). Conformist learning in nine-spined sticklebacks' foraging decisions. *Biology Letters*, *6*(4), 466–468. https://doi.org/10.1098/rsbl.2009.1014

Planqué, R., Britton, N. F., & Slabbekoorn, H. (2014). On the maintenance of bird song dialects. *Journal of Mathematical Biology*, *68*(1-2), 505–531. https://doi.org/10.1007/s00285-012-0632-8

Potvin, D. A., & Clegg, S. M. (2015). The relative roles of cultural drift and acoustic adaptation in shaping syllable repertoires of island bird populations change with time since colonization. *Evolution*, *69*(2), 368–380. https://doi.org/10.1111/evo.12573

Quinn, J. L., Cole, E. F., Patrick, S. C., & Sheldon, B. C. (2011). Scale and state dependence of the relationship between personality and dispersal in a great tit population. *Journal of Animal Ecology*, *80*(5), 918–928.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rendell, L., & Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and Brain Sciences*, *24*(2), 309–324. https://doi.org/10.1017/S0140525X0100396X

Richards, A. (2015, August). *University of oxford advanced research computing*. Manual. https://doi.org/10.5281/zenodo.22558

Rivera-Gutierrez, H. F., Matthysen, E., Adriaensen, F., & Slabbekoorn, H. (2010). Repertoire sharing and song similarity between great tit males decline with distance between forest fragments. *Ethology*, *116*(10), 951–960. https://doi.org/10.1111/j.1439-0310.2010.01813.x

Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2011). Difficulties when Assessing Birdsong Learning Programmes under Field Conditions: A Re-Evaluation of Song Repertoire Flexibility in the Great Tit (M. Giurfa, Ed.). *PLoS ONE*, *6*(1), e16003. https://doi.org/10.1371/journal.pone.0016003

Rivera-Gutierrez, H. F., Pinxten, R., & Eens, M. (2012). Tuning and fading voices in songbirds: Age-dependent changes in two acoustic traits across the life span. *Animal Behaviour*, *83*(5), 1279–1283. https://doi.org/10.1016/j.anbehav.2012.03.001

Rodola, G. (2023). *Psutil (python system and process utilities)*. https://psutil.readthedocs.io/en/latest/

Rudolph, K. (2023). *box: Write reusable, composable and modular r code* [R package version 1.1.3]. https://CRAN.R-project.org/package=box

Rupprecht, F. (2023). *httpgd: A 'HTTP' server graphics device* [R package version 1.3.1]. https://CRAN.R-project.org/package=httpgd

Schlawack, H. (2019). *Attrs*. https://www.attrs.org/en/stable/

Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2021). Variation in vocal production learning across songbirds. *Phil. Trans. R. Soc. B*, *376*(1836), 20200257. https://doi.org/10.1098/rstb.2020.0257

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1), 60. https://doi.org/10.1186/s40537-019-0197-0

Stan Development Team. (2023). Stan modeling language users guide and reference manual, version 2.18.0. http://mc-stan.org/

Suárez, D., Arribas, P., Jiménez-García, E., & Emerson, B. C. (2022). Dispersal ability and its consequences for population genetic differentiation and diversification. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1975), 20220489. https://doi.org/10.1098/rspb.2022.0489

Tchernichovski, O., Eisenberg-Edidin, S., & Jarvis, E. D. (2021). Balanced imitation sustains song culture in zebra finches. *Nat Commun*, *12*(1), 2562. https://doi.org/10.1038/s41467-021-22852-3

Tchernichovski, O., Feher, O., Fimiarz, D., & Conley, D. (2017). How social learning adds up to a culture: From birdsong to human public opinion. *Journal of Experimental Biology*, *220*(1), 124–132. https://doi.org/10.1242/jeb.142786

The pandas development team. (2023, January). Pandas-dev/pandas: Pandas. https://doi.org/10.5281/zenodo.7549438

TorchVision maintainers and contributors. (2016). TorchVision: PyTorch's computer vision library.

Ushey, K., & Wickham, H. (2023). *renv: Project environments* [R package version 1.0.0]. https://CRAN.R-project.org/package=renv

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T., & scikit-image contributors, t. (2014). Scikit-image: Image processing in Python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

Van Duyse, E., Pinxten, R., Snoeijs, T., & Eens, M. (2005). Simultaneous treatment with an aromatase inhibitor and an anti-androgen decreases the likelihood of dawn song in free-living male great tits, Parus major. *Hormones and Behavior*, *48*(2), 243–251. https://doi.org/10.1016/j.yhbeh.2005.02.013

van Kemenade, H., Woodsend, B., Hamrén, J., JustAnotherArchivist, Crall, J., O'Mahony, K., Lay, E. L.,), H. M. (, Sychev, M., Swenson, D. W., Dawborn, T., Garcia-Armas, M., Ceccon, R., Moiron, J., Górny, M., NaN-git, Naoki, I., Beasley, B., Zhou, L., … CozyDoomer. (2023, January). Ultrajson. https://doi.org/10.5281/zenodo.7510698

van Rossum, G. (1995, January). Python reference manual.

Verhulst, S., Perrins, C. M., & Riddington, R. (1997). Natal Dispersal of Great Tits in a Patchy Environment. *Ecology*, *78*(3), 864–872. https://doi.org/10.1890/0012-9658(1997)078[0864:NDOGTI]2.0.CO;2

Viciana, H. (2021). Animal culture: But of which kind? *Studies in History and Philosophy of Science Part A*, *90*, 208–218. https://doi.org/10.1016/j.shpsa.2021.10.012

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Warti, R., Kurnia, A., & Sadik, K. (2020). Evaluation of Proportional Odds and Continuation Ratio Models for Smoker in Indonesia. *Proceedings of the 1st International Conference on Statistics and Analytics, ICSA 2019, 2-3 August 2019, Bogor, Indonesia*.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Whitehead, H., & Lusseau, D. (2012). Animal social networks as substrate for cultural behavioural diversity. *Journal of Theoretical Biology*, *294*, 19–28. https://doi.org/10.1016/j.jtbi.2011.10.025

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse.

Journal of Open Source Software, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., Henry, L., Pedersen, T. L., Luciani, T. J., Decorde, M., & Lise, V. (2023). *svglite: An 'SVG' graphics device* [R package version 2.1.1]. https://CRAN.R-project.org/package=svglite

Wickham, H., Hester, J., Chang, W., & Bryan, J. (2022). *devtools: Tools to make developing r packages easier* [R package version 2.4.5]. https://CRAN.R-project.org/package=devtools

Wickham, H., & Seidel, D. (2022). *scales: Scale functions for visualization* [R package version 1.2.1]. https://CRAN.R-project.org/package=scales

Williams, H., & Lachlan, R. F. (2021). Evidence for cumulative cultural evolution in bird song. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1843), 20200322. https://doi.org/10.1098/rstb.2020.0322

Woodman, J. P., Cole, E. F., Firth, J. A., Perrins, C. M., & Sheldon, B. C. (2023). Disentangling the causes of age-assortative mating in bird populations with contrasting life-history strategies. *Journal of Animal Ecology*, *92*(5), 979–990. https://doi.org/10.1111/1365-2656.13851

Wright, W. J., Irvine, K. M., Rodhouse, T. J., & Litt, A. R. (2021). Spatial Gaussian processes improve multi-species occupancy models when range boundaries are uncertain and nonoverlapping. *Ecology and Evolution*, *11*(13), 8516–8527. https://doi.org/10.1002/ece3.7629

Zandberg, L., Morfi, V., George, J., Clayton, D. F., Stowell, D., & Lachlan, R. F. (2022, December). Bird song comparison using deep learning trained from avian perceptual judgments. https://doi.org/10.1101/2022.12.23.521425

Zeileis, A., Fisher, J. C., Hornik, K., Ihaka, R., McWhite, C. D., Murrell, P., Stauffer, R., & Wilke, C. O. (2020). colorspace: A toolbox for manipulating and assessing colors and palettes. *Journal of Statistical Software*, *96*(1), 1–49. https://doi.org/10.18637/jss.v096.i01

Zhu, H. (2021). *kableExtra: Construct complex table with 'kable' and pipe syntax* [R package version 1.3.4]. https://CRAN.R-project.org/package=kableExtra

SUPPLEMENTARY INFORMATION

Table S4.1. Model information

Model	Formula	Family	N	Groups
Individual Repertoires				
rep_m_1	repertoire_size ∼ 1 + immigrant + s(sampling_effort) + year + (1 | father)	cratio	301	242
rep_m_1.1	repertoire_size ∼ 1 + dispersal_distance + s(sampling_effort) + year + (1 | father)	cratio	133	105
rep_m_1.2	repertoire_size ∼ 1 + age + s(sampling_effort) + year + (1 | father)	cratio	256	205
repnov_m_1	average_frequency ∼ 1 + immigrant + s(sampling_effort) + year + (1 | father)	lognormal	300	242
repnov_m_1.1	average_frequency ∼ 1 + dispersal_distance + s(sampling_effort) + year + (1 | father)	lognormal	133	105
repnov_m_1.2	average_frequency ∼ 1 + age + s(sampling_effort) + year + (1 | father)	lognormal	256	205
Cultural Similarity				
disp_m_1	mean_dist1 ∼ 0 + natal_distance + nest_distance + year_born_diff + year + (1 | mm(father, father2))	Gaussian	8745	105
imm_m_1	mean_dist1 ∼ 0 + resident_status + year_born_diff + nest_distance + year + (1 | mm(father, father2))	Gaussian	11029	205
age_m_1	mean_dist1 ∼ 0 + natal_distance + nest_distance + year_born_diff + year + (1 | mm(father, father2))	Gaussian	8745	105
Cultural Novelty and Diversity				
nov_m_1	uniqueness ∼ 0 + prop_immigrant + mean_dispersal_distance + prop_same_birds + mean_age + s(n_current_songs) + year + gp(x, y, by = year)	lognormal	791	GP
div_m_1	diversity ∼ 0 + prop_immigrant + mean_dispersal_distance + prop_same_birds + mean_age + s(n_current_songs) + year + gp(x, y, by = year)	lognormal	791	GP
nov_m_2	uniqueness ∼ 0 + diversity + s(n_current_songs) + year + gp(x, y, by = year)	lognormal	791	GP
nov_m_2.1	uniqueness ∼ 0 + diversity + year + gp(x, y, by = year)	lognormal	791	GP
div_m_2	n_unique_current_songs ∼ 0 + prop_immigrant + mean_dispersal_distance + prop_same_birds + mean_age + s(recorded) + year + gp(x, y, by = year)	Gaussian	791	GP
div_m_2.1	n_unique_current_songs ∼ 0 + prop_immigrant + mean_dispersal_distance + prop_same_birds + mean_age + s(n_current_songs) + year + gp(x, y, by = year)	Gaussian	791	GP
Cultural Turnover				
turn_m_1	prop_shared ∼ 0 + prop_same_birds + year + gp(x, y, by = year, k = 25, c = 5/4)	hurdle lognormal	544	GP
turn_m_2	prop_shared ∼ 0 + prop_immigrant + mean_dispersal_distance + mean_age + prop_same_birds + year + gp(x, y, by = year, k = 25, c = 5/4)	hurdle lognormal	544	GP

Table S4.2. Model variable key

Variable	Description
repertoire_size	Number of distinct song types sung by an individual
immigrant	Immigrant status (hatched in the population / not hatched in the population)
sampling_effort	Total number of recordings obtained for this individual
year	Year in which the repertoire was recorded
father	ID of the male; BTO (British Trust for Ornithology) ring number
dispersal_distance	Postnatal dispersal distance, in metres, from the natal nest box to the breeding nest box
age	The age of the bird. This is exact if the bird was ringed as a pullus in the population, or approximate (based on plumage moult) otherwise
average_frequency	The mean of the frequency of each song in a bird's repertoire in a given year
mean_dist1	Average minimum Euclidean distance between the repertoires of two birds
natal_distance	Distance in metres at which two birds were born
nest_distance	Distance in metres at which two birds bred
year_born_diff	The absolute difference between the birth years of two birds
mm(...)	Multi-membership grouping term for the similarity network
resident_status	Origin of a pair of birds: both, one or neither were hatched in the population
prop_immigrant	Proportion of the birds in a neighbourhood that were not hatched inside the population
mean_dispersal_distance	Mean postnatal dispersal distance, in metres, of the birds in the neighbourhood
prop_same_birds	Proportion of the birds in a neighbourhood that were already there the year before
mean_age	Mean age of the birds in a neighbourhood
n_current_songs	Absolute number of song types in a neighbourhood (not unique song types), which correlates with neighbourhood size but we use to further adjust for the fact that some birds sing more songs
gp(x, y)	2D Gaussian process to model spatial dependency in the data
n_unique_current_songs	Number of unique song types in a neighbourhood
uniqueness	The uniqueness of a bird's repertoire, quantified as 1 minus the logarithm of the mean frequency of the songs in its repertoire for a given year
recorded	Number of birds recorded singing in a neighbourhood, which is linearly correlated with neighbourhood size
prop_shared	Proportion of the song types in a neighbourhood that were already present. Reported as 1-prop_shared: 'song turnover'

Table S4.3. Model estimates

Model	Hypothesis	Estimate[a]	Evid. Ratio	Post. Prob
		Individual Repertoires		
rep_m_1	immigrant > 0	0.239 [-0.098, 0.593]	6.963	0.874
rep_m_1.1	dispersal distance < 0	-0.201 [-0.443, 0.045]	10.111	0.910
rep_m_1.2	age > 0	0.064 [-0.108, 0.241]	2.701	0.730
repnov_m_1	non-immigrant < 0	-0.049 [-0.2, 0.1]	2.401	0.706
repnov_m_1.1	dispersal distance > 0	0.203 [0.088, 0.316]	741.857	0.999
repnov_m_1.2	age > 0	-0.017 [-0.093, 0.058]	0.540	0.351

Continued on next page

Table S4.3. Model estimates (Continued)

Model	Hypothesis	Estimate[a]	Evid. Ratio	Post. Prob
Cultural Similarity				
disp_m_1	natal distance > 0	0.001 [0, 0.002]	22.529	0.958
disp_m_1	nest distance > 0	-0.005 [-0.006, -0.004]	Inf	1
imm_m_1	both resident-both immigrant < 0	0.002 [-0.005, 0.01]	0.438	0.304
imm_m_1	both resident-one immigrant < 0	0.002 [-0.002, 0.006]	0.331	0.248
age_m_1	age difference 0-1 > 0	0.002 [0, 0.003]	12.289	0.925
age_m_1	age difference 0-2 > 0	0.004 [0.002, 0.006]	Inf	1
age_m_1	age difference 0-3 > 0	0.006 [0.003, 0.008]	1999	1
age_m_1	age difference 0-4+ > 0	0.01 [0.006, 0.013]	Inf	1
Cultural Novelty and Diversity				
nov_m_1	mean dispersal distance < 0	-0.018 [-0.023, -0.012]	Inf	1
nov_m_1	proportion immigrant < 0	0.001 [-0.005, 0.006]	0.752	0.429
nov_m_1	mean age < 0	0.012 [0.005, 0.019]	399	0.998
nov_m_1	individual turnover < 0	0.001 [-0.005, 0.008]	1.733	0.634
div_m_1	mean dispersal distance < 0	-0.005 [-0.01, 0]	26.65	0.964
div_m_1	proportion immigrant < 0	0.002 [-0.004, 0.007]	0.442	0.306
div_m_1	mean age < 0	0.021 [0.014, 0.027]	Inf	1
div_m_1	individual turnover < 0	-0.002 [-0.008, 0.005]	0.446	0.309
nov_m_2	diversity > 0	0.713 [0.629, 0.795]	Inf	1
nov_m_2.1	diversity > 0	-0.099 [-0.216, 0.018]	0.086	0.080
div_m_2	mean dispersal distance < 0	-0.658 [-0.999, -0.315]	570.429	0.998
div_m_2	proportion immigrant < 0	0.469 [0.1, 0.837]	62.492	0.984
div_m_2	mean age < 0	1.045 [0.576, 1.495]	Inf	1
div_m_2	individual turnover < 0	0.204 [-0.291, 0.683]	3.077	0.755
div_m_2.1	mean dispersal distance < 0	0.019 [-0.213, 0.249]	0.824	0.452
div_m_2.1	proportion immigrant < 0	0.072 [-0.168, 0.312]	2.259	0.693
div_m_2.1	mean age < 0	0.928 [0.599, 1.245]	Inf	1
div_m_2.1	individual turnover < 0	-0.031 [-0.349, 0.279]	0.726	0.421
Cultural Turnover				
turn_m_1	individual turnover > 0	0.072 [0.054, 0.09]	Inf	1
turn_m_2	mean dispersal distance < 0	-0.022 [-0.039, -0.006]	79	0.988
turn_m_2	proportion immigrant < 0	-0.051 [-0.065, -0.037]	Inf	1
turn_m_2	mean age < 0	-0.044 [-0.063, -0.026]	3999	1
turn_m_2	individual turnover < 0	0.047 [0.028, 0.066]	Inf	1

