Plasticity, personality and stability: hidden levels of variation in bird song

Dissertation
Fakultät für Biologie
Ludwig-Maximilians-Universität München

durchgeführt an
der Fakultät für Biologie und dem Max-Planck-Institut für Ornithologie
Seewiesen

vorgelegt von
Alexander Hutfluss
München, August 2021

Doktorvater: **Prof. Dr. Niels Dingemanse**

Erstgutachter: **Prof. Dr. Niels Dingemanse**

Zweitgutachter: **Prof. Dr. Joachim Haug**

Eingereicht am: **09.08.2021**

Tag der mündlichen Prüfung: **01.02.2022**

Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass meine Dissertation selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist.
Die vorliegende Dissertation wurde weder ganz, noch teilweise bei einer anderen Prüfungskommission vorgelegt.
Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder an einer Doktorprüfung teilzunehmen.

München, den 03.03.2022 Alexander Hutfluss
(Unterschrift)

Table of Contents

Summary 5

General Introduction 9

Hierarchical levels of variation 9
Study System 15
Thesis Outline 18

General Discussion 21

Trade-offs 23
Repeatability in the mean 26
"Repeatability" in behavioural variability 28
Conclusions 29
References 31

Appendices 47

Publication I: Great tits responding to territorial intrusions sing less but alarm more on colder days 48

Publication II: Does song overlap signal aggressiveness? An experimental study with repeated measures in free-ranging great tits 71

Manuscript I: Within-individual stability in bird song shows cross-year repeatability and is affected by food supplementation in a wild passerine bird 110

Manuscript II: Song stability neither predicts within-pair reproductive success nor extra-pair paternity gain 148

Acknowledgments 186

Author contributions 188

Addresses of co-authors 189

Fehler! Textmarke nicht definiert.

Summary

Behavioural variation is distinctly and hierarchically structured, resulting in different “levels of variation”. These levels include species, genotype and the individual level. Recent focus has been specifically on the latter, showing variation among and within individuals, both in the average behaviour and the variance of repeatedly expressed traits. Importantly, variation at any of these levels has been shown to be biologically important and can be under selection if genetically underpinned. However, their respective effects have not been studied in an overarching integrative matter. The overall aim of my PhD project was thus to examine different levels of variation in conjunction from an adaptive perspective. Using the song of great tits (*Parus major*) as an ideal behavioural context, I studied to what extent individuals adjusted their behaviour to external factors, i.e. within-individual variation or plasticity, and to what extent they show consistent differences, i.e. among-individual variation or personality. I put a specific focus on these patterns in the variance of behaviour, i.e. the stable production of song traits.

In **Chapter 1,** we investigated how temperature affected two acoustic territory defence behaviours, singing and alarm calling, that were assumed to differ in energetic costs and effectiveness. Effects of temperature on singing have been studied frequently to indirectly assess the energetic costs of bird song. However, such research often neglected other alternative behaviours. Using an 8-year dataset with over 6,000 simulated territory intrusion, we found both traits to be affected by ambient but not night temperatures. Importantly, increasing temperatures were associated with increasing song rates but decreasing alarm calling. These changes were, however, not associated with a change in overall acoustic output. When temperature was lower and energy required to maintain homeostasis higher, birds shifted to alarm calls as an alternative acoustic trait. Combined with the strong preference for singing as initial and main territory defence behaviour, this indicated differences in the energetic costs and/or effectiveness of the two song traits. This Chapter thus highlights the need to study alternative behaviours simultaneously to get a fuller insight into such behavioural patterns.

Chapter 2 explored whether song overlapping signals aggressiveness; a function that is frequently discussed but remains unresolved. Besides the general signal value, we further examined whether song overlapping represents a motivational signal or a personality-related signal based on the extent of within- and among-individual effects of aggression. Overall, we found a negative association between aggression and overlapping, driven exclusively by within-individual effects. That means that increases in aggression led to decreasing overlapping rates.

In addition, song overlapping showed only little consistent individual differences, indicating that this trait is neither personality-related nor repeatable. Finally, the majority of all birds overlapped less than expected. This contradicts the notion that overlapping signals aggression. Instead, it indicates that overlapping signals "non-engagement" or that there is a general interference avoidance.

In **Chapter 3**, we investigated patterns of variation in the stability of bird song, focusing specifically on within- and among- individual variation. Behavioural ecologists have only recently focused on the idea that the stable expression of behaviours can be biologically important. Although research on bird song has acknowledged and studied this level of variation earlier, proper quantifications of the within- and among-individual variation are scarce. Using state-of-the-art double hierarchical mixed effect models, we found consistent individual differences in song stability ("individuality in stability") for minimum frequency but not phrase length. This indicates that the minimum frequencies were more stable than the ones produced by others. Such differences could indicate that song stability functions as a quality-indicator trait and that it might be under selection if it has a genetic basis. We further found strong within-individual variation in stability ("plasticity in stability") because of seasonal variation or in response to food supplementation. Overall, the findings of this chapter strongly indicate that stability is a biologically important trait, although costs and benefits shaping this variation seem to be highly trait-specific.

Chapter 4 explored the selective potential of individual differences in song stability. Assuming that song stability is a quality-indicator trait that is costly yet beneficial to produce, stable singing should be under selection. We, therefore, explored the links between the stable production of minimum frequency and different reproductive success parameters. Importantly, we distinguished three major pathways by which stable singers could increase their fitness: 1) Increasing within-pair success, e.g. because of mating a high-quality female or occupying high-quality territories, 2) Reducing within-pair paternity losses 3) Increasing extra-pair paternity gains. Overall, we found no evidence for selection acting on song stability via any of these pathways. Instead, we found firm support for an actual absence of selection, at least for some parameters. These findings imply that song stability is either not used by females as the basis of mate choice or not a signal of male quality in the first place.

Overall, my thesis forcefully demonstrates the importance of studying different levels of behavioural variation and the insights that this approach can reveal. Besides the already established relevance of partitioning within- and among-individual variation in the mean, we demonstrate the potential importance of studying behavioural variation, especially in bird song.

We propose that studying multiple levels of variation simultaneously should become standard procedure. This could then lead to more insights into patterns of variation and mechanisms, e.g. trade-offs between different traits or variance components. Importantly, I addressed these questions in a comparably small set of song features and only within one species. However, the benefits of using similar approaches should be seen as applicable not only in animal signalling and communication, but in various other labile traits and contexts as well.

General Introduction

Hierarchical levels of variation

The primary aim of behavioural ecology is to understand the adaptive nature of behavioural variation in natural populations (Westneat & Fox, 2010). To achieve this, behavioural ecologists use a variety of methods, from mathematical modelling to experimental studies in the wild, often focusing on addressing the third of Tinbergen's four questions, the trait's adaptive function (Tinbergen, 1963). Historically, behavioural ecologists have often tested their theories by exposing the same individual to multiple conditions (i.e., within-subject designs), thus focusing on one specific variance component: within-individual variance. This source of variation stems from (reversible) within-individual plasticity or "flexibility", representing the degree by which an individual adjusts its behaviour to environmental change experienced within its lifetime (Duckworth, 2006; Sih, 2004). Importantly, despite the obvious adaptive benefits of "flexible" adjustments to micro-environmental change (C. K. Ghalambor, McKay, Carroll, & Reznick, 2007; Nussey, Wilson, & Brommer, 2007; Piersma & Drent, 2003), this form of **reversible phenotypic plasticity** is not without costs and limits (DeWitt, Sih, & Wilson, 1998), and therefore will only evolve if net beneficial. Examples of costs or limits include the development and maintenance of the sensory systems required to assess the environmental changes and necessary behavioural adjustments or the potential mismatches between optimal and realized responses (Auld, Agrawal, & Relyea, 2010). Importantly, plasticity is not the only level of variation that can be adaptive (Westneat, Wright, & Dingemanse, 2015) (see Figure 1). Contemporary behavioural ecology research instead focuses on multiple levels of variation in conjunction, such as within- *and* among-individual variation.

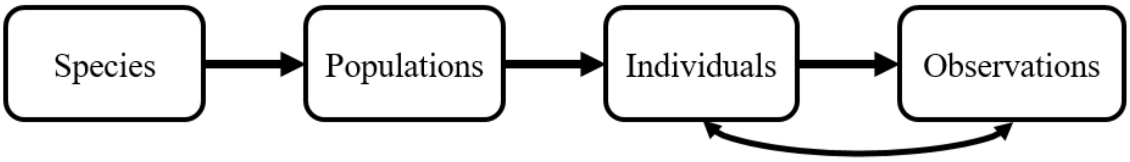


Figure 1: Hierarchical structuring of the levels of phenotypic variation: variation in behaviour exists among species, among populations of the same species, among individuals of the same population, and for repeatedly expressed behaviours, among instances. There are also interactions between levels, exemplified here between individuals and instances, where interactions exist because individuals differ in phenotypic plasticity (individual-by-environment

interaction or I×E). Such interactions can also exist among other levels (not illustrated), for example, the amount of individual variation can differ between populations.

A specific focus of contemporary research is the emergence of among-individual variation in repeatedly expressed behaviours. Those are behaviours that vary substantially within-individuals (Fig. 1). Interestingly, behavioural ecologists have long viewed among-individual variation in such behaviours as merely representing random "noise" around an adaptive population-level mean, or the "raw" material for selection to act upon (Dall, Houston, & McNamara, 2004a; Wilson, 1998). Over the last decade, however, repeatable variation among individuals has been increasingly viewed from an adaptive perspective, particularly since the meta-analytical discovery that as much as 40% of the variation in behaviour is, on average, explained by variation among rather than within individuals (Bell, Hankison, & Laskowski, 2009; Réale, Reader, Sol, McDougall, & Dingemanse, 2007). While classic behavioural ecology paradigms assumed that functionally distinct behavioural axes can be studied in isolation (Coleman & Wilson, 1998), recent studies imply that this assumption is normally invalid: individuals often differ in entire suites of correlated traits (so-called "behavioural syndromes"; Sih et al. 2004; Bell 2007) similar to how humans differ in personality (McCrae, Costa, John, Robins, & Pervin, 1999). As a result, a burgeoning sub-field of behavioural ecology focuses on understanding individuality under the flagship name of "**animal personality**“ research (Réale et al., 2010). Over the past two decades, research on animal personality has generated a suite of novel adaptive theory, explaining why also repeatable among-individual variation might be adaptive and maintained by selection (Dingemanse and Wolf 2010; reviewed by Wolf and Weissing 2010).

Adaptive personality research has, over the past few years, also led to a suite of adaptive models predicting that natural selection may also maintain a mix of individuals differing in plasticity, "flexibility" or "responsiveness" (Wolf et al. 2008; McNamara et al. 2009; Wolf et al. 2011), i.e. the interaction between levels illustrated by double-headed arrows in Figure 1. Evolutionary biologists have long recognized that phenotypic plasticity can only evolve when the population exhibits heritable individual variation in plasticity (Nussey et al., 2007). As above, contemporary behavioural ecology theory now suggests that variation in this form of individuality (called **individual plasticity**; or individual by environment interactions, I×E) can also be adaptive. Individual plasticity can exist in the context of social interactions ("social responsiveness") or responses to non-social environmental factors (Wolf et al. 2008; Wolf et al. 2011; Dingemanse and Wolf 2013; Snell-Rood 2013). In addition, some adaptive theory predicts that selection can favour the integration of plasticity and personality (Max Wolf et al., 2008): individuals with high levels of behavioural expression may, for example, also be less plastic (Mathot, Wright, Kempenaers, & Dingemanse, 2012). This phenomenon is called **personality-related plasticity**, and has been documented in birds (Dingemanse et al. 2012), fish (Dingemanse et al. 2012), and mammals. For example, aggressive mice do not adjust their behaviour to social context, whereas non-aggressive conspecifics do (Natarajan, De Vries, Saaltink, De Boer, & Koolhaas, 2009).

Among and within individual variation (i.e. personality and plasticity), and their interaction (personality-related plasticity), can be studied in conjunction when one applies a reaction norm framework to the study of behaviour (Niels J. Dingemanse, Kazem, Réale, & Wright, 2010a). Doing so, views behaviour as a "norm of reaction", in its simplest form represented by a linear regression describing the phenotype as a function of an environmental gradient. The intercept of this line is representing an individual's response in the average environment (achieved by mean-centring the environmental gradient), whereas the slope represents the strength in unit response to unit change in the environment (i.e. plasticity; Figure 2A).

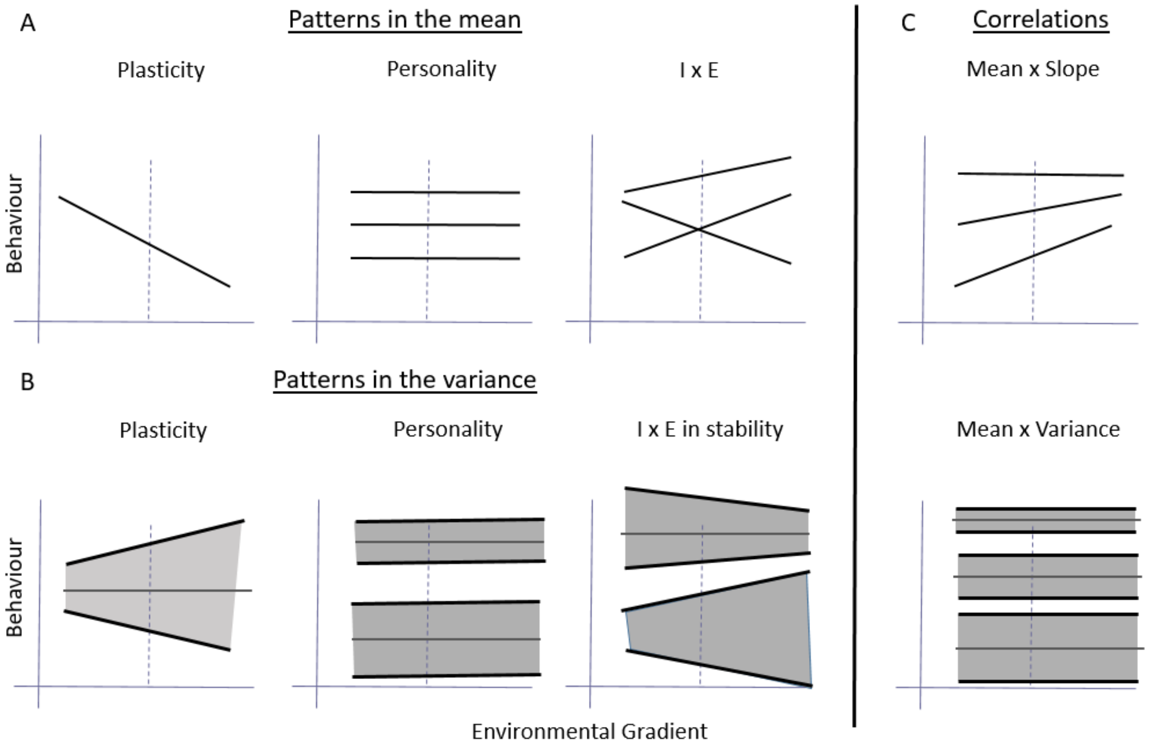


Figure 2: Schematic overview of the different levels of variation. A: variation in average behaviour, where black lines indicate the mean trait expression for different individuals (lines).

Left: representing plastic changes along an environmental gradient where all individuals are plastic but share exactly the same reaction norm. Middle: No plasticity but individual differences. Right: Individuals differ in the strength of their plastic response. B: Variation in stability, where grey lines indicate an individual's mean trait level and black lines and grey areas delimit the individual's residual within-individual variance (RWV, i.e., residual 'stability'). Left: Plasticity in stability, but no individual differences. Middle: Individuality in stability but no plasticity. Right: Individuality in overall stability and in the plastic changes in stability. C: Correlations between different levels of variation. Top: Personality × Plasticity correlations, here indicating that individuals with higher trait levels are less plastic. Bottom: Personality × Stability, here indicating that individuals with higher average trait levels are also producing a more stable phenotype.

More recently, behavioural ecologists have focused on the notion that an individual's **stability** or predictability, represented by its level of residual-within individual variation ("RWV"), and the ability to plastically adjust its stability, can also be adaptive (Cleasby & Nakagawa, 2011; Westneat et al., 2015). In general, RWV measures how much the observations of phenotypes deviate from an individual's average reaction norm. Classic behavioural ecology theory has long implied such heterogeneous variance can be adaptive but empirical work focussed so far on specific "variance sensitivity" contexts (Stephens, Brown, & Ydenberg, 2008); in other biological contexts, it has been the norm to assume homogeneous residual variation (Cleasby, Nakagawa, & Schielzeth, 2015). Variance sensitivity means that individuals assess differences in underlying variance, e.g. in food availability, or the consequences of this variance and alter their behaviour and thus their residual variance (Westneat, Schofield, & Wright, 2013). A prime example of variance-sensitivity is foraging, where parents adjust provisioning patterns in response to food availability and nestling demands (Mathot et al., 2017; Westneat, Mutzel, Bonner, & Wright, 2017; Westneat et al., 2013). Besides this concept, the idea of the potential adaptive nature of stability is now increasingly applied to other contexts. For example, animals might show decreased predictability in anti-predator behaviour, e.g. shelter emergence or movement patterns, during high-risk situations, as this increases survival rates when predators are present (Brembs, 2011; Briffa, 2013; Jones, Jackson, & Ruxton, 2011). In non-threat situations, behavioural predictability should be plastically increased to avoid deviations from the optimal response (Jones & Godin, 2010; Ydenberg, 1986). More generally, individuals should show increased stability in contexts that involve and benefit from the exchange of reliable information about an individual's quality or behaviour, e.g. male-male competition and female choice (Sakata, Hampton, & Brainard, 2008; Schuett, Tregenza, & Dall, 2010). Specifically, if stable trait expression is costly, individuals might use behavioural stability to assess a conspecific's quality or expected behaviour, e.g. aggression or parental care (Royle, Schuett, & Dall, 2010). All three mechanisms strongly imply that the ability to adjust behavioural stability, i.e. "**plasticity in stability**" should be selectively advantageous.

Similar to the patterns of variation in the mean, plasticity in stability has been of key interest in contemporary studies of RWV. In addition, individuals can be repeatable in RWV, even after accounting for plasticity and personality. In short, some individuals are predictably more stable/predictable than others. Among-individual variation in stability, or "**individual stability**", has primarily been reported in humans (Hoffman, 2007; MacDonald, Nyberg, & Bäckman, 2006). Over the past decade, however, individual plasticity is also increasingly studied in non-human models (Biro & Adriaenssens, 2013; Hertel, Royauté, Zedrosser, & Mueller, 2021; Jolles, Briggs, Araya-Ajoy, & Boogert, 2019; Stamps, Briffa, & Biro, 2012; Westneat et al., 2013). Importantly, behavioural stability can also be heritable, implying that it can evolve in response to selection (Martin et al. 2017; Prentice et al. 2020). A framework for the existence and maintenance of individual differences in stability has long been lacking, and I therefore borrow here from concepts developed to understand variation in the mean phenotypes. For example, directional selection should be expected to reduce genetic and individual variation in RWV. A likely explanation for the existence of individual differences are costs associated with plasticity in stability, similar to those in other forms of plasticity (Auld et al., 2010). Such costs could lead to among-individual variation if individuals differ in their capabilities or resource availability, such that only some can bear the associated costs (Schuett et al., 2010). Alternatively, individual differences might arise if all individuals can bear the costs, but the cost-benefit ratios differ depending on other components of an individual's phenotype. Such costs might be specifically relevant in contexts influenced by sexual selection and intense and individuals interact repeatedly (Bergmüller & Taborsky, 2010; Royle et al., 2010; Westneat et al., 2015). A final explanation is that this variation is maintained by trade-offs with life-history traits (Bridger, Bonner, & Briffa, 2015; Mulder, Gienapp, & Visser, 2016). To date, few studies appropriately have quantified the existence, let alone the amount, of among-individual variation in stability; my PhD-dissertation seeks to address this caveat.

Accumulating evidence for repeatable individual differences and reversible plasticity in stability implies that behavioural stability can represent a biologically important trait in its own right. In the hierarchy of variation sketched above (Figure 1), this insight implies that stability in one trait may interact with stability in other traits, or with other levels of variation (Westneat et al., 2015). Revealing the nature of trait correlations is important because trait correlations affect evolutionary responses to selection (Lande & Arnold, 1983). For example, stability in one costly trait could be negatively correlated with stability or mean expression of other costly traits (Sakata et al., 2008; Schuett et al., 2010); such patterns of covariance can mitigate the erosion of variation in stability induced by directional selection (Mangel & Stamps, 2001). Alternatively, individuals might have to balance strength/vigour (mean) and precision/skill (variance) of specific behaviours, e.g. visual or acoustic performances in the context of sexual selection (J. Byers, Hebets, & Podos, 2010; Lane & Briffa, 2020). Interestingly, mean and variance of the same trait might also be correlated. For example, positive correlations can occur when "individual quality" (attributable to genetic make-up or silver spoon effects) enables high-quality individuals to express behaviours with higher average levels and stability (Holveck, Vieira de Castro, Lacklan, ten Cate, & Riebel, 2008; Van De Pol, Bruinzeel, Heg, Van Der Jeugd, & Verhulst, 2006). In such cases, selection on one component of a trait (mean) can cause indirect selection on another (variance); simultaneous selection on both could also speed up evolution in response to selection (Schluter, 1996). Finally, high mean levels of traits such as activity, aggression or exploration may characterize individuals that are less reactive to environmental cues and thus behave more predictably (Stamps et al. 2012; Mitchell et al. 2016; He et al. 2017; but see Hertel et al. 2021). Overall, such trait correlations should be studied because they might represent important clues on the mechanisms causing and maintaining individual differences in stability.

The principal aim of this thesis was to address whether behaviour can be understood from an adaptive perspective at multiple (interacting) levels. Behavioural stability has gained attention only recently, and is thus a particular focus of this thesis. Forcefully addressing such questions requires model systems, behavioural traits, and contexts for which clear biological hypotheses can be phrased and the necessary data can be collected. I investigated the existence of individuality and plasticity in stability in wild great tits (*Parus major*). A prime example for a context that perfectly fits the simultaneous study of plasticity, personality and stability, is bird song. In many species song is essential during mate choice (Klappert & Reinhold, 2003; Suter, Ermacora, Rieille, & Meyer, 2009) and territory defence (Amrhein & Lerch, 2010), and thus is a target of both sexual and natural selection. Songs are repeatedly expressed throughout an individual's life, i.e. both within *and* across years. This facilitates ongoing collection of data from the same individuals over substantial time periods, a key prerequisite for firmly examining among- and within-individual differences in conjunction. Furthermore, the omnipresence and its biological importance have led to a massive body of literature on many aspects of bird song. There are thus many indications that bird song varies at *all* of the aforementioned levels of variation, making bird song an ideal model.

Study System

Study Species

I used great tits as a suitable model species to study the different levels of behavioural variation. Great tits represent one of the most studied animals in behavioural ecology, especially in the wild. Several reasons might well explain this popularity. First, great tits are territorial passerines that readily breed in nest boxes, enabling the monitoring and repeated measuring of the same individuals. Second, great tits are a common species in Europe and parts of Asia and North Africa, allowing the gathering of large sample sizes. In combination, these two factors allow for experimental manipulations in the wild (Abbey-Lee, Kaiser, Mouchet, & Dingemanse, 2016; Saggese, Korner-Nievergelt, Slagsvold, & Amrhein, 2011; Sánchez-Virosta et al., 2020) as well as large-scale descriptive studies (Chapter 1 – 4). Third, great tits can also be kept in captivity where they will show natural behaviours, allowing their study under controlled conditions (Drent, Van Oers, & Van Noordwijk, 2003; O'Shea, Serrano-Davies, & Quinn, 2017). Finally, this suitability and historic usage as study species has led to an extensive knowledge about this species, from physiology to breeding ecology and cognitive abilities (Morand-Ferron, Hamblin, Cole, Aplin, & Quinn, 2015; O'Shea et al., 2017; Ouyang, Sharp, Quetting, & Hau, 2013). This knowledge represents an informational fundament on which current, more detailed and complex, questions can be addressed.

Over the last decades, great tits have also been an important species for the study of behavioural variation. Historic studies have focused on variation at the population level, e.g. in life-histories, foraging, or singing (Gosler, 1996; Krebs, Ashcroft, & Webber, 1978; Perrins & McCleery, 1989; Pettifor, Perrins, & McCleery, 1988). More recently, focus has shifted towards variation at the individual level, i.e. within- and among-individual variation. For the within-individual level, i.e. phenotypic plasticity, great tits are often used to study the impacts of climate change on breeding behaviour. Specifically, research has focused on shifts in the time of breeding and reproductive investment in response to elevated spring temperatures and changed temporal patterns of food availability (Charmantier et al., 2008; Verhagen, Tomotani, Gienapp, & Visser, 2020). For the among-individual level, i.e. animal personality, great tits are even more impactful as they have been used in one of the earliest studies on the ecological and evolutionary importance of consistent individual differences (Niels J. Dingemanse, Both, Drent, Van Oers, & Van Noordwijk, 2002). Since this early work on exploratory behaviour, such among-individual differences have been found in various other traits in great tits (Fucikova et al. 2009; Naguib et al. 2010; Araya-Ajoy and Dingemanse 2014; Chapter 1, Strauß et al. 2020, Chapter 2, Hutfluss et al. 2021, Chapter 3 & 4). Furthermore, these differences have been revealed to be integrated into behavioural syndromes with other traits such as dominance or hormonal profiles (Niels Jeroen Dingemanse & de Goede, 2004; van Oers, Buchanan, Thomas, & Drent, 2011). Finally, individual differences in behaviour have been shown to not only apply to the average behavioural expression but to the degree of plasticity as well (Araya-Ajoy & Dingemanse, 2017a; Niels J. Dingemanse & Wolf, 2013).

Besides this extensive knowledge of breeding ecology and behavioural variation, great tits have also commonly been used to study the function and adaptive potential of specific acoustic traits and song in general (Lambrechts & Dhondt, 1987; McGregor, Krebs, & Perrins, 1981; Naguib et al., 2010). Great tit songs can be characterized by the repeated and highly stereotypic expression of specific song types. Song types can be defined as sequences of two to four distinct notes or elements, with each individual knowing a repertoire of several distinct song types. Similar to most passerines, the song in great tits plays an essential role in male-male competition and female choice (Amrhein & Lerch, 2010; Slagsvold, Dale, & Sætre, 1994; Stearns, 1989; Suter et al., 2009). Individual producing “high-quality” songs, e.g. by singing at higher rates, lower frequencies or higher stability, are expected to be of a higher underlying quality and have increased fitness (Bradbury & Vehrencamp, 1998; Cramer, 2013b; Martín-Vivaldi, Palomino, Soler, & Martínez, 2008; McGregor et al., 1981; Rytkonen, Orell, Koivula, & Welling, 2008). At the same time, the production of such attractive signals is considered to be costly, either in terms of energy required or via the potentially harmful responses of competitors (Gil & Gahr, 2002). This cost-benefit aspect of singing should lead to both phenotypic plasticity optimizing it and individual differences in song traits if individuals differ in the ability to produce these traits or resolve potential trade-offs differently. While plastic changes in response to factors affecting the cost-benefit ratio are well reported (Thomas 1999; Godfrey and Bryant 2000; Barnett and Briskie 2007; Pohl et al. 2012; Chapter 1, Strauß et al. 2020), proper estimates of individual differences are limited to specific song traits (Naguib et al. 2010; Araya-Ajoy and Dingemanse 2014; Chapter 1, Strauß et al. 2020; Chapter 2, Hutfluss et al. 2021). Finally, the stereotypic nature of great tit songs makes them ideally suited for the study of song stability (Rivera-Gutierrez et al. 2011, Chapter 3; Chapter 4).

General Field Methods

All data within this thesis were collected as part of a long-term, large-scale field study performed in southern Bavaria, between Starnberg and Herrsching. Established in 2010, 12 nest box populations were monitored each breeding season following a standardized protocol. Each population contained 50 nest boxes arranged in a 50m grid, resulting in a total of 600 boxes (550 after 2017). Starting at the beginning of April, we inspected all nest boxes (bi)weekly to determine the lay date (date first egg was laid, backcalculated assuming one egg per day), clutch size and onset of incubation. Several days before the estimated date, we visited the respective boxes daily to observe the precise hatching date. When the nestlings were 7-10 days old, adult birds were caught in the nest box using spring traps. Their exploratory behaviour was then immediately tested using an open field cage test. After that we measured morphological and further behavioural traits, ringed all unringed individuals to assign unique identities and took blood samples for paternity analysis (see Chapter 4). When the nestlings were 14 days old, we performed the same measurements on them except the open field cage test. Around the age of 21 days, we visited the nest boxes to measure the final number of fledglings.

Song Recordings

In order to acquire repeated measurements of singing behaviour, we performed four simulated territory intrusions (STI) for all breeding males. Two STIs were performed during the egg-laying phase and two during the incubation phase. During these tests, we presented the resident male with a taxidermic great tit model as a visual stimulus and the playback of a conspecific song as acoustic stimulus. Both were placed 1 m in front of the nest box, the model on a 1.2 m pole, the speaker (Foxpro Shockwave) on the ground. Models (n = 23) and playbacks (n = 195) were randomly assigned. From 2010 to 2016, we used 21 songs found in the Xeno-Canto database as playbacks. In 2017, we used 13 songs recorded before the breeding season from the voluntary singing of 13 different local birds. For 2018 and 2019, we added 169 playbacks based on recordings made during STIs performed in 2017. Playbacks incorporated natural levels of variation in terms of strophe number and duration, which was later accounted for in all statistical analyses.

Tests started after the focal male entered a 15m radius around the nest box. After that, the observer located at a 15m distance from the box counted the number of songs and alarm calls the male produced for three minutes. We further estimated the minimum approach distance to the intruder during the three-minute period. From 2017 to 2019, the observer also recorded the song with a windscreen-covered directional microphone (Sennheiser ME66/K6) and an attached recorder (TASCAM DR-05, recording 44.1 kHz, 16-bit WAV files). Individuals that did not enter the radius within 15 min were scored “non-responsive”. Specific acoustic measurements, e.g. song overlap or minimum frequency, were extracted after each field season using semi-automated measurements and standardized protocols (for details, see Chapters 2 and 3).

Thesis Outline

To address the main questions of my thesis, I conducted four studies focusing on different levels of variation, song features and questions of current relevance. All studies were based on long-term observational data specifically collected to address these questions. What is unique about this thesis is that it studies multiple levels of variation in bird song in conjunction. I focus on three key levels in particular:

1. Within-individual variation in mean behaviour **(Chaper 1):** Phenotypic plasticity has been the focus of song research for decades (Marler & Slabbekoorn, 2004). Birds plastically adjust their singing behaviour in response to a variety of external factors, both social and non-social (Jablonszky et al., 2021; Ripmeester, Kok, van Rijssel, & Slabbekoorn, 2010; Strauß et al., 2020; Thomas, 1999). Social factors include, for example, the number or behaviour of conspecifics (Aubin, Mathevon, Silva, Vielliard, & Sebe, 2004; Jablonszky et al., 2021; Ripmeester et al., 2010; Vergne, Avril, Martin, & Mathevon, 2007). Stronger competition due to higher breeding densities or quality of the opponent lead to corresponding changes in various song features due to changes in motivation or the costs of lost paternity or territories (Jablonszky et al., 2021; Penteriani, 2003). Non-social factors, on the other hand, often influence internal cost-benefit trade-offs or trade-offs with other traits. Under the common assumption that singing is costly, factors such as food availability and temperature will affect the amount of energy and time available for singing (Barnett & Briskie, 2007; Strain & Mumme, 1988; Strauß et al., 2020; Thomas, 1999). I studied behavioural plasticity in a study documenting how temperature affects vocalization behaviour and physical aggression expressed during simulated territory intrusions. I asked whether night temperature negatively influences the number of vocalizations produced in the subsequent morning, and whether negative effects on singing are associated with corresponding changes in other vocalisations (alarming).
2. Within- and among-individual variation in average behaviour **(Chapter 2)**: Besides plasticity, birds also show consistent individual differences in numerous song traits (Naguib, Diehl, van Oers, & Snijders, 2019; Paul, Thompson, & Foote, 2020; Zsebők, Moskát, & Bán, 2017). Such repeatable differences may be caused by genetic or early life

differences and are necessary for song traits to function as reliable signals of quality, for example, as a mating partner or competitor (Beani & Dessì-Fulgheri, 1995; Houtmann, 1992). Quality varies primarily among individuals, certainly when determined genetically. Song traits can, however, also function as motivational signals. Those are signals that can vary both among- and within-individuals (Guilford & Dawkins, 1995; Ripmeester, de Vries, & Slabbekoorn, 2007). Because variation in singing is rarely attributable to one of the two exclusively, studying among- and within-individual variation in conjunction is essential to understand the signalling values of song. Territorial interactions form an extremely suitable context for addressing such questions. Many traits expressed in aggressive contexts are argued to transfer information on the sender's motivation or underlying ability to fight and are thus expected to covary with physical aggression (Anderson, Searcy, Hughes, & Nowicki, 2012; Anderson, Searcy, Peters, & Nowicki, 2008; Naguib, Altenkamp, & Griessmann, 2001). However, studies rarely explicitly examine variation in such traits at multiple levels; few have also studied whether signal values are consistent over extended periods of time (Akçay, Campbell, & Beecher, 2014; Nowicki, Searcy, Krueger, & Hughes, 2002). I investigated the existence of repeatable among-individual variation and reversible plasticity within a single study, while focusing on whether song overlapping represents a signal of aggressiveness. This demonstrates the utility of the joint study of among- and within-individual variance, as doing so enables distinguishing between multiple functional explanations for song overlapping, specifically, motivation versus individual quality.

3. Within- and among-individual variation in residual within-individual variation **(Chapter 3)**: Birds have frequently been used to study behavioural stability in acoustic communication (Botero et al. 2009; De Kort et al. 2009; Sakata and Vehrencamp 2012; Cramer 2013b). Stereotypic song production, especially over extended time spans, requires the precise and repeated activity of syringeal and respiratory muscles and is thus likely to be energetically costly (Grava, Grava, & Otter, 2012; Lambrechts & Dhondt, 1988; Sakata & Vehrencamp, 2012; Suthers & Zollinger, 2004). As individuals can differ in their ability to pay these costs, song stability might serve as a quality indicator. Furthermore, individuals might plastically downregulate stability in contexts where these costs outweigh the benefits. Few studies have, however, quantified within- and among-individual variation in stability in conjunction, perhaps because appropriate statistical approaches for doing so have long been missing. Such quantifications are necessary when addressing whether song stability is a biologically distinct trait in itself. It also allows studying its adaptive potential. I therefore investigated whether individuals were both plastic and repeatable in their song stability. I

used experimental manipulations (food supplementation) to test whether song stability is a costly trait under selection.

4. Fitness consequences of variance components, illustrated using bird song **(Chapter 4)**: Previous chapters focused on testing for the existence of variation in bird song among and within-individuals, focusing both on behavioural mean and variance (stability). This chapter takes the next step and asks whether variation in stability in bird song is under selection. Although bird song is thought to act primarily in the context of sexual selection, natural and sexual selection are intertwined (Catchpole & Slater, 2003; Marler & Slabbekoorn, 2004). Females in many species choose their social and mating partners based on specific song traits (Catchpole, 1987; Marler & Slabbekoorn, 2004; W A Searcy & Andersson, 1986). These traits correlate with or signal the underlying quality of a male, e.g. in terms of size, condition or behaviour (Buchanan & Catchpole, 2000; Gil & Gahr, 2002; J. P. Martin, Doucet, Knox, & Mennill, 2011; Spencer, Buchanan, Goldsmith, & Catchpole, 2003). Consequently, “good” singers might not only be able to convince females to mate with them, they might also occupy better territories, raise more fledglings and might even survive for longer (Lord, Waas, Innes, & Whittingham, 2001; Rivera-Gutierrez, Pinxten, & Eens, 2010; Soma & Garamszegi, 2011). In addition, reproductive success in many species is not limited to the social partner, but also extends to other conspecific females. Females have been hypothesized to modify within- and extra-pair paternity activities based on their mate’s singing capacities (Hasselquist et al. 1996; Byers 2007; but see Garamszegi 2004). These links between song traits and reproductive success are suggested for song stability as well, but evidence is extremely scarce (B. E. Byers, 2007). Therefore, I investigated how song stability affects male reproductive success, focusing on successful fledglings with the social partner as well as the extra-pair paternity gained in other and lost in the focal individual’s nest.

General Discussion

The amount of phenotypic variation in animal behaviour that can be observed in nature is astounding. Although this variation can seem random at first glance, it is characterized by a deeply hierarchical structure, from variation among species and individuals to variation within single individuals for traits like behaviours (Westneat et al., 2015). Each of these levels can act as the raw material of selection, at least if they are, to some extent, genetically determined. Therefore, quantifications of behavioural variation at any level are of key importance to understand the adaptive potential of behaviour and the patterns of selection shaping and maintaining it. Historically, behavioural ecologists have studied specific levels largely in isolation. In the last decades, research has instead highlighted the importance of studying levels of variation in conjunction. This change has also been facilitated by the development of complex study designs and statistical methods (Cleasby et al., 2015; Niels J. Dingemanse & Dochtermann, 2013). This approach is essential, because the different levels do not exist independently of each other but are often linked with key adaptive consequences. Prime examples of such adaptive links are personality-related variation in plasticity, or plasticity in variance sensitivity (Mathot et al., 2012; Stamps, 2016; Stephens et al., 2008; Westneat et al., 2013).

Studying the multilevel nature of animal behaviour has become one of the key research fields in behavioural ecology over the last decades. However, not all research areas have adopted this approach equally, limiting our understanding of existence and relevance of specific levels of variation and their interactions for various contexts. Neglecting specific variance components, e.g. among-individual variation, may prevent us from gaining general insights into biological mechanisms and patterns of behavioural variation. For example, individuals of an invasive species will show habituation (within-individual variation) when settling in new habitats, but individuals at the invasion front will differ strongly from the ones further behind (among-individual variation) (Brown, Phillips, & Shine, 2014; Duckworth & Badyaev, 2007). Neglecting specific variance components becomes even more problematic when they interact with other levels, e.g. when the degree of plasticity depends on the personality of an individual (Niels J. Dingemanse, Kazem, Réale, & Wright, 2010b; Mathot et al., 2012; Stamps, 2016). In such cases, focusing on a single level would cause biased, if not erroneous, findings. Surprisingly, the field of bird song is one of the research areas that falls short of adopting these approaches, although intuitively it seems to be predestined to benefit from such studies. Song is arguably a prime example of a labile trait that is repeatedly expressed at various timescales and that is adjusted to biotic and abiotic environmental factors (Marler & Slabbekoorn, 2004). At the same time, song traits are used to transfer information from the sender to a receiver, e.g. the identity or quality as a competitor or potential mate (Gil & Gahr, 2002). To reliably signal this information, individuals would need to be consistent in their singing, regardless of the environment (Sakata & Vehrencamp, 2012). For reliable and honest signals, individual differences in quality should thus be reflected by individual differences in the respective acoustic traits. Quantifying the levels at which song varies is thus essential to our understanding of signal reliability, trade-offs and, ultimately, the adaptive potential of song. For example, focusing only on reversible plasticity (within-individual variation) in song cannot reveal whether song traits signal differences in individual quality, and whether they could be under selection. Similarly, links between mean and stability of a trait imply that focusing on only one of the two can lead to biased results if any effect is only a by-product of the correlation with the neglected variance component. However, comparably few studies have attempted to distinguish different levels of variation in song and vice versa have used song as a suitable candidate behaviour to study mechanisms and importance of different levels of behavioural variation.

My dissertation fills this gap by applying to bird song state-of-the-art variance-partitioning methods commonly used in other contexts. Throughout the chapters of this thesis, I focused on a specific selection of key song traits. All traits were chosen as they were assumed to have important signalling functions and potential fitness consequences. Starting with plasticity in singing behaviour, each chapter adds layers of complexity, ending with the realized fitness consequences of residual within-individual variation in Chapter 4. The nuanced meanings and implications of the results of my studies are discussed separately in each chapter. In the following, I will embed the results of my dissertation in a more general framework and highlight specific properties and findings present in multiple chapters with relevance to the study of behavioural variation.

Table 1: Variance components addressed and quantified in each chapter. X's in brackets indicate that the respective level of variation was quantified or accounted for but was not the main focus of the study or further discussed.

	Chapter 1	Chapter 2	Chapter 3	Chapter 4
Behavioural means				
Within individuals	x	x	(x)	(x)
Among individuals	(x)	x	(x)	x
Behavioural variance				
Within individuals			x	(x)
Among individuals			x	x

Trade-offs

Following the optimality perspective from which animal behaviour can be studied, directional selection is expected to favour single behavioural optima within populations. Such selection pressures should, therefore, erode behavioural variation at many levels, but specifically among individuals (Dall, Houston, & McNamara, 2004b; Schuett et al., 2010; Sih, Bell, Johnson, & Ziemba, 2004). The existence of consistent individual differences in many traits, therefore, leads to another essential question in behavioural ecology: What are the mechanisms that maintain this behavioural variation? Potentially the most universally applicable explanation origins from one of the key paradigms of behavioural ecology. Animals constantly face constraints and trade-offs limiting their range of behavioural expression (Houston & McNamara, 1999; McNamara & Houston, 1986; Roff & Fairbairn, 2007). Selection will then favour the optimal resolution of such trade-offs based on the cost-benefit ratio of the respective behaviour. Importantly, neither costs nor benefits are fixed, but instead vary depending on external and internal factors such as environmental factors, condition or context (Cameron K. Ghalambor, Angeloni, & Carroll, 2010). Physical aggression, for example, might be less advantageous when resources are abundant, competition low, or the opponent considerably stronger (Cain & Ketterson, 2013; Haigh, O'Riordan, & Butler, 2017; Knell, 2009; Songvorawit, Butcher, & Chaisuekul, 2018; Wu, Whiting, Fu, & Qi, 2019). Therefore, even for a single trait, cost-benefit ratios are constantly changing, requiring repeated adjustments to optimize behavioural responses and ultimately leading to the evolution and maintenance of reversible plasticity (Beaman, White, & Seebacher, 2016). In addition, individuals can differ in the way they resolve such trade-offs (Roff & Fairbairn, 2007; Williams, 1966). Such differences might arise when individuals differ in the ability or tendency to express certain behaviours (Biro et al., 2018; Niels J. Dingemanse & Wolf, 2013). When defined by a genetic basis, selection can maintain these differences, e.g. when temporal or spatial variation in selection favours different trade-off resolutions or when different resolutions have equal fitness outcomes (Niels J. Dingemanse, Both, Drent, & Tinbergen, 2004; Mangel & Stamps, 2001; Reale & Dingemanse, 2010; Réale et al., 2010).