[a] Estimates are Medians and 95% Credible Intervals

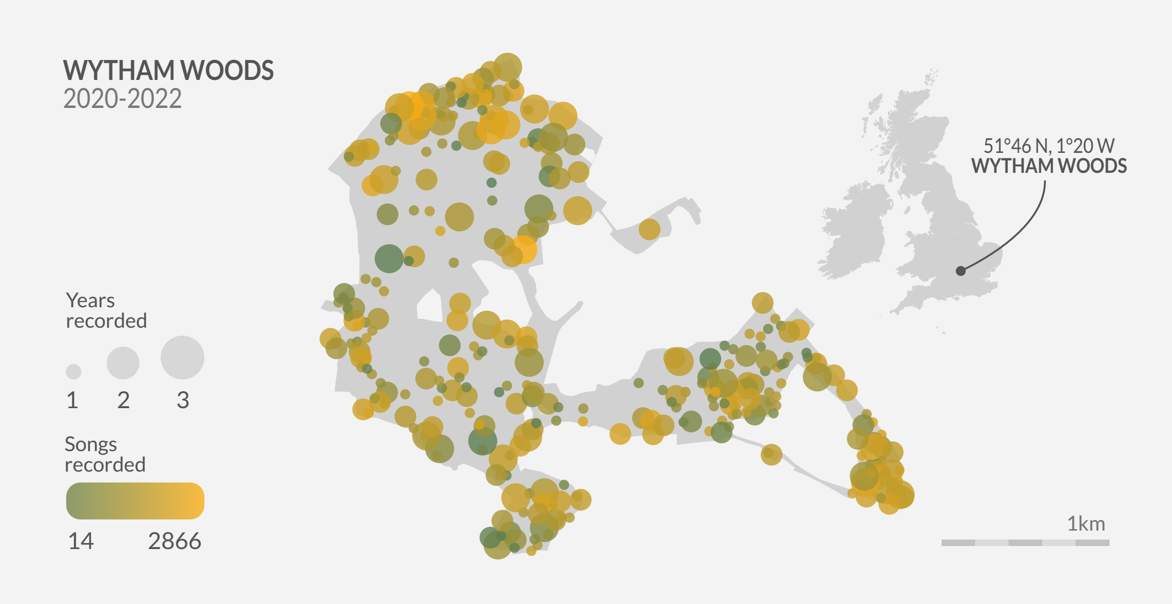


Figure S4.1. Map of the study site and sampling locations.
This study was conducted in Wytham Woods, a 385-hectare semi-natural woodland surrounded by farmland. Data was collected during the breeding seasons of 2020, 2021, and 2022 by regularly checking 1018 nest boxes, documenting information such as breeding pair identities, clutch initiation and hatching dates, clutch size, and fledgling details according to standardized protocols, and recording the songs of the birds in the population using 60 AudioMoth acoustic loggers. The map shows the location of the nest boxes where we recorded song repertoires.

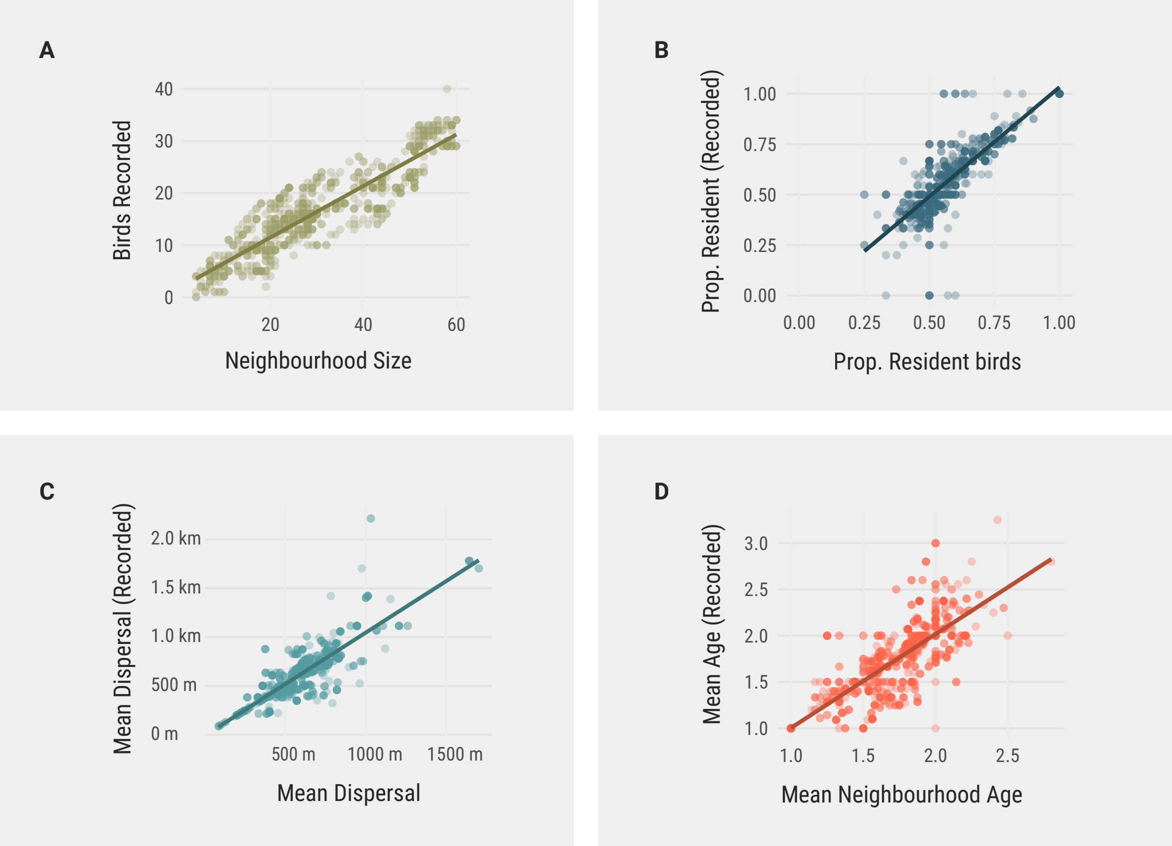


Figure S4.2. Demographic characteristics of recorded birds compared to those of all birds in the neighbourhood.
Comparison between the actual neighbourhood properties and neighbourhood properties estimated from birds for which we have song recordings. (A) Neighbourhood size (number of individuals) and number of individuals for which we have song recordings in that same neighbourhood. (B) Proportion of resident birds calculated from monitoring data and only from those birds with song recordings. Residents are birds that were ringed as nestlings in the population. (C) Mean dispersal distance of the birds in a neighbourhood calculated from monitoring data and only from birds with song recordings. (D) Mean age of birds in a neighbourhood calculated from monitoring data and only from birds with song recordings.

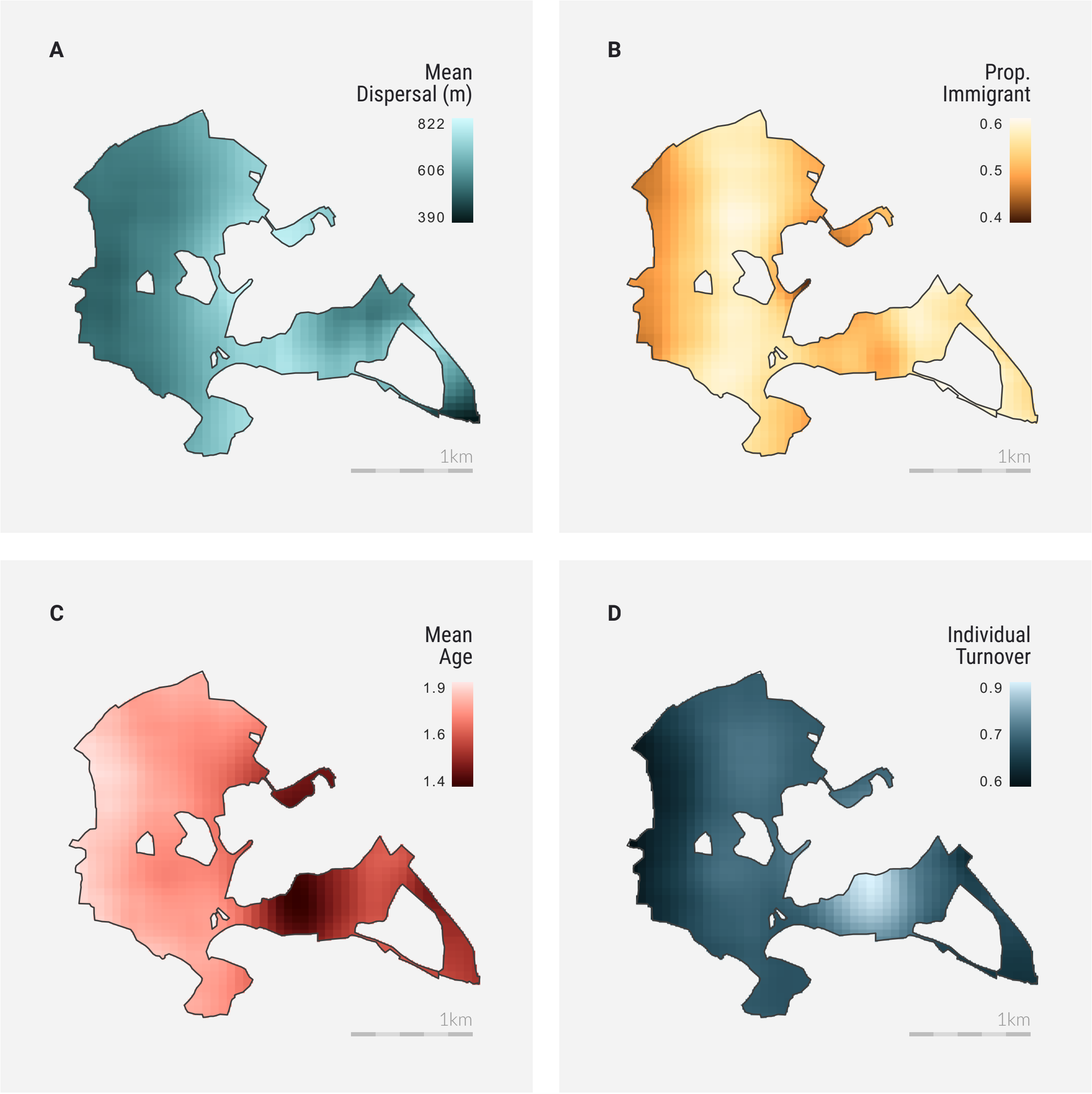


Figure S4.3. Spatial distribution of the neighbourhood-level predictor variables in the study.
(A) Mean natal dispersal distance, or the mean distance between the natal nest box and the breeding site for all birds in the neighbourhood. (B) Proportion of immigrant birds in the neighbourhood. (C) Mean age of birds in the neighbourhood. (D) Individual turnover, or the proportion of birds that were not already in a neighbourhood in the preceding year.

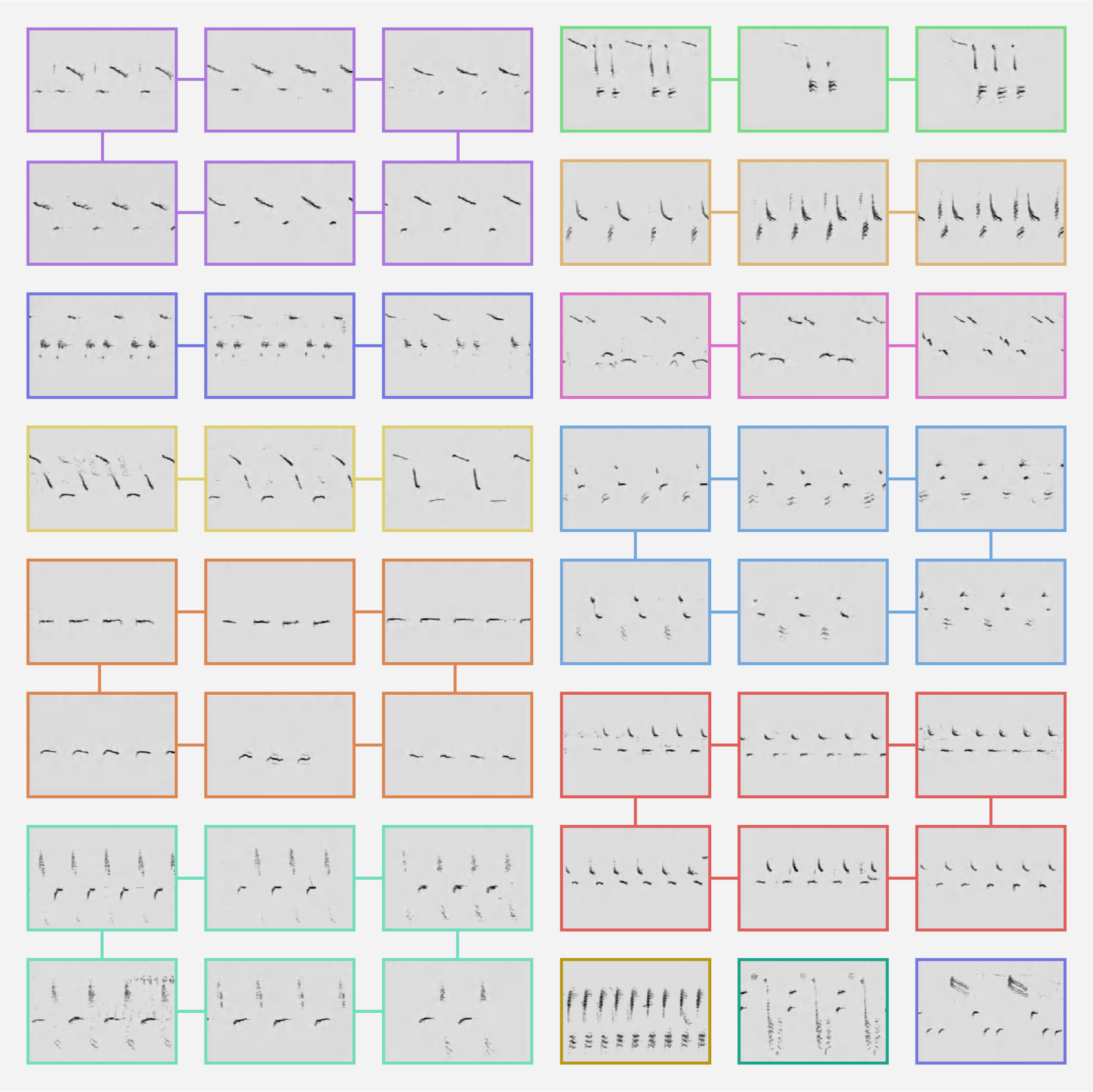


Figure S4.4. Examples of song type clusters in the study population.
Colours and connected lines represent the same song type cluster sung by different birds. Some song types are sung by many birds, while others are unique to a single bird. The clustering process is based on song similarity derived from a deep metric learning model and a manual categorization process following McGregor and Krebs (McGregor & Krebs, 1982b).