The trade-off between energy or time allocated to alternative activities is likely one of the most fundamental trade-offs in behavioural ecology (Boyd & Hoelzel, 2002; Goldstein, 1988). Each behaviour will consume parts of the finite amounts of time and energy available to an individual, although the required amounts can differ substantially among behaviours (e.g. Wolf and Hainsworth 1971; Ladds et al. 2018). Bird song is certainly a prime example of a trait for which these assumed energetic and temporal costs are of key importance as they could ensure the reliability of song as a signal of quality (Gil & Gahr, 2002). However, direct measures of the energy consumed during singing reveal a rather mixed picture of the energetic costs (Eberhardt, 1994; Gaunt, Bucher, Gaunt, & Baptista, 1996; Oberweger & Goller, 2001; Ward & Slater, 2005). Importantly, these measures have, in birds, almost exclusively been collected in the lab; evidence from the field is largely absent. In the end, the energy demands in the wild where the environment is less controlled might be higher and even if they remain small, in many species considerable time spent singing might lead to a considerable cumulative energy consumption. Therefore, measuring energetic costs indirectly via the effects of environmental factors assumed to affect energetic reserves on song traits is commonplace (Garson and Hunter 1979; Reid 1987; Barnett and Briskie 2007; Chapter 1, Strauß et al. 2020). Increased food availability, for example, can reduce the time (and energy) required for foraging, leading to more resources for singing (Barnett & Briskie, 2007; Thomas, 1999), whereas lower temperatures increase the energy consumed to maintain homeostasis, resulting in less available energy (Reinertsen, 1983; Reinertsen & Haftorn, 1986). Consistent with these findings, chapter 1 forcefully demonstrated that ambient temperatures influence song rate, and support its classification as a classic performance trait that represents accumulating energy consumption.

What makes Chapter 1 unique is the fact that we studied the simultaneous impacts of temperature on two distinct acoustic traits, song and alarm rates. Animals can behave in more than one way to achieve a specific aim, and each strategy comes with its own costs and benefits (e.g. Taborsky et al. 2008). Especially during energetically demanding periods, for example food scarcity or low temperatures, overall resource availability becomes limited. This change in access to energy sources can also lead to the presumable "best" behaviour becoming too costly (Abrams, 1993). As a consequence, animals might decide to switch to cheaper behaviours with better cost-benefit ratios under such conditions or to balance different traits to achieve the optimal compromise. Classic examples can be found in trade-offs between future and current reproduction, trill rate and bandwidth as performance limits in bird song, and between different foraging strategies (Chittka, Dyer, Bock, & Dornhaus, 2003; Podos, 1997; Stearns, 1989). However, studies on bird song often study such trade-offs focusing on the energetic costs of a single trait, or mechanical and temporal constraints of multiple traits, whereas energetic costs and the differences in the effectiveness of multiple acoustic traits remain largely neglected. Chapter 1 fills this gap by showing that different acoustic traits are affected differently by temperature, indicating differences in their energy demands. Considering the extreme prevalence of singing during territory defence, this further indicates a difference in effectiveness between singing and alarming. Overall, this study thus shows that energetic trade-offs are not restricted to single traits, but instead can affect the balance between alternative behavioural responses.

Moreover, trade-offs also occur between different levels of the same trait, with mean-variance correlations probably representing the best-known example (Briffa, Bridger, & Biro, 2013; Highcock & Carter, 2014; Prentice et al., 2020; Rönnegård, Felleki, Fikse, Mulder, & Strandberg, 2010). Mean-variance correlations are often considered to be a mathematical or methodological artefact, e.g. when higher average values are associated with stronger measurement errors leading to a higher variance (Taylor, 1961; Viswanathan, 2005; Westneat et al., 2015). An increasing number of studies, however, indicates that behavioural variance can be a biologically meaningful trait, making mean-variance associations insightful in terms of biology (Mitchell, Beckmann, & Biro, 2021; Mitchell et al., 2016; Westneat et al., 2015). For bird song, positive associations between mean and variance might indicate that the ability to produce high-quality song traits is linked to a higher neural or motor-control skills and thus lower variance (Gil & Gahr, 2002; Lambrechts, 1997; Sakata & Vehrencamp, 2012). This so called "big-car-big-house" phenomenon assumes that high quality individuals can afford multiple costly behaviours at the same time, in this case the stable production of low frequency songs (Betini & Norris, 2012; Rivera-Gutierrez et al., 2010). On the contrary, negative correlations might indicate trade-offs, whereby individuals either try to produce song traits of a desired average level with low stability, or produce these song traits with high stability but at average levels that are not as beneficial (Duckworth, Potticary, & Badyaev, 2018; Sih & Del Giudice, 2012). The results of Chapter 3, demonstrating positive mean-variance correlations, rather indicate the former as individuals singing lower minimum frequencies and thus at the assumingly “better” average trait level, were also singing more stable. Importantly, our findings cannot show whether the detected pattern was an artefact or representing interesting biological. Nonetheless, this finding demonstrates that studying correlations between different levels of variation can reveal interesting biological patterns and even potential mechanisms behind them.

Repeatability in the mean

Consistent individual differences have been a hot topic in behavioural ecology for the last decades (Bell et al., 2009; Réale et al., 2010; Max Wolf & Weissing, 2012). One of the key metrics used to quantify such individual differences is repeatability, i.e. the proportion of the total phenotypic variance attributable to variation among individuals (Hayes & Jenkins, 1997). High values indicate that most of the variation is due to variation among individuals, as it is common for morphological traits, whereas lower values indicate higher flexibility or within-individual variation, which is typical for labile traits such as behaviours. Using this standardized measure does not only give an easily understandable measure of individual differences it also enables comparisons between studies, species and behaviours. In addition to that, repeatability is assumed to represent the upper limit of heritability, thereby functioning as a first and more easily acquired indicator of genetic differences and bases of behaviours, and thus of the adaptive potential of a trait (Falconer & Mackay, 1996). Overall, behaviour shows an average repeatability of roughly 35% but the estimates strongly differ among behavioural categories (e.g. courtship), taxa or the study context (laboratory vs wild) (Bell et al., 2009). Although birds are one of the most frequently studied taxa in the field of animal personality, repeatability estimates of specific song traits are not readily available; an issue my thesis helped address.

Song rate is one of the most frequently studied song traits, and one for which repeatability has been quantified in a number of species and contexts (Murphy et al. 2008; Cain and Langmore 2015; Snijders et al. 2015; Zsebők et al. 2017; Chapter 1, Strauß et al. 2020). Overall, these estimates show moderate to reasonably high values ranging from 0.15 to 0.45, although most values were closer to the lower boundary. Interestingly, repeatability seemed to be independent of the context, i.e. song rates during the dawn chorus were not more or less repeatable compared to song rates during active territory disputes. Consequently, the estimates included in this thesis (Chapter 1 & 2) are not only representing a valuable replication, which is often missing in ecological studies (Kelly, 2006). They also indicate that individual differences were not more pronounced during high threat situations such as territory intrusions. This could have been expected if the potential costs of losing and the motivation to sing at the limit are higher than during the dawn chorus. Another important point is that these individual differences can be found although song rate is an extremely plastic trait, influenced by a variety of environmental factors, such as temperature (Garson and Hunter 1979; Thomas 1999; Chapter 1, Strauß et al. 2020), or breeding stage (Chapters 1 & 2)(Araya-Ajoy & Dingemanse, 2014). This strongly suggests that song rates can reliably signal properties of the individual such as the territory quality, aggression, or the male's underlying genetic quality.

Alarm calls might be an even more frequently produced vocalization type than songs, but repeatability estimates are very rare, making it difficult to interpret their signal value (Araya-Ajoy & Dingemanse, 2014). Alarm calls are usually associated with the presence of a predator (Hope, 1980; Klump & Shalter, 1984), but at least in great tits they are also frequently produced during territorial disputes (Araya-Ajoy & Dingemanse, 2014, 2017b). Certainly, songs are the preferred way of defending territories or starting aggressive encounters in many species. However, alarm calls can represent an alternative type of vocal response that might become specifically important when the costs of singing increase, e.g. when energy resources are low. This potential signal value is further supported by repeatable individual differences detected for alarm rate. These differences could, however, also be a by-product of the negative correlation with the repeatable song rate. In the first case, it remains to be revealed whether the signal value of alarm calls is different or lower compared to songs and whether they remain an acoustic signal produced by supposedly lower quality males.

Song overlapping is argued to be a signal of aggressiveness, although findings are mixed and its actual signal value is strongly disputed (Helfer & Osiejuk, 2015; William A. Searcy & Beecher, 2009, 2011). Contrary to performance traits such as song rate, song overlapping as other aggressive signals is assumed to be a motivational signal. As such, repeatability would be expected to be rather low, as all individuals should be equally able to produce it and any variation in this trait should depend on the willingness to actually escalate a fight (Guilford & Dawkins, 1995). However, this trait has also been suggested to serve as a signal of quality, which should result in a considerable amount of among-individual variation (Bischoff, Tschirren, & Richner, 2009; Kunc, Amrhein, & Naguib, 2006). Although the distinction between motivation signals and index signals is not strictly defined and is arguably somewhat subjective, partitioning the phenotypic variation in specific traits can give valuable insights. However, within- and among-individual variation have, to our knowledge, not been measured for this trait, thus preventing us from making solid inferences. In Chapter 2, we found a very low repeatability (~1%) for song overlapping, indicating that variation in this trait mostly occurs within, rather than among individuals. This result further seems to contradict the potential as a quality signal, as individual differences in quality should then manifest in individual differences in song overlapping. Likewise, physical aggression has been shown to be repeatable and if song overlapping would signal aggression, such repeatable differences should exist as well (Hutfluss et al. 2021).

"Repeatability" in behavioural variability

Behavioural stability itself and its associated patterns of variation have gained the focus of behavioural ecologists only recently thus leaving many questions unaddressed. Therefore, the core question of this thesis was to first establish whether individuals are repeatable in song stability. The stable production of acoustic traits has been studied often compared to behavioural stability in other behaviours (Chapter 3; Mitchell et al. 2021). Despite this interest, reliable estimates of consistent individual differences in song stability are extremely rare (Taff et al. 2012; Cramer 2013a; Cramer 2013b; Kubli and MacDougall-Shackleton 2014; Byers et al. 2015; Iserbyt et al. 2017, Chapter 3). This is surprising as song stability has been linked to various fitness indicators and has been argued to be a signal of male quality (Bartsch, Hultsch, Scharff, & Kipper, 2016; Botero et al., 2009; De Kort et al., 2009). Despite an expected directional selection towards more stable songs, we would thus expect individuals to consistently differ in stability. While repeatability in the stricter sense cannot be calculated for stability, analogous estimates such as the coefficients of variation for predictability (CVp) might be interpreted in similar ways, for example as an upper boundary of heritability. Consequently, such estimates of within- and among-individual variation are necessary to better understand the signal and adaptive value of song stability.

In a recent study, Mitchell et al. (2021) published a meta-analytical summary of the existing estimates of among-individual variation in stability. Overall, the CVp of behaviour 0.27, compared to a CVp for physiological traits. Similar to the repeatability of behaviours, there was substantial variation among the different estimates, ranging from almost 0 to 0.76. Importantly, there are still very few estimates of variation in stability, and these are coming from an even smaller sample of working groups leading to clustered results in terms of study species and candidate behaviours. Contrary to many other research areas in behavioural ecology, birds are extremely underrepresented. Furthermore, studies so far strongly focused on activity and boldness, whereas there is not a single CVp estimate for song (Mitchell et al., 2021). This is surprising given the suitability of song for the study of stability and the clear biological hypotheses that can be formulated for this behaviour.

As a consequence, the results of Chapter 3 are the first estimates of among-individual variation in residual within-individual variation for bird song. Our estimates for the two key acoustic parameters studied, minimum frequency and phrase length, are well within the range reported by Mitchell et al. (0.15 and 0.12), but certainly more towards the lower range. This is not surprising, given the fact that studies on wild animals reveal smaller CVps than studies in the laboratory. Beyond this, comparisons with other studies on behavioural stability need to be done carefully, due to the lack of song trait estimates or other mating behaviours in the literature.

An important alternative explanation for the comparably low CVp values could be our unique consideration of the notion that individuals can show pseudo-repeatability. Pseudo-repeatability represents an overestimation of among-individual variance when individuals plastically adjust to environmental factors that are individually repeatable (Niels J. Dingemanse & Dochtermann, 2013; Westneat, Hatch, Wetzel, & Ensminger, 2011). This is, for example, a common problem for territorial species that live for more than a year, such as passerines (Araya-Ajoy & Dingemanse, 2017b; Mitchell, Dujon, Beckmann, & Biro, 2020; Niemelä & Dingemanse, 2017; Zsebők, Herczeg, et al., 2017). Factors associated with specific territories, e.g. food availability or predation, will differ among territories but stay relatively constant within years, or even among breeding seasons. Testing individuals repeatedly without accounting for this, e.g. by considering the within-individual among-year variation, can then lead to within-individual variation to be considered as among-individual variation. At least in our study, neglecting this variance component would have inflated among-individual variation in stability substantially. This might apply to some of the existing studies on behavioural stability as well, especially to those studying vertebrates in the wild. Notably, one of the activity studies with a similar CVp to ours also accounted for pseudo-repeatability, revealing that observations within a year were more similar than across years and that long-term repeatability would have been strongly overestimated without consideration (Mitchell et al., 2021). Another interesting pattern is that most studies with substantially stronger individuality in stability measured activity and boldness, whereas similar estimates came from studies on metabolism and other types of behaviours (Mitchell et al., 2021). Individuals characterized as active or bold are known to be less reactive and plastic in response to environmental factors (He et al., 2017; Mitchell et al., 2016; Stamps et al., 2012; but see Hertel et al., 2021). In addition, both behaviours are known to be repeatable, thereby leading to individual differences in stability due to the individual differences in average behaviour (Biro, Beckmann, & Stamps, 2010).

Conclusions

My dissertation includes both valuable replications of former studies and findings as well as novel insights into the adaptive value of different levels of variation by applying state-of-the-art statistical analyses to the context of bird song. Overall, the results of my work have several biological implications relevant for future work. First, plasticity in response to environmental factors is not restricted to single, independent traits. Instead, studying within-individual variation in multiple traits simultaneously can result in a deeper understanding of the signal values and costs of different traits. Second, partitioning major variance components, namely within- and among-individual variation in the average behaviour, can reveal potential mechanisms behind the function of specific animals signals. Finally and most importantly, behavioural stability may well represent a biologically meaningful trait, consistently differing among individuals. However, this individuality in stability is not linked to reproductive success and might thus not be under selection. Future studies should, therefore, investigate patterns of variation and selection for the stability of other signalling traits and what mechanisms might cause this variation if selection is indeed absent or working via different pathways than reproductive success.

References

Abbey-Lee, R. N., Kaiser, A., Mouchet, A., & Dingemanse, N. J. (2016). Immediate and carry-over effects of perceived predation risk on communication behavior in wild birds. *Behavioral Ecology*, *27*(3), 708–716. https://doi.org/10.1093/beheco/arv210

Abrams, P. A. (1993). Optimal traits when there are several costs: the interaction of mortality and energy costs in determining foraging behavior. *Behavioral Ecology*, *4*(3), 246–259. https://doi.org/10.1093/beheco/4.3.246

Akçay, Ç., Campbell, S. E., & Beecher, M. D. (2014). Individual differences affect honest signalling in a songbird. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1775), 20132496. https://doi.org/10.1098/rspb.2013.2496

Amrhein, V., & Lerch, S. (2010). Differential effects of moving versus stationary territorial intruders on territory defence in a songbird. *Journal of Animal Ecology*, *79*, 82–87. https://doi.org/10.1111/j.1365-2656.2009.01629.x

Anderson, R. C., Searcy, W. A., Hughes, M., & Nowicki, S. (2012). The receiver-dependent cost of soft song: A signal of aggressive intent in songbirds. *Animal Behaviour*, *83*(6), 1443–1448. https://doi.org/10.1016/j.anbehav.2012.03.016

Anderson, R. C., Searcy, W. A., Peters, S., & Nowicki, S. (2008). Soft Song in Song Sparrows: Acoustic Structure and Implications for Signal Function. *Ethology*, *114*(7), 662–676. https://doi.org/10.1111/j.1439-0310.2008.01518.x

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2014). Characterizing behavioural 'characters': an evolutionary framework. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*, 20132645. https://doi.org/10.1098/rspb.2013.2645

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2017a). Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. *Journal of Animal Ecology*, *86*, 227–238. https://doi.org/10.1111/1365-2656.12621

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2017b). Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. *Journal of Animal Ecology*, *86*(2), 227–238. https://doi.org/10.1111/1365-2656.12621

Aubin, T., Mathevon, N., Silva, M. L. Da, Vielliard, J. M. E., & Sebe, F. (2004). How a simple and stereotyped acoustic signal transmits and individual information the song of the White-browed Warbler Basileuterus lecoblepharus. *Anais Da Academia Brasileira de Ciências*, *76*(2), 335–344.

Auld, J. R., Agrawal, A. A., & Relyea, R. A. (2010). Re-evaluating the costs and limits of adaptive phenotypic plasticity. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1681), 503–511. https://doi.org/10.1098/rspb.2009.1355

Barnett, C. A., & Briskie, J. V. (2007). Energetic state and the performance of dawn chorus in silvereyes (Zosterops lateralis). *Behavioral Ecology and Sociobiology*, *61*, 579–587. https://doi.org/10.1007/s00265-006-0286-x

Bartsch, C., Hultsch, H., Scharff, C., & Kipper, S. (2016). What is the whistle all about? A study on whistle songs, related male characteristics, and female song preferences in common nightingales. *Journal of Ornithology*, *157*(1), 49–60.

https://doi.org/10.1007/s10336-015-1245-y

Beaman, J. E., White, C. R., & Seebacher, F. (2016, March 1). Evolution of Plasticity: Mechanistic Link between Development and Reversible Acclimation. *Trends in Ecology and Evolution*. Elsevier Ltd. https://doi.org/10.1016/j.tree.2016.01.004

Beani, L., & Dessì-Fulgheri, F. (1995). Mate choice in the grey partridge, Perdix perdix: Role of physical and behavioural male traits. *Animal Behaviour*, *49*(2), 347–356. https://doi.org/10.1006/anbe.1995.0047

Bell, A. M. (2007). Future directions in behavioural syndromes research. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1611), 755–761. https://doi.org/10.1098/rspb.2006.0199

Bell, A. M., Hankison, S. J., & Laskowski, K. L. (2009). The repeatability of behaviour: a meta-analysis. *Animal Behaviour*, *77*(4), 771–783. https://doi.org/10.1016/J.ANBEHAV.2008.12.022

Bergmüller, R., & Taborsky, M. (2010). Animal personality due to social niche specialisation. *Trends in Ecology & Evolution*, *25*(9), 504–511. https://doi.org/10.1016/J.TREE.2010.06.012

Betini, G. S., & Norris, D. R. (2012). The relationship between personality and plasticity in tree swallow aggression and the consequences for reproductive success. *Animal Behaviour*, *83*(1), 137–143. https://doi.org/10.1016/j.anbehav.2011.10.018

Biro, P. A., & Adriaenssens, B. (2013). Predictability as a personality trait: consistent differences in intraindividual behavioral variation. *The American Naturalist*, *182*(5), 621–629. https://doi.org/10.1086/673213

Biro, P. A., Beckmann, C., & Stamps, J. A. (2010). Small within-day increases in temperature affects boldness and alters personality in coral reef fish. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1678), 71–77. https://doi.org/10.1098/rspb.2009.1346

Biro, P. A., Garland, T., Beckmann, C., Ujvari, B., Thomas, F., & Post, J. R. (2018). Metabolic scope as a proximate constraint on individual behavioral variation: Effects on personality, plasticity, and predictability. *American Naturalist*, *192*(2), 142–154. https://doi.org/10.1086/697963

Bischoff, L. L., Tschirren, B., & Richner, H. (2009). Long-term effects of early parasite exposure on song duration and singing strategy in great tits. *Behavioral Ecology*, *20*(2), 265–270. https://doi.org/10.1093/beheco/arp012

Botero, C. A., Rossman, R. J., Caro, L. M., Stenzler, L. M., Lovette, I. J., de Kort, S. R., & Vehrencamp, S. L. (2009). Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Animal Behaviour*, *77*(3), 701–706. https://doi.org/10.1016/j.anbehav.2008.11.020

Boyd, I. L., & Hoelzel, A. (2002). Energetics: consequences for fitness. In *Marine Mammal Biology: an evolutionary approach* (pp. 247–277). Blackwell Publishing Inc.

Bradbury, J. W., & Vehrencamp, S. L. (1998). *Principles of animal communication.*

Brembs, B. (2011). Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society B:*

Biological Sciences, *278*(1707), 930–939. https://doi.org/10.1098/rspb.2010.2325

Bridger, D., Bonner, S. J., & Briffa, M. (2015). Individual quality and personality: bolder males are less fecund in the hermit crab *Pagurus bernhardus*. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1803), 20142492. https://doi.org/10.1098/rspb.2014.2492

Briffa, M. (2013). Plastic proteans: reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, *9*(5), 20130592. https://doi.org/10.1098/rsbl.2013.0592

Briffa, M., Bridger, D., & Biro, P. A. (2013). How does temperature affect behaviour? Multilevel analysis of plasticity, personality and predictability in hermit crabs. *Animal Behaviour*, *86*(1), 47–54. https://doi.org/10.1016/J.ANBEHAV.2013.04.009

Brown, G. P., Phillips, B. L., & Shine, R. (2014). The straight and narrow path: The evolution of straight-line dispersal at a cane toad invasion front. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1795). https://doi.org/10.1098/rspb.2014.1385

Buchanan, K. L., & Catchpole, C. K. (2000). Song as an indicator of male parental effort in the sedge warbler. *Proceedings of the Royal Society B: Biological Sciences*, *267*(1441), 321–326. https://doi.org/10.1098/rspb.2000.1003

Byers, B. E. (2007). Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behavioral Ecology*, *18*(1), 130–136. https://doi.org/10.1093/beheco/arl058

Byers, B. E., Akresh, M. E., & King, D. I. (2015). A proxy of social mate choice in prairie warblers is correlated with consistent, rapid, low-pitched singing. *Behavioral Ecology and Sociobiology*, *69*(8), 1275–1286. https://doi.org/10.1007/s00265-015-1940-y

Byers, J., Hebets, E., & Podos, J. (2010). Female mate choice based upon male motor performance. *Animal Behaviour*, *79*(4), 771–778. https://doi.org/10.1016/j.anbehav.2010.01.009

Cain, K. E., & Ketterson, E. D. (2013). Costs and Benefits of Competitive Traits in Females: Aggression, Maternal Care and Reproductive Success. *PLoS ONE*, *8*(10), e77816. https://doi.org/10.1371/journal.pone.0077816

Cain, K. E., & Langmore, N. E. (2015). Female and male song rates across breeding stage: Testing for sexual and nonsexual functions of female song. *Animal Behaviour*, *109*, 65–71. https://doi.org/10.1016/j.anbehav.2015.07.034

Catchpole, C. K. (1987, April 1). Bird song, sexual selection and female choice. *Trends in Ecology and Evolution*. Elsevier Current Trends. https://doi.org/10.1016/0169-5347(87)90165-0

Catchpole, C. K., & Slater, P. J. B. (2003). *Bird song: biological themes and variations*. Cambridge University Press.

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E. B., & 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

Chittka, L., Dyer, A. G., Bock, F., & Dornhaus, A. (2003). Bees trade off foraging speed for accuracy. *Nature*, *424*(6947), 388. https://doi.org/10.1038/424388a

Cleasby, I. R., & Nakagawa, S. (2011). Neglected biological patterns in the residuals. *Behavioral Ecology and Sociobiology*, *65*(12), 2361–2372. https://doi.org/10.1007/s00265-011-1254-7

Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, *6*(1), 27–37. https://doi.org/10.1111/2041-210X.12281

Coleman, K., & Wilson, D. S. (1998). Shyness and boldness in pumpkinseed sunfish: individual differences are context-specific. *Animal Behaviour*, *56*(4), 927–936. https://doi.org/https://doi.org/10.1006/anbe.1998.0852

Cramer, E. R. A. (2013a). Measuring consistency: spectrogram cross-correlation versus targeted acoustic parameters. *Bioacoustics*, *22*(3), 247–257. https://doi.org/10.1080/09524622.2013.793616

Cramer, E. R. A. (2013b). Physically Challenging Song Traits, Male Quality, and Reproductive Success in House Wrens. *PLoS ONE*, *8*(3), e59208. https://doi.org/10.1371/journal.pone.0059208

Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004a). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, *7*(8), 734–739. https://doi.org/10.1111/J.1461-0248.2004.00618.X@10.1111/(ISSN)1461-0248.JOINT-DYNAMICS

Dall, S. R. X., Houston, A. I., & McNamara, J. M. (2004b). The behavioural ecology of personality: consistent individual differences from an adaptive perspective. *Ecology Letters*, *7*(8), 734–739. https://doi.org/10.1111/j.1461-0248.2004.00618.x

De Kort, S. R., Eldermire, E. R. B., Valderrama, S., Botero, C. A., & Vehrencamp, S. L. (2009). Trill consistency is an age-related assessment signal in banded wrens. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1665), 2315–2321. https://doi.org/10.1098/rspb.2009.0127

DeWitt, T. J., Sih, A., & Wilson, D. S. (1998). Costs and limits of phenotypic plasticity. *Trends in Ecology and Evolution*, *13*(2), 77–81. https://doi.org/10.1016/S0169-5347(97)01274-3

Dingemanse, N. J., Barber, I., Wright, J., & Brommer, J. E. (2012). Quantitative genetics of behavioural reaction norms: genetic correlations between personality and behavioural plasticity vary across stickleback populations. *Journal of Evolutionary Biology*, *25*(3), 485–496. https://doi.org/10.1111/j.1420-9101.2011.02439.x

Dingemanse, Niels J., Both, C., Drent, P. J., & Tinbergen, J. M. (2004). Fitness consequences of avian personalities in a fluctuating environment. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1541), 847–852. https://doi.org/10.1098/rspb.2004.2680

Dingemanse, Niels J., Both, C., Drent, P. J., Van Oers, K., & Van Noordwijk, A. J. (2002). Repeatability and heritability of exploratory behaviour in great tits from the wild. *Animal Behaviour*, *64*(6), 929–938. https://doi.org/10.1006/anbe.2002.2006

Dingemanse, Niels J., Bouwman, K. M., van de Pol, M., van Overveld, T., Patrick, S. C., Matthysen, E., & Quinn, J. L. (2012). Variation in personality and behavioural plasticity across four populations of the great tit Parus major. *Journal of Animal Ecology*, *81*(1), 116–126. https://doi.org/10.1111/j.1365-2656.2011.01877.x

Dingemanse, Niels J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse, Niels J., Kazem, A. J. N., Réale, D., & Wright, J. (2010a). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89. https://doi.org/10.1016/j.tree.2009.07.013

Dingemanse, Niels J., Kazem, A. J. N., Réale, D., & Wright, J. (2010b). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89. https://doi.org/10.1016/J.TREE.2009.07.013

Dingemanse, Niels J., & Wolf, M. (2010). Recent models for adaptive personality differences: a review. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 3947–3958. https://doi.org/10.1098/rstb.2010.0221

Dingemanse, Niels J., & Wolf, M. (2013). Between-individual differences in behavioural plasticity within populations: Causes and consequences. *Animal Behaviour*, *85*(5), 1031–1039. https://doi.org/10.1016/j.anbehav.2012.12.032

Dingemanse, Niels Jeroen, & de Goede, P. (2004). The relation between dominance and exploratory behavior is context-dependent in wild great tits. *Behavioral Ecology*, *15*(6), 1023–1030. https://doi.org/10.1093/beheco/arh115

Drent, P. J., Van Oers, K., & Van Noordwijk, A. J. (2003). Realized heritability of personalities in the great tit (Parus major). *Proceedings of the Royal Society B: Biological Sciences*, *270*(1510), 45–51. https://doi.org/10.1098/rspb.2002.2168

Duckworth, R. A. (2006). Behavioral correlations across breeding contexts provide a mechanism for a cost of aggression. *Behavioral Ecology*, *17*(6), 1011–1019. https://doi.org/10.1093/beheco/arl035

Duckworth, R. A., & Badyaev, A. V. (2007). Coupling of dispersal and aggression facilitates the rapid range expansion of a passerine bird. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(38), 15017–15022. https://doi.org/10.1073/pnas.0706174104

Duckworth, R. A., Potticary, A. L., & Badyaev, A. V. (2018). On the Origins of Adaptive Behavioral Complexity: Developmental Channeling of Structural Trade-offs. In *Advances in the Study of Behavior* (Vol. 50, pp. 1–36). Academic Press Inc. https://doi.org/10.1016/bs.asb.2017.10.001

Eberhardt, L. S. (1994). Oxygen consumption during singing by male Carolina wrens (Thryothorus ludovicianus). *The Auk*, *111*(1), 124–130.

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Pearson Education India.

Fucikova, E., Drent, P. J., Smits, N., & van Oers, K. (2009). Handling Stress as a Measurement of Personality in Great Tit Nestlings (Parus major). *Ethology*, *115*(4), 366–374. https://doi.org/10.1111/j.1439-0310.2009.01618.x

Garamszegi, L. Z. (2004). Extrapair paternity and the evolution of bird song. *Behavioral Ecology*, *15*(3), 508–519. https://doi.org/10.1093/beheco/arh041

Garson, P. J., & Hunter, M. L. (1979). Effects of temperature and time of year on the singing behaviour of wrens troglodytes troglodytes and great tits parus major. *Ibis*, *121*, 481–487.

Gaunt, A. S., Bucher, T. L., Gaunt, S. L. L., & Baptista, L. F. (1996). Is singing costly? *Auk*, *113*(3), 718–721. https://doi.org/10.2307/4089004

Ghalambor, C. K., McKay, J. K., Carroll, S. P., & Reznick, D. N. (2007). Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, *21*(3), 394–407. https://doi.org/10.1111/j.1365-2435.2007.01283.x

Ghalambor, Cameron K., Angeloni, L. M., & Carroll, S. P. (2010). Behavior as phenotypic plasticity." Evolutionary behavioral ecology. In *Evolutionary behavioral ecology* (pp. 90–107).

Gil, D., & Gahr, M. (2002). The honesty of bird song: multiple constraints for multiple traits. *Trends in Ecology & Evolution*, *17*(3), 133–141. https://doi.org/10.1016/S0169-5347(02)02410-2

Godfrey, J. D., & Bryant, D. M. (2000). State-dependent behaviour and energy expenditure: an experimental study of European robins on winter territories. *Journal of Animal Ecology*, *69*(2), 301–313. https://doi.org/10.1046/j.1365-2656.2000.00393.x

Goldstein, D. L. (1988). Estimates of Daily Energy Expenditure in Birds: The Time-Energy Budget as an Integrator of Laboratory and Field Studies. *American Zoologist*, *28*, 829–844.

Gosler, A. G. (1996). Environmental and Social Determinants of Winter Fat Storage in the Great Tit Parus major. *The Journal of Animal Ecology*, *65*(1), 1. https://doi.org/10.2307/5695

Grava, T., Grava, A., & Otter, K. A. (2012). Vocal performance varies with habitat quality in black-capped chickadees (Poecile atricapillus). *Behaviour*, *149*(1), 35–50. https://doi.org/10.1163/156853912X625854

Guilford, T., & Dawkins, M. S. (1995). What are conventional signals? *Animal Behaviour*, *49*(6), 1689–1695. https://doi.org/10.1016/0003-3472(95)90090-X

Haigh, A., O'Riordan, R., & Butler, F. (2017). Variations in aggression and activity levels amongst squirrels inhabiting low and high density areas. *Ecological Research*, *32*(6), 931–941. https://doi.org/10.1007/s11284-017-1506-8

Hasselquist, D., Bensch, S., & Von Schantz, T. (1996). Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature*, *381*(6579), 229–232. https://doi.org/10.1038/381229a0

Hayes, J. P., & Jenkins, S. H. (1997). Individual Variation in Mammals. *Journal of Mammalogy*, *78*(2), 274–293. https://doi.org/10.2307/1382882

He, R., Pagani-Núñez, E., Chevallier, C., & Barnett, C. R. A. (2017). To be so bold: boldness is repeatable and related to within individual behavioural variability in North Island robins. *Behavioural Processes*, *140*, 144–149. https://doi.org/10.1016/j.beproc.2017.04.014

Helfer, B., & Osiejuk, T. S. (2015). It Takes All Kinds in Acoustic Communication: A New Perspective on the Song Overlapping Phenomenon. *Ethology*, *121*(4), 315–326. https://doi.org/10.1111/eth.12356

Hertel, A. G., Royauté, R., Zedrosser, A., & Mueller, T. (2021). Biologging reveals individual variation in behavioural predictability in the wild. *Journal of Animal Ecology*, *90*(3), 723–737. https://doi.org/10.1111/1365-2656.13406

Highcock, L., & Carter, A. J. (2014). Intraindividual Variability of Boldness Is Repeatable across Contexts in a Wild Lizard. *PLoS ONE*, *9*(4), e95179. https://doi.org/10.1371/journal.pone.0095179

Hoffman, L. (2007). Multilevel Models for Examining Individual Differences in Within-Person Variation and Covariation Over Time. *Multivariate Behavioral Research*, *42*(4), 609–629. https://doi.org/10.1080/00273170701710072

Holveck, M.-J., Vieira de Castro, A. C., Lacklan, R. F., ten Cate, C., & Riebel, K. (2008). Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behavioral Ecology*, *19*(6), 1267–1281.

Hope, S. (1980). Call Form in Relation to Function in the Steller's Jay. *The American Naturalist*, *116*(6), 788–820. https://doi.org/10.1086/283670

Houston, A. I., & McNamara, J. M. (1999). *Models of adaptive behaviour: an approach based on state*. Cambridge University Press.

Houtmann, A. M. (1992). *Sexual selection in the zebra finch (Poephila guttata).*

Hutfluss, A., Rohr, V. A., Scheidt, S., Steinbichl, L., Bermúdez-Cuamatzin, E., Slabbekoorn, H., & Dingemanse, N. J. (2021). Does song overlap signal aggressiveness? An experimental study with repeated measures in free-ranging great tits. *Animal Behaviour*, *179*, 199–211. https://doi.org/10.1016/j.anbehav.2021.07.006

Iserbyt, A., Eens, M., Baetens, W., Vermeulen, A., & Müller, W. (2017). Within- and between-individual (co)variance partitioning reveals limited pleiotropic effects of testosterone on immune function, sexual signaling, and parental investment. *Behavioral Ecology and Sociobiology*, *71*(4), 1–12. https://doi.org/10.1007/s00265-017-2308-2

Jablonszky, M., Zsebők, S., Laczi, M., Nagy, G., Vaskuti, É., & Garamszegi, L. Z. (2021). The effect of social environment on bird song: listener-specific expression of a sexual signal. *Behavioral Ecology*, 1–12. https://doi.org/10.1093/beheco/araa132

Jolles, J. W., Briggs, H. D., Araya-Ajoy, Y. G., & Boogert, N. J. (2019). Personality, plasticity and predictability in sticklebacks: bold fish are less plastic and more predictable than shy fish. *Animal Behaviour*, *154*, 193–202. https://doi.org/10.1016/j.anbehav.2019.06.022

Jones, K. A., & Godin, J. G. J. (2010). Are fast explorers slow reactors? Linking personal type and anti-predator behavior. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1681), 625–632. https://doi.org/10.1098/rspb.2009.1607

Jones, K. A., Jackson, A. L., & Ruxton, G. D. (2011). Prey jitters; protean behaviour in grouped prey. *Behavioral Ecology*, *22*(4), 831–836.

Kelly, C. D. (2006, September 19). Replicating empirical research in behavioral ecology: How and why it should be done but rarely ever is. *Quarterly Review of Biology*. The University of Chicago Press . https://doi.org/10.1086/506236

Klappert, K., & Reinhold, K. (2003). Acoustic preference functions and sexual selection on the male calling song in the grasshopper Chorthippus biguttulus. *Animal Behaviour*, *65*(1), 225–233. https://doi.org/http://dx.doi.org/10.1006/anbe.2002.2034

Klump, G. M., & Shalter, M. D. (1984). Acoustic Behaviour of Birds and Mammals in the Predator Context; I. Factors Affecting the Structure of Alarm Signals. II. The Functional Significance and Evolution of Alarm Signals. *Zeitschrift Für Tierpsychologie*, *66*(3), 189–226. https://doi.org/10.1111/j.1439-0310.1984.tb01365.x

Knell, R. J. (2009). Population density and the evolution of male aggression. *Journal of Zoology*, *278*(2), 83–90. https://doi.org/10.1111/j.1469-7998.2009.00566.x

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

Kubli, S. P., & MacDougall-Shackleton, E. A. (2014). Developmental timing of signals affects information content: song complexity but not consistency reflects innate immune strategy in male song sparrows. *The American Naturalist*, *183*(5), 660–670. https://doi.org/10.1086/675757

Kunc, H. P., Amrhein, V., & Naguib, M. (2006). Vocal interactions in nightingales, Luscinia megarhynchos: more aggressive males have higher pairing success. *Animal Behaviour*, *72*(1), 25–30. https://doi.org/10.1016/J.ANBEHAV.2005.08.014

Ladds, M. A., Salton, M., Hocking, D. P., McIntosh, R. R., Thompson, A. P., Slip, D. J., & Harcourt, R. G. (2018). Using accelerometers to develop time-energy budgets of wild fur seals from captive surrogates. *PeerJ*, *6*, e5814. https://doi.org/10.7717/peerj.5814

Lambrechts, M. M. (1997). Song frequency plasticity and composition of phrase versions in great tits Parus major. *Ardea*, *85*(1), 99–109.

Lambrechts, M. M., & Dhondt, A. A. (1987). Differences in singing performance between male great tits. *Ardea*, *75*(1), 43–52.

Lambrechts, M. M., & Dhondt, A. A. (1988). The anti-exhaustion hypothesis: a new hypothesis to explain song performance and song switching in the great tit. *Animal Behaviour*, *36*, 327–334.

Lande, R., & Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution*, *37*(6), 1210. https://doi.org/10.2307/2408842

Lane, S. M., & Briffa, M. (2020). The role of spatial accuracy and precision in hermit crab contests. *Animal Behaviour*, *167*, 111–118. https://doi.org/10.1016/j.anbehav.2020.07.013

Lord, A., Waas, J. R., Innes, J., & Whittingham, M. J. (2001). Effects of human approaches to nests of northern New Zealand dotterels. *Biological Conservation*, *98*(2), 233–240. https://doi.org/10.1016/S0006-3207(00)00158-0

MacDonald, S. W. S., Nyberg, L., & Bäckman, L. (2006, August 1). Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*. Elsevier Current Trends. https://doi.org/10.1016/j.tins.2006.06.011

Mangel, M., & Stamps, J. (2001). *Trade-offs between growth and mortality and the maintenance of individual variation in growth*. *Evolutionary Ecology Research* (Vol. 3). Evolutionary Ecology, Ltd.

Marler, P., & Slabbekoorn, H. W. (2004). *Nature's music : the science of birdsong*. Elsevier Academic.

Martín-Vivaldi, M., Palomino, J. J., Soler, M., & Martínez, J. G. (2008). Song strophe-length and reproductive success in a non-passerine bird, the Hoopoe Upupa epops. *Ibis*, *141*(4), 670–679. https://doi.org/10.1111/j.1474-919x.1999.tb07375.x

Martin, J. G. A., Pirotta, E., Petelle, M. B., & Blumstein, D. T. (2017). Genetic basis of between-individual and within-individual variance of docility. *Journal of Evolutionary Biology*, *30*(4), 796–805. https://doi.org/10.1111/jeb.13048

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

Mathot, K. J., Olsen, A.-L., Mutzel, A., Araya-Ajoy, Y. G., Nicolaus, M., Westneat, D. F., … Dingemanse, N. J. (2017). Provisioning tactics of great tits (Parus major) in response to long-term brood size manipulations differ across years. *Behavioral Ecology*, *28*(6), 1402–1413. https://doi.org/10.1093/beheco/arx083

Mathot, K. J., Wright, J., Kempenaers, B., & Dingemanse, N. J. (2012). Adaptive strategies for managing uncertainty may explain personality-related differences in behavioural plasticity. *Oikos*, *121*(7), 1009–1020. https://doi.org/10.1111/j.1600-0706.2012.20339.x

McCrae, R. R., Costa, P. T., John, O. P., Robins, R. W., & Pervin, L. A. (1999). *Handbook of personality: Theory and research.* New York: NY: Guilford.