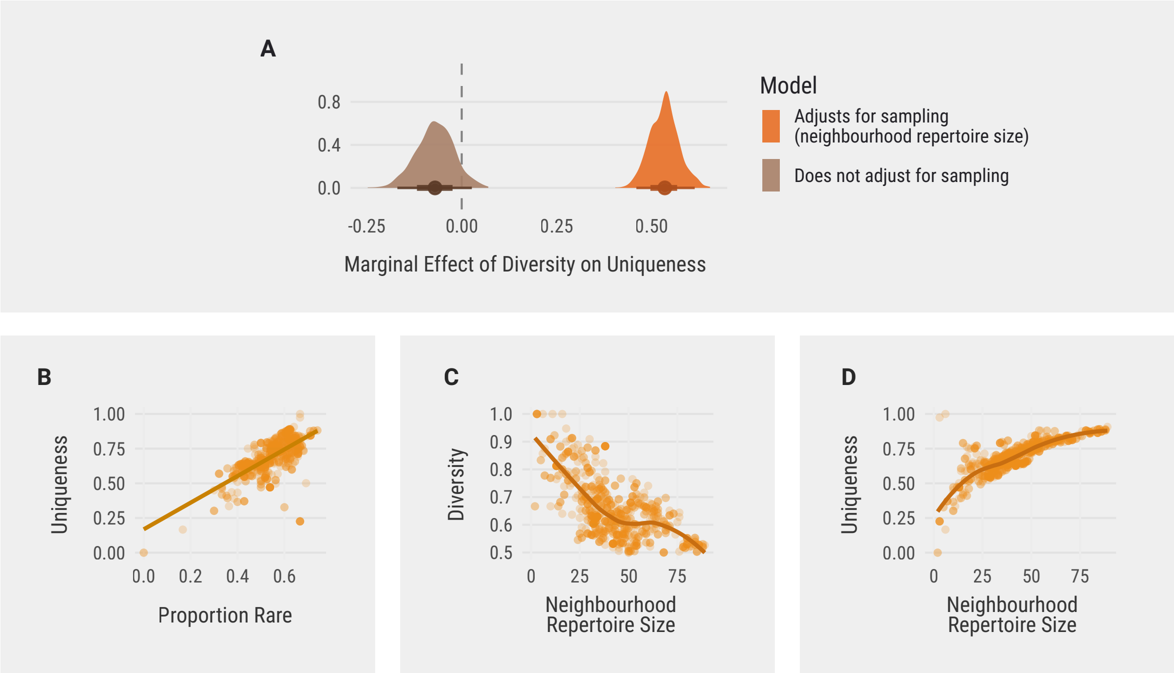


Figure S4.5. Estimates of cultural outcomes depend on the size of the neighbourhood repertoire.
(A) Marginal effect of diversity—which describes the proportion of unique songs in a neighbourhood—on uniqueness, that is, how uncommon, on average, the songs of the birds in a neighbourhood are. These two ways of characterizing cultural diversity are anti-correlated in our study site due to the effect of sampling: more frequent songs are sampled more readily, causing larger sample sizes (neighbourhoods with more birds and therefore songs) to yield lower average estimates of diversity (C) and higher average estimates of uniqueness (D), in a nonlinear manner. Once this is adjusted for, which we do by including GAM terms capturing neighbourhood song density or number of birds, diversity and uniqueness are positively correlated, as expected. (B) Our measure of cultural uniqueness (y-axis) has the advantages of being continuous and not using an arbitrary cut-off, but is nonetheless correlated with definitions traditionally used in the literature, such as ’songs shared by fewer than 4 birds (McGregor & Krebs, 1982b), here on the x-axis.

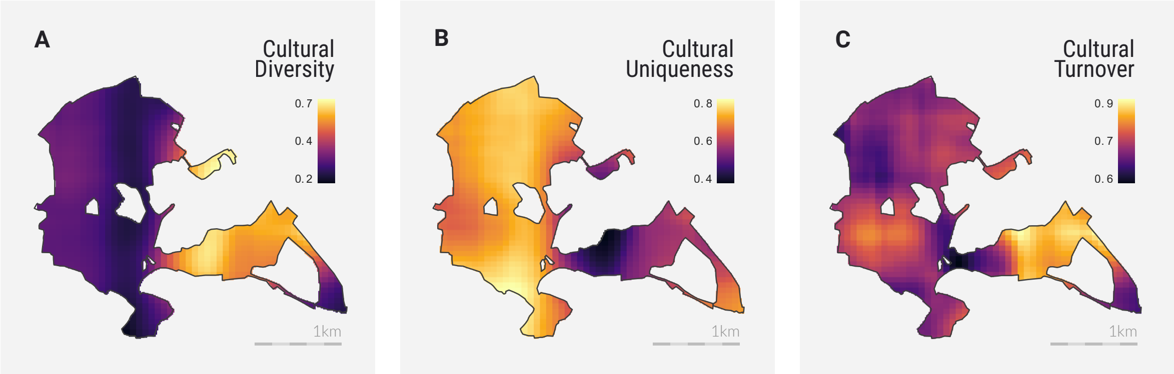


Figure S4.6. Spatial distribution of the neighbourhood-level cultural variables in the study.
(A) Relative diversity, or the proportion of unique songs in a neighbourhood. (B) Uniqueness, or how uncommon, on average, the songs of the birds in a neighbourhood are. (C) Rate of song cultural turnover, or the proportion of unique song types in a given year that were not already present in the same neighbourhood the preceding year. As described in Figure S4.5, (A) and (B) are anti-correlated due to the effect of sampling, but once this is adjusted for, neighbourhoods with more cultural diversity also tend to have more unique songs, as expected.

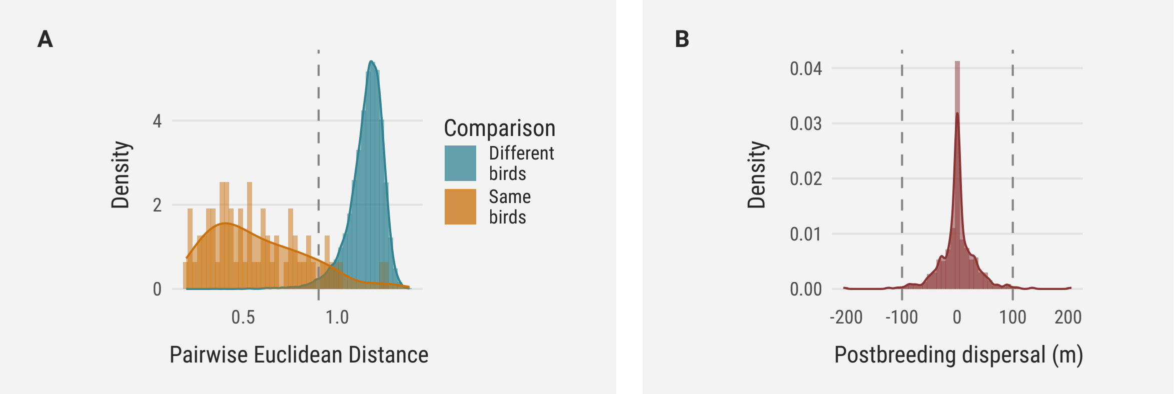


Figure S4.7. Thresholds used during the process of reidentifying individual birds based on their songs. (A) Shows the distribution of the acoustic distances between the same song type sung by the same known bird in different years, in orange, and the minimum pairwise distance between different birds and years. The x-intercept of the vertical line = 0.9. (B) Shows the distribution of the change in distance from the natal nest box to the breeding site in different years for birds that bred more than once. Adult birds have high nest site fidelity, which we used as a further constraint when reidentifying them from their songs.

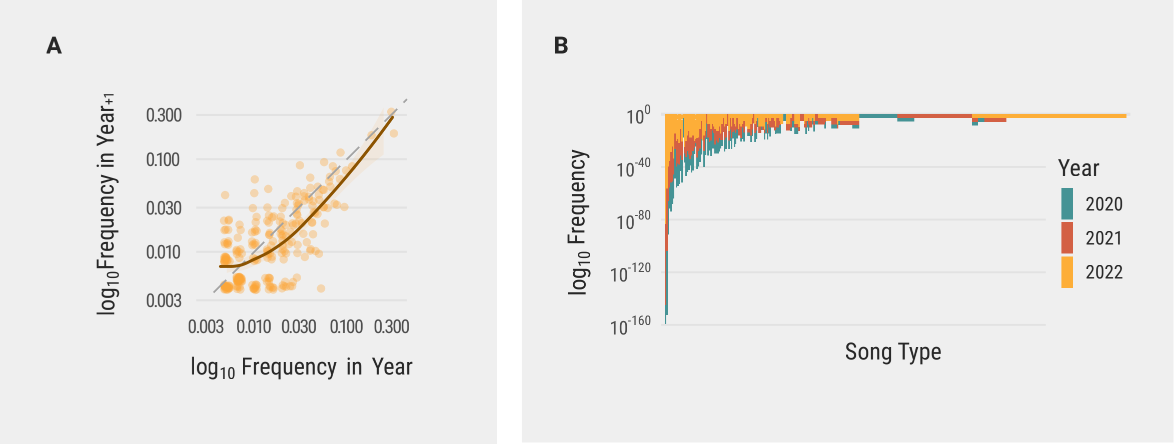


Figure S4.8. Song frequencies and their relationship with abundance in the following year. (A) The abundance of a song type in a given year predicts its abundance in the following year, with higher variance around rare songs. (B) Histogram showing the frequency of individual song types in the study.

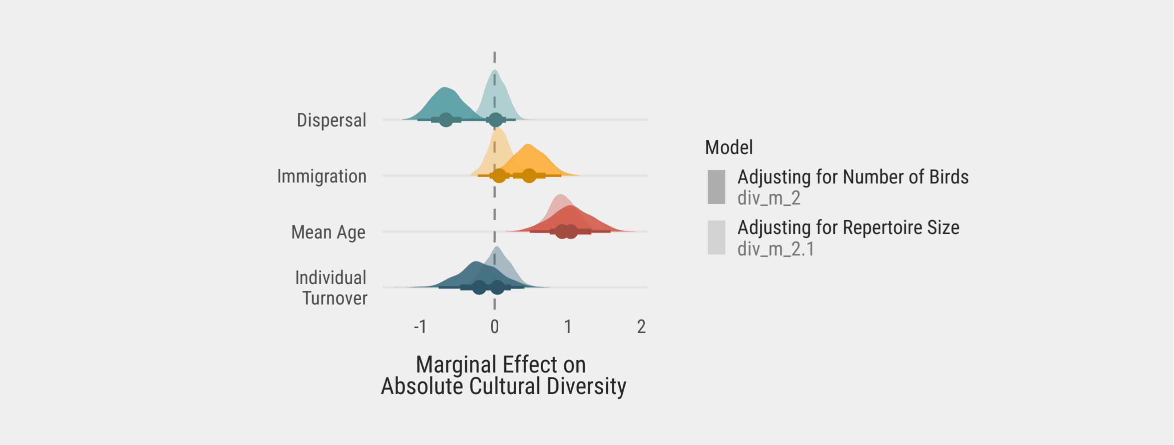


Figure S4.9. Effect of demographic variation on absolute cultural diversity within neighbourhoods. To explore how the number of individuals and their repertoire sizes within a neighbourhood affect the absolute number of different song types (as opposed to the relative diversity reported in Figure 4.2), we fit two models: one adjusting for the nonlinear effect of the number of individuals (higher opacity fill, corresponding to model *div_m_2*), and a second adjusting for the nonlinear effect of the number of song types, including repeated variants (lower opacity fill, *div_m_2.1*). See Table S4.1 for full model specifications.

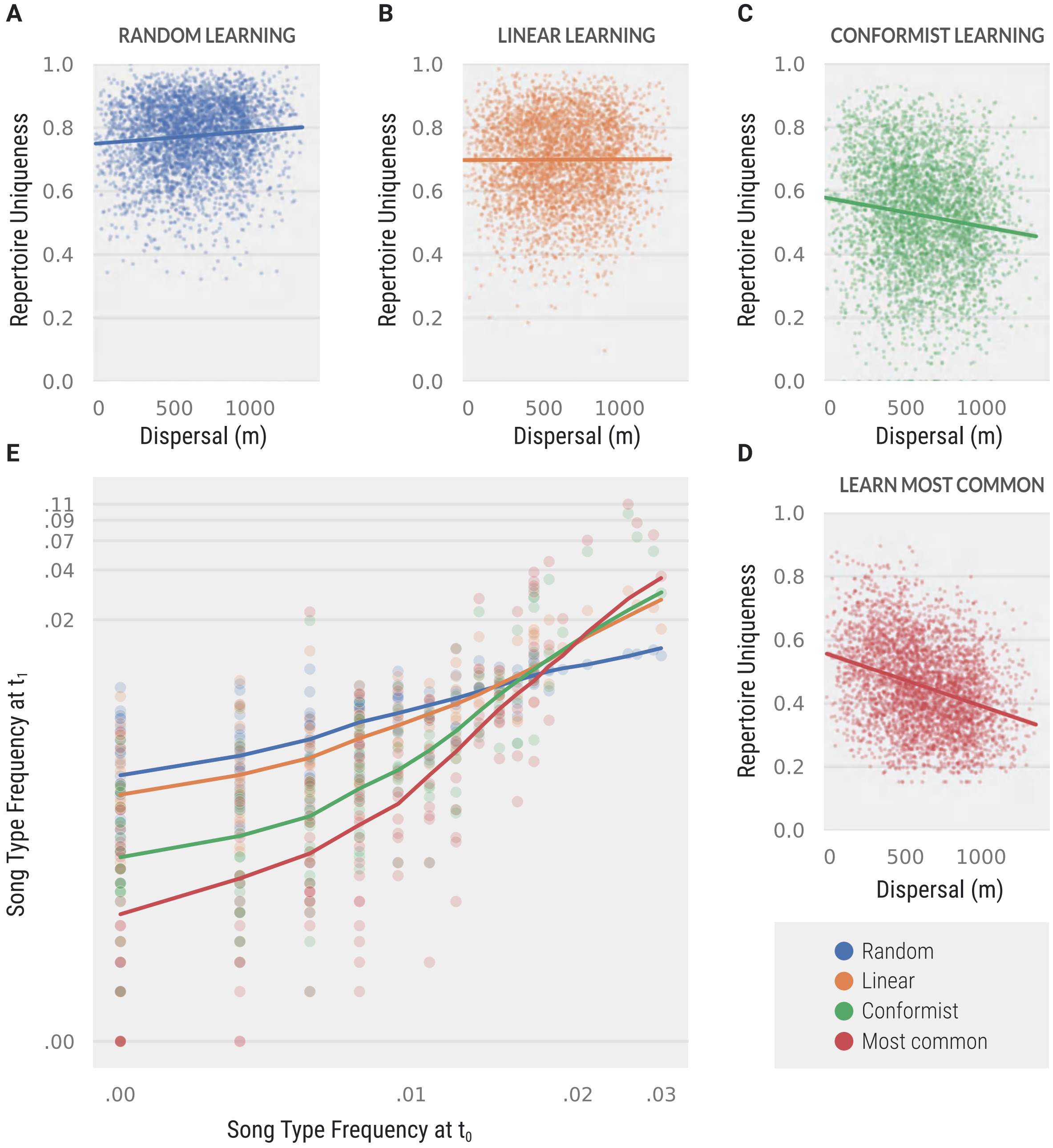


Figure S4.10. Simulation of the effect of natal dispersal on repertoire uniqueness.
We simulate the relationship between pre-breeding bird movement and the uniqueness of songs in their repertoires (relative to the population). We initialize 200 birds in a 1500 x 1500 square, each capable of singing 4 songs selected from a pool of 200 song types. Birds do not initially move. New birds are born and move based on a log-normal distribution parametrized to represent realistic dispersal behaviour in our population. Each bird can learn the songs it hears within a 200 m radius as it moves. At the end of their movement, a bird's crystallized repertoire is determined by its learning mechanism: (A) random learning of songs, (B) linearly frequency-dependent learning, (C) positively frequency-dependent learning, or (D) learn the most popular songs (strong conformism). The simulation is repeated n times per learning strategy, and we record the average uniqueness of songs in each bird's repertoire, which is a transformation of the average frequency of the bird's songs, as well as the distance that each bird has moved. The results show that the relationship between dispersal and repertoire uniqueness depends on the learning mechanism, and that the effect of dispersal detected in our study might be expected to arise if being exposed to a larger number of songs influences learning in a nonlinear frequency-dependent manner.