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/10.1086/283811

McNamara, J. M., & Houston, A. I. (1986). The Common Currency for Behavioral Decisions. *The American Naturalist*, *127*(3), 358–378. https://doi.org/10.1086/284489

McNamara, J. M., Stephens, P. A., Dall, S. R. ., & Houston, A. I. (2009). Evolution of trust and trustworthiness: social awareness favours personality differences. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1657), 605–613. https://doi.org/10.1098/rspb.2008.1182

Mitchell, D. J., Beckmann, C., & Biro, P. A. (2021). Understanding the unexplained: The magnitude and correlates of individual differences in residual variance. *Ecology and Evolution*, ece3.7603. https://doi.org/10.1002/ece3.7603

Mitchell, D. J., Dujon, A. M., Beckmann, C., & Biro, P. A. (2020). Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity. *Behavioral*

Ecology, 31(1), 222–231.

Mitchell, D. J., Fanson, B. G., Beckmann, C., & Biro, P. A. (2016). Towards powerful experimental and statistical approaches to study intraindividual variability in labile traits. *Royal Society Open Science, 3*(10). https://doi.org/10.1098/rsos.160352

Morand-Ferron, J., Hamblin, S., Cole, E. F., Aplin, L. M., & Quinn, J. L. (2015). Taking the Operant Paradigm into the Field: Associative Learning in Wild Great Tits. *PLOS ONE, 10*(8), e0133821. https://doi.org/10.1371/journal.pone.0133821

Mulder, H. A., Gienapp, P., & Visser, M. E. (2016). Genetic variation in variability: Phenotypic variability of fledging weight and its evolution in a songbird population. *Evolution*, *70*(9), 2004–2016. https://doi.org/10.1111/evo.13008

Murphy, M. T., Sexton, K., Dolan, A. C., & Redmond, L. J. (2008). Dawn song of the eastern kingbird: an honest signal of male quality? *Animal Behaviour*, *75*(3), 1075–1084. https://doi.org/10.1016/j.anbehav.2007.08.020

Naguib, M., Altenkamp, R., & Griessmann, B. (2001). Nightingales in space: song and extra-territorial forays of radio tagged song birds. *Journal of Ornithology, 142*, 306–312.

Naguib, M., Diehl, J., van Oers, K., & Snijders, L. (2019). Repeatability of signalling traits in the avian dawn chorus. *Frontiers in Zoology*, *16*(1), 27. https://doi.org/10.1186/s12983-019-0328-7

Naguib, M., Kazek, A., Schaper, S. V., Van Oers, K., & Visser, M. E. (2010). Singing Activity Reveals Personality Traits in Great Tits. *Ethology*, *116*(8), no-no. https://doi.org/10.1111/j.1439-0310.2010.01791.x

Natarajan, D., De Vries, H., Saaltink, D. J., De Boer, S. F., & Koolhaas, J. M. (2009). Delineation of violence from functional aggression in mice: An ethological approach. *Behavior Genetics*, *39*(1), 73–90. https://doi.org/10.1007/s10519-008-9230-3

Niemelä, P. T., & Dingemanse, N. J. (2017). Individual versus pseudo-repeatability in behaviour: Lessons from translocation experiments in a wild insect. *Journal of Animal Ecology*, *86*(5), 1033–1043. https://doi.org/10.1111/1365-2656.12688

Nowicki, S., Searcy, W. A., Krueger, T., & Hughes, M. (2002). Individual variation in response to simulated territorial challenge among territory-holding song sparrows. *Journal of Avian Biology*, *33*(3), 253–259. https://doi.org/10.1034/j.1600-048X.2002.330307.x

Nussey, D. H., Wilson, A. J., & Brommer, J. E. (2007). The evolutionary ecology of individual phenotypic plasticity in wild populations. *Journal of Evolutionary Biology*, *20*(3), 831–844. https://doi.org/10.1111/j.1420-9101.2007.01300.x

O’Shea, W., Serrano-Davies, E., & Quinn, J. L. (2017). Do personality and innovativeness influence competitive ability? An experimental test in the great tit. *Behavioral Ecology*, *28*(6), 1435–1444. https://doi.org/10.1093/beheco/arx104

Oberweger, K., & Goller, F. (2001). The metabolic cost of birdsong production. *The Journal of Experimental Biology*, *204*, 3379–3388.

Ouyang, J. Q., Sharp, P., Quetting, M., & Hau, M. (2013). Endocrine phenotype, reproductive success and survival in the great tit, Parus major. *Journal of Evolutionary Biology*, *26*(9), 1988–1998. https://doi.org/10.1111/jeb.12202

Paul, N., Thompson, M. J., & Foote, J. R. (2020). Characterising the flight song: repeatable individual variation of Ovenbird song features. *Bioacoustics*. https://doi.org/10.1080/09524622.2020.1718552

Penteriani, V. (2003). Breeding density affects the honesty of bird vocal displays as possible indicators of male/territory quality. *Ibis*, *145*(3), E127–E135. https://doi.org/10.1046/j.1474-919X.2003.00173.x

Perrins, C. M., & McCleery, R. H. (1989). Laying dates and clutch size in the great tit. *The Wilson Bulletin*, 236–253.

Pettifor, R. A., Perrins, C. M., & McCleery, R. H. (1988). Individual optimization of clutch size in great tits. *Nature*, *336*(6195), 160–162. https://doi.org/10.1038/336160a0

Piersma, T., & Drent, J. (2003). Phenotypic flexibility and the evolution of organismal design. *Trends in Ecology and Evolution*, *18*(5), 228–233. https://doi.org/10.1016/S0169-5347(03)00036-3

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution*, *51*(2), 537–551. https://doi.org/10.1111/j.1558-5646.1997.tb02441.x

Pohl, N. U., Leadbeater, E., Slabbekoorn, H., Klump, G. M., & Langemann, U. (2012). Great tits in urban noise benefit from high frequencies in song detection and discrimination. *Animal Behaviour*, *83*(3), 711–721. https://doi.org/http://dx.doi.org/10.1016/j.anbehav.2011.12.019

Prentice, P. M., Houslay, T. M., Martin, J. G. A., & Wilson, A. J. (2020). Genetic variance for behavioural ‘predictability’ of stress response. *Journal of Evolutionary Biology*, *33*(5), 642–652. https://doi.org/10.1111/jeb.13601

Reale, D., & Dingemanse, N. J. (2010). Selection and evolutionary explanations for the maintenance of personality differences. In *The evolution of personality and individual differences* (pp. 527–557). Oxford University Press.

Réale, D., Garant, D., Humphries, M. M., Bergeron, P., Careau, V., & Montiglio, P. O. (2010). Personality and the emergence of the pace-of-life syndrome concept at the population level. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4051–4063. https://doi.org/10.1098/rstb.2010.0208

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Reid, M. L. (1987). Costliness and reliability in the and singing vigour of Ipswich sparrows. *Animal Behaviour*, *35*, 1735–1743.

Reinertsen, R. E. (1983). Nocturnal hypothermia and its energetic significance for small birds living in the arctic and subarctic regions. A review. *Polar Research*, *1*(3), 269–284. https://doi.org/10.1111/j.1751-8369.1983.tb00743.x

Reinertsen, R. E., & Haftorn, S. (1986). Different metabolic strategies of northern birds for nocturnal survival. *Journal of Comparative Physiology B*, *156*(5), 655–663.

https://doi.org/10.1007/BF00692743

Ripmeester, E. A. P., de Vries, A. M., & Slabbekoorn, H. (2007). Do Blackbirds Signal Motivation to Fight with Their Song? *Ethology*, *113*(11), 1021–1028. https://doi.org/10.1111/j.1439-0310.2007.01398.x

Ripmeester, E. A. P., Kok, J. S., van Rijssel, J. C., & Slabbekoorn, H. (2010). Habitat-related birdsong divergence: A multi-level study on the influence of territory density and ambient noise in European blackbirds. *Behavioral Ecology and Sociobiology*, *64*(3), 409–418. https://doi.org/10.1007/s00265-009-0857-8

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). Songs differing in consistency elicit differential aggressive response in territorial birds. *Biology Letters*, *7*(3), 339–342. https://doi.org/10.1098/rsbl.2010.0962

Roff, D. A., & Fairbairn, D. J. (2007). The evolution of trade-offs: where are we? *Journal of Evolutionary Biology*, *20*(2), 433–447. https://doi.org/10.1111/j.1420-9101.2006.01255.x

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of residual variance - Estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution*, *42*(1), 1–10. https://doi.org/10.1186/1297-9686-42-8

Royle, N. J., Schuett, W., & Dall, S. R. X. (2010). Behavioral consistency and the resolution of sexual conflict over parental investment. *Behavioral Ecology*, *21*(6), 1125–1130.

Rytkonen, S. O., Orell, M. I., Koivula, K. T., & Welling, P. P. (2008). Song Rate Correlates With Paternal Care and Survival in Willow Tits: Advertisement of Male Quality? *Behaviour*, *134*(11–12), 891–904. https://doi.org/10.1163/156853997x00214

Saggese, K., Korner-Nievergelt, F., Slagsvold, T., & Amrhein, V. (2011). Wild bird feeding delays start of dawn singing in the great tit. *Animal Behaviour*, *81*(2), 361–365. https://doi.org/10.1016/j.anbehav.2010.11.008

Sakata, J. T., Hampton, C. M., & Brainard, M. S. (2008). Social Modulation of Sequence and Syllable Variability in Adult Birdsong. *Journal of Neurophysiology*, *99*(4), 1700–1711. https://doi.org/10.1152/jn.01296.2007

Sakata, J. T., & Vehrencamp, S. L. (2012, January 15). Integrating perspectives on vocal performance and consistency. *Journal of Experimental Biology*. The Company of Biologists Ltd. https://doi.org/10.1242/jeb.056911

Sánchez-Virosta, P., Espín, S., Ruiz, S., Panda, B., Ilmonen, P., Schultz, S. L., … Eeva, T. (2020). Arsenic-related oxidative stress in experimentally-dosed wild great tit nestlings. *Environmental Pollution*, *259*, 113813. https://doi.org/10.1016/j.envpol.2019.113813

Schluter, D. (1996). Adaptive radiation along genetic lines of least resistance. *Evolution*, *50*(5), 1766–1774. https://doi.org/10.1111/j.1558-5646.1996.tb03563.x

Schuett, W., Tregenza, T., & Dall, S. R. X. (2010). Sexual selection and animal personality.

Biological Reviews, *85*(2), 217–246. https://doi.org/10.1111/j.1469-185X.2009.00101.x

Searcy, W A, & Andersson, M. (1986). Sexual Selection and the Evolution of Song. *Annual Review of Ecology and Systematics*, *17*(1), 507–533. https://doi.org/10.1146/annurev.es.17.110186.002451

Searcy, William A., & Beecher, M. D. (2009, December). Song as an aggressive signal in songbirds. *Animal Behaviour*. https://doi.org/10.1016/j.anbehav.2009.08.011

Searcy, William A., & Beecher, M. D. (2011). Continued scepticism that song overlapping is a signal. *Animal Behaviour*, *81*(1), e1–e4. https://doi.org/10.1016/j.anbehav.2010.10.013

Sih, A. (2004). A behavioral ecological view of phenotypic plasticity. In *Phenotypic Plasticity: Functional and Conceptual Approaches* (pp. 112–125).

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*(7), 372–378. https://doi.org/10.1016/J.TREE.2004.04.009

Sih, A., Bell, A. M., Johnson, J. C., & Ziemba, R. E. (2004, September 19). Behavioral syndromes: An integrative overview. *Quarterly Review of Biology*. The University of Chicago Press. https://doi.org/10.1086/422893

Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: A behavioural ecology perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society. https://doi.org/10.1098/rstb.2012.0216

Slagsvold, T., Dale, S., & Sætre, G.-P. (1994). Dawn singing in the great tit (parus major): mate attraction, mate guarding, or territorial defence? *Behaviour*, *131*(1–2), 115–138.

Snell-Rood, E. C. (2013). An overview of the evolutionary causes and consequences of behavioural plasticity. *Animal Behaviour*, *85*(5), 1004–1011. https://doi.org/10.1016/j.anbehav.2012.12.031

Snijders, L., van Rooij, E. P., Henskens, M. F. A., van Oers, K., & Naguib, M. (2015). Dawn song predicts behaviour during territory conflicts in personality-typed great tits. *Animal Behaviour*, *109*, 45–52. https://doi.org/10.1016/j.anbehav.2015.07.037

Soma, M., & Garamszegi, L. Z. (2011). Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. *Behavioral Ecology*, *22*(2), 363–371. https://doi.org/10.1093/beheco/arq219

Songvorawit, N., Butcher, B. A., & Chaisuekul, C. (2018). Resource Holding Potential and the Outcome of Aggressive Interactions between Paired Male Aegus chelifer chelifer (Coleoptera: Lucanidae) Stag Beetles. *Journal of Insect Behavior*, *31*(4), 347–360. https://doi.org/10.1007/s10905-018-9683-z

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

Stamps, J. A. (2016). Individual differences in behavioural plasticities. *Biological Reviews*, *91*(2), 534–567. https://doi.org/10.1111/brv.12186

Stamps, J. A., Briffa, M., & Biro, P. A. (2012). Unpredictable animals: individual differences in intraindividual variability (IIV). *Animal Behaviour*, *83*(6), 1325–1334. https://doi.org/10.1016/J.ANBEHAV.2012.02.017

Stearns, S. C. (1989). Trade-Offs in Life-History Evolution. *Functional Ecology*, *3*(3), 259. https://doi.org/10.2307/2389364

Stephens, D. W., Brown, J. S., & Ydenberg, R. C. (Eds.). (2008). *Foraging: behavior and ecology*. University of Chicago Press.

Strain, J. G., & Mumme, R. L. (1988). Effects of Food Supplementation, Song Playback, and Temperature on Vocal Territorial Behavior of Carolina Wrens. *The Auk*, *105*(1), 11–16.

Strauß, A. F. T., Hutfluss, A., & Dingemanse, N. J. (2020). Great tits responding to territorial intrusions sing less but alarm more on colder days. *Ethology*, *126*(4), 445–454. https://doi.org/10.1111/eth.12989

Suter, S. M., Ermacora, D., Rieille, N., & Meyer, D. R. (2009). A distinct reed bunting dawn song and its relation to extrapair paternity. *Animal Behaviour*, *77*, 473–480. https://doi.org/10.1016/j.anbehav.2008.11.002

Suthers, R. A., & Zollinger, S. A. (2004). Producing Song: The Vocal Apparatus. *Annals of the New York Academy of Sciences*, *1016*(1), 109–129. https://doi.org/10.1196/annals.1298.041

Taborsky, M., Oliveira, R. F., & Brockmann, H. J. (2008). The evolution of alternative reproductive tactics: concepts and questions. In *Alternative reproductive tactics: an integrative approach*.

Taff, C. C., Steinberger, D., Clark, C., Belinsky, K., Sacks, H., Freeman-Gallant, C. R., … Whittingham, L. A. (2012). Multimodal sexual selection in a warbler: Plumage and song are related to different fitness components. *Animal Behaviour*, *84*(4), 813–821. https://doi.org/10.1016/j.anbehav.2012.07.002

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature*, *189*(4766), 732–735. https://doi.org/10.1038/189732a0

Thomas, R. J. (1999). Two tests of a stochastic dynamic programming model of daily singing routines in birds. *Animal Behaviour*, *57*(2), 277–284. https://doi.org/10.1006/ANBE.1998.0956

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift Für Tierpsychologie*, *20*(4), 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Van De Pol, M., Bruinzeel, L. W., Heg, D., Van Der Jeugd, H. P., & Verhulst, S. (2006). A silver spoon for a golden future: Long-term effects of natal origin on fitness prospects of oystercatchers (Haematopus ostralegus). *Journal of Animal Ecology*, *75*(2), 616–626. https://doi.org/10.1111/j.1365-2656.2006.01079.x

van Oers, K., Buchanan, K. L., Thomas, T. E., & Drent, P. J. (2011). Correlated response to selection of testosterone levels and immunocompetence in lines selected for avian personality. *Animal Behaviour*, *81*(5), 1055–1061. https://doi.org/10.1016/j.anbehav.2011.02.014

Vergne, A. L., Avril, A., Martin, S., & Mathevon, N. (2007). Parent--offspring communication in the Nile crocodile Crocodylus niloticus: do newborns' calls show an individual signature? *Naturwissenschaften, 94*(1), 49–54. https://doi.org/10.1007/s00114-006-0156-4

Verhagen, I., Tomotani, B. M., Gienapp, P., & Visser, M. E. (2020). Temperature has a causal and plastic effect on timing of breeding in a small songbird. *Journal of Experimental Biology, 223*(8). https://doi.org/10.1242/jeb.218784

Viswanathan, M. (2005). *Measurement error and research design.* Sage.

Ward, S., & Slater, P. J. B. (2005). Heat transfer and the energetic cost of singing by canaries Serinus canaria. *Journal of Comparative Physiology A, 191*(10), 953–964. https://doi.org/10.1007/s00359-005-0022-4

Westneat, D. F., & Fox, W. C. (Eds.). (2010). *Evolutionary behavioral ecology*. Oxford University Press.

Westneat, D. F., Hatch, M. I., Wetzel, D. P., & Ensminger, A. L. (2011). Individual variation in parental care reaction norms: Integration of Personality and plasticity. *American Naturalist*, *178*(5), 652–667. https://doi.org/10.1086/662173

Westneat, D. F., Mutzel, A., Bonner, S., & Wright, J. (2017). Experimental manipulation of brood size affects several levels of phenotypic variance in offspring and parent pied flycatchers. *Behavioral Ecology and Sociobiology, 71*(6), 1–13. https://doi.org/10.1007/s00265-017-2315-3

Westneat, D. F., Schofield, M., & Wright, J. (2013). Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology, 24*(3), 598–604. https://doi.org/10.1093/beheco/ars207

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743. https://doi.org/10.1111/brv.12131

Williams, G. C. (1966). Natural Selection, the Costs of Reproduction, and a Refinement of Lack’s Principle. *The American Naturalist*, *100*(916), 687–690. https://doi.org/10.1086/282461

Wilson, D. S. (1998). Adaptive individual differences within single populations. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353*(1366), 199–205. https://doi.org/10.1098/rstb.1998.0202

Wolf, L. L., & Hainsworth, F. R. (1971). Time and Energy Budgets of Territorial Hummingbirds. *Ecology, 52*(6), 980–988. https://doi.org/10.2307/1933803

Wolf, M., Van Doorn, G. S., & Weissing, F. J. (2011). On the coevolution of social responsiveness and behavioural consistency. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1704), 440–448. https://doi.org/10.1098/rspb.2010.1051

Wolf, Max, Van Doorn, G. S., & Weissing, F. J. (2008). Evolutionary emergence of responsive and unresponsive personalities. *Proceedings of the National Academy of Sciences of the United States of America, 105*(41), 15825–15830. https://doi.org/10.1073/pnas.0805473105

Wolf, Max, Van Doorn, G. S., & Weissing, F. J. (2011). On the coevolution of social responsiveness and behavioural consistency. *Proceedings of the Royal Society B:*

Biological Sciences, *278*(1704), 440–448. https://doi.org/10.1098/rspb.2010.1051

Wolf, Max, & Weissing, F. J. (2010). An explanatory framework for adaptive personality differences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 3959–3968. https://doi.org/10.1098/rstb.2010.0215

Wolf, Max, & Weissing, F. J. (2012). Animal personalities: Consequences for ecology and evolution. *Trends in Ecology and Evolution*, *27*(8), 452–461. https://doi.org/10.1016/j.tree.2012.05.001

Wu, Y., Whiting, M. J., Fu, J., & Qi, Y. (2019). The driving forces behind female-female aggression and its fitness consequence in an Asian agamid lizard. *Behavioral Ecology and Sociobiology*, *73*(6). https://doi.org/10.1007/s00265-019-2686-8

Ydenberg, R. C. (1986). The economics of fleeing from predators. *Advances in the Study of Behavior*, *16*(C), 229–249.

Zsebők, S., Herczeg, G., Blázi, G., Laczi, M., Nagy, G., Szász, E., … Garamszegi, L. Z. (2017). Short- and long-term repeatability and pseudo-repeatability of bird song: sensitivity of signals to varying environments. *Behavioral Ecology and Sociobiology*, *71*(10), 154. https://doi.org/10.1007/s00265-017-2379-0

Zsebők, S., Moskát, C., & Bán, M. (2017). Individually distinctive vocalization in Common Cuckoos (Cuculus canorus). *Journal of Ornithology*, *158*(1), 213–222. https://doi.org/10.1007/s10336-016-1376-9

Appendices

Publication I: Great tits responding to territorial intrusions sing less but alarm more on colder days

Aurelia F.T. Strauß, Alexander Hutfluss & Niels J. Dingemanse

Published in Behavioural Ecology, 2020

https://doi.org/10.1111/eth.12989

Received: 27 August 2019 | Revised: 30 October 2019 | Accepted: 31 October 2019
DOI: 10.1111/eth.12989

RESEARCH PAPER

ethology WILEY

Great tits responding to territorial intrusions sing less but alarm more on colder days

Aurelia F. T. Strauß | **Alexander Hutfluss** | **Niels J. Dingemanse**

Department Biology II, Behavioural Ecology, Ludwig-Maximilians-Universität (LMU) Munich, Planegg-Martinsried, Germany

Correspondence
Alexander Hutfluss, Behavioural Ecology, Department Biology II, Ludwig-Maximilians-Universität (LMU) Munich, 82152 Planegg-Martinsried, Germany.
Email: alexander.hutfluss@gmx.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: DI 1694/1-1

Editor: Wolfgang Goymann

Abstract

Bird song transmits information required to defend territories and attract mates. These functions contribute to fitness by affecting survival and reproductive success. Singing is also costly due to physiological costs. We used observational data to evalu- ate support for the hypothesis that lower temperatures result in decreased singing behaviour in wild great tits due to increased energy consumption during cold condi- tions required for thermoregulation. More than 6,500 simulated territorial intrusions were performed over an 8-year period in twelve nest box populations of great tits *Parus major* south of Munich, Germany. We measured song rate as well as the number of alarm calls and the aggressive response of territorial males to a simulated territo- rial intrusion. We found a decrease in song rate with decreasing current temperature, but also a concurrent increase in the number of alarm calls. Night temperature did not affect these acoustic traits. We conclude that warmer conditions allow birds to choose more energetically expensive (yet functionally superior) activities during ter- ritorial intrusions, thereby facilitating avoidance of physical aggressiveness during territorial intrusions.

KEYWORDS

alarm call, energy consumption, simulated territorial intrusions, song, temperature, trade-off

1 | INTRODUCTION

Acoustic communication is the active transmission of information from sender to receiver (Bright, 1985; Simmons, Popper, & Fay, 2005). Such information can be specific to properties of the sender, such as its age, size or identity, and, ultimately, vocalisations can affect the receiver's behaviour (Aubin, Mathevon, Silva, Vielliard, & Sebe, 2004; Simmons et al., 2005; Vergne, Avril, Martin, & Mathevon, 2007). Acoustic communication has generally been asso- ciated with various life-history traits including survival (Zuberbühler, 2001), mate attraction (Klappert & Reinhold, 2003; Suter, Ermacora, Rieille, & Meyer, 2009) and territory defence (e.g. Amrhein & Lerch, 2010). Many birds use song to defend territories, not only against takeovers, but also to discourage extra-pair mating attempts made by other males (Naguib, Altenkamp, & Griessmann, 2001; Slagsvold, Dale, & Sætre, 1994).

There is considerable variation (both among and within-indi- viduals) in singing behaviour. This variation exists, in part, because singing also carries costs (Gil & Gahr, 2002). First, singing increases predation risk because vocalising individuals are more traceable and often sing from a position preferred for sound transmission that is consequently relatively exposed (Bright, 1985; Simmons et al., 2005). As a result, birds might reduce song output or choose suboptimal singing locations when perceiving increases in preda- tion levels (Abbey-Lee, Mathot, & Dingemanse, 2016; Campos, Bander, Raksi, & Blumstein, 2009; Møller, Nielsen, & Garamszegi, 2006). Second, singing is energetically demanding as it increases physiological activity (Hasselquist & Bensch, 2008; Oberweger & Goller, 2001; Ward & Slater, 2005). Inspiration, expiration and the contraction of muscles coordinate the production of acoustic

---

Strauß and Hutfluss should be considered joint first authors.

signals. Consequently, the amplitude of muscle contractions during singing is significantly higher than during resting (Suthers, Goller, & Pytte, 1999). Ultimately, prolonged singing can lead to exhaustion. Indeed, metabolic rate (i.e. oxygen consumption and heat transfer) is increased during singing compared with resting (Oberweger & Goller, 2001; Ward & Slater, 2005). Third, singing and foraging (as well as other behaviours) are mutually exclusive activities (Strain & Mumme, 1988; Thomas et al., 2003). Consequently, animals have to choose how much of their limited time they invest in each be- haviour. Due to these time budgets, physiological costs and limited energy reserves, singing must constantly be traded off with other energetically demanding activities.

Environmental factors can affect the relative balance between the aforementioned costs and benefits of singing. Such factors include weather conditions, altering sound transmission (Bright, 1985) or environmental and anthropogenic noise, masking sig- nals and leading to altered song characteristics such as shifted frequency (Bright, 1985; Pohl, Leadbeater, Slabbekoorn, Klump, & Langemann, 2012). Food availability represents another envi- ronmental component that may increase accessible energy, and reduce time required for foraging, thereby positively affecting in- vestments made in singing (Barnett & Briskie, 2007; Reid, 1987; Strain & Mumme, 1988; Thomas, 1999). Importantly, the amount of required and available resources, that is both food and energy reserves, varies as a function of environmental conditions such as temperature. Current temperatures are strongly linked to ther- moregulatory costs (Broggi et al., 2007; Dubois, Hallot, & Vézina, 2016; Kendeigh, 1944). Higher energetic demands during periods of lower temperatures can result in a decrease in body weight and fat reserves, and potentially hypothermia (Reinertsen, 1983; Reinertsen & Haftorn, 1986), which consequently increases time required for foraging (Dubois et al., 2016; Reid, 1987). Ultimately, this leads to a trade-off between foraging and singing, especially in colder conditions, when singing activity is consequently de- creased (Garson & Hunter, 1979; Godfrey & Bryant, 2000; Reid, 1987; Thomas, 1999). Such temperature effects on singing can ei- ther be caused by immediate effects of the current temperature (Gottlander, 1987; Strain & Mumme, 1988) or by carry-over ef- fects of temperatures during preceding nights (Garson & Hunter, 1979; Reid, 1987; Thomas, 1999). Importantly, these two environ- mental factors are usually strongly correlated (Garson & Hunter, 1979), which may make it difficult to tease apart their respective effects (Naguib, Diehl, van Oers, & Snijders, 2019), particularly in studies with small sample sizes.

Most studies of acoustic communication in this context focus solely on singing. However, acoustic signals can come in the forms of either singing *or* alarming (Araya-Ajoy & Dingemanse, 2014; Morton, 1977 and citations therein). While songs are often elab- orate and complex signals, alarm calls are in general short and simple, following stereotypic patterns (Marler, 2004). Calls have been observed and studied in various contexts, such as begging (Thielcke, 1976; Wilkinson, 1980), foraging (Bugnyar, Kijne, & Kotrschal, 2001; Williams, 1969) and predator defence (Hope, 1980; Klump & Shalter, 2010). However, even though calls have been reported to be produced during territory defence in some species (Morton, 1977 and citations therein), quantifications of the occurrence and their importance in relation to singing are rare (Araya-Ajoy & Dingemanse, 2014; Araya-Ajoy et al., 2016). In wild great tits, alarms are frequently produced during territorial intru- sions, while being negatively correlated with singing (Araya-Ajoy & Dingemanse, 2014). That is, birds that sing much call little and vice versa; within-individual up-regulations in singing are asso- ciated with within-individual down-regulations of alarm calling, likely due to time or energy allocation trade-offs (Araya-Ajoy & Dingemanse, 2014). Furthermore, direct comparisons between singing and alarming within species are missing, but cross-species comparison implies that alarming (and other similar calls) is ener- getically less costly than singing (compare Eberhardt, 1994 and Jurisevic, Sanderson, & Baudinette, 1999). In fact, some studies were unable to report significant energetic costs associated with alarming (Chappell, Zuk, Kwan, & Johnsen, 1995; Horn, Leonard, & Weary, 1995).

A key question is whether environmental conditions that af- fect energetic trade-offs, such as temperature, influence acoustic behaviours chosen for territory defence. Specifically, we ask how singing, alarming (which has been neglected in the context of ter- ritory defence) and the relationship between these two types of acoustic behaviours are altered by temperature. We propose here two scenarios for the effects of temperature, positing a key role for energy demands associated with each behaviour. In both sce- narios, two trade-offs are assumed: a temporal trade-off, which means more time spent on one behaviour results in less time for the other and an energetic trade-off, which means that colder temperatures increase thermoregulatory costs leaving less energy reserves for vocalisations. If alarming and singing do not differ in energetic costs, we predict that the number of vocalisations in- creases with increased temperatures without causing a shift from alarming towards singing or *vice versa* (Scenario 1). If alarming is, by contrast, less costly than singing, we expect birds to vocalise at a similar rate, whereas singing is down-regulated, but alarming up-regulated with lower temperatures (Scenario 2). These predic- tions are made assuming that any acoustic response to a territo- rial intrusion is preferable to not responding at all, which seems a reasonable assumption as non-responsive birds may lose their territory and purely behavioural responses without any type of vocalisation are extremely rare (<1%).

Here, we investigated support for these two alternative sce- narios by using a longitudinal (8 years) observational data set of simulated territorial intrusions conducted in twelve nest box pop- ulations of great tits. We hypothesised (a) that birds would sing less with lower temperatures due to depleted energy reserves and lower foraging efficiency and (b) that birds would instead alarm more (assuming lower energetic costs of this behaviour compared with singing; see above), thereby supporting Scenario 2. While evaluating the support for our hypotheses, we also investigated whether effects of temperature were attributable to current versus night temperature, which was uniquely possible because of the large sample size of our study. Finally, assuming birds replenish energy reserves during foraging in the morning hours, we predicted (c) that any effect of night temperature on acoustic output would diminish with increasing time between sunrise and a focal behavioural test.

2 | METHODS

2.1 | Study Populations and Territorial Intrusions

We studied twelve nest box populations located in southern Germany, each containing 50 nest boxes distributed in a grid with 50-m distance between adjacent boxes. Nest boxes were checked (bi-)weekly from the beginning of April till August and lay date (back-calculated assuming one egg laid per day), and clutch size was recorded. Seven or ten days after the eggs hatched, the parents were caught, morphologically measured and colour-ringed to allow individual identification (for further details see Nicolaus et al., 2015).

2.2 | Territorial intrusion tests

From 2010 onwards, each male was subjected to four simulated territorial intrusions during its first breeding attempt (defined as attempts initiated within 30 days after the first egg of the year in all plots found; Nicolaus, Both, Ubels, Edelaar, & Tinbergen, 2009). Two tests were performed during egg-laying (1 and 3 days after the first egg was observed) and two during incubation (1 and 3 days after clutch incubation was confirmed). All tests were conducted between 7h00 a.m. and 12h00 p.m.; the specific time was semi-randomly assigned, that is the exact test time was not randomly predefined, but depended on working hours, duration and location of preceding tests. The test consisted of both a visual (in the form of a taxidermic model of a male great tit) and an acoustic stimulus (in form of a playback of a great tit song). The model and speaker (2010–2016: Samsung U5 Digital Audio Player connected to a Radioshack Mini Amplifier; 2017: Foxpro Shockwave speakers) were placed 1 m in front of the subject's nest box. Following previous studies (Derryberry, 2007; Logue & Gammon, 2004; Searcy, Anderson, & Nowicki, 2006), the speaker was placed on the ground, in part because great tits often engage in conflicts on the ground (personal observation; Falls, Krebs, & McGregor, 1982), while the model was placed on a 1.2 m wooden pole for visibility. Model and playback were chosen randomly from an available set of 23 models and 34 playbacks. Song stimuli were either recordings of great tits of German and Dutch populations from Xeno-Canto (see http://www.xeno-canto.org/) or local great tits recorded outside our study areas before the breeding season 2017. Playbacks were not standardised to contain the same number of songs or to have the same overall length, but rather represented natural singing behaviour and variation within song lengths and rates. Following the onset of a focal test, that is when the male entered a 15-m radius (horizontal) around the nest box, we recorded the focal male's behaviour for a period of 3 min. Considering that the closest neighbours within our plots were usually breeding within 50 m of distance, this radius was used to ensure that any bird scored was in fact the inhabitant of the nest box the test was performed at, which observations of colour-banded birds confirm. Subjects not arriving within 15 min or not entering the radius were scored non-responsive. The observer, positioned at a distance of 15 m from the nest box, counted the number of alarms and songs and estimated the minimum distance to the model ("approach distance"); we have previously shown that the latter behaviour represents an appropriate proxy for willingness to engage in physical attacks (Araya-Ajoy & Dingemanse, 2014). Songs were defined as longer and more complex vocalisations containing repeated sequences of notes. All vocalisations not categorised as songs, were summarised as alarms, defined by short length, simplicity compared with songs, and a monosyllabic structure (Marler & Slabbekoorn, 2004). All observers were thoroughly trained to reliably spot and sex great tits, and categorise and count male vocalisations. For further details on the test procedure, see Araya-Ajoy and Dingemanse (2014, 2017).

2.3 | Temperature measurements

Both night temperature (Garson & Hunter, 1979; Reid, 1987) and ambient temperature (Garson & Hunter, 1979; Thomas, 1999) have been shown to predict song duration and rate; however, these two temperature variables are highly correlated (Garson & Hunter, 1979; Naguib et al., 2019); their respective effects could be teased apart because of the large sample size of this study. Effects of both variables were modelled to assess temperature effects on song rates. We downloaded hourly temperatures for each day from the weather station in Rothenfeld (47°58024″N, 11°13024″E; Agrarmeteorologie Bayern, www.lfl-design3.bayern.de/agm/).

To evaluate whether temperatures within our study area differed substantially between populations and nest boxes, iButton temperature loggers (Thermochron iButton, iButtonLink Technology) were placed on all occupied nest boxes in the breeding season 2017. The loggers were placed on the outside bottom of the nest boxes and measured temperatures during the breeding season in a 30-min interval. All iButtons were tested afterwards in a controlled and common environment to ensure that differences between loggers were not due to measurement errors caused by malfunctions. After correcting for potential deviations, the collected temperature data were analysed using mixed-effect models, to examine the sources of variation in temperatures (see Table S1). Random intercepts were thus fitted for nest box (n = 201) and plot (n = 12) to account for spatial differences in temperatures, day (n = 65) and time of day (n = 1,440) to account for temporal temperature patterns, cardinal direction of each nest box (alignment of the nest box n = 8) and iButton ID (n = 200) to account for repeated measurements (i.e. pseudo-replication). This analysis showed that in our study area temperatures did not differ substantially among plots or nest boxes (see Table

S1) and that temperatures within the plots were highly correlated with the temperatures collected by the weather station (Pearson's product-moment correlation (mean [95% credible interval]: 0.919 [0.919, 0.920]). Minimum and average night temperatures between the day-specific time of sunset and sunrise were collected, but due to the extremely tight correlation between these two proxies for night temperatures (Pearson's product-moment correlation (mean [95% credible intervals]: 0.98 [0.97, 1.00]), we pragmatically per- formed our analysis using minimum night temperature. Night and current temperature (i.e. temperature measured during the assay) were less tightly correlated (Pearson's product-moment correlation (mean [95% credible intervals]: 0.537 [0.512, 0.560]; we attempted to further lessen the autocorrelation between these two variables in our statistical analyses by expressing minimum night temperature as a deviation from the current temperature during each observation.

2.4 | Statistical analyses

To investigate how temperature affected acoustic responses to intruders, we focused on three traits recorded during the territorial intrusion tests: (a) the number of songs, (b) the number of alarms and (c) the sum of songs and alarms (i.e. vocalisation count). As approach distance covaries with alarming and singing in our populations (Araya-Ajoy & Dingemanse, 2014), we additionally examined 4) the minimum distance to the taxidermic model as a measure of physical aggression. For all analyses, we used univariate general(ised) mixed- effects models, with the aforementioned traits fitted as response variables. We expected a priori that birds tested later during the day had more time to refill energy storages. All models thus fitted cur- rent temperature, the difference between current temperature and minimum night temperature, the time (in hours) between sunrise and the test (“time of day”), as well as the two-way-interactions between the two temperature variables and the time from sunrise (both vari- ables were population-mean centred). We also fitted the time until the focal subject entered the test radius (in seconds from the start of the playback) as a covariate to control for any bias that might result from birds responding later being exposed to the playback for longer. Furthermore, to account for previously observed seasonal changes in responses to artificial intruders (Araya-Ajoy & Dingemanse, 2014, 2017), breeding stage (egg-laying vs. incubating) and test sequence (first vs. second test within breeding stage) were added as cate- gorical fixed effects. Random intercepts were fitted for: individual (*n* = 813) to account for repeated measurements (i.e. pseudo-replica- tion), nest box (*n* = 501) and plot (*n* = 12) to account for spatial varia- tion, date (*n* = 263) and year identity (*n* = 8) to account for temporal variation (see Figures S1 and S2), observer (*n* = 52), and playback song (*n* = 34) and taxidermic model (*n* = 23) to account for differ- ent responses induced by features of specific playbacks or models. We also fitted random intercepts for the unique combination of plot and year, called “Plot-year” to account for plot-specific year effects (*n* = 96; see Abbey-Lee et al., 2018; Araya-Ajoy & Dingemanse, 2017; Araya-Ajoy et al., 2016 for further explanation). Models (1), (2) and (3) were parametrised assuming a Poisson error distribution, while in model (4), the response variable (distance) was square-root transformed and modelled assuming Gaussian errors. Tests scored as “non-responsive” were not considered in our analysis. Furthermore, only males of known identity were included in our analyses.

All statistical analyses were performed using R 3.2.4 (R Core Team, 2018). Generalised linear mixed-effects models were ap- plied using the “glmer” function (package lme4; Bates et al., 2015). All non-Gaussian models were tested for overdispersion; if neces- sary, we fitted an observation-level random effect (“Observation ID” in our statistical tables) to account for it. Model fit was visually assessed based on the distribution of residuals. We used the “sim” function (package arm) to simulate the posterior distributions of the model parameters based on 2,000 simulations (Gelman & Hill, 2006). The statistical significance of fixed effects was inferred from the 95% credible intervals (CI) associated with the mean parameter esti- mate (β). We consider an effect to be “significant” in the frequentist's sense when the 95% CI did not overlap zero (Nakagawa & Cuthill, 2007) and we describe such results as showing “strong support” for the predictions. For certain analyses, we further compared parame- ters among predictor variables estimated within the same model; for such cases, we compared the overlap of their 84% CIs as this avoids over-conservative interpretations caused by compilation of model- ling error (Han & Dingemanse, 2017; Julious, 2004).

This study was approved by the Regierung Oberbayern (permit number ROB-55.2-2532.Vet_02-17-215) in accordance with the ASAB/ABS Guidelines for the use of animals in research. Our experiments were designed to minimise subject discomfort.

3 | RESULTS

We performed 6,596 simulated territorial intrusions over the course of 8 years (2010–2017). In 4,476 tests (67.9%), the focal male en- tered the test radius and its behaviour was scored. In 3,387 (75.7%) of these tests, the identity of the focal male was known. These birds responded acoustically in 96.8% of the simulated territorial intru- sions (*n* = 3,278). Singing was observed in 80.6% and alarming in 41.0% of tests with known birds. Behavioural responses without any vocalisations did occur extremely rarely (<1%).

3.1 | Acoustic traits

Birds did not change their vocalisation count in response to any of the examined predictors but arrival time (Table 1). Specifically, birds responding later produced fewer vocalisations (effect of arrival time male: β = −.109, CI = −0.134, −0.083), an effect caused by it producing both fewer songs (β = −.071, CI = −0.114, −0.026) and fewer alarm calls (β = −.164, CI = −0.306, −0.022). Birds sang more in warmer conditions (main effect of current temperature β = .023, CI = 0.009, 0.036; Figure 1a), increased song rate over the season (effect of breeding context: β = .198, CI = 0.153, 0.243) and with

TABLE 1 Estimated effect sizes and 95% credible intervals (CIs) for predictors of vocalisation count (n = 3,278), number of songs (n = 3,334), number of alarms (n = 3,318) and minimum distances (n = 3,340) shown by great tits in response to simulated territorial intrusions

	Vocalisation count[a]	Song rate	Alarm rate	Minimum distance[b]
Fixed effects	β (95% CI)	β (95% CI)	β (95% CI)	β (95% CI)
Intercept[c]	2.919 (2.839, 2.997)	2.055 (1.928, 2.176)	−1.383 (−1.554, −1.209)	0.047 (−0.11, 0.197)
Current temperature (CT)	−0.007 (−0.015, 0.001)	0.023 (0.009, 0.036)	−0.081 (−0.12, −0.041)	0.005 (−0.008, 0.018)
Night temperature (NT)[d]	0.003 (−0.005, 0.01)	0.005 (−0.009, 0.019)	−0.017 (−0.058, 0.023)	0.013 (−0.001, 0.026)
Time of day	−0.01 (−0.037, 0.016)	0.009 (−0.038, 0.053)	−0.205 (−0.358, −0.058)	0.026 (−0.018, 0.071)
Interaction (CT * Time)	−0.001 (−0.008, 0.005)	−0.007 (−0.018, 0.005)	−0.026 (−0.063, 0.01)	0.003 (−0.008, 0.013)
Interaction (NT * Time)	0.004 (−0.003, 0.01)	−0.004 (−0.017, 0.009)	−0.008 (−0.048, 0.034)	0.003 (−0.008, 0.015)
Arrival time male	−0.109 (−0.134, −0.083)	−0.071 (−0.114, −0.026)	−0.164 (−0.306, −0.022)	0.047 (0.005, 0.087)
Breeding context	−0.022 (−0.049, 0.005)	0.198 (0.153, 0.243)	−0.914 (−1.049, −0.776)	0.285 (0.244, 0.329)
Test sequence	−0.012 (−0.039, 0.013)	0.072 (0.029, 0.116)	−0.273 (−0.407, −0.136)	0.062 (0.022, 0.102)
Random effects	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)	σ^2 (95% CI)
Individual	0.042 (0.038, 0.047)	0.179 (0.161, 0.198)	1.344 (1.21, 1.488)	0.248 (0.225, 0.272)
Nest box	0.007 (0.006, 0.008)	0.019 (0.016, 0.023)	0 (0, 0)	0.119 (0.105, 0.135)
Date	0.005 (0.004, 0.006)	0 (0, 0)	0 (0, 0)	0.015 (0.012, 0.017)
Plot-year	0.003 (0.002, 0.004)	0.01 (0.008, 0.013)	0 (0, 0)	0.018 (0.013, 0.024)
Observer	0.031 (0.023, 0.04)	0.058 (0.04, 0.08)	0 (0, 0)	0.074 (0.055, 0.095)
Playback	0 (0, 0.001)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Model	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0.001)
Population	0.006 (0.002, 0.012)	0.01 (0.003, 0.019)	0 (0, 0)	0.023 (0.008, 0.043)
Year	0 (0, 0.001)	0.007 (0.002, 0.015)	0 (0, 0)	0.024 (0.006, 0.052)
Residual	0.053 (0.049, 0.057)	0.121 (0.108, 0.136)	1.607 (1.47, 1.746)	1.155 (1.103, 1.211)
Observation ID[e]	0.476 (0.456, 0.495)	1.443 (1.382, 1.502)	10.098 (9.613, 10.593)	–

[a] Sum of the number of songs and calls during a territorial intrusion test.

[b] Minimum approach distance of the focal bird to the dummy during a territorial intrusion test.

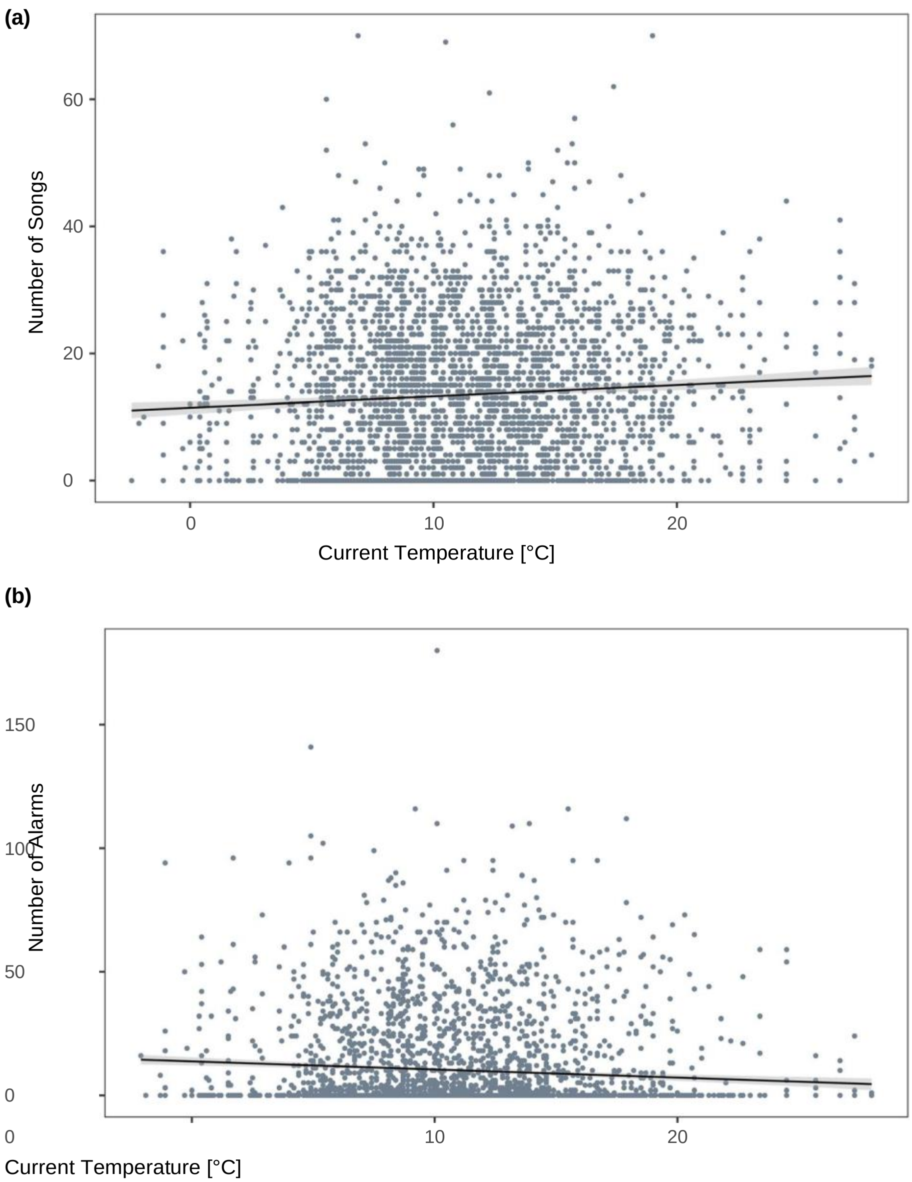
[c] Reference category; estimate for average current temperature, average night temperature, average time of day, average arrival time male and for tests during sequence 1 in the egg-laying phase.

[d] Night temperature was expressed as deviation of minimum night temperature from current temperature.

[e] Observation-level random effect to account for overdispersion.

repeated testing (effect of test sequence: β = .072, CI = 0.029, 0.116), whereas night temperature did not affect singing (main effect of night temperature: β = .005 CI = −0.009, 0.019). The 84% credible intervals did not overlap between the main effects of night versus current temperature (for a rational of comparing 84% CIs between parameters from the same model; see Han & Dingemanse, 2017), implying that their effects were distinct (non-overlapping; effect of current temperature: β = .023, 84% CI = 0.016, 0.030; effect of night temperature: β = .005, 84% CI = −0.002, 0.012). Time of day did not affect how much birds sang (main effect of time of day: β = .009, CI = −0.038, 0.053; Figure S3a). Opposite to expectations, the effects of current and night temperature were both not a function of time of day: the increase in song rates in warmer conditions was not lessened for tests conducted earlier in the morning (effect of the interaction term current temperature × time: β = −.007, CI = −0.018, 0.005), neither did birds sing more later the day following comparably colder nights (effect of the interaction term night temperature × time: β = −.004, CI = −0.017, 0.009; Table 1).

Though birds sang more (see above), they alarmed less in warmer conditions (main effect of current temperature: β = −.081, CI = −0.12, −0.041; Figure 1b). By contrast, colder nights did not cause a similar pattern (main effect of night temperature: β = −.017 CI = −0.058, 0.023). The 84% CIs of the effects of the two temperature variables did not overlap, implying again that their effects were distinct (main effect of current temperature: β = −.081, 84% CI = −0.101, −0.062; main effect of night temperature: β = −.017, 84% CI = −0.037, 0.005). Birds alarmed less later in the morning (main effect of time of day: β = −.205, CI = −0.358, −0.058; Figure S3b). Furthermore, alarming



FI G U R E 1 The effect of current temperature on (a) song (n = 3,334) and
(b) alarm rate (n = 3,318) expressed by territorial great tits during simulated territorial intrusions. Black lines represent regression lines and grey areas 95% confidence intervals. Graph based on raw data

decreased from egg-laying to incubation (effect of breeding context: β = −.914, CI = −1.049, −0.776) and between consecutive tests (effect of test sequence: β = −.273, CI = −0.407, −0.136). Time of day did not alter the effects of the two temperature variables on alarming (effect of the interaction current temperature × time: β = −.026, CI = −0.063, 0.01; effect of the interaction night temperature × time: β = −.008, CI = −0.048, 0.034; Table 1).

3.2 | Minimum distance

Neither current nor night temperature nor time of day affected approach distance (Table 1). By contrast, birds increased the distance to which they approached the intruder when responding later (effect of arrival time male: β = .047, CI = 0.005, 0.087) as well as later in the season (effect of breeding context: β = .285, CI = 0.244, 0.329) and with later test sequences (β = .062, CI = 0.022, 0.102), echoing our earlier analyses (Araya-Ajoy & Dingemanse, 2014, 2017) (Table 1).

4 | DISCUSSION

This study examined the links between temperature and the acoustic response behaviours of territorial great tit males to a male intruder using long-term observational data. We evaluated support for the hypothesis that birds would either shift from singing to alarming in colder conditions, assuming higher energetic costs for singing, or would decrease their overall acoustic output, indicating no difference in energy demands between the two types of vocalisation. Great tits indeed sang less and alarmed more in colder current temperature conditions, but the vocalisation count was not affected. Interestingly, night temperature did not affect acoustic responsiveness in any trait examined. Follow-up analyses demonstrated that these temperature-related changes in behaviour resulted from both among- and within-year within-individual plasticity in response to macro-temporal (among-year) and micro-temporal (within-year) variation in current temperature (Texts S1 and S2 and Table S2). Contrary to expectations, increasing the time between sunrise and the measurement of their acoustic responsiveness to a simulated intrusion did not affect the impact of current and night temperature for either singing or alarming. Instead, birds generally alarmed less later in the morning. Finally, the level of physical aggression (measured as minimum approach distance to the taxidermic model; Araya-Ajoy & Dingemanse, 2014) was not affected by either current or night temperature. In conclusion, our study shows that great tits prefer to respond to intruders by singing, but that they shift from singing to alarming in colder conditions. This finding supports the idea of alarms being less costly, therefore supporting our Scenario 2.

4.1 | Temperature effects on the acoustic responsiveness to intruders

Consistent with previous studies positing a trade-off between singing and other activities (e.g. foraging), we found a positive effect of current temperature on song rate (Gillooly & Ophir, 2010; Gottlander, 1987; Ophir, Schrader, & Gillooly, 2010; Reid, 1987; Strain & Mumme, 1988). Colder conditions require more energy for thermoregulation (Broggi et al., 2007; Dubois et al., 2016) leading to reduced body mass and fat reserves (Dubois et al., 2016; Moiron, Mathot, & Dingemanse, 2018; Reid, 1987). At the same time, singing is energetically demanding (Eberhardt, 1994; Oberweger & Goller, 2001; Ward & Slater, 2005). Consequently, in colder conditions, when energy reserves are depleted and foraging is difficult, birds might be unable to sustain high song rates. High temperatures could lead to a similar effect, when physiologically demanding behaviours lead to hyperthermia and therefore need to be reduced or stopped (Guillemette et al., 2016). Temperature might thus have a non-linear effect on singing, where the number of songs is reduced under both, cold and warm conditions. However, the lack of a quadratic pattern in our data implied that his mechanism was not supported, perhaps because temperatures during the behavioural tests never exceeded 20°C (see Text S3, Table S3 and Figure S1).

Interestingly, great tits have been shown to respond to an intruder not only by singing but also by alarming (Araya-Ajoy & Dingemanse, 2014). Alarms might offer an alternative to songs, especially when costs of singing increase, because they are by definition shorter and simpler than songs (Marler, 2004), therefore, potentially imposing lower production costs; an assumption that is in line with previous findings based on species comparisons implying that alarming (vs. singing) is associated with relatively smaller energetic costs (Chappell et al., 1995; Horn et al., 1995; Jurisevic et al., 1999). In fact, we found that, contrary to singing, alarm rates decreased with increasing temperatures. Because the vocalisation count was not affected by temperature, this finding indicates a shift from singing to alarming when the environment is colder. Follow-up analyses demonstrated that the observed temperature-related shifts were attributable to within-individual behavioural plasticity (see Texts S1 and S1 and Table S2). Overall, our study thus indicates that, in response to temperature, birds shift their acoustic response behaviour to the shorter and simpler alarms, away from songs, the most frequent and supposedly more efficient signal. We assume such a difference in efficiency, because the majority of birds responded to an intruder with singing highlighting the importance of this behaviour in the context of territory defence (Amrhein & Lerch, 2010). Furthermore, our previous work showed that relatively aggressive birds using alarming as a response to intrusions also gain relatively little extra-pair paternity (Araya-Ajoy et al., 2016), implying that alarming has a different signal value, potentially signalling a lack of reserves required for using song as a response. Furthermore, especially if alarming is associated with lower costs, shifting from singing to alarming might enable birds to save energy reserves for later. Despite the imminent necessity to spend energy to defend one's territory, this extra energy might be needed later, given that colder conditions are likely to persist. Nevertheless, responding with alarming should be preferable to not responding acoustically at all, considering that no response could lead to a territory loss. Indeed, in extremely few cases (3.2%; *n* = 3,278 observations), the response to an intruder was totally non-acoustic (i.e. birds entering the 15-min radius without vocalising). Thus, when environmental conditions reduce energy available for acoustic communication, individuals seem to shift to cheaper means of acoustic communication, despite thereby signalling reduced competitive ability.

Contrary to previous findings, night temperature did not affect the number of songs or calls (Garson & Hunter, 1979; Thomas, 1999). Furthermore, we also expected birds to replenish energetic shortages after colder nights by increasing early-morning foraging, as we have recently demonstrated for this population during winter (Moiron et al., 2018). This should help mitigate the expected detrimental effects of cold nights on song rate, particularly later during the day. The lack of an interactive effect of night temperature and time of day was therefore unexpected. However, these two findings make sense if night temperature only affects singing around sunrise, that is during the dawn chorus (Reid, 1987; Thomas, 1999). Importantly, all tests in this study were performed later during the day, perhaps providing the birds enough time to replenish their energy reserves. This would also explain why the correlation between night and day temperature in our study was much lower than reported in other studies (Garson & Hunter, 1979; Naguib et al., 2019). We did not detect a significant interaction between time and current temperature, which, on first sight, contradicts the idea that more time before an intrusion would allow birds to replenish their energy reserves and sing at higher rates (Thomas, 1999). We offer the following explanations. First, given that temperatures in subsequent periods are usually strongly correlated, colder conditions normally persist throughout the day. Therefore, birds may save (re-) filled energy reserves for later to compensate for fast depletion due to increased metabolic costs and decreased foraging efficiency (Avery & Krebs, 2008). Second, other costly song characteristics (such as song complexity or amplitude) that we did not measure may have increased over the day (Franz & Goller, 2003; but see Ward, Speakman, & Slater, 2003) rather than song rate. Indeed, our finding that birds can plastically alternate between singing and alarming already implies that such changes in “acoustic structure” occur in our populations. This explanation also seems plausible because alarm rates did show time of day effects, potentially because this acoustic signal is relatively fixed in length and structure (Marler, 2004) and thus has little scope for exhibiting plasticity in its structure.

An individual's level of physical aggressiveness, measured as its minimum approach distance, was not affected by temperature. While previous studies have shown positive links between tempera- ture and aggression (González-Gómez, Ricote-Martinez, Razeto- Barry, Cotorás, & Bozinovic, 2011), others showed the opposite (Fisher, Poulin, Todd, & Brigham, 2004); therefore, a null result is not unexpected. Apparently, ecological conditions that remain unquanti- fied moderate the association between aggressiveness and tempera- ture. Nevertheless, given our finding that alarm rates increased after colder nights, we would have expected increased aggressiveness (i.e. shorter minimum distances) as an attempt to compensate for the less efficient transmission of alarm calls (Ryan, 1988). Furthermore, we also expected an effect on approach distance because alarming and minimum distance are negatively correlated both within- and among-individuals in our population (Araya-Ajoy & Dingemanse, 2014). Previously, we hypothesised that these negative within-indi- vidual correlations resulted from all three traits responding to varia- tion in the same environmental factor(s) (Araya-Ajoy & Dingemanse, 2014). Our study demonstrates that this is clearly not (fully) the case: variation in temperature only underpins the negative correlation be- tween the two acoustic traits. Thus, the negative covariance with physical aggressiveness (minimum approach distance), indicative of animals approaching closer singing less and alarming more, must be caused by another environmental effect uncorrelated with current temperature. Ultimately, these findings might imply, in contrast to previous suggestions (Araya-Ajoy & Dingemanse, 2014), that sing- ing, alarming and minimum distance may ultimately represent a qua- si-independent (rather than a single overarching) functional unit as the acoustic traits are proximately underpinned by more of the same environmental factors. Similarly, our finding that songs are plasti- cally adjusted to within-year changes in temperature, while alarms are not, implies that the two acoustic traits do not vary as a function of the same environmental factors.

5 | CONCLUSION

We analysed effects of temperature on acoustic communication traits using data from more than 6,500 standard territorial intrusion tests conducted over an 8-year period. We demonstrate that birds show phenotypic plasticity in response to within- and among-year changes in temperature, plastically down-regulating singing versus up-regulating alarm rating with decreasing current temperature. Though we did not measure energetic costs of singing and alarming directly, overall, our findings are in line with proposed differences in energetic production costs and signal efficiency between sing- ing and alarming. Furthermore, they are consistent with the notion that great tits switch to assumed energetically cheaper acoustic re- sponses (alarming as opposed to singing) when faced with energetic shortfalls. Consequently, current temperature affects territory- defence strategies. Future research should address whether sing- ing (rather than alarming) is more effective in avoiding territory takeovers following territorial intrusions, whether energetic costs of singing are indeed larger than those for alarming in this species and whether acoustic (vs. non-acoustic) responses are indeed favoured by natural selection.

ACKNOWLEDGEMENTS

A.F.T.S., A.H. and N.J.D. were supported by the German Science Foundation (grant no. DI 1694/1-1). N.J.D. was supported by the Max Planck Society. We thank all past members of the “Evolutionary Ecology of Variation” group at the Max Planck Institute for Ornithology and the “Behavioural Ecology” group at the LMU, field assistants and students for help in data collection. All procedures complied with guidelines from the District Government of Upper Bavaria (Regierung von Oberbayern) for Animal Care.

CONFLICT OF INTEREST

None.

ORCID

Alexander Hutfluss https://orcid.org/0000-0001-8514-6452

REFERENCES

Abbey-Lee, R. N., Mathot, K. J., & Dingemanse, N. J. (2016). Behavioral and morphological responses to perceived predation risk: A field ex- periment in passerines. *Behavioral Ecology*, *27*(3), 857–864. https://doi.org/10.1093/beheco/arv228

Abbey-Lee, R. N., Uhrig, E. J., Zidar, J., Favati, A., Almberg, J., Dahlbom, J., … Løvlie, H. (2018). The influence of rearing on behavior, brain monoamines, and gene expression in three-spined stickle- backs. *Brain, Behavior and Evolution*, *91*(4), 201–213. https://doi.org/10.1159/000489942

Amrhein, V., & Lerch, S. (2010). Differential effects of moving versus stationary territorial intruders on territory defence in a songbird. *Journal of Animal Ecology*, 79, 82–87. https://doi.org/10.1111/j.1365-2656.2009.01629.x

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2014). Characterizing behavioural ‘characters’: An evolutionary framework. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1776), 20132645. https://doi.org/10.1098/rspb.2013.2645

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2017). Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. *Journal of Animal Ecology*, *86*(2), 227–238. https://doi.org/10.1111/1365-2656.12621

Araya-Ajoy, Y. G., Kuhn, S., Mathot, K. J., Mouchet, A., Mutzel, A., Nicolaus, M., … Dingemanse, N. J. (2016). Sources of (co)variation in alternative siring routes available to male great tits (*Parus major*). *Evolution*, *70*(10), 2308–2321. https://doi.org/10.1111/evo.13024

Aubin, T., Mathevon, N., Da Silva, M. L., Vielliard, J. M. E., & Sebe, F. (2004). How a simple and stereotyped acoustic signal transmits and individual information the song of the White-browed Warbler Basileuterus lecoblepharus. *Anais Da Academia Brasileira De Ciências*, *76*(2), 335–344.

Avery, M. I., & Krebs, J. R. (2008). Temperature and foraging success of Great Tits *Parus major* hunting for spiders. *Ibis*, *126*(1), 33–38. https://doi.org/10.1111/j.1474-919X.1984.tb03661.x

Barnett, C. A., & Briskie, J. V. (2007). Energetic state and the perfor- mance of dawn chorus in silvereyes (*Zosterops lateralis*). *Behavioral*

Ecology and Sociobiology, 61(4), 579–587. https://doi.org/10.1007/s00265-006-0286-x

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … Green, P. (2015). Package 'lme4'. *Convergence*, *12*(1), 2.

Bright, M. (1985). *Animal language*. Ithaca, NY: Cornell Univ. Press. Broggi, J., Hohtola, E., Koivula, K., Orell, M., Thomson, R. L., & Nilsson, J.-Å. (2007). Sources of variation in winter basal metabolic rate in the great tit. *Functional Ecology*, *21*(3), 528–533. https://doi.org/10.1111/j.1365-2435.2007.01255.x

Bugnyar, T., Kijne, M., & Kotrschal, K. (2001). Food calling in ravens: Are yells referential signals? *Animal Behaviour*, *61*(5), 949–958. https://doi.org/10.1006/ANBE.2000.1668

Campos, D. P., Bander, L. A., Raksi, A., & Blumstein, D. T. (2009). Perch exposure and predation risk: A comparative study in pas- serines. *Acta Ethologica*, *12*(2), 93–98. https://doi.org/10.1007/ s10211-009-0061-x

Chappell, M. A., Zuk, M., Kwan, T. H., & Johnsen, T. S. (1995). Energy cost of an avian vocal display: Crowing in red junglefowl. *Animal Behaviour*, *49*(1), 255–257. https://doi.org/10.1016/0003-3472(95)80177-4

Derryberry, E. P. (2007). Evolution of bird song affects signal efficacy: An experimental test using historical and current signals. *Evolution*, *61*(8), 1938–1945. https://doi.org/10.1111/j.1558-5646.2007.00154.x

Dubois, K., Hallot, F., & Vézina, F. (2016). Basal and maximal metabolic rates differ in their response to rapid temperature change among avian species. *Journal of Comparative Physiology B*, *186*(7), 919–935. https://doi.org/10.1007/s00360-016-1001-5

Eberhardt, L. S. (1994). Oxygen consumption during singing by male Carolina wrens (*Thryothorus ludovicianus*). *The Auk*, *111*(1), 124–130. https://doi.org/10.2307/4088511

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

Fisher, R. J., Poulin, R. G., Todd, L. D., & Brigham, R. M. (2004). Nest stage, wind speed, and air temperature affect the nest defence behaviours of burrowing owls. *Canadian Journal of Zoology*, *82*(5), 707–713. https://doi.org/10.1139/z04-035

Franz, M., & Goller, F. (2003). Respiratory patterns and oxygen consumption in singing zebra finches. *Journal of Experimental Biology*, *206*(6), 967–978. https://doi.org/10.1242/jeb.00196

Garson, P. J., & Hunter, M. L. (1979). Effects of temperature and time of year on the singing behaviour of wrens troglodytes troglodytes and great tits *Parus major*. *Ibis*, *121*(4), 481–487. https://doi.org/10.1111/ j.1474-919X.1979.tb06688.x

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge, UK: Cambridge University Press.

Gil, D., & Gahr, M. (2002). The honesty of bird song: Multiple constraints for multiple traits. *Trends in Ecology and Evolution*, *17*(3), 133–141. https://doi.org/10.1016/S0169-5347(02)02410-2

Gillooly, J. F., & Ophir, A. G. (2010). The energetic basis of acoustic communication. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1686), 1325–1331. https://doi.org/10.1098/rspb.2009.2134

Godfrey, J. D., & Bryant, D. M. (2000). State-dependent behaviour and energy expenditure: An experimental study of European robins on winter territories. *Journal of Animal Ecology*, *69*(2), 301–313. https://doi.org/10.1046/j.1365-2656.2000.00393.x

González-Gómez, P. L., Ricote-Martinez, N., Razeto-Barry, P., Cotorás, I. S., & Bozinovic, F. (2011). Thermoregulatory cost affects ter- ritorial behavior in hummingbirds: A model and its application. *Behavioral Ecology and Sociobiology*, *65*(11), 2141–2148. https://doi.org/10.1007/s00265-011-1222-2

Gottlander, K. (1987). Variation in the song rate of the male pied flycatcher *Ficedula hypoleuca*: Causes and consequences. *Animal Behaviour*, *35*(4), 1037–1043. https://doi.org/10.1016/ S0003-3472(87)80160-4

Guillemette, M., Woakes, A. J., Larochelle, J., Polymeropoulos, E. T., Granbois, J.-M., Butler, P. J., … Portugal, S. J. (2016). Does hyper- thermia constrain flight duration in a short-distance migrant? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1704), 20150386. https://doi.org/10.1098/rstb.2015.0386

Han, C. S., & Dingemanse, N. J. (2017). Protein deprivation decreases male survival and the intensity of sexual antagonism in southern field crickets *Gryllus bimaculatus*. *Journal of Evolutionary Biology*, *30*(4), 839–847. https://doi.org/10.1111/jeb.13052

Hasselquist, D., & Bensch, S. (2008). Daily energy expenditure of singing great reed warblers *Acrocephalusarundinaceus*. *Journalof Avian Biology*, *39*(4), 384–388. https://doi.org/10.1111/j.0908-8857.2008.04427.x

Hope, S. (1980). Call form in relation to function in the Steller's Jay. *The American Naturalist*, *116*(6), 788–820. https://doi.org/10.1086/283670

Horn, A. G., Leonard, M. L., & Weary, D. M. (1995). Oxygen consumption during crowing by roosters: Talk is cheap. *Animal Behaviour*, *50*(5), 1171–1175. https://doi.org/10.1016/0003-3472(95)80033-6

Julious, S. A. (2004). Using confidence intervals around individual means to assess statistical significance between two means. *Pharmaceutical Statistics*, *3*(3), 217–222. https://doi.org/10.1002/pst.126

Jurisevic, M. A., Sanderson, K. J., & Baudinette, R. V. (1999). Metabolic rates associated with distress and begging calls in birds. *Physiological and Biochemical Zoology*, *72*(1), 38–43. https://doi. org/10.1086/316636

Kendeigh, S. C. (1944). Effect of air temperature on the rate of energy metabolism in the English sparrow. *The Journal of Experimental Zoology*, *96*(1), 1–16. https://doi.org/10.1002/jez.1400960102

Klappert, K., & Reinhold, K. (2003). Acoustic preference functions and sexual selection on the male calling song in the grasshopper *Chorthippus biguttulus*. *Animal Behaviour*, *65*(1), 225–233. https://doi.org/10.1006/anbe.2002.2034

Klump, G. M., & Shalter, M. D. (2010). Acoustic behaviour of birds and mammals in the predator context; I. Factors affecting the structure of alarm signals. II. The functional significance and evolution of alarm signals. *Zeitschrift Für Tierpsychologie*, *66*(3), 189–226. https://doi.org/10.1111/j.1439-0310.1984.tb01365.x

Logue, D. M., & Gammon, D. E. (2004). Duet song and sex roles during territory defence in a tropical bird, the black-bellied wren. *Thryothorus Fasciatoventris*. *Animal Behaviour*, *68*(4), 721–731. https://doi.org/10.1016/J.ANBEHAV.2003.10.026

Marler, P. (2004). Bird calls: Their potential for behavioral neurobiology. *Annals of the New York Academy of Sciences*, *1016*(1), 31–44. https://doi.org/10.1196/annals.1298.034

Marler, P. R., & Slabbekoorn, H. (2004). *Nature's music: The science of birdsong*. San Diego, CA: Elsevier Academic.

Moiron, M., Mathot, K. J., & Dingemanse, N. J. (2018). To eat and not be eaten: Diurnal mass gain and foraging strategies in wintering great tits. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1874), 20172868. https://doi.org/10.1098/rspb.2017.2868

Møller, A. P., Nielsen, J. T., & Garamszegi, L. Z. (2006). Song post ex- posure, song features, and predation risk. *Behavioural Ecology*, *17*(2), 155–163. https://doi.org/10.1093/beheco/arj010

Morton, E. S. (1977). On the occurrence and significance of motiva- tion-structural rules in some bird and mammal sounds. *The American Naturalist*, *111*(981), 855–869. https://doi.org/10.1086/283219

Naguib, M., Altenkamp, R., & Griessmann, B. (2001). Nightingales in space: Song and extra-territorial forays of radio tagged song birds. *Journal of Ornithology*, *142*(3), 306–312. https://doi.org/10.1007/BF01651369

Naguib, M., Diehl, J., van Oers, K., & Snijders, L. (2019). Repeatability of signalling traits in the avian dawn chorus. *Frontiers in Zoology*, *16*(1), 27. https://doi.org/10.1186/s12983-019-0328-7

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: A practical guide for biologists. *Biological Reviews*, *82*(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nicolaus, M., Both, C., Ubels, R., Edelaar, P., & Tinbergen, J. M. (2009). No experimental evidence for local competition in the nestling phase as a driving force for density-dependent avian clutch size. *Journal of Animal Ecology*, *78*(4), 828–838. https://doi.org/10.1111/j.1365-2656.2009.01535.x

Nicolaus, M., Mathot, K. J., Araya-Ajoy, Y. G., Mutzel, A., Wijmenga, J. J., Kempenaers, B., & Dingemanse, N. J. (2015). Does coping style predict optimization? An experimental test in a wild passerine bird. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1799), 20142405–20142405. https://doi.org/10.1098/rspb.2014.2405

Oberweger, K., & Goller, F. (2001). The metabolic cost of birdsong production. *The Journal of Experimental Biology*, *204*(19), 3379–3388.

Ophir, A. G., Schrader, S. B., & Gillooly, J. F. (2010). Energetic cost of calling: General constraints and species-specific differences. *Journal of Evolutionary Biology*, *23*(7), 1564–1569. https://doi.org/10.1111/j.1420-9101.2010.02005.x

Pohl, N. U., Leadbeater, E., Slabbekoorn, H., Klump, G. M., & Langemann, U. (2012). Great tits in urban noise benefit from high frequencies in song detection and discrimination. *Animal Behaviour*, *83*(3), 711–721. https://doi.org/10.1016/j.anbehav.2011.12.019

R Development Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reid, M. L. (1987). Costliness and reliability in the and singing vigour of Ipswich sparrows. *Animal Behaviour*, *35*(6), 1735–1743.

Reinertsen, R. E. (1983). Nocturnal hypothermia and its energetic significance for small birds living in the arctic and subarctic regions. A review. *Polar Research*, *1*(3), 269–284. https://doi.org/10.1111/j.1751-8369.1983.tb00743.x

Reinertsen, R. E., & Haftorn, S. (1986). Different metabolic strategies of northern birds for nocturnal survival. *Journal of Comparative Physiology B*, *156*(5), 655–663. https://doi.org/10.1007/BF00692743 Ryan, M. R. (1988). Energy, calling and selection. *American Zoologist*, *28*(3), 885–898. https://doi.org/10.1093/icb/28.3.885

Searcy, W. A., Anderson, R. C., & Nowicki, S. (2006). Bird song as a signal of aggressive intent. *Behavioral Ecology and Sociobiology*, *60*(2), 234– 241. https://doi.org/10.1007/s00265-006-0161-9

Simmons, A. M., Popper, A. N., & Fay, R. R. (2005). *Acoustic communication* (16th ed.). New York, NY: Springer.

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

Strain, J. G., & Mumme, R. L. (1988). Effects of food supplementation, song playback, and temperature on vocal territorial behavior of carolina wrens. *The Auk*, *105*(1), 11–16. https://doi.org/10.109/ auk/105.1.11

Suter, S. M., Ermacora, D., Rieille, N., & Meyer, D. R. (2009). A distinct reed bunting dawn song and its relation to extrapair paternity. *Animal Behaviour*, *77*(2), 473–480. https://doi.org/10.1016/j.anbehav.2008.11.002

Suthers, R. A., Goller, F., & Pytte, C. (1999). The neuromuscular control of birdsong. *Philosophical Transactions of the Royal Society B: Biological Sciences*,*354*(1385), 927–939. https://doi.org/10.1098/ rstb.1999.0444

Thielcke, G. (1976). *Bird sounds*. Retrieved from https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2433732

Thomas, R. J. (1999). Two tests of a stochastic dynamic programming model of daily singing routines in birds. *Animal Behaviour*, *57*(2), 277– 284. https://doi.org/10.1006/ANBE.1998.0956

Thomas, R. J., Cuthill, I. C., Goldsmith, A. R., Cosgrove, D. F., Lidgate, H. C., & Proctor, S. L. B. (2003). The trade-off between singing and mass gain in a daytime-singing bird, the European Robin. *Behaviour Brill*, *140*(3), 387–404. https://doi.org/10.2307/4536033

Vergne, A. L., Avril, A., Martin, S., & Mathevon, N. (2007). Parent–offspring communication in the Nile crocodile *Crocodylus niloticus*: Do newborns' calls show an individual signature? *Naturwissenschaften*, *94*(1), 49–54. https://doi.org/10.1007/s00114-006-0156-4

Ward, S., & Slater, P. J. B. (2005). Heat transfer and the energetic cost of singing by canaries *Serinus canaria*. *Journal of Comparative Physiology A*, *191*(10), 953–964. https://doi.org/10.1007/s00359-005-0022-4

Ward, S., Speakman, J. R., & Slater, P. J. B. (2003). The energy cost of song in the canary, *Serinus canaria*. *Animal Behaviour*, *66*(5), 893–902. https://doi.org/10.1006/anbe.2003.2250

Wilkinson, R. (1980). Calls of nestling chaffinches *Fringilla coelebs*: The use of two sound sources. *Zeitschrift Für Tierpsychologie*, *54*(4), 346– 356. https://doi.org/10.1111/j.1439-0310.1980.tb01251.x

Williams, H. W. (1969). Vocal behavior of adult California quail. *The Auk*, *86*(4), 631–659. https://doi.org/10.2307/4083452

Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, *Cercopithecus campbelli*. *Behavioral Ecology and Sociobiology*, *50*(5), 414–422. https://doi.org/10.1007/s002650100383

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Electronic Supplementary Material of Chapter 1

Supplementary Material Table S1.

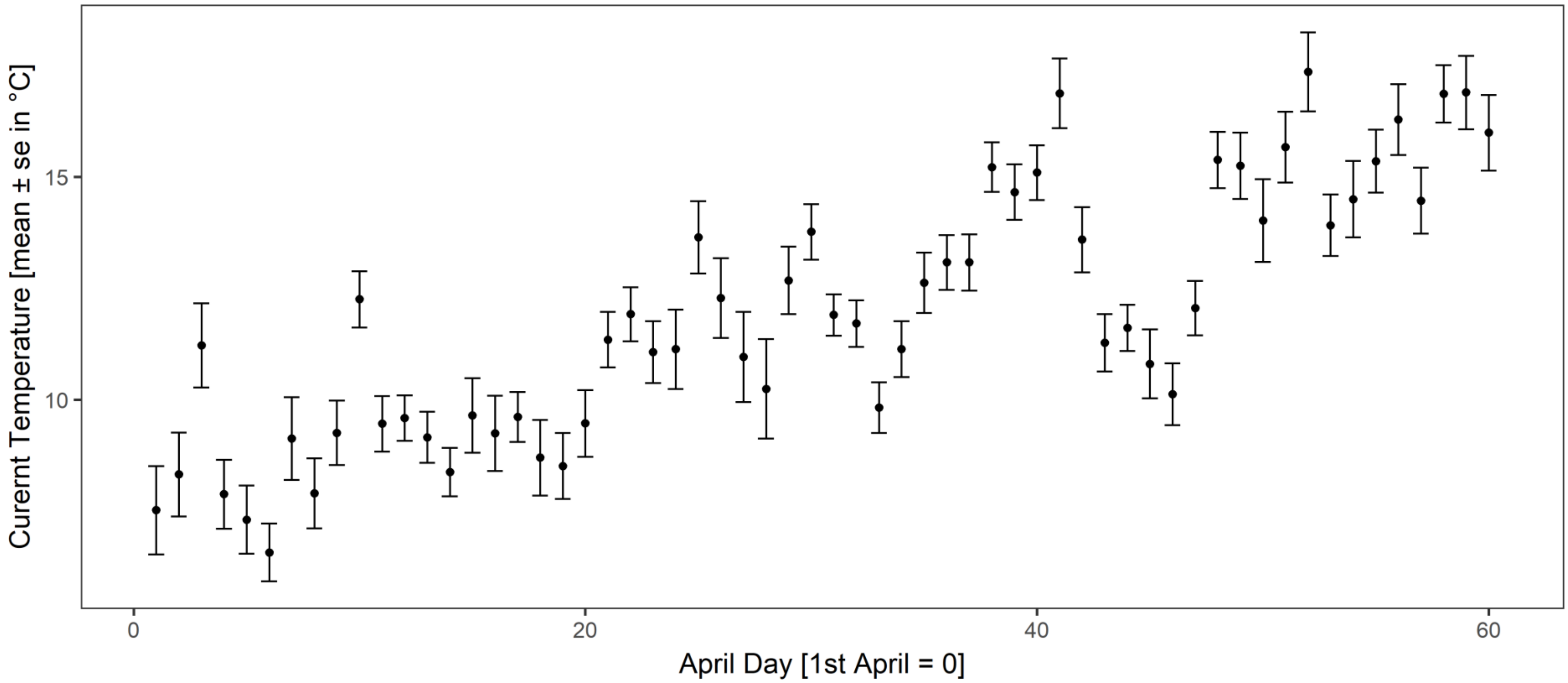
Estimated effect sizes and 95% credible intervals (CI) around the mean of temperature data from the iButtons during the breeding season 2017. A linear mixed effect model was run with following random effects: Day (n=65), time of day (n=1440), plot (n=12), nest box (n=201), cardinal direction (alignment of the nest box; n=8) and iButton ID (n=200).

Fixed Effects	**β (95% CI)**
Intercept	0.41 (0.17, 0.64)
Random Effects	**σ^2 (95% CI)**
Day	0.89 (0.81, 1.05)
Time of day[a]	0.23 (0.22, 0.24)
Plot	0.006 (0.004, 0.008)
Nest box	0.00
Cardinal direction	0.00
iButton ID	0.010 (0.008, 0.011)
Residual	0.19 (0.19, 0.19)

[a] as minutes from midnight

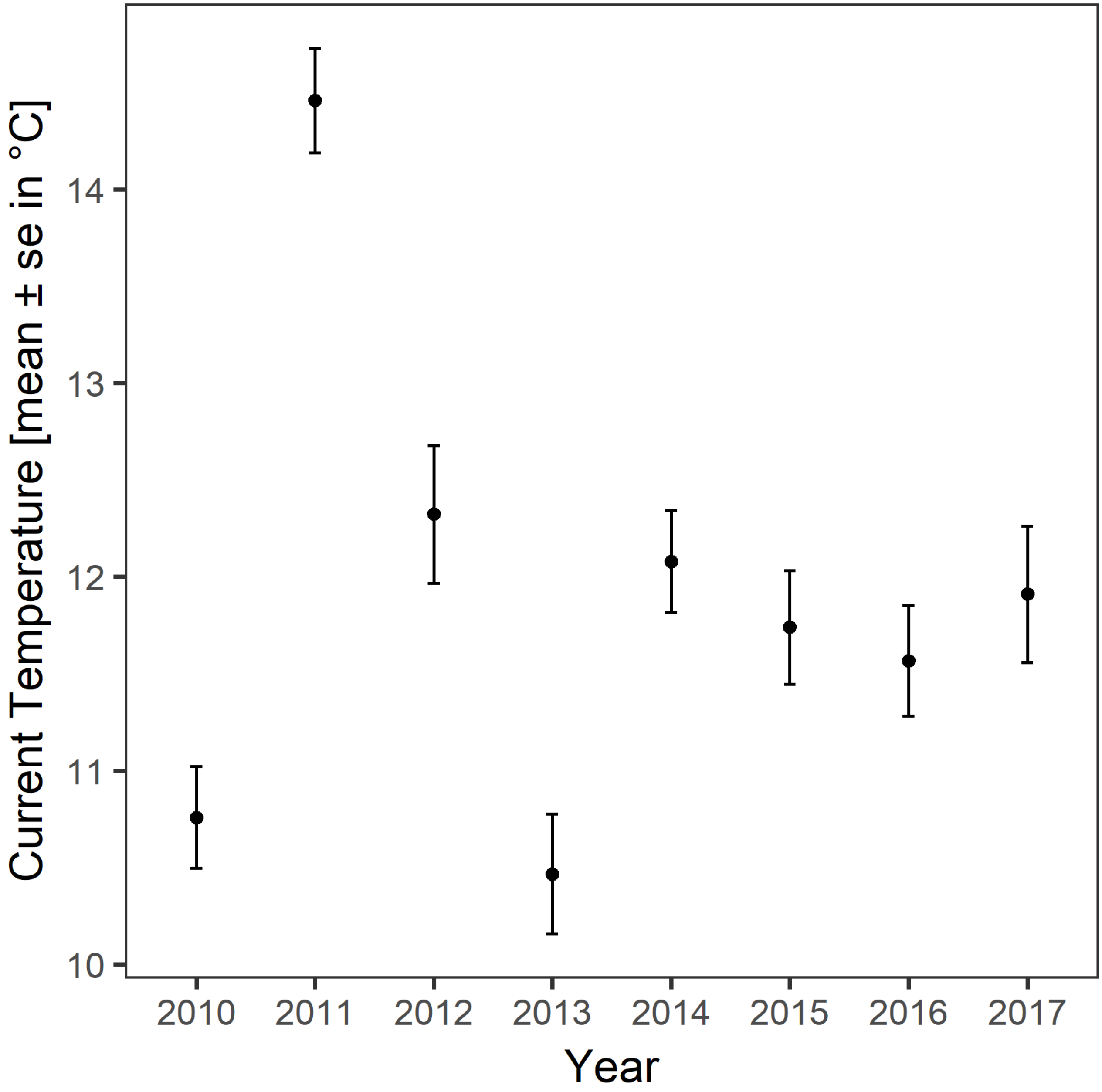
Supplementary Material Figure S1.

Current temperature over the course of the breeding season. Each point represents the mean current temperature (between 7h00AM and 12h00PM) of the 8-year study period per day. Vertical lines indicate standard errors.