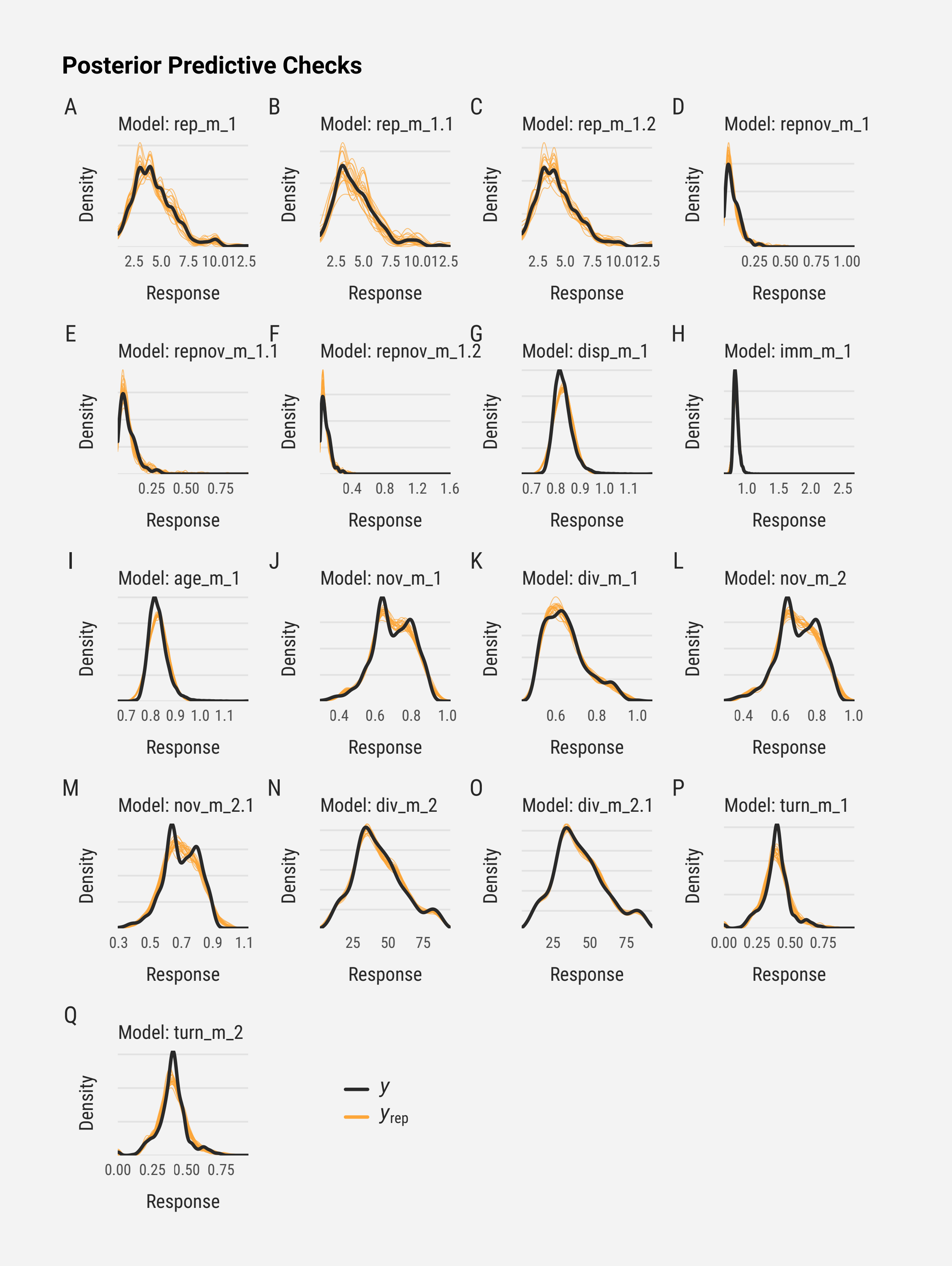
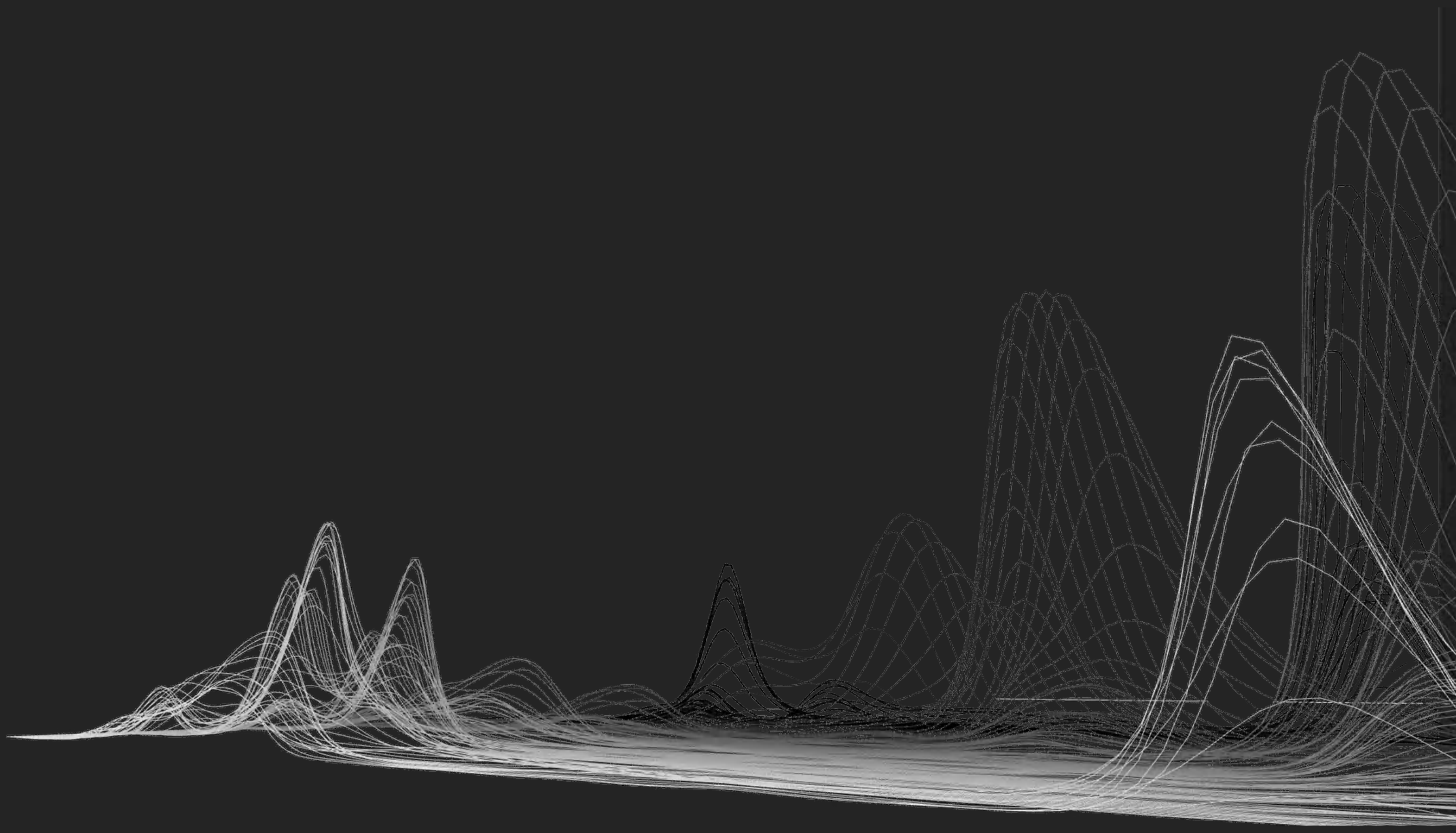


Figure S4.11. Posterior predictive checks for the main models in the study.
Comparing simulations from the posterior predictive distribution y^{rep} (thin orange lines) with the outcome y (black lines) using Kernel density estimates. The posterior predictive distribution is a distribution of possible outcomes of the model given the data and the model parameters, here used to check the fit of the model to the data.



3D mesh spectrogram of a great tit song.
Watch video animation here

5. MELODIC AND RHYTHMIC BIASES ANCHOR CULTURAL VARIATION IN GREAT TIT SONGS

Nilo Merino Recalde [1,*] and Ben C. Sheldon [1]

[1] Edward Grey Institute, Department of Biology, University of Oxford, Oxford, UK
*Corresponding author: nilo.recalde@biology.ox.ac.uk

SUMMARY

Human and avian song traditions are not static; they evolve culturally. When melodies are passed from one individual to another, be it a person or a bird, they are shaped in non-random ways by the inherent biases of learners and their interactions. If biases are shared among individuals, this process can serve to stabilize variation. For instance, while human songs exhibit considerable cross-cultural diversity in the organization of pitch and rhythm, some melodic and rhythmic patterns emerge repeatedly, and these facilitate accurate learning and transmission. Despite their many differences, both human and bird songs rely on vocal production and a process of cultural transmission, which suggests that a comparable process might also occur in bird songs. This study presents preliminary evidence that the acoustically simple yet highly diverse songs of great tits are also underpinned by non-random rhythmic and melodic structures. These structures share some common characteristics with human songs, such as the use of a limited number of melodic interval categories, small interval sizes, and isochronic rhythms. Our findings suggest that i) biases associated with vocal learning and performance can give rise to similar structural patterns across different species, and ii) melodic and rhythmic biases may help to stabilize song cultures, even in the face of high turnover.

Keywords: animal culture; bird song; demography, cultural evolution

INTRODUCTION

Music, especially in the form of song, is ubiquitous across human societies. Although some features of music are considered universal—songs are dominated by short phrases, discrete pitches with small melodic intervals, and simple rhythmic ratios (Mehr et al., 2019; Savage et al., 2015)—there exists significant variation that is specific to different cultural traditions, times, and places. This diversity emerges from a complex interplay between cognitive and motor factors common to all humans, which anchors variation, with the dynamics of exposure and preference, transmission errors, and other neutral processes that lead, over time, to cultural divergence (Savage, 2019; Tchernichovski et al., 2017; Tierney et al., 2011a; Verhoef & Ravignani, 2021)

Songs are typically passed down orally, and this process of learning and reproduction introduces changes in their structure; some random, and others directional. However, these learning and cultural transmission processes do more than just introduce minor variations in pitch or timing; they can also shape the underlying structure of melody and rhythm, pushing them toward stable musical 'cultural attractors' (Buskell, 2017; Claidière & Sperber, 2007). These attractors manifest as relatively discrete modes within the distribution of melodic and rhythmic intervals (Anglada-Tort et al., 2023; Jacoby et al., 2021; Verhoef & Ravignani, 2021), which may serve to align the perception, memory, and reproduction of information between individuals, enhancing the fidelity of cultural transmission and limiting the drift of cultural traits (Anglada-Tort et al., 2023; Falandays & Smaldino, 2022; Fehér et al., 2009; Heyes, 2018; Saldana et al., 2019; Trehub, 2015).

Recent findings suggest that similar processes may take place in the songs of certain bird species that exhibit structured, categorical organization in their songs. For example, Roeske et al. (2020) found that domestic zebra finch *Taeniopygia guttata* and thrush nightingale *Luscinia luscinia* songs had discretized isochronous rhythms. Xing et al. (2022) describe that Australian pied butcherbird *Cracticus nigrogularis* rhythms are categorically organized, and the songs of hermit thrushes *Catharus guttatus* (Doolittle et al., 2014) and musician wrens *Cyphorhinus arada* (Doolittle, 2012) use notes whose frequencies are related by small intervals in the overtone series (but see exceptions: Araya-Salas, 2012; Dobson and Lemon, 1977). This has been interpreted as evidence that some basic characteristics of human music might arise through processes that are more taxonomically widespread and not exclusive to it (Doolittle et al., 2014).

It is essential to recognize that, while human music and bird song are often compared, they are far from perfect analogues; they serve different purposes and are subject to distinct evolutionary pressures. Birdsong, in particular, is much more constrained due to factors such as sexual selection, energy costs, and the different species' ecology (Demery et al., 2021; Nowicki et al., 1998; Sierro et al., 2023; Spencer et al., 2003; Verzijden et al., 2012). Nevertheless, there are significant parallels between human and songbirds' vocal learning capabilities and the neural circuitry that underlies them, which has led to the establishment of songbirds as a model system in the study of vocal learning (Jarvis, 2019; Rouse et al., 2021). Our current understanding suggests that bird species capable of vocal learning have a genetically predetermined 'song space' (James & Sakata, 2017; Lachlan et al., 2010); within it, and paralleling human cultural variation, songs can evolve in various directions due to demographic and cultural processes. In this sense, it is reasonable to speculate that some of the same mechanisms that give rise to 'cultural attractors' biasing or enhancing the fidelity of song transmission in human cultures may also operate in bird songs. These mechanisms, as discussed by (Tchernichovski et al., 2017), likely include a compression process that leads to the transition from continuous variation to categorical syllable types—a phenomenon similar to that described in the evolution of linguistic structure (deCastro-Arrazola & Kirby, 2019; Gibson et al., 2019; Kirby, 2017; Silvey et al., 2019; Tamariz & Kirby, 2016). This transition could result from motor and perceptual biases, or be driven by the dynamics of cultural transmission itself: both of these processes could contribute to more accurate song transmission, aided by a restricted, categorical song space.

This leads to the intriguing possibility that cultural stability in bird song across larger temporal and spatial scales may not solely result from accurate learning of existing variation, as assumed in most models of song evolution (Hudson & Creanza, 2022; Lachlan et al., 2018; Pichkar et al., 2023; Youngblood & Lahti, 2022). Instead, underlying structural biases similar to those present in music and language may play an important role during the learning and production process. The presence of stable structural patterns in bird songs could shed light on the seemingly paradoxical nature of cultural change in species where song cultures exhibit polymorphism, experience significant and rapid turnover of cultural variants, and yet adhere to recognizable and stable patterns over long time periods. To explore these questions, rather than focusing on concrete acoustic features such as spectral bandwidth, fundamental frequency, or note duration, we need to move on to study the relational properties underpinning songs—their melodic and rhythmic structure. We aim to determine a) whether these structural aspects exhibit clustering within the song space of species and, if so, b) to what extent attractor points or regions of high probability can themselves vary culturally.

This chapter tackles only the first of these questions, which regards the empirical distribution of melodic and rhythmic structure in bird songs. To do this, we use a dataset comprising more than a million notes from a population of great tits, collected over three years in Wytham Woods, Oxfordshire, UK (Merino Recalde et al., 2023). While great tits may not be celebrated for their melodious tunes, with some likening their songs to a 'squeaky wheelbarrow' or 'bicycle pump', they are great for this research: their songs consist of phrases composed of one to three spectrally simple and stable notes, which restricts their combinatorial complexity and allows automated analysis. Our preliminary findings provide evidence that the melodic and rhythmic landscape of great tit songs exhibits a high degree of structure, suggestive of sources of stability beyond accurate learning of song types.

METHODS

Chapter 3 contains a detailed description of the data collection and preparation process, and Figure 3.2 shows a graphical description. Figure 5.1A contains a visual summary of the melodic interval and rhythmic ratio extraction process.

Study site and data collection

We study great tits, small birds with a short reproductive lifespan of around 1.9 years. These birds sing often during the breeding season, which lasts from March to June. They form socially monogamous pairs and defend territories around their nests (Hinde, 1952). We carried out this research Wytham Woods, in Oxfordshire, UK (51°46 N, 1°20 W). This semi-natural deciduous woodland spans approximately 385 hectares and has been under study since 1947 (Lack, 1964). Most great tits in this population breed in marked nest boxes, with individuals bearing unique British Trust for Ornithology (BTO) metal leg rings. Our data collection spanned from late March to mid-May in the breeding seasons of 2020, 2021, and 2022. Fieldworkers regularly checked the 1018 great tit nest boxes weekly before and during the egg-laying period, recording information such as breeding pair identities, clutch initiation and egg-hatching dates, clutch size, and fledgling number.

To capture male great tit vocalizations, we took advantage of the bird's behaviour during the reproductive period, when they consistently sing near their nests at dawn, before and during egg-laying (Mace, 1987). We deployed autonomous AudioMoth (Hill et al., 2019) sound recorders near nest boxes occupied by great tits, placing them on nearby trees at a consistent height and orientation (within 1 to 2 metres of the nest box and no more than 5 metres away). We used 60 AudioMoth recorders (30 in 2020), housed within custom waterproof enclosures. Our recording sessions began approximately one hour before sunrise, (sunrise times during the recording period spanned from 05:36 to 04:00 UTC). Each session consisted of seven consecutive 60-minute recordings, sampled at 48 kHz with a 16-bit depth. To cover a wide range of bird activity, we placed each recorder at a single location for at least three consecutive days before relocating it to a different nest box. This involved moving 20 recorders daily (10 in 2020) across the study site. Despite variations in song amplitude due to distance and direction, our observations didn't reveal any systematic bias in song selection.

Data processing and annotation

We processed and annotated the recordings using custom software and Python 3 scripts (van Rossum, 1995), with the open-source package `pykanto` (Merino Recalde, 2023). The tools and scripts can be accessed at github.com/nilomr/great-tit-hits-setup. In our data processing workflow, we began by segmenting songs based on visual inspection of spectrograms, aiming to isolate segments where bird notes were distinctly discernible from background noise. This manual process was performed using Sonic Visualiser (Cannam et al., 2010). To ensure that songs were correctly attributed to individual birds, we implemented measures to minimize false positives. Recordings with multiple vocalizing birds were discarded unless one bird's vocalization was significantly louder. Songs with a maximum amplitude below a threshold of -16 dB were also discarded (see Merino Recalde et al., 2023). We used normalized, band-passed mel-scale spectrogram representations for all subsequent operations (specific parameters detailed in the repository repository).

Note segmentation

Next, we segmented songs into their constituent notes using a dynamic threshold algorithm implemented in pykanto (Merino Recalde, 2023), partly based on work by Sainburg et al. (2019). After a series of morphological transformations and de-echoing, the algorithm identified spectral envelope minima as silences and divided the signal into shorter segments when needed, while discarding segments below minimum note duration. This relatively naive automated segmentation process is sensitive to factors like background noise, varying note amplitude, vegetation interference, sound direction changes, and performance variations, which can lead to missed or incorrectly delimited notes. Despite these challenges, we achieved an estimated 96.3% accuracy based on manual checks of 1048 randomly selected notes.

Song type annotation

Finally, we classified each song type in the dataset using a semi-supervised approach implemented in `pykanto` . This involved creating average unit spectrograms for each song, reducing dimensionality with UMAP (McInnes et al., 2018), and performing cluster analysis with HDBSCAN (McInnes et al., 2017) for each bird. While useful, this method can sometimes separate similar song renditions due to performance variation or background noise, potentially overestimating the repertoire size. To mitigate this, we employed `pykanto` 's interactive app to review and adjust clusters as needed Merino Recalde et al., 2023. We benefited from the consistent songs of great tits, which have relatively limited, stereotyped repertoires (1 to fewer than 15 song types).

Calculating rhythmic and melodic structure

Melodic intervals

First, for each note segment, we convert onset and offset times in seconds to frame indices, extract the corresponding audio segments and compute the Short-Time Fourier Transform (window size = 1024, giving a frequency resolution of 21.53 Hz). Then, we apply a binary frequency mask to the magnitude spectrogram to retain only the frequencies within a specified range (f_{min} to f_{max}), determined by the region of interest in the frequency domain drawn during segmentation. We then compute the mean of the masked spectrogram along the frequency axis to obtain the mean magnitude for each frequency bin and identify the one with the maximum value as the peak frequency. Great tit songs are spectrally simple, so this provides a good estimate in most cases.

After this, we calculate the interval size for each pair of adjacent note peak frequencies in the song,

$$\Delta_{ST} = 12 \cdot \log_2 \left(\frac{f_i}{f_{i+1}} \right) \qquad (5.1)$$

Where Δ_{ST} represents the semitones difference between two consecutive note frequencies, f_i is the frequency of the i-th note, and f_{i+1} is the frequency of the $i+1$-th note.

Strength of interval categories For clarity, we partition our analyses based on the spectral properties of notes in a song (see Song typology). The pure and unmodulated melodies typically preferred in studies of bird songs (Dobson & Lemon, 1977; Doolittle, 2012; Doolittle et al., 2014; Tierney et al., 2011b) show strong evidence of interval categories, which cannot be attributed to poor sampling (this is the second most common category, with n = 28,542) or an artefact of the frequency extraction process (we can resolve differences > 21.5 Hz). Although the position of the modes in the distributions derived from all other 2-note types is similar, we still need to make sure that the weaker pattern found in songs where notes are spectrally more complex (e.g., have harmonics) is not an artefact of the frequency extraction process and indeed reflects looser categories.

Rhythmic intervals

We calculated song rhythmic ratios, denoted as R, as the ratio of each interval to the total length of that interval along with the subsequent one, following Roeske et al. (2020) and Xing et al. (2022):

$$R = \frac{i_n}{i_n + i_{n+1}} \tag{5.2}$$

Here, i_n stands for the Inter-Onset Interval between the n-th note and the next note $(n+1)$. Each Inter-Onset Interval (IOI_i) is calculated from the time segmentation described in the previous section as:

$$IOI_i = Onset_{i+1} - Onset_i \tag{5.3}$$

Where i ranges from 1 to $n-1$, representing each note except the last one.

Resampling and Kernel Density Estimation

Songs have varying lengths, and levels of vocal production by birds differ greatly. To ensure that our results were not driven by an imbalance in the data, we sampled 5 intervals with replacement from each song sung by each bird, then repeated this process for each song type, so that each song type contributed 10 songs to the balanced dataset. Finally, to account for variation in repertoire size and sample size per year, we again sampled songs with replacement from each individual bird or year, depending on the analysis, during the Kernel Density Estimation (KDE) procedure.

To explore the distribution of melodic interval sizes and rhythmic ratios, following Anglada-Tort et al. (2023) we compute one-dimensional KDEs over a grid of 1000 values in the range of the data, using a bandwidth of 0.15 semitones for melodic intervals and 0.015 s for the rhythmic intervals to balance smoothing and resolution. Then, to quantify uncertainty around the mean KDE estimates, we bootstrap them by repeatedly resampling the data, recomputing the KDE, and calculating the mean and standard deviation of the density estimate across the bootstrap samples. We do this for the entire dataset, and for each in the typology of songs described below separately.

Song typology

To investigate potential variations in the motor challenges and constraints associated with different song types, we classified songs into distinct subsets based on the number of notes, spectral properties of the notes, and the presence or absence of frequency modulation, defining the following song typology (sample sizes in the balanced dataset in parentheses):

- **One note**: Songs consisting of a single note repeated rhythmically (5,161).
- **Two notes**: All songs composed of two different, alternating notes (126,947).
- **Two notes, pure and unmodulated**: Songs with two notes, each consisting of a single stable, pure, unmodulated tone (28,542).
- **Two notes, none pure**: All two-note songs where neither note is pure and unmodulated (19,394).
- **Two notes, both harmonic**: Songs featuring two notes, both of which have harmonic components (75,665).
- **Three or more notes**: Songs comprising three or more notes. There are few three-note melodies, so we did not split these by spectral or modulation properties (31,305).

Because we are extracting dominant or peak frequencies f_d instead of fundamentals f_0, there is a chance that the results for songs with notes that are not pure-tone or where f_d and f_0 do not coincide are noisier, or do not capture the target pitch as well. This typology also allows us to isolate songs made up of pure-tone notes, which have been preferentially used in prior work (Dobson & Lemon, 1977; Doolittle, 2012; Doolittle et al., 2014; Tierney et al., 2011b), and where that is not a problem.

Ethical note

All work involving birds was subject to review by the University of Oxford, Department of Zoology, Animal Welfare and Ethical Review Board (approval number: APA/1/5/ZOO/NASPA/Sheldon/TitBreedingEcology). Data collection adhered to local guidelines for the use of animals in research and all birds were caught, tagged, and ringed by BTO licence holders (NMR's licence: C/6904).

RESULTS AND DISCUSSION

We analyse a dataset of 1,161,033 notes and extract a balanced dataset of 164,055 melodic and rhythmic intervals. Our results are divided into two main sections, one focusing on melodic structure and the other on rhythm. In both cases, we find that the structure of great tit songs is not random and shares similarities with human songs and those of other singing birds and mammals. Our findings have implications for understanding cultural change and variation in bird songs, which we discuss briefly.

Melodic structure

In a series of iterated learning experiments with humans, Anglada-Tort et al. (2023) discovered that as melodies were learned and passed on, they acquired properties seen across human cultures (Brown & Jordania, 2013; Mehr et al., 2019; Savage et al., 2015). In particular, they became biased towards a small vocabulary of intervals ($\leq$7 categories, or ’peaks’), and used small interval sizes ($\leq$7 semitones, or a perfect fifth). We now present evidence that the distribution of melodic intervals in the songs of our great tit population follows a structured pattern and shares some of these characteristics with human melodies and discuss some possible reasons for this convergence. (See Figure 5.1A and Figure 5.2.)

Great tit songs favour small intervals

The mean interval size observed in great tit songs from Wytham Woods was approximately 4.37 semitones (95% CI [4.36, 4.38]). This falls in line with the interval sizes produced by human participants from India and the United States, which had a combined mean of 4.48 semitones (95% CI [4.24, 4.72]). This prevalence of small interval sizes ($\leq$7 semitones), one of a few true statistical ’universals’ in music (Savage et al., 2015), can be likely attributed to vocal range constraints: melodies transmitted orally tend to be biased toward small intervals, in contrast to those transmitted through other means or solely driven by preference Anglada-Tort et al. (2023). That the precise range is so similar to humans is interesting but presumably a coincidence driven by the particular vocal abilities of great tits. In contrast, a study analysing melodic intervals in musician wrens (Doolittle, 2012) found that the most common interval was the octave (12 semitones), which is extremely rare in great tits; and a comparison of 80 bird songs with human melodies showed that the latter have smaller average intervals on average (Tierney et al., 2011b). This difference might stem from the fact that some birds can execute large pitch jumps in their songs by independently adjusting tension in the labia of their bipartite syrinx (Suthers, 2004).

Interestingly, in the cross-cultural experiment by Anglada-Tort et al. (2023), melodic intervals produced by Indian participants were significantly smaller (3.25 [2.93, 3.57] semitones on average, 95% CI) than those produced by US participants (5.71 [5.55, 5.86] semitones). This aligns with corpus studies showing that intervals in Carnatic melodies are smaller than those in Western melodies (Bowling et al., 2012) and suggests that, even within the range dictated by motor-vocal constraints, cultural influences can still shape the general distribution of interval sizes in human music. As for great tit song, we cannot definitively conclude whether the variation in average interval sizes is similarly influenced by cultural factors, simply because our dataset, although comprehensive, is derived from a single population. However, we present qualitative evidence that, despite the high turnover in great tit songs due to population turnover and other demographic processes (as discussed in the previous chapter), the coarse-grained distribution of interval sizes remains remarkably stable across three consecutive years (see Figure S5.1). This stability is again likely to stem, at least in part, from vocal range constraints (both motor and physiological).

Skip-reversal patterns In order to further characterize the general melodic structure of great tit songs, we plot the interval size in semitones as a function of the interval’s starting frequency in Figure 5.1C. This reveals a very strong pattern known as ’skip-reversal’, where larger intervals are associated with extreme pitches—a phenomenon that has stirred debate in musicology quarters (Von Hippel, 2000) but likely arises from a simple process of regression to the mean due to motor constraints. In essence, when a melody approaches the boundaries of the singer’s comfortable vocal range, also known as the ’tessitura,’ it has little choice but to return toward the centre of the range (Tierney et al., 2008; Von Hippel, 2000; Von Hippel & Huron, 2000).

Melodic intervals are categorically organized

As we have seen, great tit songs predominantly feature small melodic intervals, which can likely be attributed to motor-vocal constraints. Next, we want to determine if the distribution of interval sizes within the vocal range of great tits is different from what we would expect by chance. To find out, we created a null distribution by shuffling the empirical distribution of frequencies and recalculating interval sizes. This null distribution represents what we would expect if birds with the same vocal range had no preference for specific interval sizes (shown as the grey discontinuous line in Figure 5.1B). Simply due to the unimodal shape of the distribution of

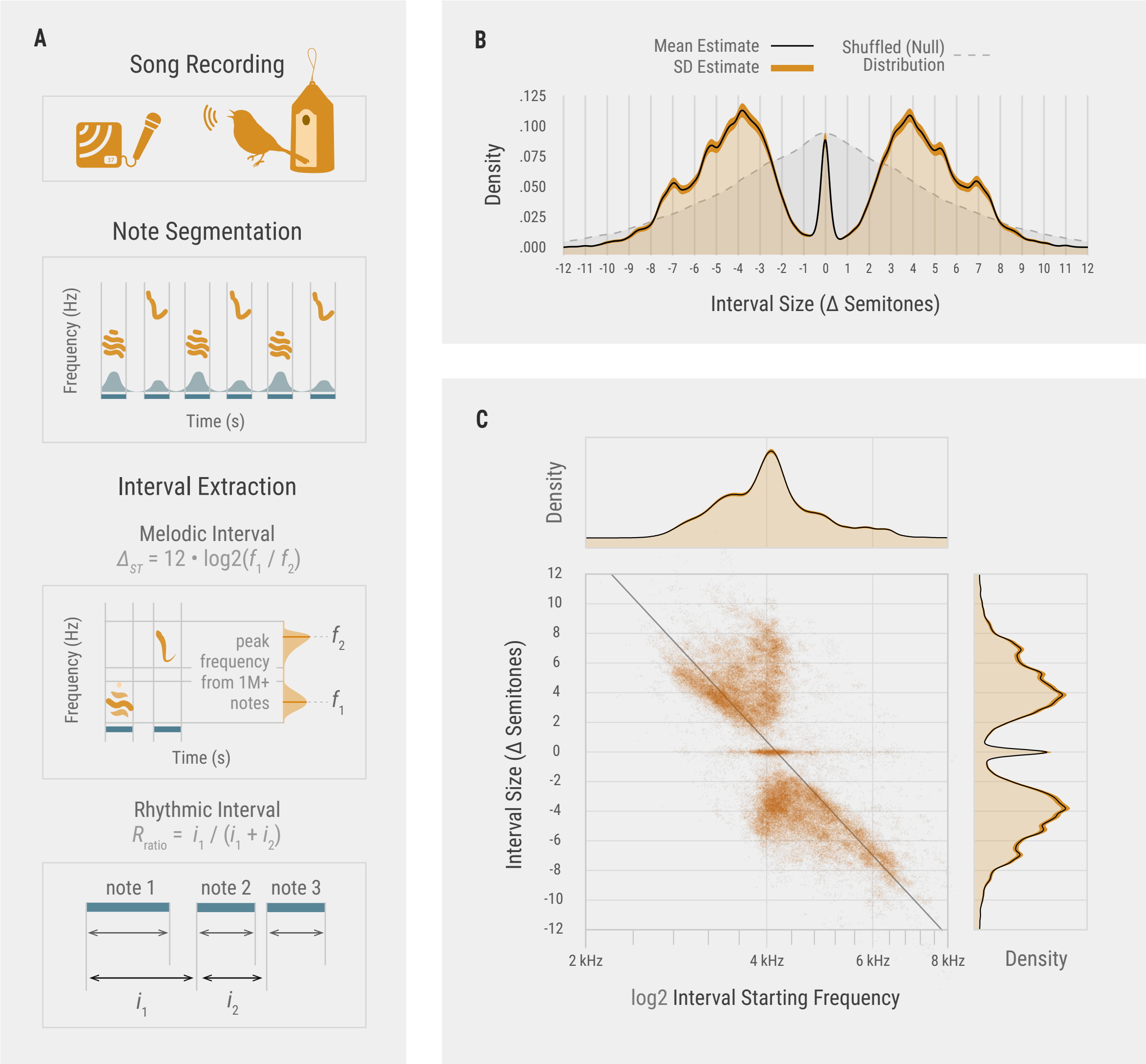


Figure 5.1. Visual description of the methods; melodic structure in great tit songs.
(A) Simplified visual summary of the process used to extract melodic and rhythmic intervals; see also Methods (B) Density estimate of the distribution of interval sizes over the entire dataset (songs consisting of two notes), with mean (black line) and SD (orange line). The discontinuous grey line represents a null distribution created by shuffling the empirical distribution of frequencies and recalculating interval sizes. It represents the expected distribution if birds with the same motor-vocal constraints did not have a preference for interval sizes within their range. (C) Great tit songs strongly show a pattern known as skip-reversal, where larger intervals are associated with extreme pitches, which has been the subject of debate in (human) musicology (Von Hippel, 2000) but likely arises from a simple process of regression to the mean.

note frequencies sung by great tits (mean frequency: 4310.05 Hz, SD: 926.5; Figure 5.1C, top panel), interval sizes in this null distribution are centred around zero. Though perhaps naive—due to communicative and perceptual pressures, we might expect birds to avoid barely perceptible changes, for example—this comparison allows us to see that the empirical distribution of melodic intervals is highly structured, and it is so in interesting ways. Figure 5.2 presents the probability density distribution estimated for each set of songs in our typology.

One-note songs The one-note case in Figure 5.2A is, by definition, tightly clustered around the unison (0 semitones), and gives a good idea of the population-level variance found when birds are trying to repeat the same note consistently.

Songs with multiple notes In cases where birds sing melodies with two or more notes (Figure 5.2B-F), we observe a noticeable dip in the smallest intervals (less than 2 semitones). This suggests that birds strongly avoid melodies with very small frequency changes. Interestingly, this pattern resembles what we find in human songs, where it's interpreted as a result of perceptual preferences (Bowling et al., 2012; Kuroyanagi et al., 2019). However, it could also relate to production difficulties, as very small intervals might be harder to sing accurately than unisons or larger leaps (Anglada-Tort et al., 2023).