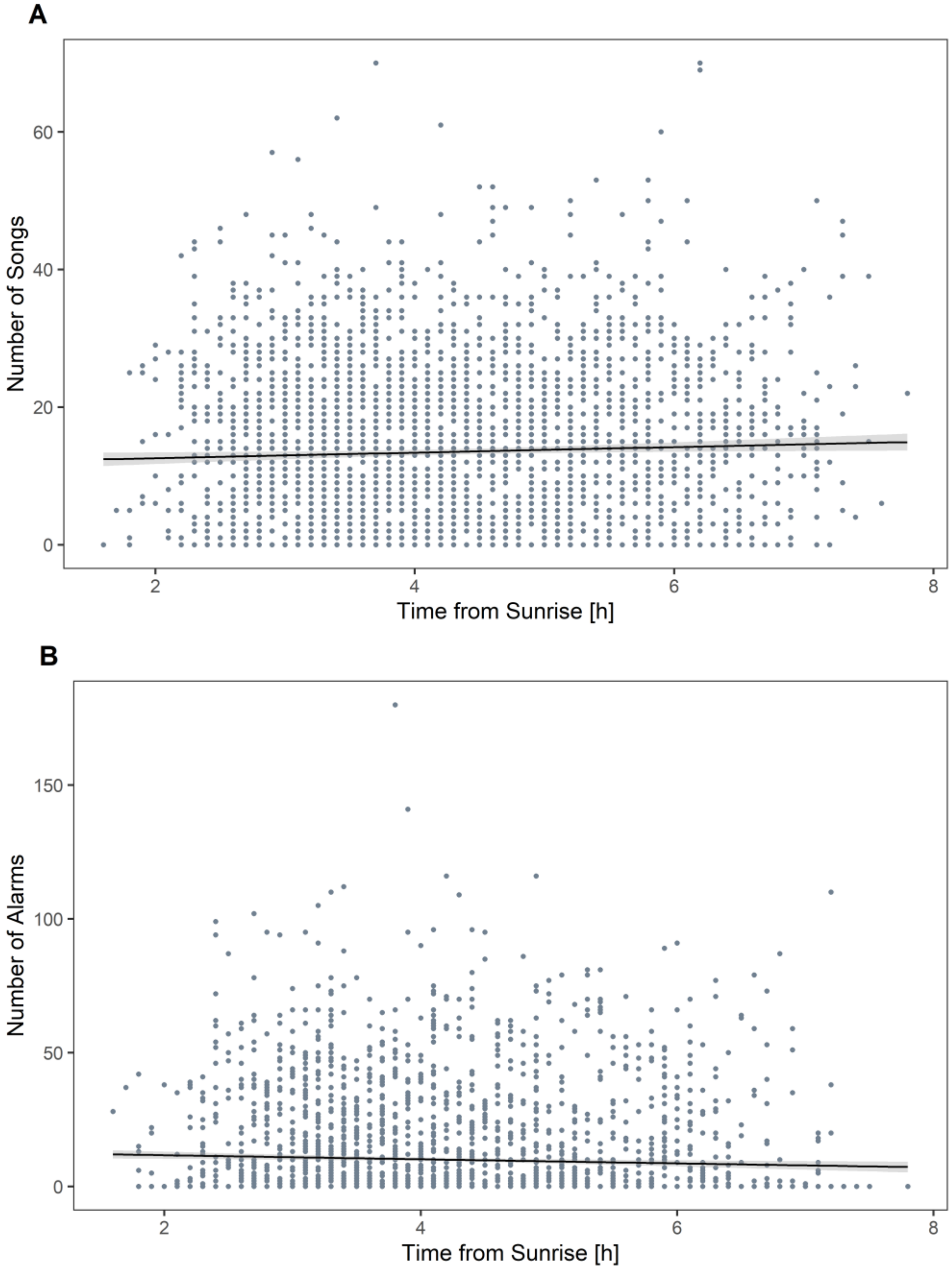
Supplementary Material Figure S2.

Mean current temperature (between 7h00AM and 12h00PM) for each of the 8 study years for the period from the 31st of March to the 30th of May. Vertical lines indicate standard errors.



Supplementary Material Figure S3.

Effect of time of day (in hours) on (A) song rate (n = 3334) and (B) alarm rate (n = 3318) expressed during territorial intrusions. Black lines represent regression lines and grey areas 95% confident intervals. Graph based on raw data.



Supplementary Material Text S1.

Birds might show plasticity in acoustic traits in response to either day-to-day or year-to-year changes in temperature. First, temperature effects may exist among years but within individuals, if individuals plastically adjust vocalisations to among-year changes in temperature. Second, effects may exist within years and within individuals if males plastically adjust their vocal response to day-to-day changes in temperature within the same year. Analyses of such effects require that temperature in our study area showed both within- and among-year variation, which is indeed the case (Supplementary Material Figures S1 and S2). To investigate whether detected effects of current temperature in our initial models (detailed above) were attributable to these two forms of within-individual plasticity, we re-ran our initial analysis after replacing the focal fixed-effect current temperature (x_{ijk}) (where i = observation, j = individual, and k= year) for two derived covariates: an individual's yearly mean temperature expressed as a deviation from the average temperature over all its assays $(\bar{x}_{jk} - \bar{x}_k)$, which examines the among-year-within-individual effect of temperature, and the deviation of each observation day's temperature from an individual's yearly mean temperature $(x_{ijk} - \bar{x}_{jk})$, which examines the within-year-within-individual effect of temperature.

Supplementary Material Text S2. Within-individual plasticity to within- and among-year changes in current temperature

The detected temperature effects can result from within-individual plasticity in response to within- or among-year variation in current temperature. We distinguished these two levels by reformulating our initial models (see Supplementary Text S2), which we achieved by partitioning our temperature variable to estimate within-individual plasticity in response to among-year-within-individual and within-year-within-individual changes in current temperature. Individuals sang more (among-year-within-individual effect of temperature: β = 0.05, CI = 0.024, 0.075) but alarmed less (β = -0.198, CI = -0.28, -0.116) in years that were relatively warm. Birds also sang more (within-year-within-individual effect of temperature: β = 0.013, CI = 0.001, 0.025), but did not change alarm rates (β = -0.036, CI = -0.074, 0.002) in warmer conditions within the same year (Table S2). Those analyses thereby demonstrated the existence of within-individual phenotypic plasticity in response to among-year and within-year changes in current temperature (particularly in the context of song rate).

We subsequently asked whether within- and among-year effects of temperature equally affected acoustic behaviour. 84% CIs of the two current temperature-variables overlapped for song (among-year-within-individual effect of temperature: β = 0.05, 84% CI = 0.037, 0.063; within-year-within-individual effect of temperature: β = 0.013, 84% CI = 0.006, 0.019), implying that the way that temperature affected singing did not differ within versus among years (Van de Pol & Wright, 2009). By contrast, the two current temperature-variables had different effects for alarming (among-year-within-individual effect of temperature: β = -0.198, 84% CI = -0.241, -0.155; within-year-within-individual effect of temperature: β = -0.036, 84% CI = -0.057, -0.017).

This demonstrated that alarming was adjusted differently to among- versus within-year changes in temperature; birds responded to the former but not to the latter.

Supplementary Material Table S2.

Estimated effect sizes and 95% credible intervals (CI) for predictors of number of songs (n = 3334) and number of alarms (n = 3318) shown by great tits in response to simulated territorial intrusions, analysing within-individual responses to current temperature.

	Song rate	**Alarm rate**
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept[†]	2.05 (1.93, 2.17)	-1.41 (-1.59, -1.25)
Current temperature		
Among-years-within-individuals	0.05 (0.02, 0.08)	-0.20 (-0.28, -0.12)
Within-individuals-within-years	0.01 (0.001, 0.03)	-0.04 (-0.07, 0.002)
Time of day	0.01 (-0.03, 0.06)	-0.22 (-0.37, -0.08)
Arrival time male	-0.07 (-0.11, -0.02)	-0.17 (-0.32, -0.03)
Breeding context	0.20 (0.16, 0.25)	-0.95 (-1.09, -0.81)
Test sequence	0.07 (0.03, 0.12)	-0.27 (-0.41, -0.13)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	0.17 (0.16, 0.19)	1.41 (1.27, 1.56)
Nest box	0.022 (0.019, 0.026)	0 (0, 0)
Date	0 (0, 0)	0 (0, 0)
Plot-year	0.01 (0.008, 0.014)	0 (0, 0)
Observer	0.06 (0.04, 0.08)	0 (0, 0)
Playback song	0 (0, 0)	0 (0, 0)
Model	0 (0, 0)	0 (0, 0)
Population	0.01 (0.004, 0.02)	0 (0, 0)
Year	0.006 (0.002, 0.01)	0 (0, 0)
Residual	0.12 (0.11, 0.14)	1.63 (1.50, 1.77)
Observation ID[‡]	1.44 (1.38, 1.50)	10.02 (9.56, 10.5)

[†] Reference category; Estimate for average current temperature variables, average time of day, average arrival time male and for tests during sequence 1 in the egg laying phase

[‡] Observation level random effect to account for overdispersion

Supplementary Material Text S3.

Birds might not only decrease their singing rates in response to low temperatures but also because of high temperatures. This pattern might occur, when physiologically demanding behaviours lead to hyperthermia and therefore need to be reduced or stopped (Guillemette et al., 2016). Temperature might thus have a non-linear effect on singing, where the number of songs is reduced under both, cold and warm conditions. To investigate this, we re-ran the models depicted in Table 1 after adding quadratic effect terms for current and night temperature, as well as the interactions of these effects with time from sunrise. We could not find evidence for non-linear patterns in the effects of temperature on any of the investigated response variables (see Supplementary Material Table S3).

Supplementary Material Table S3.

Estimated effect sizes and 95% credible intervals (CI) for predictors of number of vocalisations (n = 3278), number of songs (n = 3334) and number of alarms (n = 3318) shown by great tits in response to simulated territorial intrusions, analysing non-linear effects of temperature.

	Vocalisation count[†]	**Song rate**	**Alarm rate**	**Minimum distance**[‡]
Fixed Effects	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Intercept[§]	2.91 (2.83, 3.00)	2.02 (1.88, 2.16)	-0.98 (-1.26, -0.71)	-0.002 (-0.17, 0.17)
Current temperature (CT)	-0.01 (-0.02, 0)	0.02 (0.005, 0.03)	-0.08 (-0.11, -0.04)	0.003 (-0.01, 0.02)
Quadratic current temperature (QCT)	0 (-0.001, 0.001)	-0.001 (-0.002, 0.001)	-0.001 (-0.006, 0.003)	0 (-0.002, 0.001)
Night temperature (NT) [¶]	0.002 (-0.006, 0.01)	0.002 (-0.01, 0.02)	-0.02 (-0.06, 0.03)	0.01 (-0.003, 0.03)
Quadratic night temperature (QNT)	0 (-0.001, 0.002)	0.001 (-0.002, 0.004)	-0.002 (-0.01, 0.007)	0.003 (0, 0.006)
Arrival time male	-0.11 (-0.14, -0.08)	-0.07 (-0.12, -0.02)	-0.17 (-0.31, -0.04)	0.05 (0.01, 0.10)
Time of day	-0.002 (-0.05, 0.04)	0 (-0.07, 0.06)	-0.20 (-0.38, -0.01)	0.04 (-0.02, 0.10)
Breeding context	-0.02 (-0.05, 0.007)	0.19 (0.14, 0.24)	-0.79 (-0.92, -0.67)	0.29 (0.25, 0.33)
Test sequence	-0.009 (-0.04, 0.02)	0.08 (0.03, 0.12)	-0.23 (-0.35, -0.11)	0.06 (0.02, 0.10)
Interaction (CT*Time)	-0.001 (-0.007, 0.006)	-0.006 (-0.02, 0.006)	-0.03 (-0.06, 0.008)	0.005 (-0.006, 0.02)
Interaction (QCT*Time)	0 (-0.001, 0.002)	0.001 (0, 0.003)	0.002 (-0.003, 0.006)	0 (-0.001, 0.001)
Interaction (NT*Time)	0.003 (-0.005, 0.01)	-0.002 (-0.02, 0.01)	-0.01 (-0.06, 0.03)	0.008 (-0.005, 0.02)
Interaction (QNT*Time)	-0.001 (-0.002, 0.001)	-0.001 (-0.004, 0.001)	-0.002 (-0.009, 0.005)	-0.001 (-0.004, 0.001)

Random Effects	**σ² (95% CI)**	**σ² (95% CI)**	**σ² (95% CI)**	**σ² (95% CI)**
Individual	0.042 (0.038, 0.046)	0.18 (0.16, 0.20)	1.30 (1.17, 1.44)	0.26 (0.23, 0.28)
Nest box	0.007 (0.006, 0.009)	0.029 (0.025, 0.034)	0.39 (0.34, 0.45)	0.12 (0.10, 0.14)
Date	0.006 (0.005, 0.007)	0 (0, 0)	0.02 (0.017, 0.024)	0.008 (0.007, 0.01)
Plot-year	0.002 (0.001, 0.002)	0.013 (0.01, 0.017)	0.09 (0.06, 0.11)	0.013 (0.01, 0.017)
Observer	0.03 (0.02, 0.04)	0.06 (0.04, 0.08)	0.2 (0.14, 0.27)	0.08 (0.06, 0.10)
Playback song	0 (0, 0.001)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Model	0 (0, 0)	0 (0, 0)	0 (0, 0)	0 (0, 0)
Population	0.008 (0.003, 0.02)	0.01 (0.004, 0.02)	0 (0, 0)	0.02 (0.009, 0.05)
Year	0 (0, 0.001)	0.008 (0.002, 0.018)	0 (0, 0)	0.03 (0.007, 0.06)
Residual	0.053 (0.049, 0.057)	0.13 (0.11, 0.14)	1.30 (1.11, 1.51)	1.17 (1.12, 1.23)
Observation ID[††]	0.48 (0.46, 0.50)	1.47 (1.40, 1.53)	7.37 (7.02, 7.72)	-

[†] Sum of the number of songs and calls during a territorial intrusion test

[‡] Minimum approach distance of the focal bird to the dummy during a territorial intrusion test

[§] Reference category; Estimate for average current temperature, average night temperature, average time of day, average arrival time male and for tests during sequence 1 in the egg laying phase

[¶] Night temperature was expressed as deviation of minimum night temperature from current temperature

[††] Observation level random effect to account for overdispersion

References

Guillemette, M., Woakes, A. J., Larochelle, J., Polymeropoulos, E. T., Granbois, J.-M., Butler, P. J., … Portugal, S. J. (2016). Does hyperthermia constrain flight duration in a short-distance migrant? *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1704), 20150386. https://doi.org/10.1098/rstb.2015.0386

Publication II: Does song overlap signal aggressiveness? An experimental study with repeated measures in free-ranging great tits

Alexander Hutfluss, Veronika A. Rohr, Saray Scheidt, Linda Steinbichl, Eira Bermúdez-Cuamatzin, Hans Slabbekoorn, Niels J. Dingemanse

Published in Animal Behaviour, 2021

https://doi.org/10.1016/j.anbehav.2021.07.006

Abstract

Vocal communication is often used to signal willingness to escalate into a physical fight during territorial conflicts. In songbirds, starting to sing when an opponent already sings (song overlapping) has been suggested to signal aggressive intent (willingness to escalate). We used a multiyear data set to test whether song overlapping predicts aggressiveness in great tits, *Parus major*. Territorial males were subjected twice to a simulated territorial intrusion when their mate was in the egg-laying phase, and twice when she was incubating. Males were presented with a taxidermic mount and a noninteractive playback of a conspecific song near their nestbox. The experiment was conducted over 3 consecutive years, resulting in repeated measures for males that bred across multiple years. The estimated minimum approach distance to the intruder, a repeatable and heritable trait that predicts the likelihood of physical attack, was used as a measure of aggression. We determined the duration of song overlapping by the focal male relative to values expected by chance. Against expectations, we found that birds that overlapped were less (rather than more) aggressive. In addition, variance partitioning demonstrated that this link resulted from a within-individual effect: when birds became less aggressive from one observation to the next, they also overlapped more. There was no among-individual effect: individuals that were on average more aggressive did not, on average, overlap either more or less than others. Our results thus imply that song overlapping is linked to aggression but opposite to expectations, and not among individuals. Furthermore, the majority of birds overlapped at or below chance levels. Overall, song overlapping may not signal aggressive intent but rather ‘nonengagement’, or result from interference avoidance, allowing aggressive residents to better hear an intruder’s acoustic output during territorial intrusions.

Introduction

Individuals of many animal species exchange information through acoustic signals to moderate conflicts over mating partners or territories (Narins, Hödl, & Grabul, 2003; Sales, 1972; Waas, 1991). Physical fighting is not only energetically costly (Briffa & Elwood, 2004), but can also result in exposure to predators or cause injuries (Kelly & Godin, 2001), which may be prevented by signalling competitive abilities or motivation to fight (Galeotti, Saino, Sacchi, & Møller, 1997; Robertson, 1986; Shackleton & Ratcliffe, 2010). Despite the potential benefit of signalling, not all individuals may do so, because individuals vary in how they resolve the trade-off between costs and benefits of a fight depending on individual characteristics and condition (Maynard Smith & Price, 1973). Nevertheless, mutual assessment between competitors through acoustic signals is very common, for example in insects (Greenfield & Minckley, 2010), anurans (Arak, 1983; Reichert & Gerhardt, 2013; Robertson, 1986), mammals (Jennings, Elwood, Carlin, Hayden, & Gammell, 2012; Kitchen, Seyfarth, Fischer, & Cheney, 2003) and birds (Anderson, Searcy, Hughes, & Nowicki, 2012; Anderson, Searcy, Peters, & Nowicki, 2008; Capp & Searcy, 1991; Naguib, Altenkamp, & Griessmann, 2001). However, despite repeated calls for replication in ecological studies (Nakagawa & Parker, 2015), few studies have explicitly tested with sufficient within-study replication (e.g. across years) how aggression levels vary with the occurrence and nature of acoustic signals.

Signals used in agonistic contexts fall into two distinct categories: they convey information on either sender quality or motivation. First, vocalizations can serve as quality signals when morphological or physiological features of the sender restrict their production (Gil & Gahr, 2002). Quality signals, therefore, logically vary primarily among individuals, as individuals with insufficient capacity will be unable to, or have difficulty to, produce such signals. Examples of quality signals include vocalization rate, indicative of individual condition (Beani & Dessì-Fulgheri, 1995; Houtmann, 1992; Moller, Saino, Taramino, Galeotti, & Ferrario, 1998), or minimum frequency, indicative of size (ten Cate, Slabbekoorn, & Ballintijn, 2002). Second, vocalizations can convey a motivational message, which is not restricted by any physical features, but related to fluctuations in aggression or excitation (Guilford & Dawkins, 1995). Motivational signals vary, therefore, both among and within individuals, as any individual is able to produce them and may vary in motivational state. Some examples of motivational signals are increased twitter frequencies in European blackbirds, *Turdus merula* (Ripmeester, de Vries, & Slabbekoorn,

2007) and song type matching in song sparrows, *Melospiza melodia morphna* (Burt, Campbell, & Beecher, 2001), both associated with levels of arousal and probability of aggressive interaction.

Because patterns of variation in vocalizations can result from variation in quality (among individuals) and motivation (among and within individuals), study designs are required with repeated measures data, such that among- and within-individual effects can be teased apart (Allegue et al., 2017; Dingemanse & Dochtermann, 2013; Niemelä & Dingemanse, 2018; Westneat et al., 2020). Among-individual variation in behaviour has come to the foreground in adaptive animal personality research, which investigates the repeatable part of repeatedly expressed behaviours (Dingemanse, Kazem, Réale, & Wright, 2010; Réale, Reader, Sol, McDougall, & Dingemanse, 2007; Sih, Bell, & Johnson, 2004). Within-individual variation results instead from short-term internal fluctuations in motivation due to microenvironmental variation in context, and represents a form of reversible plasticity (Bradshaw, 1965; Dingemanse & Wolf, 2013; Piersma & Drent, 2003; Westneat, Potts, Sasser, & Shaffer, 2019). Previous research has already shown that aggressiveness is repeatable and heritable, but simultaneously exhibits reversible plasticity (Araya-Ajoy & Dingemanse, 2017; Betini & Norris, 2012; Thys, Pinxten, & Eens, 2021; Tuni, Han, & Dingemanse, 2018). Along the same lines, acoustic signals also vary among and within individuals (Amy et al. 2010; Naguib and Mennill 2010; Jacobs et al. 2014). Studies with repeated measures of aggression level and signal characteristics are required to disentangle effects of individual quality versus motivation, but are, to our knowledge, rare (Akçay, Campbell, & Beecher, 2014; Nowicki, Searcy, Krueger, & Hughes, 2002).

Birds are one of the most studied taxonomic groups in acoustic signalling research, particularly in the context of aggression (Collins, 2004; Searcy & Beecher, 2009). Birdsong serves a communication function in the context of both mate attraction (Eriksson & Wallin, 1986) and territory defence (Krebs, Ashcroft, & Webber, 1978). In territorial conflicts, acoustic signals convey information about aggressive intent, i.e. motivation to escalate into a physical fight (Cooney & Cockburn, 1995; Krebs et al., 1978). Indeed, various song traits have been hypothesized to signal aggressive intent in birds, including song type matching and frequency matching (Krebs, Ashcroft, & Orsdol, 1981; Shackleton & Ratcliffe, 2010; Vehrencamp, 2001), low-amplitude singing (Dabelsteen & Pedersen, 1990) or wide-frequency bandwidth singing with high trill rates (DuBois, Nowicki, & Searcy, 2009; Szymkowiak & Kuczyński, 2017). One of the most intriguing and controversial traits proposed to signal aggression is song overlapping (Helfer & Osiejuk, 2015; Naguib & Mennill, 2010; Searcy & Beecher, 2011).

Song overlapping occurs when an individual starts singing while its opponent is already singing (Masco, Allesina, Mennill, & Pruett-Jones, 2016), and has been reported for a range of animal taxa (Gerhardt & Huber, 2002; Naguib & Mennill, 2010; Schulz, Whitehead, Gero, & Rendell, 2008). In birds, various studies have proposed that overlapping represents a signal of aggressive intent (Brindley, 1991; Langemann, Tavares, Peake, & McGregor, 2000; Naguib & Kipper, 2006; Naguib & Mennill, 2010), positing that it varies plastically within individuals as a function of (social) context. Importantly, overlapping can also signal individual quality (Bischoff, Tschirren, & Richner, 2009; Kunc, Amrhein, & Naguib, 2006). If so, we expect repeatable among-individual variation in overlapping among repeatedly assayed individuals. The proposed functional value has been supported by studies showing that overlapping and aggression can covary positively (Brindley, 1991; Naguib & Kipper, 2006; Van Dongen, 2006). However, various other studies have also shown them to covary negatively (Akçay, Kağan, Avşar, Çabuk, & Bilgin, 2020; Osiejuk, Ratyńska, & Cygan, 2007; Vehrencamp, Hall, Bohman, Depeine, & Dalziell, 2007). These contrasting findings cast doubt on whether overlapping tendency signals aggressive intent or whether deviations from random overlap just aim at reducing signal interference and improving auditory perception (Planqué and Slabbekoorn 2008; Yang et al. 2014; Wilson et al. 2016). Altogether, the diversity of patterns and apparently contradictory findings make the generality of song overlapping as an aggressive signal an unresolved issue that requires further study (Searcy and Beecher 2009 & 2011; Helfer and Osiejuk 2015).

The great tit, *Parus major,* is a suitable model for studying the relationship between song overlapping and aggression. This territorial passerine readily breeds in nestboxes and frequently engages in territorial conflicts around the nestbox (e.g. Drent, 1987). It is a model organism in behavioural ecology (Davies, Krebs, & West, 2012), commonly used to study the role of song in territory acquisition and maintenance (Akçay et al., 2020; Dabelsteen, McGregor, Shepherd, Whittaker, & Pedersen, 1996; Langemann et al., 2000) as well as within- and among-individual variation in behaviour, including territorial aggression (Araya-Ajoy & Dingemanse, 2014, 2017). Moreover, several studies have addressed the production and perception of song variation in relation to animal personality in this species (Amy, Sprau, De Goede, & Naguib, 2010; Jacobs et al. 2014; Strauß, Hutfluss, & Dingemanse, 2019). In previous studies, we have already shown that song output during simulated intrusions was negatively correlated with aggression in great tits, while seasonal plasticity in aggressiveness was repeatable, heritable and age dependent (Araya-Ajoy & Dingemanse, 2014; 2017). The contradictory results from studies linking overlapping and aggression (Akçay et al., 2020; Langemann et al., 2000) beg the question of whether overlapping signals aggressiveness in great tits and, if so, whether it signals individual quality, motivation or both simultaneously.

In the current study, we used great tits to quantify song overlapping and aggression and tested whether these behavioural parameters covaried among and within individuals. Importantly, whether overlapping signals aggression can be studied from the sender's or the receiver's perspective, requiring different set-ups, data and analysis. We used an experimental approach with a standardized set-up of a multimodal, simulated, territorial intrusion into the subject's territory, with a taxidermic male model and a song playback (Araya-Ajoy & Dingemanse, 2014; 2017). In 3 consecutive years, we repeatedly tested the same individuals for their relative aggression levels in terms of approach distance and their vocal response behaviour and did so in two ecological contexts (during egg laying and incubation). Consequently, our set-up enabled us to study overlapping and aggression from the sender's perspective rather than addressing how birds respond to being overlapped. We had two main research questions: (1) for birds singing, and independent of song output, is song overlapping associated with aggression; (2) is song overlapping a personality-related (among-individual) signal of aggression or rather a context-dependent (within-individual) signal of motivation?

Methods

We monitored 12 nestbox plots (established in 2009) in southern Bavaria over a 3-year period (2017 - 2019); each plot was fitted with 50 nestboxes (Nicolaus et al., 2015). In each year, all boxes were checked twice a week from April onwards to determine life history traits, such as laying date (back calculated, assuming one egg per day was laid), onset of incubation (presence of an incubating female or warm eggs), clutch size and number of fledglings. Ten days after the offspring hatched, the adults were caught with a trapdoor in the box and then individually colour ringed for identification (detailed in Stuber et al., 2013).

Simulated Territorial Intrusions

As part of a long-term study, each male was subjected to four simulated territorial intrusions per year: two during the egg-laying stage (the first and the third day after the first egg was found) and two during the incubation stage (the first and the third day after warm eggs or an incubating female were observed; Fig. 1). Tests were postponed if weather conditions made natural intrusions unlikely and/or might damage our equipment (e.g. snowfall or rain). All trials were performed between 0700 and 1230 hours; the specific starting time was determined by the start of fieldwork and the duration and location of preceding tests, as tests were not performed in adjacent nestboxes.

During the test, we presented two stimuli to the focal birds: a visual stimulus (taxidermic mount of a male great tit) and an acoustic stimulus (playback of a great tit song). Both stimuli were chosen randomly from our large stock of 23 bird models and 175 song playbacks. All taxidermic mounts were placed within a green-wire mesh for protection and showed a neutral but variable posture. The model and speaker (Shockwave, Foxpro, Lewistown, PA, U.S.A.) were placed 1 m in front of the subject's nestbox. The speaker was placed on the ground (i.e. close to the model), while the model was placed on a 1.2 m wooden pole for consistent visibility across habitats. In 2017, the song stimuli consisted of just 13 recordings of spontaneous singing by local great tits recorded outside our study areas before the breeding season. In 2018 and 2019, we added another 161 playback stimuli made from songs recorded during the simulated territorial intrusions in 2017. Playbacks and speaker settings were adjusted to broadcast all playbacks at the same amplitude. Finally, playbacks were not standardized for equal song duration or overall length, as we preferred to retain variation in natural singing behaviour and hence variation in song duration and singing rate.

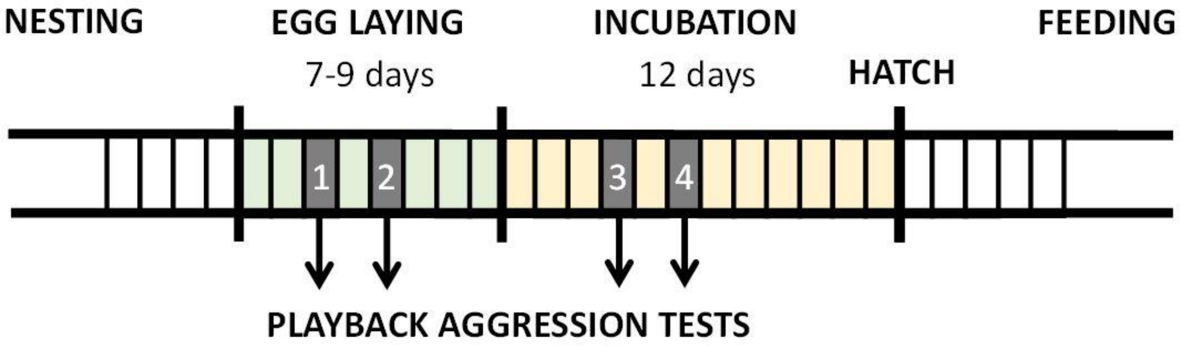


Figure 1: Typical sampling scheme. Durations of each breeding phase, for example egg laying and incubation, are approximations that differ depending on environmental conditions and individual behaviour. While the general pattern was fixed (1 day between tests within a breeding phase), tests were planned but not always executed on the marked days (day 3 and 5 of egg laying, and 4 and 6 of incubation).

Behavioural observations started once the focal male entered a 15 m radius around the nestbox. The observer, 15 m from the nestbox, subsequently counted the songs (‘song output’) and estimated the minimum distance to the model (‘approach distance’) for 3 min. We have previously shown that the latter behaviour represents an appropriate proxy for willingness to engage in physical attacks (Araya-Ajoy & Dingemanse, 2014). Subjects not entering the radius or not arriving within 15 min were scored as nonresponsive. All observers were thoroughly trained to reliably spot and identify the sex of great tits, and to categorize and count male vocalizations. For further details on the test procedure, see Araya-Ajoy and Dingemanse (2014; 2017). During the tests, the acoustic response of the focal male was recorded using a directional microphone (Sennheiser ME66/K6) covered with a windscreen and connected to a recorder (TASCAM DR-05, 44.1 kHz sampling rate, 16 bits sample size, WAV format).

Measuring Song Overlapping

We measured song output and determined the extent of overlapping (for those birds that sang) while controlling for song output. We measured overlapping by including all periods with co-occurrence of playback and responder songs, in which the responder had started to sing during the playback song and excluding those cases where overlap was caused by the playback song starting during a song of the responding bird. We compared the observed with the expected overlap duration, which we determined by randomly simulated song arrangements using the SONG package (song overlap null model generator, Masco et al., 2016). We analysed song recordings with Avisoft SASLab Pro v5.2 (Avisoft Bioacoustics, Berlin, Germany). We removed background noise below 2.0 kHz and above 8.0 kHz, by bandpass filtering our recordings between these frequencies. We normalized each file to 75% and created spectrograms for each focal song separately (FFT length = 1024 and window = FlatTop, giving a frequency resolution of 23 Hz, and overlap = 87.5%, giving a temporal resolution of 5.3 ms).

We calculated overlapping in three steps. First, we measured song duration based on start and end times, by using automatic measurements with three threshold levels (-24 dB, start -16 dB, end -12 dB). Second, we also determined the start and end times of one playback song in the recording, in the same way as with songs of the focal male and used these to calculate the temporal position of all other playback songs of the known stimulus. Third, observed and expected durations of the overlapping songs were calculated in R v3.5.1 (R Core Team, 2018) using the SONG package (Masco et al., 2016; Fig. 2). The expected level of overlapping was calculated as the mean length of overlapping of 1000 randomizations, during which the focal bird's songs were randomly rearranged with varying interval lengths between them (SampleGaps method). We then determined a chance-corrected overlapping score, calculated as the observed minus the expected total overlapping time. This score reflects whether an individual overlapped the playback more (positive values), the same (zero) or less (negative values) than expected by chance. Recordings for which any of the three steps could not be taken were excluded from the analysis. This occurred because of elevated noise levels, multiple birds singing or when it was not possible to assign songs to the playback versus the focal bird. For each recording, we also calculated whether the overlap differed significantly from chance. Here, a significant deviation from chance occurred when the observed overlapping value was in the top or bottom 2.5% of the null distribution of 1000 simulated values expected by chance.

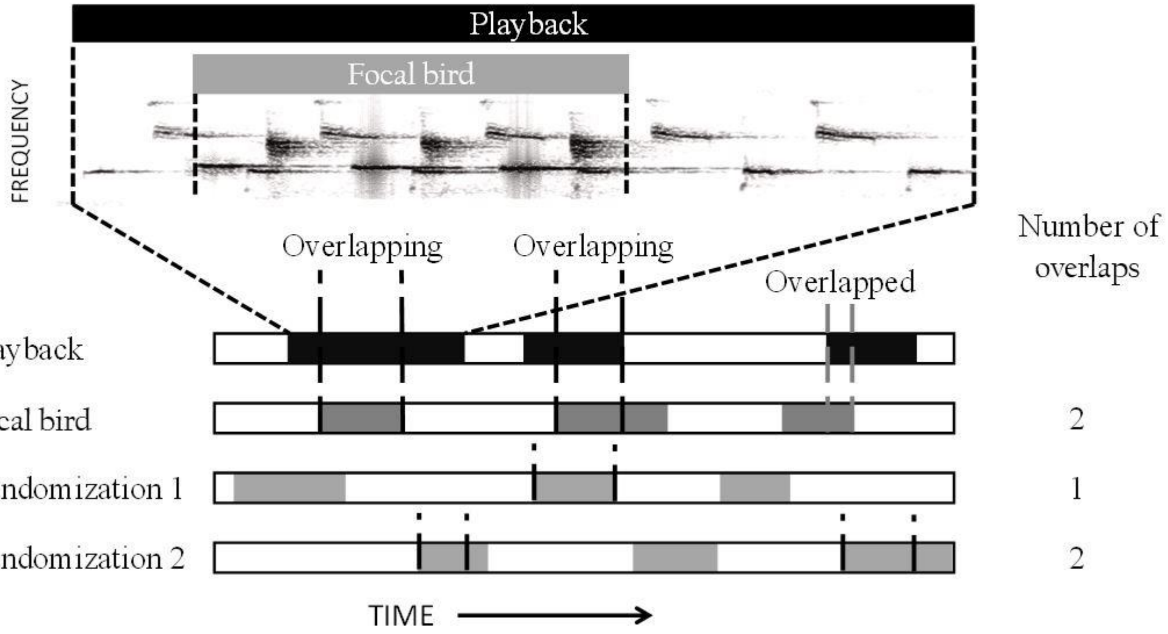


Figure 2: Comparison of playback and focal bird songs to obtain the duration of overlapping songs. Overlapped (i.e. playback song starts while focal male already singing) songs were not considered. Two example randomizations (out of 1000) are shown following an actual example of an overlapping focal bird song. Chance-corrected overlapping was then calculated as the observed length of overlapping (s) minus the mean length of overlapping of 1000 randomizations.

Statistical Analyses

We used univariate linear mixed-effect models with a Gaussian error distribution to test the links between our response variables, (1) song output and (2) chance-corrected overlapping score, and aggression. We multiplied the distance of closest approach by -1 and added the furthest distance from the model a bird showed in our data set to each score, so that the strongest response had the highest score. To account for differences between breeding seasons in overall singing behaviour, we added year (2017, 2018 or 2019) as a fixed-effect factor. Our previous work has shown that aggressiveness, singing and alarm calling can change with breeding context and with test sequence (Araya-Ajoy & Dingemanse, 2014). We thus included nest stage (0 = egg laying and 1= incubation) and test sequence (0 = first test within a stage and 1 = second test within a stage) as fixed-effect factors in all models. Additionally, because focal birds may overlap more when they sing more, we added the song output of the focal individual as a covariate in an expanded model.

As song length may also represent a signal of aggression (Nelson & Poesel, 2011; Osiejuk & Jakubowska, 2017), we added song length of the focal individual as a covariate in a separate run of the expanded model.

Each model also fitted random intercepts for individual identity (song output = 369 and chance-corrected overlapping score = 251 individuals), to estimate among-individual variation. Furthermore, random intercepts were added for playback song recording (song output = 175 and chance-corrected overlapping score = 144 playbacks) and taxidermic model (song output = 18 and chance-corrected overlapping score = 17 models), to control for variation caused by features of the two stimuli. We also added plot-year (i.e. the unique combination of plot and year, 34 levels) to account for spatiotemporal effects that may vary among plots or years, or within years among plots, such as breeding density (see Appendix). These analyses showed that plot-year explained extremely little variation in song output and chance-corrected overlapping (Appendix Table A1), implying that any factor varying at this level is unlikely to have much effect. Finally, we controlled for variation caused by the field observer (N=19) by fitting it as a random effect. In a subsequent analysis, we analysed each year separately, while keeping the model structure the same, to investigate whether effects of aggression on chance-corrected overlapping were consistent across years. The repeatabilities of chance-corrected overlapping and song output were estimated from models where aggressiveness was not fitted as a fixed-effect covariate. This was because our aim was to estimate the trait's overall repeatability rather than that while controlling for (repeatable) variation in aggression (for further discussion, see Dingemanse & Dochtermann, 2013).

In a second step, we aimed to disentangle among- and within-individual effects of aggression on singing behaviour by partitioning the overall effect into these two distinct components (Dingemanse & Dochtermann, 2013; Van de Pol & Wright, 2009). First, effects might exist among individuals if males characterized by specific aggression levels can also be characterized by specific singing characteristics, i.e. specific levels of song output or chance-corrected overlapping. This would indicate that song output and chance-corrected overlapping are personality-related signals rather than signals of short-term motivation to escalate a conflict. Second, effects might exist within individuals if males plastically adjust singing to changes in aggression, indicative of context-dependent motivation. We, thus, reran the initial models on song output and chance-corrected overlapping score after splitting the predictor aggression into two derived variables: the individual's mean aggression ($\bar{x}$) assessing the among-individual effect (β_A)

and the observation's deviation from the individual's mean $(x_i - \bar{x})$ estimating the within-individual effect (β_W). The inclusion of individuals with one data point is recommended as it improves the precision of both fixed- and random-effect estimates (Martin, Nussey, Wilson, & Réale, 2011).

All statistical analyses were conducted in R version 3.5.1 (R Core Team, 2018). Linear mixed-effects models were performed using the 'lmer' function of the package lme4. We present the mean (β) and 95% credible intervals (CIs) for each fixed and random effect parameter, derived from 2000 simulations implemented using the 'sim' function of the package arm. An effect was regarded as statistically significant when its associated 95% CI did not overlap zero (Nakagawa & Cuthill, 2007). All random effect estimates are represented as adjusted repeatabilities, defined as the proportion of the total phenotypic variation not explained by fixed effects.

Ethical Note

This study was approved by the Regierung Oberbayern (permit number ROB-55.2-2532.Vet_02-17-215) in accordance with the ASAB/ABS Guidelines for the use of animals in research. Our experiments were designed to minimize subject discomfort.

Results

General Data

Overall, we performed 2740 simulated territory intrusions of which 1338 had a male of known identity responding. We collected 481 (of 1558 recordings in total) song recordings of 251 birds of a quality sufficient for further analyses (2017: 200; 2018: 140; 2019: 141 recordings). In total, 210 birds were recorded in 1 year only, 38 birds in 2 years and three birds in 3 years. Not all birds responded (i.e. arrived within 15 min) or produced an analysable recording across all four tests in a given year: 107 males yielded one recording, 82 had two, 44 had three, 14 had four, three had five and one had seven recordings.

Of the 481 recordings, 99 (20.6%) overlapped significantly less than expected by chance, while 369 recordings (76.7%) overlapped according to chance. Overlap above chance levels thus occurred rarely (13 recordings; 2.7%). Following chance correction, overlapping scores correlated negatively with song output but not with other traits (see Table 1). A plot of the data showed that birds producing few songs overlapped the recording according to chance levels, while those producing many songs overlapped less than expected by chance (Fig. 3). Aggressiveness, song output, number of alarm calls and the number of attacks were correlated as previously recorded for this population (based on another data set; Araya-Ajoy & Dingemanse, 2014). These insights support the validity of our decision for minimum distance as a suitable measure of aggressiveness.

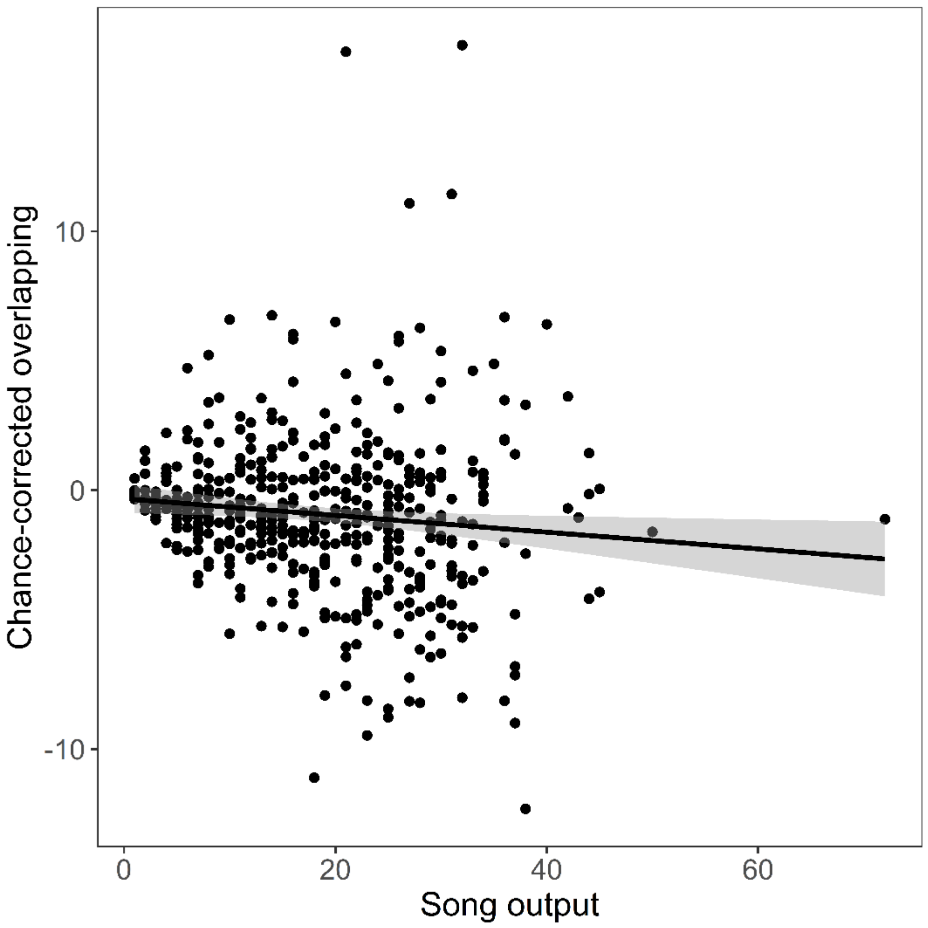


Figure 3: Song output (number of songs sung per 3 min) and chance-corrected overlapping (duration of overlapping compared with expected value) plotted against each other for all 3 years combined (N = 481). Black lines represent the regression line, while the grey areas indicate the 95% confidence interval. Graph based on raw data.