The rest of the distribution also exhibits non-random features. It reveals evidence of potential melodic interval categories, defined as local maxima or modes. These patterns may be driven by a preference for certain interval sizes or interval ratios, a phenomenon observed in human music and hypothesized to play a role in bird song (Doolittle, 2012; Doolittle et al., 2014); motor or physiological biases, or other, currently unknown processes. Some melodic intervals produced by great tits in our population cluster around specific interval categories commonly found in human music; these include intervals of 7 and 4 semitones, both ascending and descending, which correspond to what is considered the most consonant interval within the octave (perfect fifth, 3:2) and the major third, respectively. See Table 5.1. The latter, which in the harmonic series would correspond to the interval between the fourth and fifth harmonics, is the most common interval size in great tit songs.

We also see a peak around 5.2 semitones in the simpler (two-note, pure, unmodulated) songs Figure 5.2C which is not present in Figure 5.2D–F. The difference in the exact location of these peaks across the typology of songs might reflect an interaction between song syntax or production and interval categories, or, as discussed in the Equation 5.2.3, less precise measurement due to the spectral complexity of notes.

It is fascinating to consider the possibility that similar perceptual, cognitive, and production biases might shape the interval structure in both great tit and human melodies. However, even if the alignment of interval categories is coincidental, the presence of a limited number of modes (at ≈7, ≈5.2, ≈4, and ≈0 semitones; ascending and descending) within a comprehensive dataset of songs represent-

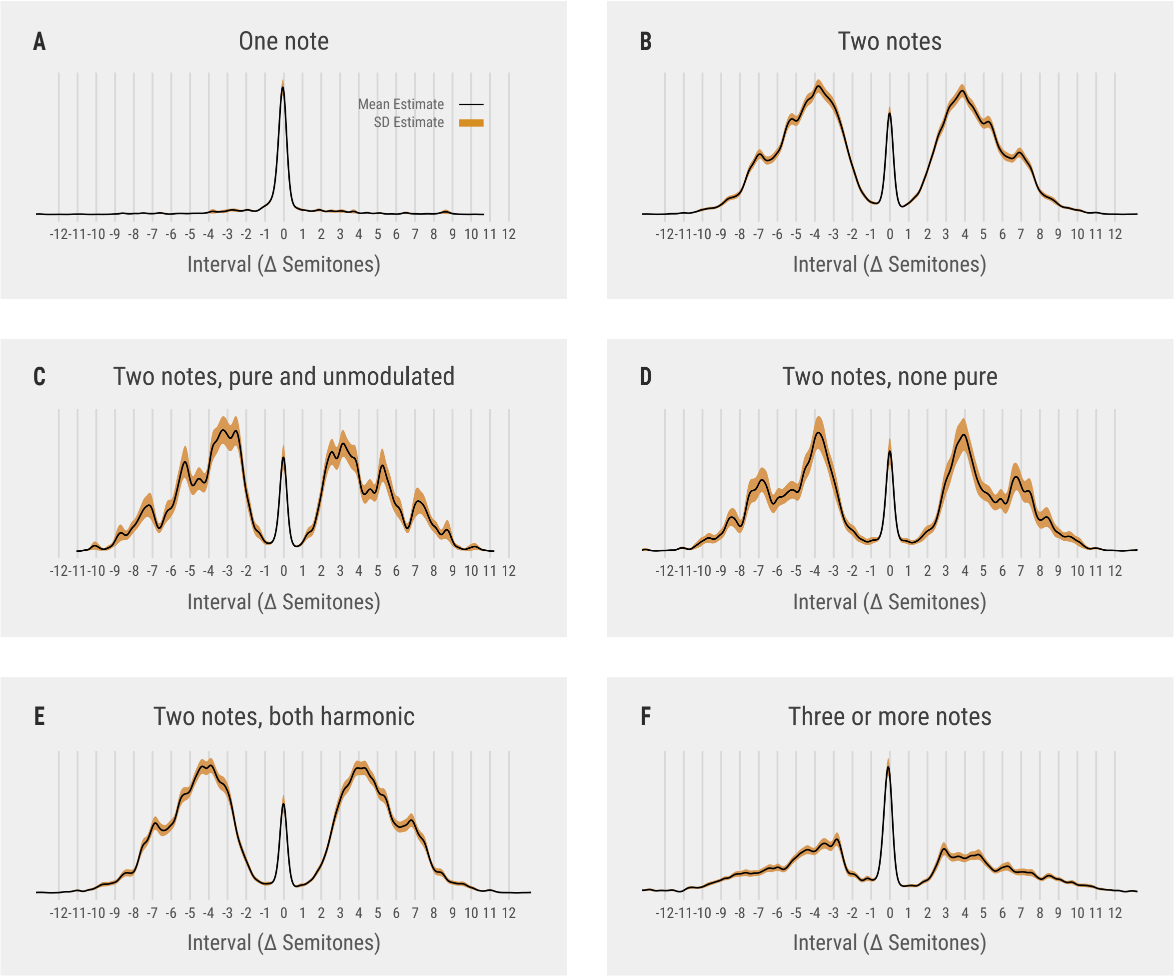


Figure 5.2. Kernel density estimates of the distribution of melodic intervals across a typology of songs. In this figure, we categorize songs into distinct types based on the number of notes they contain and their spectral properties and plot the estimated density distribution of melodic intervals, measured in semitones and spanning the octave (ascending and descending). (A) One note: Songs consisting of a single note repeated rhythmically. (B) Two notes: All songs composed of two different, alternating notes. (C) Two notes, pure and unmodulated: Songs with two notes, each consisting of a single stable, pure, unmodulated tone. (D) Two notes, none pure: All two-note songs where neither note is pure and unmodulated. (E) Two notes, both harmonic: Songs featuring two notes, both of which have harmonic components. (F) Three or more notes: Songs comprising three or more notes. There are few three-note melodies, so we did not split these by spectral or modulation properties.

Table 5.1. Name and size of intervals within the octave in the Western chromatic scale.
This table includes interval name, number of semitones, interval sizes measured in cents within equal temperament, corresponding frequency ratios with respect to just intonation, and the decimal values for those just intonation intervals in twelve-tone equal temperament (adapted from Muller, 2016).

Scale-Step Interval	Semitones	Equal Temp. (cents)	Ratio	12 TET value
Perfect Unison (U)	0	0	1:1	1
Minor Second (m2)	1	100	16:15	1.059463
Major Second (M2)	2	200	9:8	1.122462
Minor Third (m3)	3	300	6:5	1.189207
Major Third (M3)	4	400	5:4	1.259921
Perfect Fourth (P4)	5	500	4:3	1.33484
Tritone (T)	6	600	64:45	1.414214
Perfect Fifth (P5)	7	700	3:2	1.498307
Minor Sixth (m6)	8	800	8:5	1.587401
Major Sixth (M6)	9	900	5:3	1.681793
Minor Seventh (m7)	10	1000	16:9	1.781797
Major Seventh (M7)	11	1100	15:8	1.887749
Octave (P8)	12	1200	2:1	2

ing distinct song types that undergo rapid turnover, suggests that melodic structure does not merely evolve randomly due to copying errors. Instead, it raises the possibility that interval categories play a role in stabilizing and anchoring cultural variation, much like they do in human melodies, whether of motor, perceptual, or cognitive origin. More generally, there is evidence that the emergence of shared categories in an otherwise graded or continuous trait space facilitates efficient and accurate transmission (deCastro-Arrazola & Kirby, 2019; Falandays & Smaldino, 2022; Kirby, 2017; Silvey et al., 2019; Tchernichovski et al., 2017), and we suggest that a similar process might drive the patterns in our data.

From our multi-year dataset, we observe that the exact positions of interval categories are relatively stable, although not as much as the broader shape of the distribution (two main modes and a trough between them, Figure S5.1). However, adequately testing whether these fine-scale melodic biases can also shift due to cultural evolutionary processes will require longer-term data or data from a wider geographic region.

Rhythmic structure

Rhythmic patterns in songs are constructed by the temporal intervals between note onsets and can be broadly categorized into periodic or aperiodic. Periodic or repeating patterns can further be divided into isochronous—where a single interval, say, 0.5 seconds, is consistently repeated—and heterochronous, where intervals vary in duration. Heterochronous patterns, in turn, can be described by small integers (like 1:2) or more complex ratios (e.g., 16:23). In human music, rhythms often employ ratios such as 1:2, 1:3, and 2:3 but avoid ratios like 31:8 (Ravignani et al., 2014). Regular or isochronous patterns are a pervasive feature of musical rhythm, often attributed to the need for group coherence and predictability in collective performance and dancing (Kotz et al., 2018).

Furthermore, cross-cultural experiments have shed light on the prevalence of specific rhythmic templates across different societies. 1:1 and 2:1 rhythmic templates are widespread across cultures, while more complex ratio categories, such as 3:2 and 4:3, are influenced by cultural trajectories and exposure (Jacoby & McDermott, 2017; Jacoby et al., 2021; Polak et al., 2018). Combined with other laboratory experiments (Ravignani et al., 2017) showing how these can emerge from random stimuli and enhance how accurately participants can reproduce them, this may indicate that these categories emerge, at least in part, during learning and cultural transmission and serve to stabilize it.

Rhythmic abilities in other species

While the focus of most research on rhythm has rested with human music, there is increasing evidence that other animal species can both produce and discern regular rhythms (Rouse et al., 2021; van der Aa et al., 2015; Verga et al., 2022). For example, zebra finches, thrush nightingales, rock hyraxes *Procavia capensis*, and various singing primates have demonstrated isochronous rhythm production (De Gregorio et al., 2021, 2023; Demartsev et al., 2023; Norton & Scharff, 2016; Raimondi et al., 2023; Roeske et al., 2020). This suggests that more fundamental, widespread motor or perceptual processes may underlie these rhythmic abilities. Integer ratio rhythms might be more easily produced, for example, or representational constraints could affect perception and make categories characterized by simple mathematical relationships more likely to arise (Jacoby & McDermott, 2017; Ravignani et al., 2018; Roeske et al.,

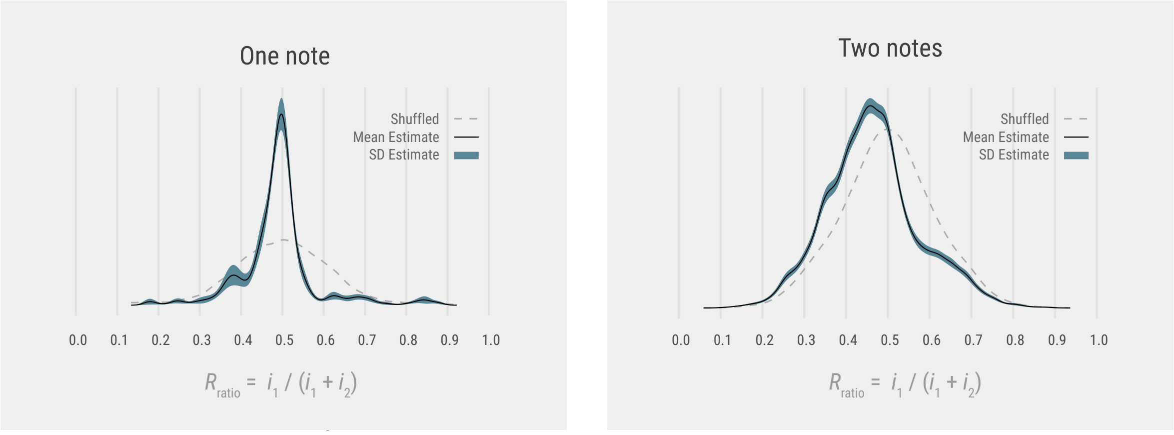


Figure 5.3. Rhythmic ratios in one and two-note songs.
Density estimate of the distribution of rhythmic ratios in one-note (left) and two-note songs (right), with mean (black line) and SD (orange line). The discontinuous grey line represents a null distribution created by shuffling the empirical distribution of inter-onset intervals and recalculating interval sizes.

2020).

Exploring Rhythmic Structure in Great Tit Songs

In this context, we examine the rhythmic structure of great tit songs. Similar to our approach with melodies, we compare the empirical distribution of rhythmic ratios with a null distribution obtained by shuffling the observed range of inter-onset intervals (the time between consecutive note onsets) and recalculating the ratios between them. The analysis reveals that rhythm in great tit songs is generally biased towards isochrony (Figure 5.3). We also investigate whether rhythmic structure remains invariant when considering other song properties, such as the number of notes in a phrase. Our findings suggest that rhythm interacts with syntax or melodic structure, as it is not consistent across all song types (Xing et al., 2022).

One-note songs For instance, the distribution of one-note songs (0 semitone intervals) is strongly organized around isochronic rhythms (identical intervals, 1:1 ratio), as evidenced by a steeply unimodal distribution perfectly centred around 0.5 in Figure 5.4A. This finding contributes to a small but growing body of evidence of this universal characteristic of human music in other species (De Gregorio et al., 2021, 2023; Demartsev et al., 2023; Raimondi et al., 2023; Roeske et al., 2020; Rouse et al., 2021; Xing et al., 2022). Indeed, we speculate that as more instances of isochrony in animal vocalizations are discovered, exceptions may become more intriguing than the norm.

Two-note songs As summarized in Figure 5.4B, two-note songs also have a unimodal distribution of rhythmic ratios, with more variance. Interestingly, they deviate slightly from an isochronous pattern, with the first interval being more often slightly shorter than the second. This pattern is reminiscent of patterns observed in Indris lemurs (De Gregorio et al., 2021) and Australian pied butcherbirds (personal observation from Xing et al., 2022). We believe there are two possible explanations for this deviation: motor constraints due to producing alternating frequencies that are not present in the single-note phrases, or a more complex, perceptually-mediated departure from expectations akin to ’tempo rubato’ (De Gregorio et al., 2021; Parncutt, 1994) or ’backbeat delay’ in music (Frane, 2017).

Songs with more than two notes In songs that feature more than two alternating notes, the rhythmic landscape becomes more complex. Peaks emerge around isochrony (1:1) and a ≈ 1:2 ratio, along with other, more complex rhythms. When great tits produce three-interval rhythms, our analysis reveals distinctive modes near the two smallest integer ratios, which are widely prevalent in human music performances and traditions (Jacoby & McDermott, 2017; Jacoby et al., 2021). This, which has also been found in thrush nightingales and indris (De Gregorio et al., 2021; Roeske et al., 2020), is again strongly suggestive of common motor, cognitive, or perceptual biases across vocalizing species.

CONCLUSION

In this study we examined the melodic and rhythmic structure of great tit songs using a large multi-year dataset of over 100,000 songs from a population in Wytham Woods, UK. We found that great tit songs favour small intervals, likely due to vocal motor constraints, and also showed patterns like avoidance of very small intervals (<2 semitones). Great tit songs exhibited structured patterns in their melodic intervals rather than random variation, with potential evidence of interval categories or modes around specific interval sizes like 0, 4, 5.2 and 7 semitones.

The songs also showed biases towards isochronous rhythms (1:1 ratios) in one-note songs, with increasing complexity in multi-note songs including emergent modes around small integer ratios like 1:2.

These results provide preliminary evidence that the melodic and rhythmic structure of great tit songs is not simply shaped by random variation, and is suggestive of underlying processes like motor constraints, perceptual biases, or cultural evolutionary dynamics. Some of these patterns, like the prevalence of small melodic intervals, multimodal intervallic distributions, and isochronous rhythm, align with observations in human music and other species, hinting at potentially shared mechanisms. However, direct comparisons should be made cautiously given the very different functions and factors shaping birdsong and human music.

The existence of relatively stable interval categories within the rapidly changing song culture of great tits (see Chapter 4) raises the possibility that such categories could play a role in anchoring or stabilizing cultural evolution, analogous to hypothesized mechanisms in human music and language. In order to test if some of these fine-grained patterns themselves evolve culturally, which would provide evidence for a cultural origin, we will need longer-term and geographically more widespread data. Our results are descriptive and cannot shed light on the underlying causes (e.g. production constraints, cognitive biases, cultural processes) of the observed distribution of intervals: further experimental and comparative studies will be needed to tease apart these factors.

DATA AND CODE

The code to replicate these analyses and figures can be found at github/nilomr/greti-song-intervals. All the data are available from osf.io/n8ac9, and documented here.

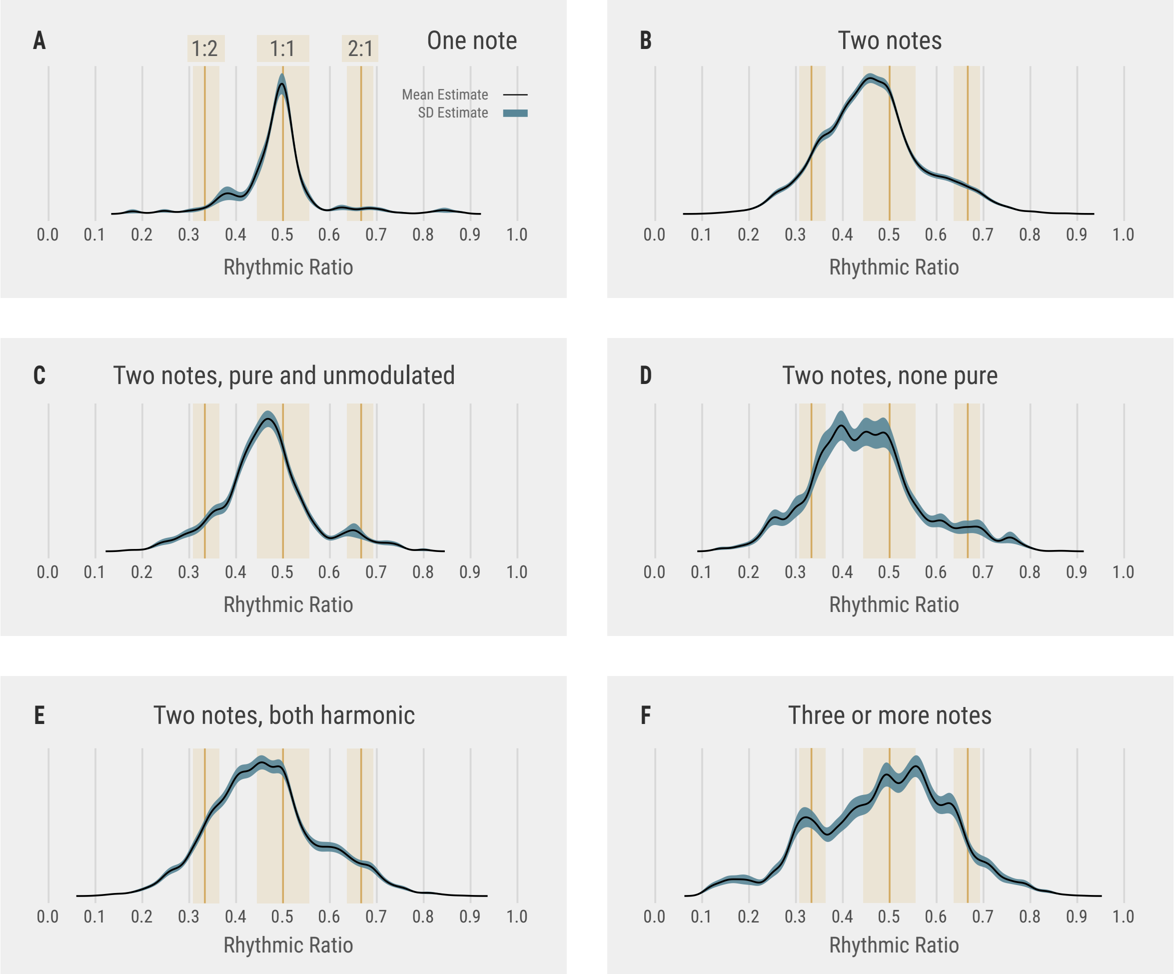


Figure 5.4. Kernel density estimates of the distribution of rhythmic ratios across a typology of songs.
We categorize songs into distinct types based on the number of notes they contain and their spectral properties and plot the estimated density distribution of rhythmic ratios, calculated as the ratio of the first inter-onset interval to the sum of the first and the second. Orange vertical lines and shaded areas represent, from left to right 1:2, 1:1, and 2:1 rhythms, and the 'on-integer' ranges as per De Gregorio et al. (2021) and Roeske et al. (2020)

ACKNOWLEDGEMENTS

This manuscript represents a very early stage of this work. We thank the members of the Computational Auditory Perception group at the Max Planck Institute for Empirical Aesthetic, Andrea Estandía, Carys Jones, and especially Manuel Anglada-Tort for a very useful discussion of preliminary ideas and results around this work. We also thank all those who have contributed to the long-term nest box study in Wytham Woods and the collection of associated data. This work was supported by a Clarendon-Mary Frances Wagley Graduate Scholarship and an EGI scholarship to Nilo Merino Recalde, and made use of the University of Oxford Advanced Research Computing facility (Richards, 2015).

AUTHOR CONTRIBUTIONS

Nilo Merino Recalde: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Ben C. Sheldon**: Supervision, Project administration, Writing – review and editing, Funding acquisition.

REFERENCES

Anglada-Tort, M., Harrison, P. M. C., Lee, H., & Jacoby, N. (2023). Large-scale iterated singing experiments reveal oral transmission mechanisms underlying music evolution. *Current Biology*, *33*(8), 1472–1486.e12. https://doi.org/10.1016/j.cub.2023.02.070

Araya-Salas, M. (2012). Is Birdsong Music? Evaluating Harmonic Intervals in Songs of a Neotropical Songbird. *Animal Behaviour*, *84*(2), 309–313. https://doi.org/10.1016/j.anbehav.2012.04.038

Bowling, D. L., Sundararajan, J., Han, S., & Purves, D. (2012). Expression of Emotion in Eastern and Western Music Mirrors Vocalization. *PLOS ONE*, *7*(3), e31942. https://doi.org/10.1371/journal.pone.0031942

Brown, S., & Jordania, J. (2013). Universals in the world's musics. *Psychology of Music*, *41*(2), 229–248. https://doi.org/10.1177/0305735611425896

Buskell, A. (2017). What are cultural attractors? *Biol Philos*, *32*(3), 377–394. https://doi.org/10.1007/s10539-017-9570-6

Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. *Proceedings of the International Conference on Multimedia - MM '10*, 1467. https://doi.org/10.1145/1873951.1874248

Claidière, N., & Sperber, D. (2007). Commentary: The role of attraction in cultural evolution. *Journal of Cognition and Culture*, *7*(1-2), 89–111. https://doi.org/10.1163/156853707X171829

De Gregorio, C., Raimondi, T., Bevilacqua, V., Pertosa, C., Valente, D., Carugati, F., Bandoli, F., Favaro, L., Lefaux, B., Ravignani, A., & Gamba, M. (2023). Isochronous singing in 3 crested gibbon species (Nomascus spp.) *Current Zoology*, zoad029. https://doi.org/10.1093/cz/zoad029

De Gregorio, C., Valente, D., Raimondi, T., Torti, V., Miaretsoa, L., Friard, O., Giacoma, C., Ravignani, A., & Gamba, M. (2021). Categorical rhythms in a singing primate. *Current Biology*, *31*(20), R1379–R1380. https://doi.org/10.1016/j.cub.2021.09.032

deCastro-Arrazola, V., & Kirby, S. (2019). The emergence of verse templates through iterated learning. *Journal of Language Evolution*, *4*(1), 28–43. https://doi.org/10.1093/jole/lzy013

Demartsev, V., Haddas-Sasson, M., Ilany, A., Koren, L., & Geffen, E. (2023). Male rock hyraxes that maintain an isochronous song rhythm achieve higher reproductive success. *Journal of Animal Ecology*, *92*(8), 1520–1531. https://doi.org/10.1111/1365-2656.13801

Demery, A.-J. C., Burns, K. J., & Mason, N. A. (2021). Bill size, bill shape, and body size constrain bird song evolution on a macroevolutionary scale. *Ornithology*, *138*(2), ukab011. https://doi.org/10.1093/ornithology/ukab011

Dobson, C. W., & Lemon, R. E. (1977). Bird song as music. *The Journal of the Acoustical Society of America*, *61*(3), 888–890. https://doi.org/10.1121/1.381345

Doolittle, E. (2012). O Canto do Uirapuru: Consonant intervals and patterns in the song of the musician wren. *journal of interdisciplinary music studies*, (6-1). https://doi.org/10.4407/jims.2013.10.003

Doolittle, E. L., Gingras, B., Endres, D. M., & Fitch, W. T. (2014). Overtone-based pitch selection in hermit thrush song: Unexpected convergence with scale construction in human music. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(46), 16616–16621. https://doi.org/10.1073/pnas.1406023111

Falandays, J. B., & Smaldino, P. E. (2022). The Emergence of Cultural Attractors: How Dynamic Populations of Learners Achieve Collective Cognitive Alignment. *Cognitive Science*, *46*(8), e13183. https://doi.org/10.1111/cogs.13183

Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, *459*(7246), 564–568. https://doi.org/10.1038/nature07994

Frane, A. V. (2017). Swing Rhythm in Classic Drum Breaks From Hip-Hop's Breakbeat Canon. *Music Perception*, *34*(3), 291–302. https://doi.org/10.1525/mp.2017.34.3.291

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How Efficiency Shapes Human Language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Heyes, C. M. (2018). *Cognitive gadgets : The cultural evolution of thinking*. Harvard University Press.

Hill, A. P., Prince, P., Snaddon, J. L., Doncaster, C. P., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, *6*, e00073. https://doi.org/10.1016/j.ohx.2019.e00073

Hinde, R. A. (1952). The Behaviour of the Great Tit (Parus Major) and Some Other Related Species. *Behaviour. Supplement*, (2), III–201.

Hudson, E. J., & Creanza, N. (2022). Modeling how population size drives the evolution of birdsong, a functional cultural trait. *Evolution*, *76*(6), 1139–1152. https://doi.org/10.1111/evo.14489

Jacoby, N., & McDermott, J. H. (2017). Integer Ratio Priors on Musical Rhythm Revealed Cross-culturally by Iterated Reproduction. *Current Biology*, *27*(3), 359–370. https://doi.org/10.1016/j.cub.2016.12.031

Jacoby, N., Polak, R., Grahn, J., Cameron, D. J., Lee, K. M., Godoy, R., Undurraga, E. A., Huanca, T., Thalwitzer, T., Doumbia, N., Goldberg, D., Margulis, E., Wong, P. C. M., Jure, L., Rocamora, M., Fujii, S., Savage, P. E., Ajimi, J., Konno, R., … McDermott, J. H. (2021). Universality and cross-cultural variation in mental representations of music revealed by global comparison of rhythm priors. *PsyArXiv*. https://doi.org/10.31234/osf.io/b879v

James, L. S., & Sakata, J. T. (2017). Learning Biases Underlie "Universals" in Avian Vocal Sequencing. *Current Biology*, *27*(23), 3676–3682. https://doi.org/10.1016/j.cub.2017.10.019

Jarvis, E. D. (2019). Evolution of vocal learning and spoken language. *Science*, *366*(6461), 50–54. https://doi.org/10.1126/science.aax0287

Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin and Review*, *24*(1), 118–137. https://doi.org/10.3758/s13423-016-1166-7

Kotz, S. A., Ravignani, A., & Fitch, W. T. (2018). The Evolution of Rhythm Processing. *Trends in Cognitive Sciences*, *22*(10), 896–910. https://doi.org/10.1016/j.tics.2018.08.002

Kuroyanagi, J., Sato, S., Ho, M.-J., Chiba, G., Six, J., Pfordresher, P., Tierney, A., Fujii, S., & Savage, P. (2019). Automatic comparison of human music, speech, and bird song suggests uniqueness of human scales. *Proceedings of the 9th International Workshop on Folk Music Analysis (FMA 2019)*, 35–40.

Lachlan, R. F., Verhagen, L., Peters, S., & ten Cate, C. (2010). Are There Species-Universal Categories in Bird Song Phonology and Syntax? A Comparative Study of Chaffinches (Fringilla coelebs), Zebra Finches (Taenopygia guttata), and Swamp Sparrows (Melospiza georgiana). *Journal of Comparative Psychology*, *124*(1), 92–108. https://doi.org/10.1037/a0016996

Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, *9*(1), 2417. https://doi.org/10.1038/s41467-018-04728-1

Lack, D. (1964). A Long-Term Study of the Great Tit (Parus major). *Journal of Animal Ecology*, *33*, 159–173. https://doi.org/10.2307/2437

Mace, R. (1987). The dawn chorus in the great tit Parus major is directly related to female fertility. *Nature*, *330*(6150), 745–746. https://doi.org/10.1038/330745a0

McInnes, L., Healy, J., & Astels, S. (2017). Hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205

McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *The Journal of Open Source Software*, *3*(29), 861.

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., Hopkins, E. J., Howard, R. M., Hartshorne, J. K., Jennings, M. V., Simson, J., Bainbridge, C. M., Pinker, S., O'Donnell, T. J., Krasnow, M. M., & Glowacki, L. (2019). Universality and diversity in human song. *Science*, *366*(6468), eaax0868. https://doi.org/10.1126/science.aax0868

Merino Recalde, N. (2023). Pykanto: A python library to accelerate research on wild bird song. *Methods in Ecology and Evolution*, *00*, 1–9. https://doi.org/10.1111/2041-210X.14155

Merino Recalde, N., Estandía, A., Pichot, L., Vansse, A., Cole, E. F., & Sheldon, B. C. (2023). A densely sampled and richly annotated acoustic dataset from a wild bird population. *bioRxiv*. https://doi.org/10.1101/2023.07.03.547484

Muller, M. (2016). *Fundamentals of music processing: Audio, analysis, algorithms, applications*. Springer International Publishing.

Norton, P., & Scharff, C. (2016). "Bird song metronomics": Isochronous organization of zebra finch song rhythm. *Frontiers in Neuroscience*, *10*(JUL). https://doi.org/10.3389/fnins.2016.00309

Nowicki, S., Peters, S., & Podos, J. (1998). Song Learning, Early Nutrition and Sexual Selection in Songbirds. *American Zoologist*, *38*(1), 179–190. https://doi.org/10.1093/icb/38.1.179

Parncutt, R. (1994). A Perceptual Model of Pulse Salience and Metrical Accent in Musical Rhythms. *Music Perception: An Interdisciplinary Journal*, *11*(4), 409–464. https://doi.org/10.2307/40285633

Pichkar, Y., Searfoss, A. M., & Creanza, N. (2023, January). Detecting cultural evolution in a songbird species using community-science data and computational modeling. https://doi.org/10.1101/2023.01.23.525255

Polak, R., Jacoby, N., Fischinger, T., Goldberg, D., Holzapfel, A., & London, J. (2018). Rhythmic Prototypes Across Cultures. *Music Perception: An Interdisciplinary Journal*, *36*(1), 1–23. https://doi.org/10.1525/mp.2018.36.1.1

Raimondi, T., Di Panfilo, G., Pasquali, M., Zarantonello, M., Favaro, L., Savini, T., Gamba, M., & Ravignani, A. (2023). Isochrony and rhythmic interaction in ape duetting. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1990), 20222244. https://doi.org/10.1098/rspb.2022.2244

Ravignani, A., Bowling, D., & Fitch, W. T. (2014). Chorusing, synchrony and the evolutionary functions of rhythm. *Frontiers in Psychology*, *5*(SEP). https://doi.org/10.3389/fpsyg.2014.01118

Ravignani, A., Delgado, T., & Kirby, S. (2017). Musical evolution in the lab exhibits rhythmic universals. *Nature Human Behaviour*, *1*(1), 1–7. https://doi.org/10.1038/s41562-016-0007

Ravignani, A., Thompson, B., Lumaca, M., & Grube, M. (2018). Why Do Durations in Musical Rhythms Conform to Small Integer Ratios? *Front Comput Neurosci*, *12*, 86. https://doi.org/10.3389/fncom.2018.00086

Richards, A. (2015, August). *University of oxford advanced research computing*. Manual. https://doi.org/10.5281/zenodo.22558

Roeske, T. C., Tchernichovski, O., Poeppel, D., & Jacoby, N. (2020). Categorical Rhythms Are Shared between Songbirds and Humans. *Current Biology*, *30*(18), 3544–3555.e6. https://doi.org/10.1016/j.cub.2020.06.072

Rouse, A. A., Patel, A. D., & Kao, M. H. (2021). Vocal learning and flexible rhythm pattern perception are linked: Evidence from songbirds. *Proceedings of the National Academy of Sciences*, *118*(29), e2026130118. https://doi.org/10.1073/pnas.2026130118

Sainburg, T., Theilman, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11605-y