Table 1: Correlation matrix between chance-corrected overlapping and the four response behaviours measured during the tests

	Chance-corrected overlapping	Aggressiveness	Song output	Alarm calling	Attacks
Aggressiveness	-0.05 (-0.15, 0.05)	-			
Song output	**-0.12 (-0.21, -0.02)**	**-0.25 (-0.30, -0.20)**	-		
Alarm calling	0.04 (005, 0.14)	**0.28 (0.23, 0.33)**	**-0.48 (-0.52, -0.44)**	-	
Attacks	0.02 (-0.12, 0.07)	**0.31 (0.26, 0.36)**	**-0.23 (-0.28, -0.18)**	**0.09 (0.04, 0.14)**	-

Overlapping: *N* = 481; other traits: *N* = 1338. Significant correlations (P <0.05) are shown in bold.

Covariance Between Aggression, Song Output and Song Overlapping

We found strong support in the raw (unpartitioned) data for a negative overall link between aggression and song output for all years combined (effect of aggressiveness, Appendix Table A1). Song output decreased with increasing aggression. This pattern was observed in all years (Fig. 4, Appendix Table A2), implying that it was not a year-specific effect but rather a general phenomenon. As observed in data collected from the same population in previous years (Araya-Ajoy & Dingemanse, 2014, 2017), we found that birds produced more songs during the incubation than the egg-laying stage (effect of nest stage Fig. 5). Song output did not change with test sequence (within-stage) and song output of the average bird did not differ between years (Appendix Table A1). Finally, song output showed significant repeatability, regardless of whether the effect of aggression was controlled for or not (see Appendix Table A1).

There was also strong support for a link between chance-corrected overlapping and aggression in the raw (unpartitioned) data (Appendix Table A1). Importantly, year-specific analyses demonstrated a lack of support for this link in 2 of the 3 years of study, indicating that the overall effect was mainly caused by one specific year (Fig. 4, Appendix Table A3). Chance-corrected overlapping did not vary with year, nest stage or test sequence (Fig. 5) but it did covary negatively with the song length of the focal bird during the recording, indicating that birds that sang for longer overlapped the playback for shorter periods (Appendix Table A1). It also covaried negatively with the number of songs the focal bird produced, indicating that birds that sang more showed less overlapping than expected (Appendix Table A1). In contrast to song output, the repeatability of overlapping was low, implying that variation in overlapping was caused by within- rather than among-individual processes (see Appendix Table A1).

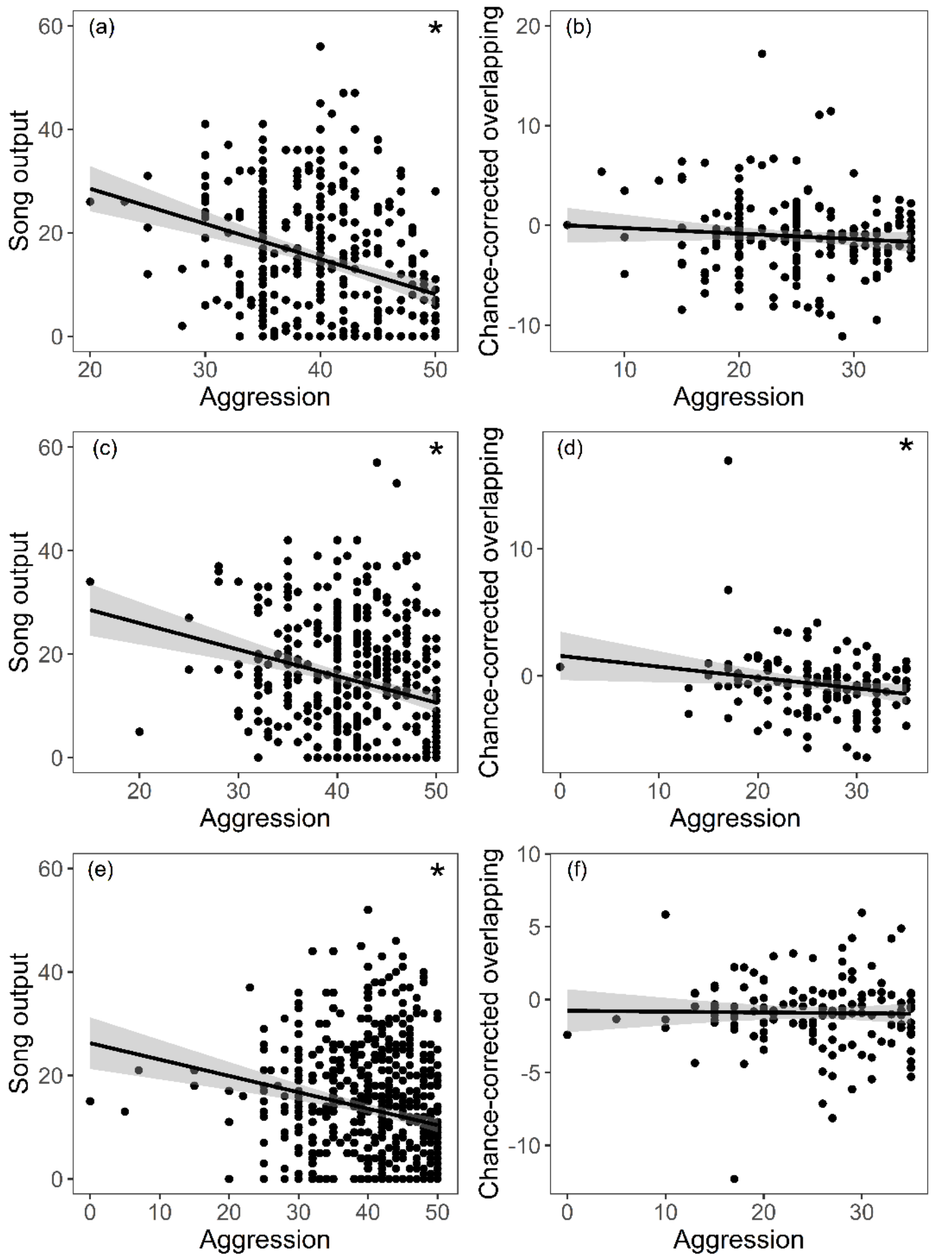


Figure 4: (a, c, e) Song output (number of songs sung per 3 min; *N* = 1338 simulated territory intrusions) and (b, d, f) chance-corrected overlapping (duration of overlapping compared with expected value; *N* = 481 recordings) plotted against the aggression score (approach distance multiplied by -1 + maximum approach distance) for each year separately. (a, b) 2017; (c, d) 2018; (e, f) 2019. Black lines represent the regression line, while the grey areas indicate the 95% credible interval. The '*' symbol in the top right corner indicates cases where patterns of nonzero

covariance were supported by the data (i.e. credible intervals not overlapping zero). All graphs based on raw data.

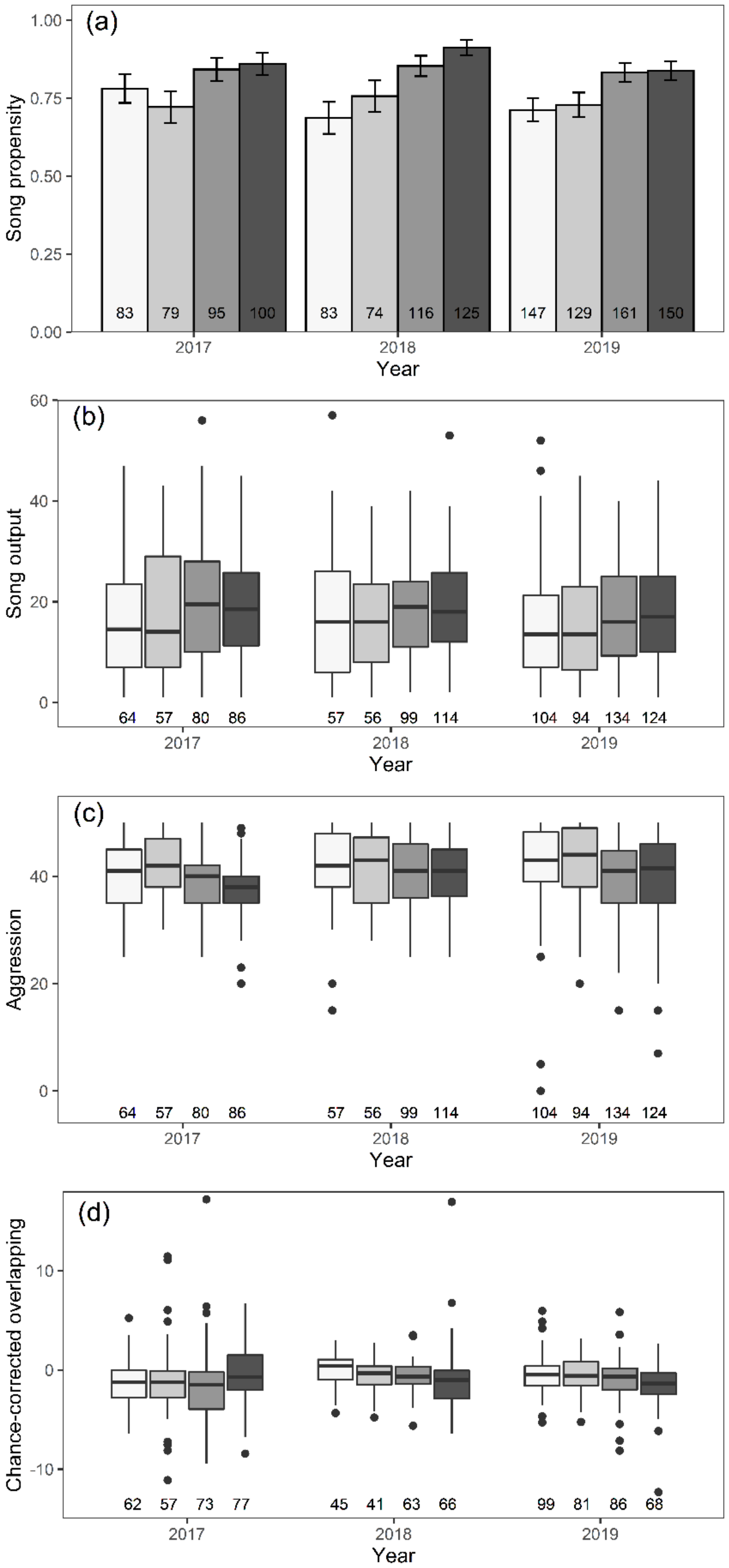

(a)
Song propensity
1.00
0.75
0.50
0.25
0.00
83
79
95
100
83
74
116
125
147
129
161
150
2017
2018
2019
Year
(b)
Song output
60
40
20
0
64
57
80
86
57
56
99
114
104
94
134
124
2017
2018
2019
Year
(c)
Aggression
40
20
0
64
57
80
86
57
56
99
114
104
94
134
124
2017
2018
2019
Year
(d)
Chance-corrected overlapping
10
0
-10
62
57
73
77
45
41
63
66
99
81
86
68
2017
2018
2019
Year

Figure 5: (a) Song propensity (singing no songs or at least one), (b) song ouput (number of songs sung per 3 min), (c) aggression score (approach distance multiplied by -1 + maximum approach distance) and (d) chance-corrected overlapping (duration of overlapping compared with expected value; negative values indicate less overlapping than expected by chance) during the egg-laying and incubation stages within each year. In (a) as song propensity is a binary trait, the average probability of singing per test sequence is shown with standard error. In (b, c, d) the box plots show the median and 25th and 75th percentiles; the whiskers indicate the values within 1.5 times the interquartile range and the circles are outliers. Within years, the first two bars/box plots represent the egg-laying phase and the last two represent the incubation phase. Numbers at the bottom show sample sizes. Graphs on song output, aggression and overlapping exclude birds without songs.

Among- versus within-individual effects of aggression

As a next step, we partitioned the patterns in the raw data (Fig. 6) into among- and within-individual effects. This demonstrated that individuals that were on average (over all their observations) more aggressive also, on average, sang less (Fig. 6, Table 2). The same negative effect was observed within individuals: reductions in aggressiveness were associated with increased song output across observations (days) of the same individual (Fig. 6, Table 2).

Applying the same partitioning of the raw data (Fig. 6) to chance-corrected overlapping demonstrated that the average aggression level of an individual did not predict its overlapping behaviour. Specifically, there was no among-individual effect of aggression on chance-corrected overlap (Fig. 6, Table 2). By contrast, changes in aggressiveness and chance-corrected overlapping correlated among observations (recordings) of the same individual, indicating that when birds upregulated their aggression from one observation to the next, they also downregulated their overlapping of the intruder’s song (Fig. 6, Table 2). Models where all random effects were excluded except the individual gave the same results (not shown).

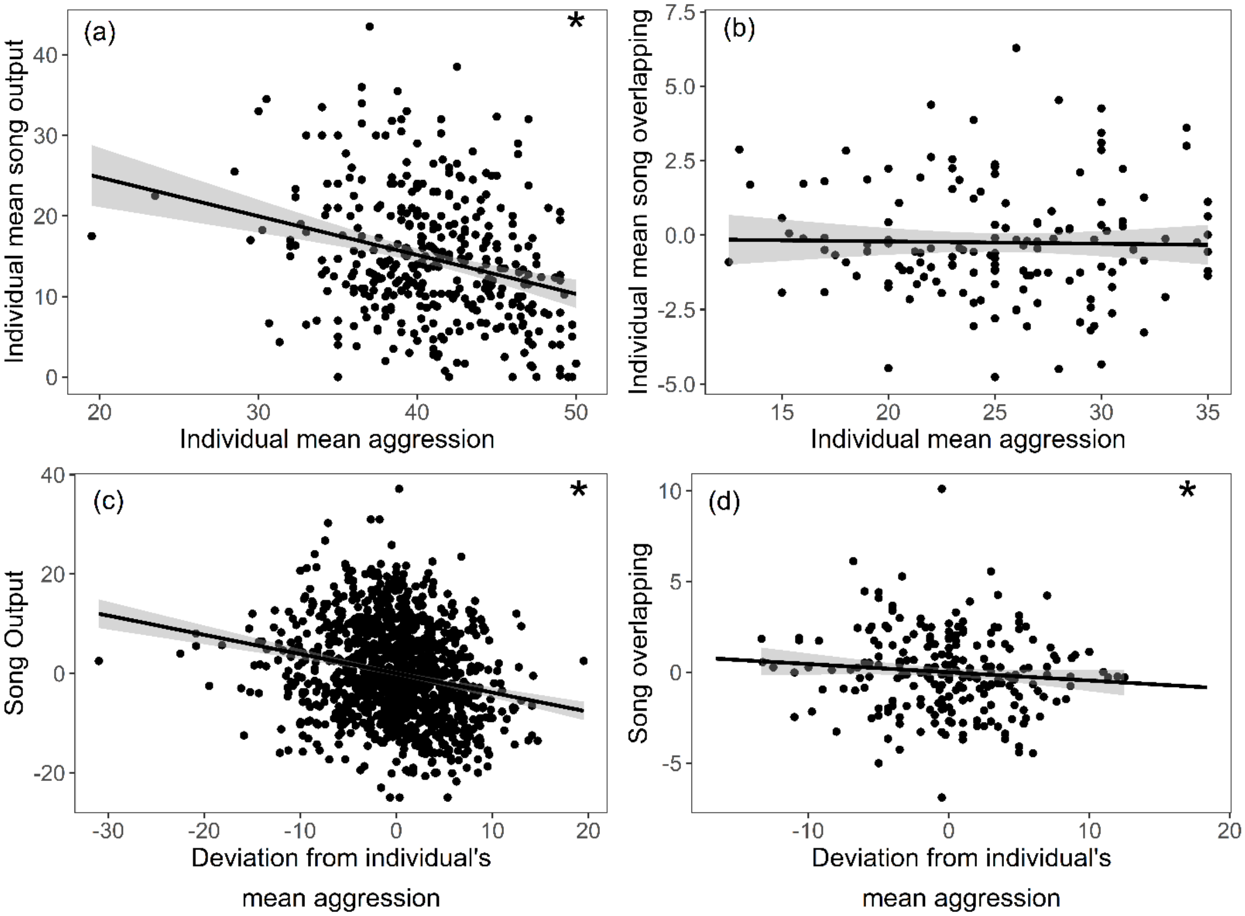


Figure 6: Visualization of (a, b) among-individual and (c, d) within-individual relationships for (a, c) song output (number of songs sung per 3 min) and aggression (approach distance multiplied by -1 + maximum approach distance) and (b, d) chance-corrected overlapping (duration of overlapping compared with expected value; negative values indicate less overlapping than expected by chance) and aggression. Among-individual relationships are visualized by plotting mean values per individual. Within-individual relationships are visualized by plotting each observation's deviation from the individual's mean value for the two focal traits. Black lines represent the regression line, while the grey areas indicate the confidence interval. The '*' symbol in the top right corner indicates cases where patterns of nonzero covariance were supported by the data (i.e. credible intervals not overlapping zero). All graphs based on raw data.

Table 2: The 95% credible interval around the estimated mean parameter estimate for fixed effects on chance-corrected overlapping and song output

	Song output N = 1338	Chance-corrected overlapping N = 477
	β/σ^2 (95% CI)	β/σ^2 (95% CI)
Fixed effects		
Intercept[a]	9.03 (6.14, 11.94)	-0.71 (-2.01, 0.65)
Aggressiveness		
Among individuals[b]	-1.70 (-2.37, -0.99)	0.12 (-0.29, 0.49)
Within individuals[c]	-1.90 (-2.48, -1.27)	-0.45 (-0.85, -0.06)
Year		
2018	0.92 (-1.40, 3.20)	0.36 (-0.37, 1.06)
2019	-0.77 (-3.02, 1.53)	0.07 (-0.61, 0.76)
Test sequence	0.69 (-0.42, 1.81)	0.20 (-0.39, 0.75)
Nest stage	2.74 (1.59, 3.91)	-0.45 (-1.03, 0.11)
Song output focal bird	-	-0.38 (-0.65, -0.09)
Song length focal bird[d]	-	-0.29 (-0.57, 0.02)
Random effects[e]		
Individual	0.153 (0.147, 0.157)	0.024 (0.021, 0.026)
Playback	0.0048 (0.0042, 0.0053)	0.002 (0.009, 0.001)
Plot-year	0.010 (0.006, 0.013)	0 (0, 0)
Observer	0.03 (0.02, 0.05)	0 (0, 0)
Taxidermic mount	0 (0, 0)	0 (0, 0)
Residual	0.80 (0.78, 0.82)	0.975 (0.973, 0.978)

β is given for fixed effects and σ^2 for random effects. CI: credible interval.

[a] Population mean for the reference of all other fixed effects, i.e. test sequence 0, nest stage 0, year 2017, for birds of average aggression (among-individual aggressiveness) and their average aggressiveness (within-individual aggressiveness).

[b] Mean aggressiveness over all records of the same individual.

[c] Deviation of each observation from an individual's mean aggressiveness.

[d] Estimate from a separate model with the same model structure, except for the exclusion of song output.

[e] Random-effect estimates are presented as adjusted repeatabilities, defined as the proportion of variation not explained by fixed effects.

Discussion

We used an experimental approach with repeated measures across seasons and multiple years to test whether great tits use song overlapping tendency as a signal of aggressive intent, and whether overlapping tendency correlates with variation in aggression among- and within-individuals. We found that (1) chance-corrected overlap correlated negatively rather than positively with aggressiveness in the data set overall and (2) this relationship existed within but not among individuals. Specifically, within-individual increases in aggression across separate intrusions were associated with within-individual decreases in chance-corrected overlapping. Altogether, our findings imply that overlapping can indeed act as a motivational signal of aggressive intent. However, the correlation between aggression and overlapping was opposite to expectations, while the majority of birds overlapped equal to or less than chance levels. Altogether, these findings imply that the role of overlapping in communicating aggression is not straightforward or prominent and that alternative adaptive mechanisms, such as interference avoidance, might explain our data.

Overlapping as a Signal of Aggression?

Our multiyear study contributes a well-replicated example to the literature (Helfer & Osiejuk, 2015; Naguib & Mennill, 2010; Searcy & Beecher, 2009). Searcy and Beecher (2009) argued that a signal of aggressive intent should (1) occur more often in agonistic than in nonagonistic interactions (context criterion), (2) correlate with other aggressive behaviours or predict a subsequent escalation (predictive criterion) and (3) elicit a differential reaction (either stronger or weaker) in receivers (response criterion). We address the literature and our own results here based on these criteria.

Context criterion

Evidence for the occurrence of overlapping varying with context can come from observational, descriptive studies, but is most compelling when derived from experimental studies. Convincing evidence could, for example, come from comparing singing interactions between low-level and high-level arousal conditions, such as before and after playback designed to cause distinct shifts in arousal. We are not aware of any great tit study that has explicitly tested for this context-dependent occurrence of overlapping, but Brindley (1991) reported an example of this in European robins, *Erithacus rubecula. Robins overlapped more in response to a non-neighbour song playback, causing relatively high arousal, than to a neighbour song playback, causing only moderate arousal. The overlapping tendency also decreased with the distance between the responding robin and the playback speaker.* Similar types of analyses would be possible in great tits but have not been performed.

Predictive criterion

Akçay et al. (2020) recently reported a series of playback experiments in which they tested the correlation between aggression and song overlapping in great tits. They found that a number of behaviours indicative of aggression, such as number of flights, closest approach distance and proportion of time spent within 1 m of the speaker (combined in a principal component) were negatively correlated with overlapping rate. Their study design was similar to ours, in being from within territory boundaries, close to the nest. However, they did not repeat their tests as often as we did, nor did they use a taxidermic mount as a visual stimulus. Despite these differences their findings are corroborated by our own results, as we did find similar patterns. Importantly, neither of the studies provides any evidence for overlapping being a signal of increased aggressive intent.

Response criterion

In a relatively early study on conflict escalation in great tits, Langemann et al. (2000) showed that male great tits exhibited response patterns in three subsequent stages of a playback procedure that differed depending on whether the intruder playback overlapped. The great tits responded most to the overlapping playback stimulus in terms of closest approach and number of flights longer than 5 m, although the birds were always exposed to overlapping last, which made playback sequence a confounding factor. In a later study, Amy et al. (2010) showed that overlapping songs induced higher song output and higher switching rates than alternating songs in their interactive playback setting. Using an exceptional set-up, Peake et al. (2001) reported a more indirect test for the response criterion. Focal individuals were eavesdropping on an interaction between two male great tits and were presented one of the two competitors afterwards. Although not all assayed behaviours were affected, when presented with the individual that had been overlapped, birds had a lower song output and made fewer song type switches. Notably, two of the three studies for this criterion found links between overlapping and acoustic traits but not with approach measures.

Specific findings on the response criterion indicated that being overlapped can also influence behavioural responses (Amy et al., 2010; Langemann et al., 2000). To account for this, we tested whether birds that were initially overlapped by the playback differed from birds that were not. We found no evidence for such behavioural differences in song overlapping, song output or aggression, indicating that our results were not biased by birds adjusting their responses to the 'behaviour' of our simulated intruder (see Appendix and Table A4).

Interference Avoidance

Based on this overview of the recent literature, we believe that overlapping could serve a signalling function, but there is no support for overlap as a signal of strong aggressive intent, at least not in great tits. Given the general lack of a positive correlation between aggression and overlapping in previous work and our own study, a more likely explanation for a deviation from random overlapping tendency may be related to withdrawal or nonescalation, or with interference avoidance (Ficken, Ficken, & Hailman, 1974; Searcy & Beecher, 2009; Vehrencamp et al., 2007). Indeed, most of our birds either did not show overlap that differed from chance levels or primarily overlapped less (rather than more) than expected by chance. This avoidance seemed to be especially pronounced for individuals that produced more or longer songs. Overall, our evidence for overlap avoidance may suggest great tits seek to increase the audibility of singing competitors, as previously suggested for other species (Wilson et al., 2016; Yang et al., 2014).

Personality or Context-dependent Motivation?

Our study adds to the picture of the signalling role of overlapping emerging from the literature over the past few decades (Akçay et al., 2020; Brindley, 1991; Langemann et al., 2000) and further represents a showcase for the benefits of repeated measurements. The first benefit concerns statistical robustness. Testing the same individuals repeatedly across and within years allowed us to assess the temporal consistency between song output, overlapping and aggression and, indirectly, whether year-specific factors affected these associations. While song output covaried negatively with aggression in all 3 years, song overlapping covaried with aggression in only 1 of these years. For song output, patterns observed in 1 year are thus representative of within-year patterns generally, whereas for overlapping, this is not the case. Consequently, variation among years in, for example, body condition, which can affect acoustic behaviours (Gil & Gahr, 2002; Gottlander, 1987), is unlikely to affect song output. Within years, we confirmed the findings of Araya-Ajoy and Dingemanse (2014) on song output, showing an increase over the breeding season, potentially indicating a shift in intruder threat, for example via changed risks of paternity loss (Araya-Ajoy et al., 2016), leading to a supposedly weaker response. Song overlapping, even though exhibiting a suggestive pattern in 2019, did not consistently change with breeding context.

Besides the statistical robustness, the second benefit of our repeated measures design is more important as it provides completely novel insights. The repeated sampling of the same individual, especially across years, allows partitioning of long-term repeatable among- (‘animal personality’) and within-individual (‘reversible plasticity’) variation in behaviour (Dingemanse & Dochtermann, 2013). Besides showing misestimated or undetected effects, caused by independent and opposing associations of the two components with a trait of interest (Moiron, Laskowski, & Niemelä, 2020; Van de Pol & Wright, 2009), this partitioning also led to valuable insights into the underlying mechanisms of aggressive signalling. While a within-individual effect indicates motivation or behavioural adjustments to environmental factors (Strauß et al., 2020), among-individual effects can result from genetic differences or early life conditions resulting in differing abilities to express certain behaviours (Bischoff et al., 2009).

We showed a significant within-individual effect of aggression on overlapping and simultaneously no evidence for an among-individual effect. This indicates that changes in song overlapping avoidance resulted from changes in aggression levels between days. This is in line with the idea that overlapping signals context-dependent motivation (DuBois et al., 2009; Ripmeester et al., 2007), rather than a personality-related tendency, signalling an individual’s average aggression or typical style of vocal interaction (Amy et al., 2010; Jacobs et al., 2014). In theory, song overlapping can, therefore, provide interacting males with information about the sender’s current motivation, while it cannot be used to predict its average aggressiveness. Importantly, the negative within-individual effect implies that the signalled motivation relates to the unwillingness to engage rather than to escalate and show physical aggression.

Conclusion

In conclusion, we have shown that great tits do not generally overlap intruder songs over and above chance levels, and that relationships between aggression and chance-corrected overlap do not imply that overlapping serves as a signal of aggressive intent but rather as a potential signal of nonengagement. Song overlapping tendency deviated from chance levels with birds decreasing rather than increasing song overlap. This suggests that interference avoidance may also be a prominent driver for overlapping tendency. Our multiyear study, with repeated testing across breeding seasons and years, clearly illustrates how long-term data collection with repeated measures can provide statistical robustness and fundamentally new insights into the ecology and evolution of birdsong. We highly recommend repeated measurements as in our study, to allow partitioning of among- and within-individual variation, which will potentially reveal more undetected patterns of biological relevance.

ACKNOWLEDGMENTS

A.H., V.A.R., S.S., L.S. and N.J.D. were supported by the German Science Foundation (grant no. DI 1694/1-1). E.B.-C. and H.S. were supported by the Institute of Biology Leiden. We thank members of the Behavioural Ecology group at the LMU, field assistants and students for help in data collection.

References

Akçay, Ç., Campbell, S. E., & Beecher, M. D. (2014). Individual differences affect honest signalling in a songbird. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1775), 20132496. https://doi.org/10.1098/rspb.2013.2496

Akçay, Ç., Kağan, P. Y., Avşar, A., Çabuk, D., & Bilgin, C. C. (2020). Song overlapping, noise and territorial aggression in great tits. *Behavioral Ecology*, 808733. https://doi.org/10.1101/808733

Allegue, H., Araya-Ajoy, Y. G., Dingemanse, N. J., Dochtermann, N. A., Garamszegi, L. Z., Nakagawa, S., … Westneat, D. F. (2017). Statistical Quantification of Individual Differences (SQuID): an educational and statistical tool for understanding multilevel phenotypic data in linear mixed models. *Methods in Ecology and Evolution*, *8*(2), 257–267. https://doi.org/10.1111/2041-210X.12659

Amy, M., Sprau, P., De Goede, P., & Naguib, M. (2010). Effects of personality on territory defence in communication networks: A playback experiment with radio-tagged great tits. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 277, pp. 3685–3692). Royal Society. https://doi.org/10.1098/rspb.2010.0598

Anderson, R. C., Searcy, W. A., Hughes, M., & Nowicki, S. (2012). The receiver-dependent cost of soft song: A signal of aggressive intent in songbirds. *Animal Behaviour*, *83*(6), 1443–1448. https://doi.org/10.1016/j.anbehav.2012.03.016

Anderson, R. C., Searcy, W. A., Peters, S., & Nowicki, S. (2008). Soft Song in Song Sparrows: Acoustic Structure and Implications for Signal Function. *Ethology*, *114*(7), 662–676. https://doi.org/10.1111/j.1439-0310.2008.01518.x

Arak, A. (1983). Sexual selection by male-male competition in natterjack toad choruses. *Nature*, *306*(5940), 261–262. https://doi.org/10.1038/306261a0

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2014). Characterizing behavioural 'characters': an evolutionary framework. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*, 20132645. https://doi.org/10.1098/rspb.2013.2645

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2017). Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. *Journal of Animal Ecology*, *86*, 227–238. https://doi.org/10.1111/1365-2656.12621

Araya-Ajoy, Y. G., Kuhn, S., Mathot, K. J., Mouchet, A., Mutzel, A., Nicolaus, M., … Dingemanse, N. J. (2016). Sources of (co)variation in alternative siring routes available to male great tits (Parus major). *Evolution*, *70*(10), 2308–2321. https://doi.org/10.1111/evo.13024

Beani, L., & Dessì-Fulgheri, F. (1995). Mate choice in the grey partridge, Perdix perdix: Role of physical and behavioural male traits. *Animal Behaviour*, *49*(2), 347–356. https://doi.org/10.1006/anbe.1995.0047

Betini, G. S., & Norris, D. R. (2012). The relationship between personality and plasticity in tree swallow aggression and the consequences for reproductive success. *Animal Behaviour*, *83*(1), 137–143. https://doi.org/10.1016/j.anbehav.2011.10.018

Bischoff, L. L., Tschirren, B., & Richner, H. (2009). Long-term effects of early parasite exposure on song duration and singing strategy in great tits. *Behavioral Ecology*, *20*(2), 265–270. https://doi.org/10.1093/beheco/arp012

Bradshaw, A. D. (1965). Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics*, *13*(1), 115–155.

Briffa, M., & Elwood, R. W. (2004). Use of energy reserves in fighting hermit crabs. *Proceedings of the Royal Society B: Biological Sciences*, *271*(1537), 373–379. https://doi.org/10.1098/rspb.2003.2633

Brindley, E. L. (1991). Response of European robins to playback of song: neighbour recognition and overlapping. *Animal Behaviour*, *41*(3), 503–512. https://doi.org/10.1016/S0003-3472(05)80853-X

Burt, J. M., Campbell, S. E., & Beecher, M. D. (2001). Song type matching as threat: A test using interactive playback. *Animal Behaviour*, *62*(6), 1163–1170. https://doi.org/10.1006/anbe.2001.1847

Capp, M. S., & Searcy, W. A. (1991). Acoustical communication of aggressive intentions by territorial male bobolinks. *Behavioral Ecology*, *2*(4), 319–326. https://doi.org/10.1093/beheco/2.4.319

Collins, S. (2004). Vocal fighting and flirting: the functions of birdsong. In P. Marler & H. Slabbekoorn (Eds.), *Nature's music* (pp. 39–79). Academic Press.

Cooney, R., & Cockburn, A. (1995). Territorial defence is the major function of female song in the superb fairy-wren, Malurus cyaneus. *Animal Behaviour*, *49*(6), 1635–1647. https://doi.org/10.1016/0003-3472(95)90086-1

Dabelsteen, T., McGregor, P. K., Shepherd, M., Whittaker, X., & Pedersen, S. B. (1996). Is the Signal Value of Overlapping Different from That of Alternating during Matched Singing in Great Tits? *Journal of Avian Biology*, *27*(3), 189. https://doi.org/10.2307/3677221

Dabelsteen, T., & Pedersen, S. B. (1990). Song and information about aggressive responses of blackbirds, Turdus merula: evidence from interactive playback experiments with territory owners. *Animal Behaviour*, *40*(6), 1158–1168. https://doi.org/10.1016/S0003-3472(05)80182-4

Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An introduction to behavioural ecology*. John Wiley & Sons.

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Dingemanse, N. J., Kazem, A. J. N., Réale, D., & Wright, J. (2010). Behavioural reaction norms: animal personality meets individual plasticity. *Trends in Ecology & Evolution*, *25*(2), 81–89. https://doi.org/10.1016/J.TREE.2009.07.013

Dingemanse, N. J., & Wolf, M. (2013). Between-individual differences in behavioural plasticity within populations: Causes and consequences. *Animal Behaviour*, *85*(5), 1031–1039. https://doi.org/10.1016/j.anbehav.2012.12.032

Drent, P. J. (1987). The importance of nestboxes for territory settlement, survival and density of the great tit. *Ardea*, *75*(1), 59–71.

DuBois, A. L., Nowicki, S., & Searcy, W. A. (2009). Swamp sparrows modulate vocal performance in an aggressive context. *Biology Letters*, *5*(2), 163–165. https://doi.org/10.1098/rsbl.2008.0626

Eriksson, D., & Wallin, L. (1986). Male bird song attracts females - a field experiment. *Behavioral Ecology and Sociobiology*, *19*(4), 297–299. https://doi.org/10.1007/BF00300645

Ficken, R. W., Ficken, M. S., & Hailman, J. P. (1974). Temporal pattern shifts to avoid acoustic interference in singing birds. *Science*, *183*(4126), 762–763. https://doi.org/10.1126/science.183.4126.762

Galeotti, P., Saino, N., Sacchi, R., & Møller, A. P. (1997). Song correlates with social context, testosterone and body condition in male barn swallows. *Animal Behaviour*, *53*(4), 687–700. https://doi.org/10.1006/anbe.1996.0304

Gerhardt, C. H., & Huber, F. (2002). *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. University of Chicago PressChicago, IL. Retrieved from https://www.press.uchicago.edu/ucp/books/book/chicago/A/bo3634687.html

Gil, D., & Gahr, M. (2002). The honesty of bird song: multiple constraints for multiple traits. *Trends in Ecology & Evolution*, *17*(3), 133–141. https://doi.org/10.1016/S0169-5347(02)02410-2

Gottlander, K. (1987). Variation in the song rate of the male pied flycatcher Ficedula hypoleuca: causes and consequences. *Animal Behaviour*, *35*(4), 1037–1043. https://doi.org/10.1016/S0003-3472(87)80160-4

Greenfield, M. D., & Minckley, R. L. (2010). Acoustic Dueling in Tarbush Grasshoppers: Settlement of Territorial Contests via Alternation of Reliable Signals. *Ethology*, *95*(4), 309–326. https://doi.org/10.1111/j.1439-0310.1993.tb00480.x

Guilford, T., & Dawkins, M. S. (1995). What are conventional signals? *Animal Behaviour*, *49*(6), 1689–1695. https://doi.org/10.1016/0003-3472(95)90090-X

Helfer, B., & Osiejuk, T. S. (2015). It Takes All Kinds in Acoustic Communication: A New Perspective on the Song Overlapping Phenomenon. *Ethology*, *121*(4), 315–326. https://doi.org/10.1111/eth.12356

Houtmann, A. M. (1992). *Sexual selection in the zebra finch (Poephila guttata).*

Jacobs, C. G. C., van Overveld, T., Careau, V., Matthysen, E., Adriaensen, F., & Slabbekoorn, H. (2014). Personality-dependent response to field playback in great tits: Slow explorers can be strong responders. *Animal Behaviour*, *90*, 65–71. https://doi.org/10.1016/j.anbehav.2014.01.016

Jennings, D. J., Elwood, R. W., Carlin, C. M., Hayden, T. J., & Gammell, M. P. (2012). Vocal rate as an assessment process during fallow deer contests. *Behavioural Processes*, *91*(2), 152–158. https://doi.org/10.1016/j.beproc.2012.07.003

Kelly, C. D., & Godin, J. G. J. (2001). Predation risk reduces male-male sexual competition in the Trinidadian guppy (Poecilia reticulata). *Behavioral Ecology and Sociobiology*, *51*(1), 95–100. https://doi.org/10.1007/s002650100410

Kitchen, D. M., Seyfarth, R. M., Fischer, J., & Cheney, D. L. (2003). Loud calls as indicators of dominance in male baboons (Papio cynocephalus ursinus). *Behavioral Ecology and Sociobiology*, *53*(6), 374–384. https://doi.org/10.1007/s00265-003-0588-1

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

Krebs, J. R., Ashcroft, R., & Orsdol, K. Van. (1981). Song matching in the great tit Parus major L. *Animal Behaviour*, *29*(3), 918–923. https://doi.org/10.1016/S0003-3472(81)80029-2

Kunc, H. P., Amrhein, V., & Naguib, M. (2006). Vocal interactions in nightingales, Luscinia megarhynchos: more aggressive males have higher pairing success. *Animal Behaviour*, *72*(1), 25–30. https://doi.org/10.1016/J.ANBEHAV.2005.08.014

Langemann, U., Tavares, J. P., Peake, T. M., & McGregor, P. K. (2000). Response of Great Tits to Escalating Patterns of Playback. *Behaviour*. Brill. https://doi.org/10.2307/4535717

Martin, J. G. A., Nussey, D. H., Wilson, A. J., & Réale, D. (2011). Measuring individual differences in reaction norms in field and experimental studies: A power analysis of random regression models. *Methods in Ecology and Evolution*, *2*(4), 362–374. https://doi.org/10.1111/j.2041-210X.2010.00084.x

Masco, C., Allesina, S., Mennill, D. J., & Pruett-Jones, S. (2016). The Song Overlap Null model Generator (SONG): A new tool for distinguishing between random and non-random song overlap. *Bioacoustics*, *25*(1), 29–40. https://doi.org/10.1080/09524622.2015.1079734

Moiron, M., Laskowski, K. L., & Niemelä, P. T. (2020). Individual differences in behaviour explain variation in survival: a meta-analysis. *Ecology Letters*, *23*(2), 399–408. https://doi.org/10.1111/ele.13438

Moller, A. P., Saino, N., Taramino, G., Galeotti, P., & Ferrario, S. (1998). Paternity and multiple signaling: Effects of a secondary sexual character and song on paternity in the barn swallow. *American Naturalist*, *151*(3), 236–242. https://doi.org/10.1086/286114

Naguib, M., Altenkamp, R., & Griessmann, B. (2001). Nightingales in space: song and extra-territorial forays of radio tagged song birds. *Journal of Ornithology*, *142*, 306–312.