Saldana, C., Fagot, J., Kirby, S., Smith, K., & Claidière, N. (2019). High-fidelity copying is not necessarily the key to cumulative cultural evolution: A study in monkeys and children. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2019.0729

Savage, P. E. (2019). Cultural evolution of music. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0221-1

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences*, *112*(29), 8987–8992. https://doi.org/10.1073/pnas.1414495112

Sierro, J., de Kort, S. R., & Hartley, I. R. (2023). Sexual selection for both diversity and repetition in birdsong. *Nat Commun*, *14*(1), 3600. https://doi.org/10.1038/s41467-023-39308-5

Silvey, C., Kirby, S., & Smith, K. (2019). Communication increases category structure and alignment only when combined with cultural transmission. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2019.104051

Spencer, K. A., Buchanan, K. L., Goldsmith, A. R., & Catchpole, C. K. (2003). Song as an honest signal of developmental stress in the zebra finch (Taeniopygia guttata). *Hormones and Behavior*, *44*(2), 132–139. https://doi.org/10.1016/S0018-506X(03)00124-7

Suthers, R. A. (2004). How birds sing and why it matters. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 272–295). https://doi.org/10.1016/B978-012473070-0/50012-8

Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, *8*, 37–43. https://doi.org/10.1016/j.copsyc.2015.09.003

Tchernichovski, O., Feher, O., Fimiarz, D., & Conley, D. (2017). How social learning adds up to a culture: From birdsong to human public opinion. *Journal of Experimental Biology*, *220*(1), 124–132. https://doi.org/10.1242/jeb.142786

Tierney, A. T., Russo, F. A., & Patel, A. D. (2008). Empirical comparisons of pitch patterns in music, speech, and birdsong. *The Journal of the Acoustical Society of America*, *123*(5_Supplement), 3721–3721. https://doi.org/10.1121/1.2935184

Tierney, A. T., Russo, F. A., & Patel, A. D. (2011a). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences*, *108*(37), 15510–15515. https://doi.org/10.1073/pnas.1103882108

Tierney, A. T., Russo, F. A., & Patel, A. D. (2011b). The motor origins of human and avian song structure. *Proceedings of the National Academy of Sciences*, *108*(37), 15510–15515. https://doi.org/10.1073/pnas.1103882108

Trehub, S. E. (2015). Cross-cultural convergence of musical features. *Proceedings of the National Academy of Sciences*, *112*(29), 8809–8810. https://doi.org/10.1073/pnas.1510724112

van der Aa, J., Honing, H., & ten Cate, C. (2015). The perception of regularity in an isochronous stimulus in zebra finches (Taeniopygia guttata) and humans. *Behavioural Processes*, *115*, 37–45. https://doi.org/10.1016/j.beproc.2015.02.018

van Rossum, G. (1995, January). Python reference manual.

Verga, L., Sroka, M. G. U., Varola, M., Villanueva, S., & Ravignani, A. (2022). Spontaneous rhythm discrimination in a mammalian vocal learner. *Biology Letters*, *18*(10), 20220316. https://doi.org/10.1098/rsbl.2022.0316

Verhoef, T., & Ravignani, A. (2021). Melodic Universals Emerge or Are Sustained Through Cultural Evolution. *Frontiers in Psychology*, *12*.

Verzijden, M. N., ten Cate, C., Servedio, M. R., Kozak, G. M., Boughman, J. W., & Svensson, E. I. (2012). The impact of learning on sexual selection and speciation. *Trends in Ecology and Evolution*, *27*(9), 511–519. https://doi.org/10.1016/j.tree.2012.05.007

Von Hippel, P. (2000). Redefining Pitch Proximity: Tessitura and Mobility as Constraints on Melodic Intervals. *Music Perception*, *17*(3), 315–327. https://doi.org/10.2307/40285820

Von Hippel, P., & Huron, D. (2000). Why Do Skips Precede Reversals? The Effect of Tessitura on Melodic Structure. *Music Perception: An Interdisciplinary Journal*, *18*(1), 59–85. https://doi.org/10.2307/40285901

Xing, J., Sainburg, T., Taylor, H., & Gentner, T. Q. (2022). Syntactic modulation of rhythm in Australian pied butcherbird song. *Royal Society Open Science*, *9*(9), 220704. https://doi.org/10.1098/rsos.220704

Youngblood, M., & Lahti, D. C. (2022). Content bias in the cultural evolution of house finch song. *Animal Behaviour*, *185*, 37–48. https://doi.org/10.1016/j.anbehav.2021.12.012

SUPPLEMENTARY INFORMATION

(Next page)

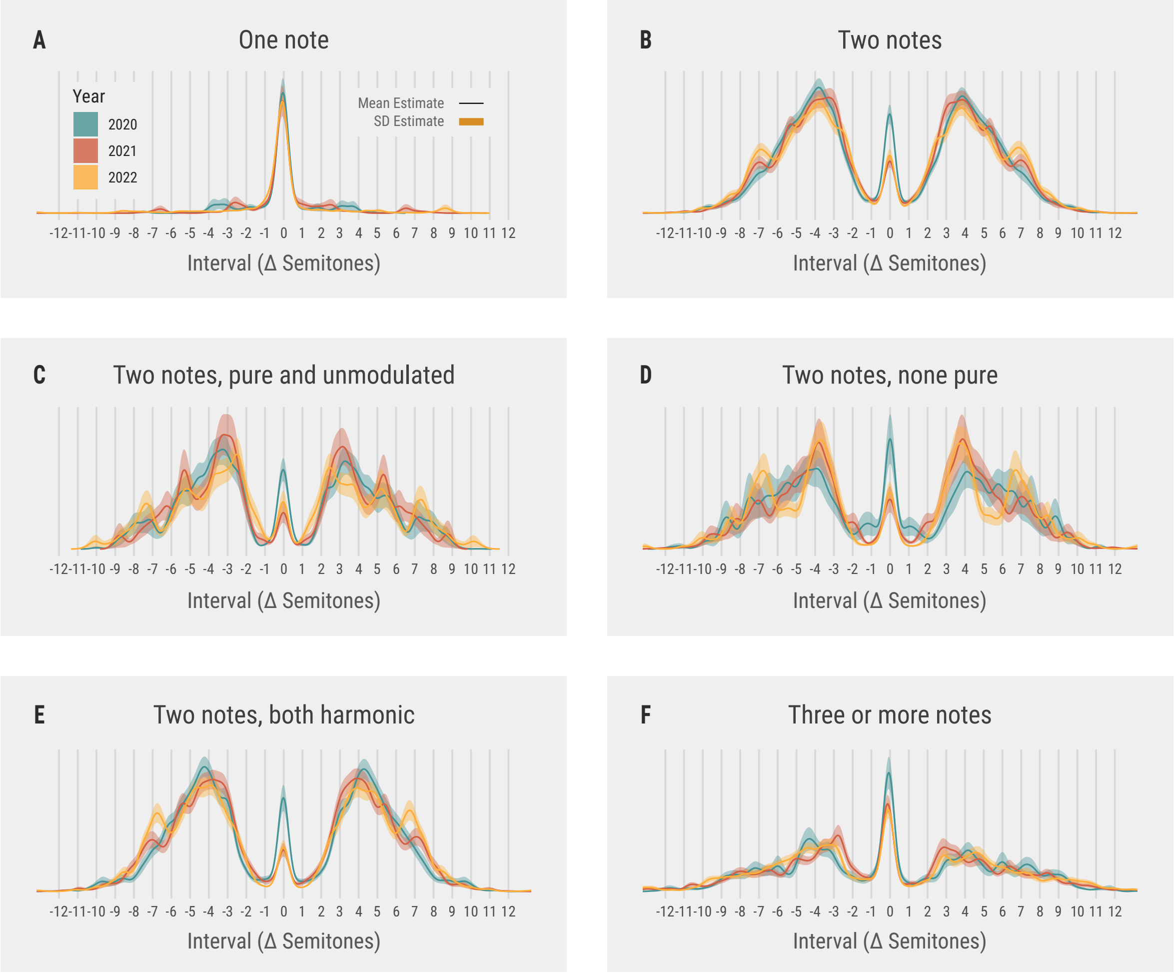


Figure S5.1. Kernel density estimates of the distribution of melodic intervals across a typology of songs in different years. In this figure, we categorize songs into distinct types based on the number of notes they contain and their spectral properties and plot the estimated density distribution of melodic intervals, measured in semitones and spanning the octave (ascending and descending). Colours represent separate years in our study.

(A) One note: Songs consisting of a single note repeated rhythmically. (B) Two notes: All songs composed of two different, alternating notes. (C) Two notes, pure and unmodulated: Songs with two notes, each consisting of a single stable, pure, unmodulated tone. (D) Two notes, none pure: All two-note songs where neither note is pure and unmodulated. (E) Two notes, both harmonic: Songs featuring two notes, both of which have harmonic components. (F) Three or more notes: Songs comprising three or more notes. There are few three-note melodies, so we did not split these by spectral or modulation properties.

6. GENERAL DISCUSSION

SUMMARY OF THE THESIS

Bird song captivates human imagination. It has influenced music, literature, and inspired extensive research into its neurobiology, production, and function. Today, the study of bird song is experiencing a renaissance driven by new technological and theoretical tools. Advances in acoustic data collection, signal processing, and machine learning have made it feasible to collect and analyse vast datasets at lower costs. At the same time, insights from cultural evolution theory provide new conceptual tools that are allowing a richer understanding of diversity and variation in bird song, understood as a cultural phenomenon. This thesis tries to take advantage of both and contribute to our understanding of vocal behaviour and cultural dynamics within natural populations of birds. It makes contributions across four areas that are of significance for the research process (chapters 2 and 3) and bird song cultures (chapters 4 and 5).

When I started this project, I, very naively, in hindsight, thought that the methodological tools I would need were available. As I built enclosures for the audio recorders and planned the first field season, I began trying different workflows around existing tools—which was difficult given how little I knew about signal processing and programming—and nothing seemed to be quite right for the task, especially given just how much data I was going to try to collect. It was not planned at first, but as I went on, some of the tools and ideas that I was learning coalesced into a software library, which is now published and has already been used by other researchers. This is introduced in Chapter 2.

I knew that the kinds of questions about cultural change and diversity that I wanted to ask required large amounts of data, beyond what has typically been the case in fields such as behavioural ecology—it is simply difficult to collect lots of high-quality bird songs from individual birds in their natural habitats, so it is rarely done. During approximately nine months in the field, I, with the invaluable help of fantastic masters students Loanne Pichot in 2021 and Antoine Vansse in 2022, collected and later annotated over 100,000 songs from hundreds of birds. As part of this process, I was also lucky to contribute to the collective long-term monitoring effort in Wytham Woods, which involved ringing and measuring traits for around 2,000 great and blue tits over three years. Now, although my personal interests lie in cultural evolution, I know that there are many open research questions—about repeatability and stability, vocal individuality, and the connections between vocal performance, diversity, and fitness, to name a few—where research is hindered by the fact that individual-level datasets of songs from wild birds tend to be small and hard to come by. So, I made publishing all the song data and associated information on individual life histories a priority, hoping that others will benefit from my effort, too. Chapter 3 consists of a detailed description of the dataset and how I generated it.

Then, I use this extensive dataset to explore two processes, or rather, combinations of processes, which influence cultural change and diversity in animal groups:

The first of these concerns the demographic composition and structure of populations, which, although in a way extrinsic to culture, has been theorized to play a central role in cultural evolution. In Chapter 4, I assess the strength of associations between variation in the demographic composition of neighbourhoods and cultural outcomes. I present evidence showing how individual-level differences in repertoire size and diversity can influence emergent group-level cultural dynamics. The results emphasize the need for empirical and modelling studies of cultural change to account for the demographic characteristics of populations, as well as their inherent heterogeneity across time and space: They shape individuals' learning opportunities, and this, in turn, can have large impacts on cultural diversity and turnover.

A second set of processes is intrinsic to the cognitive and motor abilities that support social learning and the formation of cultures. In the preceding chapter we saw that cultural turnover, the progressive replacement of song types, is very high at short temporal scales. And yet, this highly polymorphic culture is stable on larger spatial and temporal scales. In the final chapter, Chapter 5, I use ideas from the study of musical and linguistic evolution and find preliminary evidence that, despite their diversity, great tit songs use fewer rhythmic and melodic patterns than would be expected by chance. These categories likely result from motor, information-processing, or perceptual biases, and, paralleling their role in human music, they might serve to stabilize cultural change—even in the face of high turnover.

FUTURE DIRECTIONS

The week I began to get the hang of this and, for the first time, felt like I had learned enough to begin this research—that very same week I found myself facing the thesis submission deadline. That is to say that each of these chapters, and particularly the last two that present empirical findings, feels like an initial exploration. There is ample room for improvement, a few clear next steps, and some aspects of bird song culture that, while important, this thesis has not explored.

In Chapter 4, I developed a method for re-identifying birds based on their songs. This method uses deep metric learning with a visual transformer model and has proven surprisingly effective: It enabled the identification of birds that could not be physically identified but were present in the vocal dataset across different years. Given its practical applications, I intend to conduct more thorough testing and create a more streamlined pipeline to make it accessible to other researchers.

Chapter 5 is still very much a work in progress. I started thinking about this only recently, but I think it holds some potential. I plan to use data from citizen science platforms to test if rhythmic and melodic structures are stable over very large spatial scales—if they can vary, this would represent evidence that cognitive biases can be socially acquired and evolve culturally—and to develop more formal tests of these ideas.

Although we compiled a large dataset containing both song recordings and individual-level fitness metrics (see Chapter 3), I did not carry out any explicit research on the potential adaptive or selective consequences of cultural diversity or signal structure. Fitness is a complex trait influenced by many factors, such as environmental conditions, resource availability, predation pressure, territory quality, competition, and individual condition. Disentangling any specific fitness effects of cultural variation in songs from these confounding factors is extremely difficult, especially in a natural population where many variables cannot be controlled and there may be a bidirectional relationship between songs and individual fitness. Research attempting to do this often tests for correlations between a host of song characteristics and some fitness outcome, interpreting any significant results as evidence for direct selection: avoiding this problematic approach would require at the very least a full thesis of its own. So, instead, I chose to concentrate on investigating other, understudied processes shaping cultural diversity and turnover within the population, as well as the structure of the songs themselves. Why? Neutral or non-adaptive change is an important aspect of cultural evolution, and it happens to align most closely with my personal research interests. Be that as it may, exploring the potential adaptive significance of cultural diversity, turnover, and song structure—especially in the context of sexual selection—remains a logical next step for future research on this system.

The thesis presents evidence that demographic factors, such as population turnover, immigration, and age structure, can impact the diversity and pace of change in animal vocal cultures. However, cultural traits themselves may also influence demographic processes. For example, culturally transmitted migratory routes or feeding techniques may affect survival and reproductive success, shaping population growth and structure. Additionally, cultural preferences for certain mates or behaviours can influence patterns of dispersal and gene flow, further shaping demographic dynamics. Learning and song-sharing can also reinforce or modifying existing social structures by affecting mate choice, territorial behaviour, or social hierarchies, which in turn can influence the transmission and maintenance of cultural traits.

Similarly, cognitive and motor biases, such as perceptual or production constraints, are known to influence the structure and stability of cultural traits like bird songs. But, slightly annoyingly, cultural traits themselves could potentially shape or reinforce these biases. For example, which cultural variants are learned could influence the development and tuning of auditory perception or vocal motor skills in individuals while, at the same time, songs that align with existing biases may be more readily acquired and maintained, reinforcing or amplifying those biases over generations.

The potential for these recursive interactions between culture and the factors that shape it has important implications for our understanding of cultural evolution: it suggests that cultural processes cannot be studied in isolation but must be considered within a broader ecological and evolutionary context where culture may act as a source of evolutionary feedback, shaping the conditions that initially gave rise to it. Studying these feedback loops empirically will require conducting research over different timescales and using approaches different from those in this thesis, as well as integrating demographic, genetic, cognitive, and cultural data.

In this vein, an early and more humble goal of this thesis, which I never had time to fulfil (why does everything take so long?) was to integrate our knowledge about demographic processes in the Wytham great tit population with insights on structural bias in songs within the same modelling framework. This would allow us to estimate how they, together and separately, affect estimates of frequency-dependent learning biases and learning fidelity. Who knows what will come next, and if I will ever get around to doing any of this, but if you are reading this and are interested, please reach out.

* * *

Singing great tit male perched on a hawthorn in bloom. Vector Illustration. This thesis was typeset using LaTeX, developed by Leslie Lamport based on Donald Knuth's TeX. By Nilo Merino Recalde, 2023