Naguib, M., & Kipper, S. (2006). Effects of different levels of song overlapping on singing behaviour in male territorial nightingales (Luscinia megarhynchos). *Behavioral Ecology and Sociobiology*, *59*(3), 419–426. https://doi.org/10.1007/s00265-005-0066-z

Naguib, M., & Mennill, D. J. (2010). The signal value of birdsong: empirical evidence suggests song overlapping is a signal. *Animal Behaviour*, *80*(3), e11–e15. https://doi.org/10.1016/J.ANBEHAV.2010.06.001

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, *82*(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, *13*(1), 88.

https://doi.org/10.1186/s12915-015-0196-3

Narins, P. M., Hödl, W., & Grabul, D. S. (2003). Bimodal signal requisite for agonistic behavior in a dart-poison frog, Epipedobates femoralis. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(2), 577–580. https://doi.org/10.1073/pnas.0237165100

Nelson, D. A., & Poesel, A. (2011). Song length variation serves multiple functions in the white-crowned sparrow. *Behavioral Ecology and Sociobiology*, *65*(5), 1103–1111. https://doi.org/10.1007/s00265-010-1120-z

Nicolaus, M., Mathot, K. J., Araya-Ajoy, Y. G., Mutzel, A., Wijmenga, J. J., Kempenaers, B., & Dingemanse, N. J. (2015). Does coping style predict optimization? An experimental test in a wild passerine bird. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1799), 20142405–20142405. https://doi.org/10.1098/rspb.2014.2405

Niemelä, P. T., & Dingemanse, N. J. (2018, November 1). On the usage of single measurements in behavioural ecology research on individual differences. *Animal Behaviour*. Academic Press. https://doi.org/10.1016/j.anbehav.2018.09.012

Nowicki, S., Searcy, W. A., Krueger, T., & Hughes, M. (2002). Individual variation in response to simulated territorial challenge among territory-holding song sparrows. *Journal of Avian Biology*, *33*(3), 253–259. https://doi.org/10.1034/j.1600-048X.2002.330307.x

Osiejuk, T. S., & Jakubowska, A. (2017). Song duration mediates responses of territory owner in a songbird species with a small song repertoire. *Acta Ethologica*, *20*(2), 137–145. https://doi.org/10.1007/s10211-017-0257-4

Osiejuk, T. S., Ratyńska, K., & Cygan, J. P. (2007). Corn bunting (Miliaria calandra) males respond differently to alternating and overlapping playback of song. *Journal of Ethology*, *25*(2), 159–168. https://doi.org/10.1007/s10164-006-0010-3

Peake, T. M., Terry, A. M. R., McGregor, P. K., & Dabelsteen, T. (2001). Male great tits eavesdrop on simulated male-to-male vocal interactions. *Proceedings of the Royal Society B: Biological Sciences*, *268*(1472), 1183–1187. https://doi.org/10.1098/rspb.2001.1648

Piersma, T., & Drent, J. (2003). Phenotypic flexibility and the evolution of organismal design. *Trends in Ecology and Evolution*, *18*(5), 228–233. https://doi.org/10.1016/S0169-5347(03)00036-3

Planqué, R., & Slabbekoorn, H. (2008). Spectral Overlap in Songs and Temporal Avoidance in a Peruvian Bird Assemblage. *Ethology*, *114*(3), 262–271. https://doi.org/10.1111/j.1439-0310.2007.01461.x

Réale, D., Reader, S. M., Sol, D., McDougall, P. T., & Dingemanse, N. J. (2007). Integrating animal temperament within ecology and evolution. *Biological Reviews*, *82*(2), 291–318. https://doi.org/10.1111/j.1469-185X.2007.00010.x

Reichert, M. S., & Gerhardt, H. C. (2013). Gray tree frogs, Hyla versicolor, give lower-frequency aggressive calls in more escalated contests. *Behavioral Ecology and Sociobiology*, *67*(5), 795–804. https://doi.org/10.1007/s00265-013-1503-z

Ripmeester, E. A. P., de Vries, A. M., & Slabbekoorn, H. (2007). Do Blackbirds Signal Motivation to Fight with Their Song? *Ethology*, *113*(11), 1021–1028.

https://doi.org/10.1111/j.1439-0310.2007.01398.x

Robertson, J. G. M. (1986). Male territoriality, fighting and assessment of fighting ability in the Australian frog Uperoleia rugosa. *Animal Behaviour*, *34*(3), 763–772. https://doi.org/10.1016/S0003-3472(86)80060-4

Sales, G. D. (1972). Ultrasound and aggressive behaviour in rats and other small mammals. *Animal Behaviour*, *20*(1), 88–100. https://doi.org/10.1016/S0003-3472(72)80177-5

Schulz, T. M., Whitehead, H., Gero, S., & Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: insights into communication function. *Animal Behaviour*, *76*(6), 1977–1988. https://doi.org/10.1016/j.anbehav.2008.07.032

Searcy, W. A., & Beecher, M. D. (2009, December). Song as an aggressive signal in songbirds. *Animal Behaviour*. https://doi.org/10.1016/j.anbehav.2009.08.011

Searcy, W. A., & Beecher, M. D. (2011). Continued scepticism that song overlapping is a signal. *Animal Behaviour*, *81*(1), e1–e4. https://doi.org/10.1016/j.anbehav.2010.10.013

Shackleton, S. A., & Ratcliffe, L. (2010). Matched Counter-singing Signals Escalation of Aggression in Black-capped Chickadees (Parus atricapillus). *Ethology*, *97*(4), 310–316. https://doi.org/10.1111/j.1439-0310.1994.tb01049.x

Sih, A., Bell, A., & Johnson, J. C. (2004). Behavioral syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution*, *19*(7), 372–378. https://doi.org/10.1016/J.TREE.2004.04.009

Smith, M. J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*(5427), 15–18. https://doi.org/10.1038/246015a0

Strauß, A. F. T., Hutfluss, A., & Dingemanse, N. J. (2020). Great tits responding to territorial intrusions sing less but alarm more on colder days. *Ethology*, *126*(4), 445–454. https://doi.org/10.1111/eth.12989

Stuber, E. F., Araya-Ajoy, Y. G., Mathot, K. J., Mutzel, A., Nicolaus, M., Wijmenga, J. J., … Dingemanse, N. J. (2013). Slow explorers take less risk: a problem of sampling bias in ecological studies. *Behavioral Ecology*, *24*(5), 1092–1098. https://doi.org/10.1093/beheco/art035

Szymkowiak, J., & Kuczyński, L. (2017). Song rate as a signal of male aggressiveness during territorial contests in the wood warbler. *Journal of Avian Biology*, *48*(2), 275–283. https://doi.org/10.1111/jav.00969

Team, R. C. (2018). R: A Language and Environment for Statistical Computing.

ten Cate, C., Slabbekoorn, H., & Ballintijn, M. R. (2002). Birdsong and male—male competition: Causes and consequences of vocal variability in the collared dove (Streptopelia decaocto). *Advances in the Study of Behavior*, *31*(C), 31–75. https://doi.org/10.1016/S0065-3454(02)80005-5

Thys, B., Pinxten, R., & Eens, M. (2021). Long-term repeatability and age-related plasticity of female behaviour in a free-living passerine. *Animal Behaviour*, *172*, 45–54. https://doi.org/10.1016/j.anbehav.2020.11.021

Tuni, C., Han, C. S., & Dingemanse, N. J. (2018). Multiple biological mechanisms result in correlations between pre- and post-mating traits that differ among versus within individuals and genotypes. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1885), 20180951. https://doi.org/10.1098/rspb.2018.0951

Van de Pol, M., & Wright, J. (2009). A simple method for distinguishing within-versus between-subject effects using mixed models. *Behaviour*, *77*(3), 753. https://doi.org/10.1016/j.anbehav.2008.11.006

Van Dongen, W. F. D. (2006). Variation in singing behaviour reveals possible functions of song in male golden whistlers. *Behaviour*, *143*(1), 57–82.

Vehrencamp, S. L. (2001). Is song-type matching a conventional signal of aggressive intentions? *Proceedings of the Royal Society B: Biological Sciences*, *268*(1476), 1637–1642. https://doi.org/10.1098/rspb.2001.1714

Vehrencamp, S. L., Hall, M. L., Bohman, E. R., Depeine, C. D., & Dalziell, A. H. (2007). Song matching, overlapping, and switching in the banded wren: the sender's perspective. *Behavioral Ecology*, *18*(5), 849–859. https://doi.org/10.1093/beheco/arm054

Waas, J. R. (1991). Do little blue penguins signal their intentions during aggressive interactions with strangers? *Animal Behaviour*, *41*(3), 375–382. https://doi.org/10.1016/S0003-3472(05)80838-3

Wesołowski, T. (2011, June). Reports from nestbox studies: A review of inadequacies. *Acta Ornithologica*. https://doi.org/10.3161/000164511X589866

Westneat, D. F., Araya-Ajoy, Y. G., Allegue, H., Class, B., Dingemanse, N. J., Dochtermann, N., … Schielzeth, H. (2020). Collision between biological process and statistical analysis revealed by mean-centering. *Journal of Animal Ecology*. https://doi.org/doi:10.1111/1365-2656.13360

Westneat, D. F., Potts, L. J., Sasser, K. L., & Shaffer, J. D. (2019, June 1). Causes and Consequences of Phenotypic Plasticity in Complex Environments. *Trends in Ecology and Evolution*. Elsevier Ltd. https://doi.org/10.1016/j.tree.2019.02.010

Wilson, D. R., Ratcliffe, L. M., & Mennill, D. J. (2016). Black-capped chickadees, Poecile atricapillus, avoid song overlapping: Evidence for the acoustic interference hypothesis. *Animal Behaviour*, *114*, 219–229. https://doi.org/10.1016/j.anbehav.2016.02.002

Yang, X. J., Ma, X. R., & Slabbekoorn, H. (2014). Timing vocal behaviour: Experimental evidence for song overlap avoidance in Eurasian wrens. *Behavioural Processes*, *103*, 84–90. https://doi.org/10.1016/j.beproc.2013.11.011

Electronic Supplementary Material of Chapter 2

Appendix

Effect of breeding density

As discussed by Wesołowski (2011), nestbox populations differ a lot from natural populations in a variety of environmental aspects, such as predation risk and breeding density. These differences can then result in changes in behavioural expression and thus potentially bias comparisons and conclusions drawn from nestbox studies. In our populations, breeding densities, given as breeding pairs/ha, vary considerably between plots (range 1.37–3.46), between years (range 1.69–2.87) and within plots between years (range 0.48–5.02). Even though these differences could affect the behaviours measured in this study (for example, breeding density correlates positively with aggression), we are certain that they do not bias our findings. This is mainly because we accounted for spatiotemporal patterns by adding ‘plot-year’ to all our analyses. By doing so, we corrected for potential variation induced by differences within plots between years. As breeding density mostly varies at this level and because plot-year explained little variation, we can conclude that breeding density is extremely unlikely to affect our response behaviours in any way.

Table A1: The 95% credible interval (CI) around the mean for effects influencing song output and song overlapping

	Song output *N* = 1338	Chance-corrected overlapping *N* = 483
	β/σ^2 (95% Cl)	β/σ^2 (95% Cl)
Fixed effects		
Intercept[a]	9.15 (6.05, 12.20)	-0.75 (-2.03, 0.56)
Aggressiveness[b]	-2.51 (-3.15, -1.83)	-0.38 (-0.67, -0.08)
Year 2018	0.86 (-1.50, 3.31)	0.41 (-0.24, 1.05)
Year 2019	-0.91 (-3.15, 1.40)	0.09 (-0.60, 0.77)
Test sequence	0.68 (-0.42, 1.80)	0.21 (-0.35, 0.74)
Nest stage	2.72 (1.54, 3.90)	-0.44 (-1.02, 0.15)
Song output focal bird	-	-0.38 (-0.65, -0.09)
Song length focal bird[c]	-	-0.28 (-0.57, 0.001)
Random effects[d]		
Individual	0.153 (0.147, 0.159)	0.023 (0.021, 0.024)
Playback	0.0048 (0.0042, 0.0054)	0.001 (0.001, 0.001)
Plot-year	0.010 (0.006, 0.013)	0 (0, 0)
Observer	0.03 (0.02, 0.05)	0 (0, 0)
Model	0.0005 (0.0002, 0.0008)	0 (0, 0)
Residual	0.80 (0.82, 0.77)	0.976 (0.975, 0.979)
Adjusted repeatability[e]	0.16 (0.15, 0.17)	0.01 (0.01, 0.01)

β is given for fixed effects and σ^2 for random effects.

[a] Reference category; estimate of overlapping length index during test sequence 0, nest stage 0, year 2017 and for attacking birds (aggressiveness = 0).

[b] Aggressiveness measured as approach distance multiplied by -1 + maximum approach distance.

[c] Estimate from a separate model with the same model structure, except for the exclusion of song output.

[d] Random-effect estimates are presented as adjusted repeatabilities, defined as the proportion of variation not explained by fixed effects.

[e] Based on the same model structure but excluding the effects of aggression and focal song number (for overlapping).

Effect of being overlapped

We used a noninteractive playback as a stimulus, which inadvertently resulted in some birds being overlapped by the playback while others were not. We tested whether being overlapped affected their behaviour with a post hoc analysis. Each recording was assigned to one of three categories based on the overlapping of the first song: (1) the focal bird's first song neither overlapped nor was it overlapped ('no overlap'); (2) the focal bird's first song actively overlapped the playback ('focal overlapped stimulus'); (3) the focal bird's first song was overlapped by the playback ('stimulus overlapped focal'). We pragmatically limited this overlap categorization to the very first song, to reduce the number of categories and assumptions. The model was our base model expanded with overlapping category added as a categorical fixed effect. Compared to 'no overlap' birds, birds that actively overlapped the intruder showed higher chance-corrected overlapping (see Table A4). This is what one would expect if overlapping is consistent within an observation. Importantly, chance-corrected overlapping did not differ between 'no overlap' birds and those that were overlapped by the playback. This latter finding suggested that being overlapped did not affect the focal bird's behaviour. This conclusion was also warranted because we did not find differences between the three categories in song output or aggression. Furthermore, accounting for overlap of the first song did not alter the reported links between aggression and chance-corrected overlapping, or between aggression and song output.

Table A2: The 95% credible interval (CI) around the mean for effects influencing the song output for the 3 study years separately

	Song output 2017 $N = 356$	2018 $N = 398$	2019 $N = 584$
	β/σ^2 (95% Cl)	β/σ^2 (95% Cl)	β/σ^2 (95% Cl)
Fixed effects			
Intercept[a]	10.72 (5.57, 15.77)	8.43 (3.40, 13.57)	8.23 (4.20, 12.30)
Aggressiveness[b]	-4.72 (-6.20, -3.25)	-3.30 (-4.56, -2.14)	-1.35 (-2.18, -0.53)
Test sequence	0.08 (-2.17, 2.37)	1.24 (-0.83, 3.38)	0.67 (-1.02, 2.27)
Nest stage	1.91 (-0.41, 4.09)	2.98 (0.74, 5.15)	2.74 (1.10, 4.42)
Random effects[c]			
Individual	0.216 (0.204, 0.223)	0.107 (0.098, 0.112)	0.210 (0.202, 0.216)
Playback	0.02 (0.01, 0.04)	0.041 (0.037, 0.043)	0 (0, 0)
Plot-year	0 (0, 0)	0.02 (0.01, 0.03)	0.005 (0.003, 0.008)
Taxidermic mount	0.009 (0.004, 0.01)	0 (0, 0)	0.009 (0.004, 0.02)
Observer	0 (0, 0)	0.03 (0.007, 0.05)	0.06 (0.03, 0.09)
Residual	0.75 (0.72, 0.78)	0.81 (0.77, 0.85)	0.72 (0.67, 0.76)

β is given for fixed effects and σ^2 for random effects.

[a] Reference category; estimate of overlapping length index during test sequence 0, nest stage 0 and for attacking birds (aggressiveness = 0).

[b] Aggressiveness measured as approach distance multiplied by -1 + maximum approach distance.

[c] Random-effect estimates are presented as adjusted repeatabilities, defined as the proportion of variation not explained by fixed effects.

Table A3: The 95% credible interval (CI) around the mean for effects influencing the song overlapping for the 3 study years separately

	Chance-corrected overlapping 2017 *N* = 200	 2018 *N* = 142	 2019 *N* = 141
	β/σ^2 (95% Cl)	β/σ^2 (95% Cl)	β/σ^2 (95% Cl)
Fixed effects			
Intercept[a]	-2.41 (-4.65, -0.20)	0.78 (-1.19, 2.79)	1.03 (-0.86, 2.92)
Aggressiveness[b]	-0.46 (-1.04, 0.14)	-0.62 (-1.09, -0.18)	-0.15 (-0.52, 0.21)
Test sequence	1.03 (0.02, 2.08)	-0.27 (-1.08, 0.51)	-0.41 (-1.22, 0.34)
Nest stage	-0.12 (-1.15, 1,00)	-0.63 (-1.46, 0.18)	-0.89 (-1.69, -0.10)
Bird song length	-0.25 (-0.74, 0.28)	-0.11 (-0.54, 0.33)	-0.58 (-0.99, -0.17)
Random effects[c]			
Individual	0 (0, 0)	0.09 (0.08, 0.10)	0.154 (0.146, 0.154)
Playback	0 (0, 0)	0.42 (0.41, 0.42)	0 (0, 0)
Plot-year	0 (0, 0)	0 (0, 0)	0 (0, 0)
Taxidermic mount	0 (0, 0)	0.010 (0.006, 0.015)	0 (0, 0)
Observer	0 (0, 0)	0 (0, 0)	0.08 (0.02, 0.14)
Residual	~1.00 (~1.00, ~1.00)	0.48 (0.47, 0.51)	0.77 (0.83, 0.70)

β is given for fixed effects and σ^2 for random effects.

[a] Reference category; estimate of overlapping length index during test sequence 0, nest stage 0 and for attacking birds (aggressiveness = 0).

[b] Aggressiveness measured as approach distance multiplied by -1 + maximum approach distance.

[c] Random-effect estimates are presented as adjusted repeatabilities, defined as the proportion of variation not explained by fixed effects.

Table A4: The 95% credible interval (CI) around the mean for the effects of the focal individual's first song being overlapped, actively overlapping or not showing any overlap with the playback

	Chance-corrected overlapping *N*= 337	Song output *N*= 337	Aggression *N* = 337
	β/σ^2 (95% CI)	β/σ^2 (95% CI)	β/σ^2 (95% CI)
Fixed effects			
Intercept[a]	-1.19 (-2.51, 0.11)	15.67 (11.33, 20.04)	29.45 (26.73, 32.28)
First song active overlap	1.36 (0.72, 2.03)	0.04 (-2.19, 2.30)	-0.54 (-1.91, 0.76)
First song overlapped by playback	0.29 (-0.40, 0.97)	-1.00 (-3.46, 1.25)	-1.37 (-2.73, 0.03)
Aggressiveness[b]	-0.36 (-0.66, -0.08)	-1.72 (-2.72, -0.74)	-
Test sequence	0.16 (-0.37, 0.71)	0.37 (-1.42, 2.12)	0.50 (-0.58, 1.58)
Nest stage	-0.29 (-0.87, 0.30)	1.32 (-0.67, 3.22)	-2.54 (-3.69, -1.32)
Bird song length	-0.32 (-0.60, -0.04)	-	-
Random effects[c]			
Individual	0.036 (0.033, 0.039)	0.016 (0.014, 0.017)	0.13 (0.13, 0.14)
Playback	0.007 (0.006, 0.008)	0.010 (0.009, 0.0011)	0.03 (0.03, 0.04)
Plot-year	0 (0, 0)	0 (0, 0)	0.03 (0.02, 0.04)
Taxidermic mount	0 (0, 0)	0 (0, 0)	0.003 (0.002, 0.004)
Observer	0 (0, 0)	0.01 (0.006, 0.02)	0.08 (0.05, 0.10)
Residual	0.96 (0.95, 0.96)	0.96 (0.95, 0.97)	0.72 (0.68, 0.76)

β is given for fixed effects and σ^2 for random effects.

[a] Reference category; estimate of the response variable during test sequence 0, nest stage 0, for attacking birds (aggressiveness = 0) and for birds whose first song was not overlapped nor actively overlapped the playback.

[b] Aggressiveness measured as approach distance multiplied by -1 + maximum approach distance.

[c] Random-effect estimates are presented as adjusted repeatabilities, defined as the proportion of variation not explained by fixed effects.

Manuscript I: Within-individual stability in bird song shows cross-year repeatability and is affected by food supplementation in a wild passerine bird

Alexander Hutfluss, Eira Bermúdez-Cuamatzin, Alexia Mouchet, Mark Briffa, Hans Slabbekoorn, Niels J. Dingemanse

Unpublished manuscript

Abstract

1. Theory predicts that populations should exhibit within- and among-individual variation in "behavioural stability", the residual within-individual variation remaining after controlling for personality and reversible plasticity.

2. Bird song represents a good example of behaviour where stability may be costly yet beneficial because stable singing can be a signal of quality favoured by sexual selection. Assuming energetic costs, ecological variation (e.g. in food availability) should result in both within- and among-individual variation in stability. Longitudinal studies are required to estimate both variance components while avoiding environmental confounds and pseudo-repeatability.
3. We monitored 12 nest box plots of great tits *Parus major* over a three-year period during the breeding season. We recorded male songs during simulated territory intrusions, twice during their mate's laying stage, and twice during incubation. Each preceding winter, we manipulated winter food availability. Assuming that stability is costly, we expected food-supplemented males to sing more stable songs. We also expected males to sing more stable songs early in the breeding season when their social mate is fertile and paternity not decided.
4. We found strong support for plasticity in stability of a key song characteristic: minimum frequency. Males were plastic because they adjusted song stability to breeding stage and other individual-specific environmental effects. Food-supplemented males sang more stable songs, though this experimental effect was present only in the first study year. Finally, males also exhibited cross-year repeatability in song stability.

5. We found dissimilar stability patterns for another song characteristic, phrase length, which was neither repeatable nor affected by food treatment. This implies that trait-specific ecological costs and benefits shape variation in stability. The occurrence of long-term repeatability in aspects of song stability implies that future studies should address its heritability and how natural and sexual selection maintains this variation in the wild.

Introduction

Animals exhibit within- and among-individual variation in behavioural stability, the residual within-individual variation remaining after accounting for variation in average behaviour and reversible plasticity. This component of behaviour is also called predictability", or "intra-individual variation" (Stamps et al., 2012; Westneat et al., 2015), and has been shown to be repeatable, heritable, and evolve in response to selection (Martin et al., 2017; Prentice et al., 2020; Stamps et al., 2012; Westneat et al., 2013). Unpredictable behaviour can be favoured as anti-predator adaptation (Brembs, 2011; Briffa, 2013). In other situations, predictable behaviour can be favoured, for example, when representing a costly signal of quality in mating contexts (Sakata et al., 2008; Schuett et al., 2010). Repeatable variation in stability may be maintained by trade-offs resolved differently by individuals, such as social/energetic costs vs. signal value in communication contexts (Grava et al., 2013; Rivera-Gutierrez et al., 2010), or strength vs. precision/skill in agonistic contexts (Reichert & Carl Gerhardt, 2012).

Animal signals are suitable candidate behaviours for testing hypotheses concerning behavioural stability (Patricelli & Hebets, 2016; Sakata & Vehrencamp, 2012; Westneat et al., 2015). Stereotypic and stable signal production is, for example, a common characteristic of acoustic communication (Amorim et al., 2011; Bouchet et al., 2012; Röhr et al., 2020). The repeated and stable expression of signals can improve information transmission when conditions hinder communication through distortion or attenuation (Brown & Handford, 2000). Both intra- and interspecific variation may exist, and facilitate individual or species recognition (Bouchet et al., 2012; Röhr et al., 2020). Finally, where acoustic stability has energetic, social or survival costs, it may provide receivers with cues of sender quality (Bartsch et al., 2016; Botero et al., 2009).

As a unifying statistical framework has emerged only recently (Cleasby et al., 2015), past studies of behavioural stability used a variety of approaches. Simulation studies imply that studies of behavioural stability require particular sampling designs and statistical approaches to avoid biased biological conclusions (Cleasby et al. 2015; Westneat et al. 2015). For example, a classic approach has been to incorporate indices of stability (e.g., coefficients of variation; Table S1) in statistical analyses. Such "statistics-on-statistics" approaches generate highly biased estimates, which can be avoided by approaches developed for directly estimating "variance in variance", like double hierarchical generalized linear mixed models (DHGLMs) (Cleasby et al., 2015).

Another general problem is that biased sampling designs can lead to biased estimates of among- and within-individual variance (Dingemanse & Dochtermann, 2013; Westneat et al., 2011). Upward biased estimates of among-individual variance (leading to “pseudo-repeatability”) occur when individuals plastically adjust their behaviour to environmental factors that are themselves individually repeatable. For example, in territorial species, individuals will often experience environments with strong individual-specific temporal autocorrelations (Araya-Ajoy & Dingemanse, 2017; Mitchell et al., 2020; Niemelä & Dingemanse, 2017). This occurs when the environmental conditions in the same territory are more similar within than among years. Repeatedly sampling the same individuals for behavioural stability over short *and* long time periods thus allows separating transitory (pseudo-repeatability) from genuine repeatability, and help avoid such biases (Araya-Ajoy et al., 2015).

Bird song is one of the best-studied examples of acoustic communication and is widely used to study behavioural stability (Table S1). The stereotypic, ongoing production of acoustic signals requires repeated activity of many syringeal and respiratory muscles and is energetically costly (Grava et al., 2012; Lambrechts & Dhondt, 1988; Sakata & Vehrencamp, 2012; Suthers & Zollinger, 2004). Song stability represents a fitness indicator trait as it predicts various fitness proxies: stable singers are dominant (Botero et al., 2009; De Kort et al., 2009) and sire more extra-pair offspring (Botero et al., 2009; Byers, 2007; Taff et al., 2012). Consequently, we expect individuals to adaptively upregulate song stability when the benefits of maintaining stability outweigh its costs. For example, if singing stably indicates “good genes”, social mates may not engage in extra-pair fertilisations, and thus, males may primarily sing stable songs prior to rather than following clutch completion. Similarly, if energetically costly, song stability should increase with access to supplementary food. We thus generally expect reversible plasticity in song stability. We also expect repeatable among-individual variation owing to individual variation in “quality”, e.g., due to silver spoon effects (i.e., variation in early-life conditions; Holveck et al., 2008; Van De Pol et al., 2006).

Minimum frequency and phrase length are regarded as the two most suitable acoustic parameters for song stability research, particularly in great tits (*Parus major*). Minimum frequency is repeatable and under selection (Halfwerk et al., 2011; Zollinger et al., 2017). In some species, birds producing stable minimum frequencies have been found to be larger (Bartsch et al., 2016) and produce more extra-pair offspring (Byers, 2007). Furthermore, minimum frequency is plastic and can vary, for example, among songs (Cardoso & Atwell, 2011), among phrases within songs (Cardoso & Atwell, 2011), among elements within song types and phrases (Marler & Isaac, 1960), and among song types (Logue et al., 2007; Slabbekoorn & den Boer-Visser, 2006). Great tits sang their lowest songs more often just prior to the fertilization of eggs (Halfwerk et al. 2011) and great tits with lower minimum frequencies also produced more stable songs (Lambrechts, 1997). Minimum frequency may thus be a direct target of sexual selection (via female preference for low-frequency songs). It can also be an indirect target if selection targets stability instead (Cramer, 2013b). Characterizing links between behavioural mean and variance is therefore important (Martin et al., 2017). Similarly, previous work has implied that selection may also favour stability in phrase length (Lambrechts & Dhondt, 1986; Weary, 1989). Indeed, great tits producing stable phrase lengths, (i.e., that do not "drift"; Poesel & Kempenaers, 2000) have been reported to be dominant and have higher reproductive success (Lambrechts & Dhondt, 1986). In addition, fatigue is suggested to negatively influence the stable production of phrase lengths within individual great tits, suggesting stability is costly (Lambrechts & Dhondt, 1986, 1988). Studying stability in multiple traits is essential to understand whether or not findings can be generalized; an important objective in ecology (Carter et al., 2013; Nakagawa & Parker, 2015).

In the current study, we investigated individual repeatability in stability of two song traits within and among years, and the environmental factors moderating their stability. We repeatedly recorded songs of the same individuals, elicited experimentally during simulated territory intrusions. We expected both among-year and within-year repeatability if individuality and plasticity act in concert. We expected seasonal plasticity in song stability as the balance of its costs and benefits should change seasonally, with males up-regulating song stability during their mate's laying stage. Both predictions assume that song stability is costly, which we tested experimentally using supplementary feeding experiments implemented in each winter preceding our recording sessions. We expected that winter food affected song stability by influencing body condition at the onset of the breeding season. We also investigated potential mechanisms by which individuals vary song stability.

Methods

Data were collected over three years (2017-2019) in 12 nest box plots established between Starnberg and Herrsching (Bavaria, Germany) in 2009 (Nicolaus et al., 2015). Each plot contained 50 boxes in a regular grid (50 m apart). Boxes were monitored from April-July to assess key life-history parameters (detailed by Nicolaus et al. 2015). We caught breeders in their box when their offspring were 10-12 days old, measured morphological and behavioural traits and ringed those not previously banded.

We manipulated food availability from July through March. We placed feeders baited with sunflower seeds in half of the plots (“treated”), while supplementary food was not provided in the remaining plots (“control”). We randomly assigned plot treatment; half of the plots subsequently switched treatment across years in a stratified random design to ensure each plot received each treatment at least once over the three years. In treated plots, we placed a feeder at four locations to avoid resource monopolisation. We estimated natural winter food availability by counting beech seeds (the species’ main winter food source; van Balen, 1980), in four 30-cm sections arranged in a transect under ten European beeches (*Fagus sylvatica*) in the study area. Counts showed a distinct binary pattern, and natural winter food availability was thus categorized as “no seeds” (2018) vs. “seeds” (“mast years”; 2017 and 2019). This pattern was expected as mast years are typically followed by years without seeds (Tinbergen et al., 1985; van Balen, 1980).

Simulated territory intrusions

We acquired repeated song recordings by subjecting each male in each year to four simulated territorial intrusions during its first breeding attempt: two times during egg-laying (1 and 3 days after the first egg was observed) and two times during incubation (1 and 3 days after clutch incubation was confirmed). We used a taxidermic model of a conspecific male as visual stimulus and played back a conspecific song as acoustic stimulus. We placed the model on a 1.2 m pole (for consistent visibility across territories) and a speaker (Foxpro Shockwave) on the ground, both 1 m in front of the box. For further details, see Strauß et al. (2020).

We randomly assigned models (n=23) and playbacks (n=174). In 2017, we used 13 songs as playback stimuli from 13 different locally recorded birds. In 2018 and 2019, we added 161 playback song stimuli from recordings made during the experiments in 2017. To produce natural variation in strophe number and duration, we only used song recordings with uninterrupted strophes (plus subsequent silent intervals) when constructing playback stimuli. We adjusted playback stimuli and speaker volumes to a standardized equalize playback amplitude. After the focal male entered a 15-m radius around the box, we recorded its behaviour and song for three min. The observer (positioned at 15 meters from the box) recorded with a windscreen-covered directional microphone (Sennheiser ME66/K6) attached to a recorder (TASCAM DR-05, recording 44.1 kHz, 16-bit WAV files). Subjects not arriving within 15 min were scored “non-responsive” (Araya-Ajoy and Dingemanse 2014; 2017).

Song analysis

Acoustic parameters were extracted using Avisoft SAS Lab Pro (Avisoft Bioacoustics, Specht 2002) software. We applied “Time Domain FIR Filters” to filter background noise outside the species-specific frequency range used for song production (2.0-8.0 kHz). We subsequently changed amplitudes by normalizing files to 75%. Concurrently, we produced a spectrogram for each strophe (FFT length: 1024, Overlap: 87.5%, Window: Hamming) to check for and exclude low quality (e.g. noisy) recordings. We extracted spectral and temporal features of each element using the automatic measurements function (three thresholds: -24 dB, start: -16 dB, end: -12 dB, hold time: 10 ms, minimum length: 30 ms). For each element, we recorded start time, end time, and at both time points peak frequency of which we took the lower as our measurement of minimum frequency (Figure 1). We excluded elements overlapping spectrally and temporally with the readily identifiable playback stimulus songs. Observers scoring recordings were trained by scoring >2800 measurements of sixteen recordings; unknown to the observer, those consisted of two duplicates of eight files in random order. Observers started scoring real data when their between- and within- observer repeatability was ≥0.9 (see Table S2). Each phrase was assigned to one of 53 song types defined for this population. Song types were beforehand classified by the number of elements per repeated phrase, element structure within phrase, which can be tonal, frequency modulated or resemble harmonics, and element order.

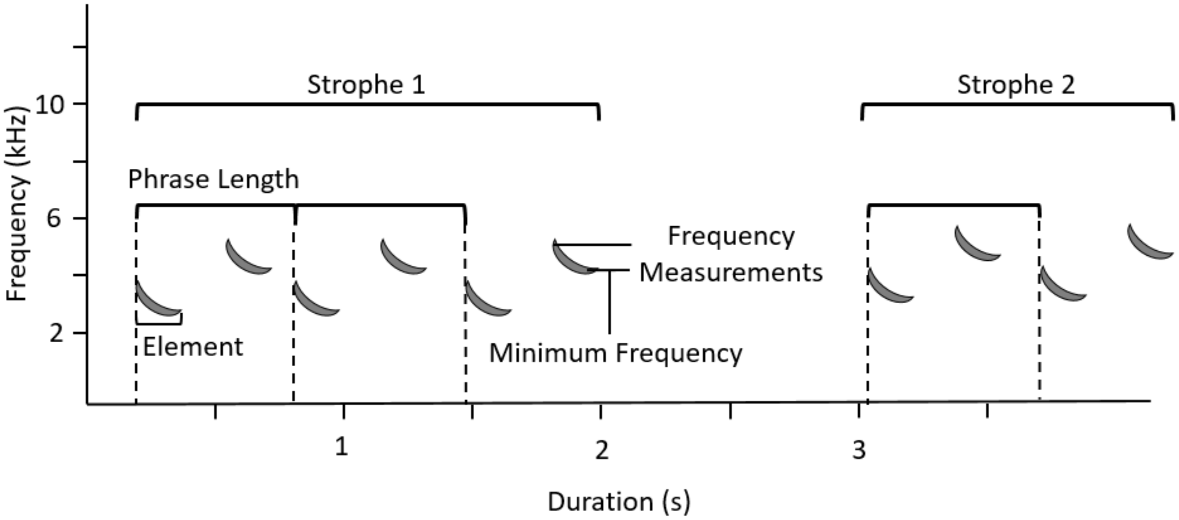


Figure 1: *Illustration of typical song components. We show two schematic songs (strophes) of one song type within a single recording. The first strophe consists of three, the second of two, phrases; each phrase contains two elements (grey). Phrase length was defined as time elapsed between the starts of a focal and the subsequent phrase (dotted lines).*

Statistical analysis

We used univariate double hierarchical generalized linear mixed models to simultaneously estimate random and fixed effects associated with "mean" and "variance" parts (Lee & Nelder, 2006; Rönnegård et al., 2010). We analysed variation in (1) minimum frequency (the lower of the two peak frequencies measured) and (2) phrase length (time elapsed between the onsets of subsequent phrases; by nature missing for the strophe's last phrase). Previous studies analysed lowest elements per strophe/phrase (e.g. Bermúdez-Cuamatzin et al., 2011). Our study question did not warrant this data selection; we thus analysed variation among all elements (within and among phrases) as an integrative measure of spectral variability. We also conducted the same analyses using a single minimum frequency per phrase/strophe which did not alter the nature of the results (not shown).

We ran three models per response variable. All fitted random intercepts for individual (n=273 levels) and recording identity (n=585 levels) for both the "mean" and "variance" parts. Following Araya-Ajoy et al. (2015), we included random intercepts for each combination of individual and year identity ("bird-year"; n=334 levels). This separated transitory (repeatable within-year) from genuine (repeatable among-year) individual effects. Variance among bird-years is attributable to transitory effects caused by within-individual responses to factors that are stable within-individuals within but not across years; such effects thus lead to the appearance of repeatable differences (i.e., pseudo-repeatability) associated with single-year repeated measures datasets. Variance among birds instead reflects whether an individual's year-specific mean is repeatable across years (i.e., long-term repeatability). Finally, we used the repeated measures data (for elements, or phrases) collected within the same recording to estimate whether average levels of behaviour differed between recordings within birds within years. This was achieved by additionally fitting random intercepts for recording identity (n=585 levels). All models included the unique combination of plot and year (plot-year; n=34 levels) as a random effect for the "mean" and "variance" parts; fitting this dummy effect avoided pseudo-replicated estimates for food-supplementation treatment, which varied among plot-years. We performed explorative analyses to investigate effects of observer (n=10 levels) and playback stimulus (n=158 levels), fitted as additional random effects for both model parts. Neither observer nor playback stimulus effects were important (Table S3) and therefore removed in further analyses. All models included year, fitted as a fixed (rather than random) effect because it contained few levels (n=3 years). Models estimated the covariance between "mean" and "variance" among birds, bird-years, and recordings. Among birds, this covariance estimates whether an individual's average behaviour correlates with its behavioural stability.

We expanded this **Base model** in two steps. First, we included nest stage (egg-laying vs. incubation), test number (first vs. second test within nest stage), and supplementary feeding (treated vs. control) as fixed effects. Previous work implied that supplementary feeding is not effective in mast-years (Perdeck et al., 2000). We therefore also fitted a year-treatment interaction. All fixed effects were fitted to both model parts. We expanded this **Environmental model** by including random intercepts for song type (which varied within *and* among individuals; n=53 levels) for both model parts, while additionally fitting the covariance between the "mean" and "variance" parts at this level (**Song type model**). This expansion tested whether variation in mean and/or residual variance was attributable to song type. Specifically, we tested whether birds would differ in stability because they differ in song types used and whether they alter their stability by switching between the produced types; this would be so if variance estimates shrank substantially when including song type.

We collected 124,921 data points (number of elements). To manage computational time, we analysed only the first two phrases per strophe. This resulted in 21,578 and 20,618 data points for the minimum frequency and phrase length, respectively. For minimum frequency, we ran the base model also on the full database, but this did not yield different results compared to the reduced dataset (Table S4). We, therefore, used the reduced dataset throughout in what follows.

Model set-up

All analyses were performed in R v.3.6.1 (R Core Team 2019). DHGLMs were implemented using the “brms” package (Bürkner, 2017). Each DHGLM was run with default priors for all parameters. See Text S1 and Table S5for model fitting procedures, performance, and diagnostics statistics. We present mean (β for fixed effects in mean part, γ for fixed effects in variance part, and δ for random effects) and 95% credible intervals (CIs) for all parameters. We regarded effects “significant” in the frequentist’s sense when 95% CIs did not overlap zero (Nakagawa & Cuthill, 2007). We regarded 95% CIs that both overlapped zero but showing substantial skew in one direction as “moderate support” for an effect. We viewed variation attributable to random effects supported when their lower 95% CI did not include zero.

Results

We conducted 2,740 simulated territory intrusions; 1,582 resulted in an acoustic recording. 586 recordings from 273 males were suitable for analyses. These held 6,544 songs, and 21,578 analysable elements (106,663 for the full database, Table S4). Birds sang (mean±SE) 23.9±0.6 songs, with 8.1±0.2 phrases. Males contributed on average, 2.45 recordings (range: 1-9) to the overall analyses. Fifty-two individuals were recorded in two years, five in all three years.

Sources of variation in behavioural mean

For the mean part of our models (Table 1a), patterns of variation differed between the two song traits. For minimum frequency, CIs did not include zero for any random effect except plot-year (Table 1a). In brief, minimum frequencies were more similar within vs. among recordings of the same individual, suggesting individual-specific environmental effects among an individual's recordings within years (recording ID effect). Average values over all elements within recordings were also more similar for recordings taken in the same year (bird-year effect). This implied the existence of individual-specific environmental effects, varying among rather than within years (i.e., pseudo-repeatability). Finally, an individual's grand mean annual value was repeatable among years (individual ID effect, Fig. 1a). This implied genuine (long-term) repeatable differences. For phrase length, by contrast, there was support for neither bird-year nor individual ID effects (owing to CIs including zero) (Table 2a). Thus, phrase length showed neither pseudo- nor genuine repeatability (Fig. 1b). Supported were recording ID effects, implying day-to-day variation in phrase length. Finally, neither trait showed spatiotemporal variation (plot-year ID).

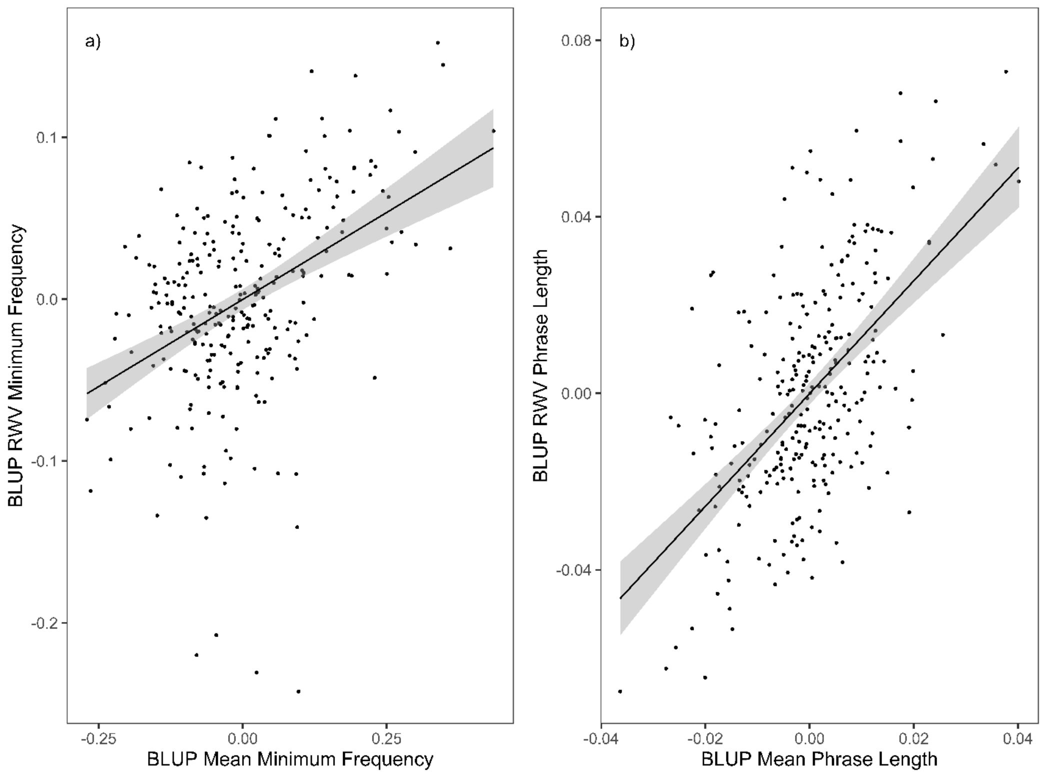


Figure 2: Visualisation of individual mean and stability for minimum frequency (a) and phrase length (b). We plot here the relationship between individual-specific best linear unbiased predictors (BLUPs) for mean and residual within-individual variance (RWV) for (a) minimum frequency and (b) phrase length. The black line indicates a linear regression line (grey area represents 95% CI). BLUPs were extracted from base models and averaged for all posterior samples.

Our **environmental models** (Table S6a) moderately supported a seasonal change in minimum frequency (effect of nest stage: β = -0.08, 95% CI: -0.17, 0.01): birds produced lower frequencies when their mate was incubating. A shortening of phrase length with repeated exposure was strongly supported (effect of test number within nest stage: β = -0.01, 95% CI: -0.03, -0.00). Minimum frequency did not vary with test number, and phrase length did not vary with nest stage. Food supplementation affected neither trait nor were treatment effects year-specific (Table S6a).

Minimum frequency and phrase length both varied among song types (**song type models**; Table S7): certain song types were shorter and had, on average, lower minimum frequencies than others. There was more variance among song types than among levels of other random effects. Importantly, inclusion of song type neither reduced random nor fixed effect estimates (Table S7).

Sources of variation in behavioural stability

For the variance part of our models (Table 1b), patterns of variation again differed between the song traits. Stability in minimum frequency exhibited both short- and long-term repeatability owing to 95% CIs for bird-year and individual ID not including zero. Though stability in phrase length showed some short-term repeatability (bird-year effect), as above, it was not repeatable long-term (95% CIs for individual ID effects included zero).

Against expectations, our models strongly supported a pattern where birds became more (rather than less) stable during the incubation stage (both traits; **environmental models**; Table S6b). Food supplementation caused birds to sing songs with higher stability in minimum frequency, but not in phrase length. This treatment effect was year-specific, with strong support for a negative treatment effect on residual variance only in 2017 (food treatment effect; Table S6b). In the following two years, the effect of food supplementation was more positive in comparison (2018 × food treatment, 2019 × food treatment; Table S6b), which meant that food supplementation then did not affect stability.

Stability varied among song types (both traits; **song type models**; Table S7): certain song types had higher spectral (minimum frequency) and temporal (phrase length) stability. Most variance was among recordings or song types, depending on the specific trait, followed by bird-years and individuals with very similar estimates. Importantly, the inclusion of song type did not reduce any estimate of variance in stability (Table S7); variation in stability within- and among-individuals was thus not attributable to variation in song type.

Correlations between mean and stability

For both song traits, mean and variance were positively correlated among recordings: when individuals increased minimum frequency (or phrase length) across days within years, they also became less stable (Recording ID correlation) (Table 1c). Mean-variance correlations were estimated with great uncertainty at the individual and Bird-Year ID level, and therefore not interpretable. Nevertheless, mean-variance relationships were of similar magnitudes across most levels (Table 1c), suggesting that individuals producing higher minimum frequencies may also have produced less stable songs. Mean-variance correlations among song types were also supported: song types of higher frequencies (or with longer phrase lengths) were less stable (Table S7).

Table 1: Effect size estimates and 95% credible intervals (CI) for predictors affecting mean (a) and residual variance (b) for minimum frequency and phrase length, and correlations between mean and variance (c) for our Base models.

(a) Means	Minimum frequency N = 21578	Phrase length N = 20618
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept	-0.01 (-0.10, 0.09)	0.45 (0.43, 0.46)
Year 2018	-0.04 (-0.17, 0.09)	0.01 (-0.01, 0.03)
Year 2019	-0.04 (-0.19, 0.10)	0.02 (-0.00, 0.04)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0.23 (0.07, 0.32)	0.03 (0.01, 0.04)
Bird-year ID	0.11 (0.01, 0.27)	0.01 (0.00, 0.03)
Recording ID	0.47 (0.42, 0.51)	0.07 (0.07, 0.08)
Plot-year ID	0.05 (0.00, 0.13)	0.01 (0.00, 0.02)
(b) Residual variances		
Fixed Effects	**φ (95% CI)**	**φ (95% CI)**
Intercept	-0.43 (-0.52, -0.35)	-3.86 (-3.99, -3.73)
Year 2018	0.03 (-0.08, 0.14)	0.06 (-0.11, 0.24)
Year 2019	0.05 (-0.07, 0.17)	-0.07 (-0.26, 0.12)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0.14 (0.01, 0.25)	0.10 (0.00, 0.24)
Bird-year ID	0.17 (0.02, 0.27)	0.13 (0.01, 0.28)
Recording ID	0.38 (0.34, 0.42)	0.76 (0.71, 0.82)
Plot-year ID	0.05 (0.00, 0.13)	0.07 (0.00, 0.18)
(c) Mean-variance Correlations (a × b)	**r (95% CI)**	**r (95% CI)**
Individual ID	0.37 (-0.77, 0.95)	0.28 (-0.85, 0.97)
Bird-year ID	0.25 (-0.90, 0.97)	-0.11 (-0.92, 0.96)
Recording ID	0.44 (0.33, 0.55)	0.22 (0.12, 0.31)

Discussion

Our multi-year study demonstrated that great tits exhibited both behavioural plasticity and cross-year repeatability in key aspects of song stability. Phenotypic plasticity occurred because individuals changed song stability over the season. Individuals also adjusted their stability to individual-specific environmental factors (of unknown origin) that were more stable within than across years. This finding is important because it would lead to pseudo-repeatability in one-year studies, underlining the need for multi-year longitudinal sampling designs to study individual difference in stability of song and other labile traits known to be affected by yearly environmental fluctuations. Furthermore, supplementary feeding made birds sing more stable songs. This provides some suggests support for the idea that song stability is costly to produce. Food supplementation, however, increased stability only in one of three years, implying that costs associated with stability vary with other ecological factors (yet to be determined). Importantly, song stability was not only plastic but simultaneously showed cross-year repeatability. We expected this because song stability has long been thought to represent a heritable fitness indicator trait. All of those patterns of variation, notably, characterized only one of the two song traits under study, implying that biological processes shaping song stability are trait-specific. Selection thus appears to have moulted expressions of song stability in specific traits rather than an overall expression of song stability in general.

Plasticity in behavioural stability

A considerable number of studies have observed stability across contexts or manipulated environmental conditions to test whether patterns of phenotypic plasticity in behavioural stability match those predicted by adaptive theory. Various studies focused on variance-sensitive foraging, testing the idea that selection can favour individuals that accept risks during foraging when faced with demanding or unfavourable conditions (Mathot et al., 2017; Westneat et al., 2013, 2017). Mathot et al. (2017), for example, showed experimentally that brood demand affects the residual within-individual variance in parental provisioning. Individuals have also been shown to plastically adjust behavioural stability to predation (risk) or temperature (Briffa, 2013; Briffa et al., 2013). Though costly, e.g. due to potential mismatches between optimal and expressed phenotype, unpredictable behaviour might be net beneficial when it reduces predator success. Finally, increasing stability over an individual's lifetime (Biro & Adriaenssens, 2013; De Kort et al., 2009) can be adaptive when high behavioural variance facilitates learning optimal behaviour, in early life-stages or in novel environments (Brembs, 2011).

In the context of bird song, our study also demonstrated that stability exhibits substantial plasticity, as males adjusted their behavioural stability over both short and long periods. Some of the temporal variation, for example, could be attributed to the breeding cycle: individuals increased stability with breeding stage. This result did not match our *a priori* predictions, because we expected that benefits of stability would peak prior to egg laying, when great tits are also known to sing their lowest-frequency song types more often during dawn-chorus singing close to their female in the nest box (Halwerk et al. 2011). However, instead of an increase we found a decrease in stability over the season. Perhaps stability is more important as a signal in contexts where the stakes are lower, but when birds still strongly rely on singing. Indeed, great tits respond to intrusions during incubation preferentially by singing, while they respond more with alarm calling and physical aggression during egg laying (Araya-Ajoy & Dingemanse, 2014; Strauß et al., 2020).

We also detected substantial amounts of unexplained within-individual variation in song stability, resulting from phenotypic plasticity to environmental factors not revealed in our study. Within years, birds may respond to food availability or micro-climatic factors, which can both affect energy reserves, and hence the ability to produce high-quality songs (Barnett & Briskie, 2007; Strauß et al., 2020; Thomas, 1999). Among years, song stability may vary with territory quality, breeding density or age (Botero et al., 2009; De Kort et al., 2009; Grava et al., 2012). Interestingly, any plastic shift in stability occurred without changing song type: song types differed in length, minimum frequency, and stability (see also Logue et al., 2007; Slabbekoorn & den Boer-Visser, 2006), but controlling for their effects did not explain any variation in song stability. Birds, therefore, did not appear to use song type switches to change the stability of their songs. Rather, when birds sang more stably they did so while essentially singing the same song types.

Our experimental manipulations revealed a year-specific increase in stability in birds experiencing the food supplementation, though only for one year and only for minimum frequency. Our food treatment might have increased body condition in this particular year, enabling birds to produce more stable songs during the breeding season. If so, this would represent an experimental demonstration of within-individual plasticity in behavioural stability. We are, however, unable to investigate this hypothesized pathway further, as we did not measure body weight at the time of song recording. Importantly, this experimental effect may also result from processes other than phenotypic plasticity, such as a non-random settlement pattern. If stable singers are more competitive (Botero et al., 2009), they might have priority of access to the high-density environments created by our manipulation (Mouchet et al. In Progress). This explanation would concern an among-individual effect and would be in line with previous work failing to detect within-individual effects of food manipulations (Kriengwatana et al., 2014).

Importantly, the positive effect of our food supplementation treatment was only present in the first of two beech mast years. Failure to find the effect in the second mast year suggests that factors other than natural winter food availability mediated the food supplementation effects. More pronounced positive effects of food supplementation are, for example, expected in cold years with snow covering natural food supplies. Ultimately, this casts doubt on whether song stability can generally act as quality signal: supportive evidence was absent in most (two out of the three) years. Interestingly, previous manipulations in our population have also revealed year-specific effects on behavioural stability, albeit in a foraging rather than communication context (Mathot et al. 2017). This implies that ecological factors changing across years may generally play an important role in moderating shifts in behavioural stability within wild populations.

Individual differences in behavioural stability

We found long-term repeatable differences in stability in minimum frequency, implying that individuals were predictably different in song stability with certain individuals producing more stable songs than others. Given that about half of all adults in our populations survive from one year to the next (Bauchau & Van Noordwijk, 1995), this finding implies that individuals were repeatable over biologically relevant timespans. Previous studies did not measure song stability over multiple years and did not account for variation within individuals among years; nonetheless, several assumed that individual measurements were repeatable (Table S1). Indeed, individual repeatability is required for song stability to serve as a signal of ‘individual quality’, potentially shaped by early-life conditions and/or genetic make-up (Bartsch et al., 2016; Holveck et al., 2008). As repeatability sets an upper limit to heritability (Falconer & Mackay, 1996), analogous estimates for individuality in stability shed light on its potential to evolve. Assuming that song stability is costly, it may reflect individuality in ability to pay the energetic, neurological or coordinative costs associated with stable singing (Sakata & Vehrencamp, 2012; Suthers & Zollinger, 2004).

Importantly, our data suggested that stability could be costly and under selection for minimum frequency but not for phrase length, implying that different aspects of song (stability) might differ in evolvability (Houle, 1992).

Very few studies have demonstrated repeatability of behavioural stability, though the subject is receiving increasing attention (Biro & Adriaenssens, 2013; He et al., 2017; Hertel et al., 2021; Highcock & Carter, 2014; Jolles et al., 2019; Mitchell et al., 2016; Prentice et al., 2020; Stamps et al., 2012). Surprisingly little formal theory exists on ecological mechanisms explaining why individual differences in stability might persist despite selection. For signalling behaviour, individual variation in stability might persist if traded-off with other costly traits (Sakata et al., 2008; Schuett et al., 2010). This can create shallow fitness landscapes where selection cannot easily erode standing variation (e.g. Mangel & Stamps, 2001). Important in this context are mean-variance correlations (Briffa et al., 2013; Highcock & Carter, 2014; Prentice et al., 2020; Rönnegård et al., 2010). Those can result from trade-offs between accuracy (mean) and precision (variance) (Duckworth et al., 2018; Sih & Del Giudice, 2012) or stem from auto-correlations (i.e. variance scales with the mean, called Taylor's law ;Taylor, 1961). The latter explanation is mechanistic (rather than functional) and fits our study: when animals produced higher frequencies, or longer phrase lengths, they were also less stable.

A potential functional explanation for our findings is, however, that "individual quality" attributable to genetic make-up or silver spoon effects explains why some birds can sing both low-frequency and stable songs. If both traits are favoured, the pattern of correlation assists evolution (Schluter, 1996), and directional sexual selection may thus erode standing variation (Lande & Arnold, 1983). The persistence of repeatable individual variation may then either concern a mutation-selection balance (Desai & Fisher, 2007) or a trade-off with other traits. For example, some personality types are generally less reactive or plastic in response to environmental factors and might thus behave more predictably compared to others (He et al., 2017; Mitchell et al., 2016; Stamps et al., 2012; but see Hertel et al., 2021). Ultimately, future studies should focus on identifying such trait correlations, and study whether selection on one component can lead to indirect selection and evolution of another (Martin et al., 2017; Prentice et al., 2020), or constrain adaptive evolution altogether. However, unpredictable behaviour may also have evolved through fluctuating selection (e.g. to mitigate predation risk). Understanding the evolutionary maintenance of individual variation in stability will therefore require more field studies on the quantitative genetics of trait stability taking patterns of fluctuating selection into account (Martin et al., 2017; Prentice et al., 2020).

Conclusion

We used a multi-year longitudinal sampling design based on which we conclude that male great tits show cross-year repeatability of song stability. Song stability was also plastic within individuals, exemplified by seasonal plasticity and experimental food treatment effects. Our sampling design revealed that great caution is required in interpreting repeatability estimates from single-year studies. Our study also demonstrated that repeatability and plasticity in stability can differ among traits, suggesting special attention is required again in generalizing results from one trait to another. Finally, our study implies that song stability may not always represent a reliable proxy for "individual quality".

ACKNOWLEDGEMENTS

A.H., A.M. and N.J.D. were supported by the German Science Foundation (grant no. DI 1694/1-1), E.B.-C. and H.S. by the Institute of Biology (Leiden University), and M.B. by the School of Biological and Marine Sciences (University of Plymouth). We thank all (past) members of the "Behavioural Ecology" group at LMU for help in data collection and discussion. All procedures complied with guidelines from the Regierung Oberbayern for Animal Care. This study was approved by the Regierung Oberbayern (permit number ROB-55.2-2532.Vet_02-17-215) in accordance with the ASAB/ABS Guidelines for the use of animals in research (www.sciencedirect.com/science/article/pii/S0003347215004613). No conflicts of interest apply.

A.H., N.J.D., M.B., E.B.-C. and H.S developed ideas and methods, A.H. and A.M collected the data, A.H. and N.J.D. analysed the data. All authors contributed critically to the drafts and gave final approval for publication.

References

Amorim, M. C. P., Simões, J. M., Almada, V. C., & Fonseca, P. J. (2011). Stereotypy and variation of the mating call in the Lusitanian toadfish, Halobatrachus didactylus. *Behavioral Ecology and Sociobiology*, *65*(4), 707–716. https://doi.org/10.1007/s00265-010-1072-3

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2014). Characterizing behavioural 'characters': an evolutionary framework. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*, 20132645. https://doi.org/10.1098/rspb.2013.2645

Araya-Ajoy, Y. G., & Dingemanse, N. J. (2017). Repeatability, heritability, and age-dependence of seasonal plasticity in aggressiveness in a wild passerine bird. *Journal of Animal Ecology*, *86*(2), 227–238. https://doi.org/10.1111/1365-2656.12621

Araya-Ajoy, Y. G., Mathot, K. J., & Dingemanse, N. J. (2015). An approach to estimate short-term, long-term and reaction norm repeatability. *Methods in Ecology and Evolution*, *6*(12), 1462–1473. https://doi.org/10.1111/2041-210X.12430

Barnett, C. A., & Briskie, J. V. (2007). Energetic state and the performance of dawn chorus in silvereyes (Zosterops lateralis). *Behavioral Ecology and Sociobiology*, *61*, 579–587. https://doi.org/10.1007/s00265-006-0286-x

Bartsch, C., Hultsch, H., Scharff, C., & Kipper, S. (2016). What is the whistle all about? A study on whistle songs, related male characteristics, and female song preferences in common nightingales. *Journal of Ornithology*, *157*(1), 49–60. https://doi.org/10.1007/s10336-015-1245-y

Bauchau, V., & Van Noordwijk, A. J. (1995). Comparison of survival estimates obtained from three different methods of recapture in the same population of the great tit. *Journal of Applied Statistics*, *22*(5–6), 1031–1038. https://doi.org/10.1080/02664769524775

Bermúdez-Cuamatzin, E., Ríos-Chelén, A. A., Gil, D., & Garcia, C. M. (2011). Experimental evidence for real-time song frequency shift in response to urban noise in a passerine bird. *Biology Letters*, *7*(1), 36–38. https://doi.org/10.1098/rsbl.2010.0437

Biro, P. A., & Adriaenssens, B. (2013). Predictability as a personality trait: consistent differences in intraindividual behavioral variation. *The American Naturalist*, *182*(5), 621–629. https://doi.org/10.1086/673213

Botero, C. A., Rossman, R. J., Caro, L. M., Stenzler, L. M., Lovette, I. J., de Kort, S. R., & Vehrencamp, S. L. (2009). Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Animal Behaviour*, *77*(3), 701–706. https://doi.org/10.1016/j.anbehav.2008.11.020

Bouchet, H., Blois-Heulin, C., Pellier, A. S., Zuberbühler, K., & Lemasson, A. (2012). Acoustic variability and individual distinctiveness in the vocal repertoire of red-capped mangabeys (Cercocebus torquatus). *Journal of Comparative Psychology*, *126*(1), 45–56. https://doi.org/10.1037/a0025018

Brembs, B. (2011). Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1707), 930–939. https://doi.org/10.1098/rspb.2010.2325

Briffa, M. (2013). Plastic proteans: reduced predictability in the face of predation risk in hermit crabs. *Biology Letters*, *9*(5), 20130592. https://doi.org/10.1098/rsbl.2013.0592

Briffa, M., Bridger, D., & Biro, P. A. (2013). How does temperature affect behaviour?

Multilevel analysis of plasticity, personality and predictability in hermit crabs. *Animal Behaviour*, *86*(1), 47–54. https://doi.org/10.1016/J.ANBEHAV.2013.04.009

Brown, T. J., & Handford, P. (2000). Sound Design for Vocalizations: Quality in the Woods, Consistency in the Fields. *The Condor*, *102*(1), 81–92. https://doi.org/10.1093/condor/102.1.81

Bürkner, P.-C. (2017). Advanced Bayesian Multilevel Modeling with the R Package brms. *R Journal*, *10*(1), 395–411. http://arxiv.org/abs/1705.11123

Byers, B. E. (2007). Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behavioral Ecology*, *18*(1), 130–136. https://doi.org/10.1093/beheco/arl058

Cardoso, G. C., & Atwell, J. W. (2011). On the relation between loudness and the increased song frequency of urban birds. *Animal Behaviour*, *82*(4), 831–836. https://doi.org/10.1016/j.anbehav.2011.07.018

Carter, A. J., Feeney, W. E., Marshall, H. H., Cowlishaw, G., & Heinsohn, R. (2013). Animal personality: what are behavioural ecologists measuring? *Biological Reviews*, *88*(2), 465–475. https://doi.org/10.1111/brv.12007

Cleasby, I. R., Nakagawa, S., & Schielzeth, H. (2015). Quantifying the predictability of behaviour: statistical approaches for the study of between-individual variation in the within-individual variance. *Methods in Ecology and Evolution*, *6*(1), 27–37. https://doi.org/10.1111/2041-210X.12281

Cramer, E. R. A. (2013a). Physically Challenging Song Traits, Male Quality, and Reproductive Success in House Wrens. *PLoS ONE*, *8*(3), e59208. https://doi.org/10.1371/journal.pone.0059208

Cramer, E. R. A. (2013b). Measuring consistency: spectrogram cross-correlation versus targeted acoustic parameters. *Bioacoustics*, *22*(3), 247–257. https://doi.org/10.1080/09524622.2013.793616

De Kort, S. R., Eldermire, E. R. B., Valderrama, S., Botero, C. A., & Vehrencamp, S. L. (2009). Trill consistency is an age-related assessment signal in banded wrens. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1665), 2315–2321. https://doi.org/10.1098/rspb.2009.0127

Desai, M. M., & Fisher, D. S. (2007). Beneficial Mutation–Selection Balance and the Effect of Linkage on Positive Selection. *Genetics*, *176*(3), 1759–1798. https://doi.org/10.1534/genetics.106.067678

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: mixed-effect modelling approaches. *Journal of Animal Ecology*, *82*(1), 39–54. https://doi.org/10.1111/1365-2656.12013

Duckworth, R. A., Potticary, A. L., & Badyaev, A. V. (2018). On the Origins of Adaptive Behavioral Complexity: Developmental Channeling of Structural Trade-offs. In *Advances in the Study of Behavior* (Vol. 50, pp. 1–36). Academic Press Inc. https://doi.org/10.1016/bs.asb.2017.10.001

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitative Genetics*. Pearson Education India.

Grava, T., Grava, A., & Otter, K. A. (2012). Vocal performance varies with habitat quality in black-capped chickadees (Poecile atricapillus). *Behaviour*, *149*(1), 35–50. https://doi.org/10.1163/156853912X625854

Grava, T., Grava, A., & Otter, K. A. (2013). Habitat-induced changes in song consistency affect perception of social status in male chickadees. *Behavioral Ecology and Sociobiology*, *67*(10), 1699–1707. https://doi.org/10.1007/s00265-013-1580-z

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

He, R., Pagani-Núñez, E., Chevallier, C., & Barnett, C. R. A. (2017). To be so bold: boldness is repeatable and related to within individual behavioural variability in North Island robins. *Behavioural Processes*, *140*, 144–149. https://doi.org/10.1016/j.beproc.2017.04.014

Hertel, A. G., Royauté, R., Zedrosser, A., & Mueller, T. (2021). Biologging reveals individual variation in behavioural predictability in the wild. *Journal of Animal Ecology*, *90*(3), 723–737. https://doi.org/10.1111/1365-2656.13406

Highcock, L., & Carter, A. J. (2014). Intraindividual Variability of Boldness Is Repeatable across Contexts in a Wild Lizard. *PLoS ONE*, *9*(4), e95179. https://doi.org/10.1371/journal.pone.0095179

Holveck, M.-J., Vieira de Castro, A. C., Lacklan, R. F., ten Cate, C., & Riebel, K. (2008). Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behavioral Ecology*, *19*(6), 1267–1281.

Houle, D. (1992). Comparing evolvability and variability of quantitative traits. *Genetics*, *130*(1).

Jolles, J. W., Briggs, H. D., Araya-Ajoy, Y. G., & Boogert, N. J. (2019). Personality, plasticity and predictability in sticklebacks: bold fish are less plastic and more predictable than shy fish. *Animal Behaviour*, *154*, 193–202. https://doi.org/10.1016/j.anbehav.2019.06.022

Kriengwatana, B., Wada, H., Schmidt, K. L., Taves, M. D., Soma, K. K., & MacDougall-Shackleton, S. A. (2014). Effects of nutritional stress during different developmental periods on song and the hypothalamic-pituitary-adrenal axis in zebra finches. *Hormones and Behavior*, *65*(3), 285–293. https://doi.org/10.1016/j.yhbeh.2013.12.013

Lambrechts, M. M. (1997). Song frequency plasticity and composition of phrase versions in great tits Parus major. *Ardea*, *85*(1), 99–109.

Lambrechts, M. M., & Dhondt, A. A. (1986). Male quality, reproduction, and survival in the great tit (Parus major). *Behavioral Ecology and Sociobiology*, *19*(1), 57–63. https://doi.org/10.1007/BF00303843

Lambrechts, M. M., & Dhondt, A. A. (1988). The anti-exhaustion hypothesis: a new hypothesis to explain song performance and song switching in the great tit. *Animal Behaviour*, *36*, 327–334.

Lande, R., & Arnold, S. J. (1983). The Measurement of Selection on Correlated Characters. *Evolution*, *37*(6), 1210. https://doi.org/10.2307/2408842

Lee, Y., & Nelder, J. A. (2006). Double hierarchical generalized linear models. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, *55*(2), 139–185. https://doi.org/10.1111/j.1467-9876.2006.00538.x

Logue, D. M., Droessler, E. E., Roscoe, D. W., Vokey, J. R., Rendall, D., & Kunimoto, R. M. (2007). Sexually antithetical song structure in a duet singing wren. *Behaviour*, *144*(3), 331–350. https://doi.org/10.1163/156853907780425749

Mangel, M., & Stamps, J. (2001). Trade-offs between growth and mortality and the maintenance of individual variation in growth. In *Evolutionary Ecology Research* (Vol. 3, Issue 5). Evolutionary Ecology, Ltd.

Marler, P., & Isaac, D. (1960). Physical Analysis of a Simple Bird Song as Exemplified by the Chipping Sparrow. *The Condor*, *62*(2), 124–135. https://doi.org/10.2307/1365677

Martin, J. G. A., Pirotta, E., Petelle, M. B., & Blumstein, D. T. (2017). Genetic basis of between-individual and within-individual variance of docility. *Journal of Evolutionary Biology, 30*(4), 796–805. https://doi.org/10.1111/jeb.13048

Mathot, K. J., Olsen, A.-L., Mutzel, A., Araya-Ajoy, Y. G., Nicolaus, M., Westneat, D. F., Wright, J., Kempenaers, B., & Dingemanse, N. J. (2017). Provisioning tactics of great tits (Parus major) in response to long-term brood size manipulations differ across years. *Behavioral Ecology, 28*(6), 1402–1413. https://doi.org/10.1093/beheco/arx083

Mitchell, D. J., Dujon, A. M., Beckmann, C., & Biro, P. A. (2020). Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity. *Behavioral Ecology*, *31*(1), 222–231.

Mitchell, D. J., Fanson, B. G., Beckmann, C., & Biro, P. A. (2016). Towards powerful experimental and statistical approaches to study intraindividual variability in labile traits. *Royal Society Open Science, 3*(10). https://doi.org/10.1098/rsos.160352

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews*, *82*(4), 591–605. https://doi.org/10.1111/j.1469-185X.2007.00027.x

Nakagawa, S., & Parker, T. H. (2015). Replicating research in ecology and evolution: Feasibility, incentives, and the cost-benefit conundrum. *BMC Biology*, *13*(1), 88. https://doi.org/10.1186/s12915-015-0196-3

Nicolaus, M., Mathot, K. J., Araya-Ajoy, Y. G., Mutzel, A., Wijmenga, J. J., Kempenaers, B., & Dingemanse, N. J. (2015). Does coping style predict optimization? An experimental test in a wild passerine bird. *Proceedings of the Royal Society B: Biological Sciences, 282*(1799), 20142405–20142405. https://doi.org/10.1098/rspb.2014.2405

Niemelä, P. T., & Dingemanse, N. J. (2017). Individual versus pseudo-repeatability in behaviour: Lessons from translocation experiments in a wild insect. *Journal of Animal Ecology, 86*(5), 1033–1043. https://doi.org/10.1111/1365-2656.12688

Patricelli, G. L., & Hebets, E. A. (2016). New dimensions in animal communication: the case for complexity. In *Current Opinion in Behavioral Sciences* (Vol. 12, pp. 80–89). Elsevier Ltd. https://doi.org/10.1016/j.cobeha.2016.09.011

Perdeck, A. C., Visser, M. E., & Van Balen, J. H. (2000). Great tit Parus major survival and the beech-crop. *Ardea*, *88*, 99–106.

Poesel, A., & Kempenaers, B. (2000). When a bird is tired from singing: a study of drift during thedawn chorus. *Etología*, *8*, 1–7.

Prentice, P. M., Houslay, T. M., Martin, J. G. A., & Wilson, A. J. (2020). Genetic variance for behavioural 'predictability' of stress response. *Journal of Evolutionary Biology, 33*(5), 642–652. https://doi.org/10.1111/jeb.13601

Reichert, M. S., & Carl Gerhardt, H. (2012). Trade-offs and upper limits to signal performance during close-range vocal competition in gray tree frogs Hyla versicolor. *American Naturalist*, *180*(4), 425–437. https://doi.org/10.1086/667575

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

Röhr, D. L., Camurugi, F., Paterno, G. B., Gehara, M., Juncá, F. A., Álvares, G. F. R., Brandão, R. A., & Garda, A. A. (2020). Variability in anuran advertisement call: a multi-level study with 15 species of monkey tree frogs (Anura, Phyllomedusidae). *Canadian Journal of Zoology*, *98*(8), 495–504. https://doi.org/10.1139/cjz-2020-0018

Rönnegård, L., Felleki, M., Fikse, F., Mulder, H. A., & Strandberg, E. (2010). Genetic heterogeneity of residual variance - Estimation of variance components using double hierarchical generalized linear models. *Genetics Selection Evolution*, *42*(1), 1–10. https://doi.org/10.1186/1297-9686-42-8

Sakata, J. T., Hampton, C. M., & Brainard, M. S. (2008). Social Modulation of Sequence and Syllable Variability in Adult Birdsong. *Journal of Neurophysiology*, *99*(4), 1700–1711. https://doi.org/10.1152/jn.01296.2007

Sakata, J. T., & Vehrencamp, S. L. (2012). Integrating perspectives on vocal performance and consistency. In *Journal of Experimental Biology* (Vol. 215, Issue 2, pp. 201–209). The Company of Biologists Ltd. https://doi.org/10.1242/jeb.056911

Schluter, D. (1996). Adaptive radiation along genetic lines of least resistance. *Evolution*, *50*(5), 1766–1774. https://doi.org/10.1111/j.1558-5646.1996.tb03563.x

Schuett, W., Tregenza, T., & Dall, S. R. X. (2010). Sexual selection and animal personality. *Biological Reviews*, *85*(2), 217–246. https://doi.org/10.1111/j.1469-185X.2009.00101.x

Sih, A., & Del Giudice, M. (2012). Linking behavioural syndromes and cognition: a behavioural ecology perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2762–2772. https://doi.org/10.1098/rstb.2012.0216

Slabbekoorn, H., & den Boer-Visser, A. (2006). Cities Change the Songs of Birds. *Current Biology*, *16*(23), 2326–2331. https://doi.org/10.1016/j.cub.2006.10.008

Stamps, J. A., Briffa, M., & Biro, P. A. (2012). Unpredictable animals: individual differences in intraindividual variability (IIV). *Animal Behaviour*, *83*(6), 1325–1334. https://doi.org/10.1016/J.ANBEHAV.2012.02.017

Strauß, A. F. T., Hutfluss, A., & Dingemanse, N. J. (2020). Great tits responding to territorial intrusions sing less but alarm more on colder days. *Ethology*, *126*(4), 445–454. https://doi.org/10.1111/eth.12989

Suthers, R. A., & Zollinger, S. A. (2004). Producing Song: The Vocal Apparatus. *Annals of the New York Academy of Sciences*, *1016*(1), 109–129. https://doi.org/10.1196/annals.1298.041

Taff, C. C., Steinberger, D., Clark, C., Belinsky, K., Sacks, H., Freeman-Gallant, C. R., Dunn, P. O., & Whittingham, L. A. (2012). Multimodal sexual selection in a warbler: Plumage and song are related to different fitness components. *Animal Behaviour*, *84*(4), 813–821. https://doi.org/10.1016/j.anbehav.2012.07.002

Taylor, L. R. (1961). Aggregation, variance and the mean. *Nature*, *189*(4766), 732–735. https://doi.org/10.1038/189732a0

Thomas, R. J. (1999). Two tests of a stochastic dynamic programming model of daily singing routines in birds. *Animal Behaviour*, *57*(2), 277–284. https://doi.org/10.1006/ANBE.1998.0956

Tinbergen, J. M., Van Balen, J. H., & Van Eck, H. M. (1985). Density dependent survival in an isolated great tit population. Kluyvers data reanalysed. *Ardea*, *73*, 38–48.

Van Balen, J. H. (1980). Population fluctuations of the great tit and feeding conditions in winter. *Ardea*, *55*(1–2), 143–164.

Van De Pol, M., Bruinzeel, L. W., Heg, D., Van Der Jeugd, H. P., & Verhulst, S. (2006). A silver spoon for a golden future: Long-term effects of natal origin on fitness prospects of oystercatchers (Haematopus ostralegus). *Journal of Animal Ecology*, *75*(2), 616–626. https://doi.org/10.1111/j.1365-2656.2006.01079.x

Weary, D. M. (1989). Categorical perception of bird song: how do great tits (Parus major) perceive temporal variation in their song? *Journal of Comparative Psychology*, *103*(4), 320–325. https://doi.org/10.1037/0735-7036.103.4.320

Westneat, D. F., Hatch, M. I., Wetzel, D. P., & Ensminger, A. L. (2011). Individual variation in parental care reaction norms: Integration of Personality and plasticity. *American Naturalist*, *178*(5), 652–667. https://doi.org/10.1086/662173

Westneat, D. F., Mutzel, A., Bonner, S., & Wright, J. (2017). Experimental manipulation of brood size affects several levels of phenotypic variance in offspring and parent pied flycatchers. *Behavioral Ecology and Sociobiology*, *71*(6), 1–13. https://doi.org/10.1007/s00265-017-2315-3

Westneat, D. F., Schofield, M., & Wright, J. (2013). Parental behavior exhibits among-individual variance, plasticity, and heterogeneous residual variance. *Behavioral Ecology*, *24*(3), 598–604. https://doi.org/10.1093/beheco/ars207

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, *90*(3), 729–743. https://doi.org/10.1111/brv.12131

Zollinger, S. A., Slater, P. J. B., Nemeth, E., & Brumm, H. (2017). Higher songs of city birds may not be an individual response to noise. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1860), 20170602. https://doi.org/10.1098/rspb.2017.0602

Electronic Supplementary Material of Chapter 3

Supplementary Material Table S1.

Song stability studies. We print the measure of stability, sample size, unit of analysis, and whether repeatability was estimated. We include only studies quantifying among-individual variation in stability.

Species	Measure[1]	No. elements/syllables/songs per measure[2]	Sample size[3]	Unit of analysis[4]	Repeated measures available	Individual differences	Interval	Study
Banded wren	SPCC	?	22	individual age	Y	N	year	De Kort et al., 2009
Banded wren	SPCC	?	4	individual	N	N	-	Cramer et al., 2011
Banded wren	SPCC	5-23	332-955	trill/song	Y	N	breeding stages, year	Vehrencamp et al., 2013
Bengalese finch	CV	?	26	individual	N	N	-	Sakata et al., 2008
Black-capped chickadee	PCA on CV	9, 7 CVs	16	individual	N	N	-	Grava et al., 2012
Brown-headed cowbird	CV	5	9	individual	N	N	-	O’Loghlen & Rothstein, 1993
Canary	CV	?	12	per recording	Y	N	days	Alward et al., 2016
Canary	SPCC	10	60	recording	Y	Y	days	Iserbyt et al., 2017
Chestnut-sided warbler	CV	10-20	36	individual	N	N	-	Byers, 2007
Common yellowthroat	SPCC and CVs	10	295	bout	Y	Y	breeding stages, year	Taff et al., 2012
Dark-eyed junco	CV	?	71	individual	N	N	-	Ferreira et al., 2016
Fify Fancy canary	SPCC	10	?	recording	Y	N	days	Müller et al., 2013
Great reed warbler	CV	10	53-62	individual year	Y	N	year	Wegrzyn et al., 2010
Great tit	SPCC	30	104	individual age	Y	N	year	Rivera-Gutierrez et al., 2012
House wren	SPCC	?	2659-4530	trill/song	Y	Y	within recording	Cramer, 2013a
House wren	SPCC, PCA on CV	>= 3	4574	trill/song	Y	Y	within day, year	Cramer, 2013b
Nightingale	SPCC	?	568	song	Y	N	within recordings	Sprau et al., 2013
Nightingale	CV	<100	20	individual	N	N	-	Bartsch et al., 2016

Prairie warbler	PCA on CV	5-43, 8 CVs	20-51	Individual breeding stage	Y	Y	breeding stages	Byers et al., 2015
Prairie warbler	PCA on CV	5-43, 3 CVs	141-318	song bout	Y	N	days	Byers et al., 2016
Skylark	SPCC and CVs	<10	?	individual and type	Y	N	days	Geberzahn & Aubin, 2014
Song sparrows	SPCC	10	22	individual	N	N	-	Schmidt et al., 2013
Song sparrow	SPCC	10	?	song, recording	Y	Y	Within recording	Kubli & MacDougall-Shackleton, 2014
Thrush nightingale	SPCC	<30	unclear	individual and type	Y	N	within recordings	Souriau et al., 2019
Tropical mockingbird	SPCC	10	12-23	individual	Y	N	breeding attempts	Botero et al., 2009
Withe-crowned sparrow	SD	5	9	individual	N	N	-	Troy Smith et al., 1995
White-crowned sparrow	CV on SPCC	SPCC: 5-10, CV: 10-45	?	individual	N	N	-	Gilman et al., 2007
Withe-crowned sparrow	similarity score and CVs	10	9-10	individual and day	Y	N	days	Meitzen et al., 2009
Zebra finch	similarity score	10	?	per recording	Y	N	within day	Olveczky et al., 2005
Zebra finch	Similarity score	10	21	individual	N	N	-	Teramitsu & White, 2006
Zebra finch	similarity score	2-14	36-96	song, recording, and individual	Y	N	months	Pytte et al., 2007
Zebra finch	repeatability	5	26	individual	N	N	-	Holveck et al., 2008
Zebra finch	SD and CV	SD:22-5441, CV: ?	?	individual treatment	?	N	-	Leblois et al., 2010
Zebra finch	PCA including SPCC	?	44	individual	N	N	-	Kriengwatana et al., 2014
Zebra finch	similarity score	5	?	recording	Y	N	weeks	Faltynek, 2017
Zebra finch	Similarity score	10	51	Individual age and context	Y	N	months	James & Sakata, 2019

[1] Measure of stability: CV = Coefficient of variation, SPCC = Spectral Cross Correlation, SD = Standard Deviation; CVs most often calculated for various spectral and temporal song traits depending on the species (trills, notes, syllables).

[2] For SPCC this is the number of syllables, trills, notes that were cross-correlated instead of the number of cross-correlations (10 notes = 45 correlations)

[3] Sample sizes as number of measures analyses were based on. ? indicate unclear numbers, while never exceeding 50

[4] In addition, measures are often averaged within song types, recordings and/or individuals further adding uncertainty

[5] "?" indicates cases where the information of interest was not mentioned in the focal paper

Supplementary Material Table S2

Table S1: Within- and among-observer correlations in two measures of interest: element starting time (a) used to calculate phrase lengths and element peak frequency (b). Black boxes in diagonal indicate within-observer repeatability, boxes below the diagonal among-observer repeatability in pairwise comparisons.

a) Element starting time									
	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Observer 6	Observer 7	Observer 8	Observer 9
Observer 1	1								
Observer 2	1	1							
Observer 3	1	1	1						
Observer 4	1	1	1	1					
Observer 5	1	1	1	1	1				
Observer 6	1	1	1	1	1	1			
Observer 7	1	1	1	1	1	1	1		
Observer 8	1	1	1	1	1	1	1	1	
Observer 9	1	1	1	1	1	1	1	1	1
b) Element peak frequency									
Observer 1	1								
Observer 2	1	1							
Observer 3	1	1	0.99						
Observer 4	1	1	1	1					
Observer 5	1	1	1	1	1				
Observer 6	1	1	1	0.99	0.99	0.97			
Observer 7	0.94	0.94	0.98	0.94	0.94	0.94	0.75		
Observer 8	1	1	1	1	1	0.97	0.86	0.85	
Observer 9	1	1	0.99	1	1	0.98	0.94	0.99	1

Supplementary Material Table S3

Table S3: Effect size estimates and 95% credible intervals (CI) for predictors affecting mean and residual variance for the initial and the base model for minimum frequency per element, as well as the correlations between mean and residual variance.

(a) Means	Minimum frequency, Initial model N = 21578	Minimum frequency, Base model N = 21578
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept	-0.01 (-0.10, 0.09)	-0.01 (-0.10, 0.09)
Year 2018	-0.04 (-0.17, 0.09)	-0.04 (-0.17, 0.09)
Year 2019	-0.04 (-0.19, 0.10)	-0.04 (-0.19, 0.10)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	**0.22 (0.05, 0.32)**	**0.23 (0.07, 0.32)**
Bird-year ID	**0.12 (0.01, 0.27)**	**0.11 (0.01, 0.27)**
Recording ID	**0.46 (0.42, 0.51)**	**0.47 (0.42, 0.51)**
Plot-year ID	0.05 (0.00, 0.13)	0.05 (0.00, 0.13)
Observer	0.05 (0.00, 0.15)	-
Playback	0.05 (0.00, 0.13)	-
(b) Residual variances		
Fixed Effects	**φ (95% CI)**	**φ (95% CI)**
Intercept[b]	-0.43 (-0.52, -0.35)	-0.43 (-0.52, -0.35)
Year 2018	0.03 (-0.08, 0.14)	0.03 (-0.08, 0.14)
Year 2019	0.05 (-0.07, 0.17)	0.05 (-0.07, 0.17)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	**0.14 (0.01, 0.26)**	**0.14 (0.01, 0.25)**
Bird-year ID	**0.17 (0.02, 0.27)**	**0.17 (0.02, 0.27)**
Recording ID	**0.38 (0.34, 0.42)**	**0.38 (0.34, 0.42)**
Plot-year ID	0.05 (0.00, 0.13)	0.05 (0.00, 0.13)
Observer	0.04 (0.00, 0.12)	-
Playback	0.04 (0.00, 0.10)	-
(c) Correlations (a x b)	**r (95% CI)**	**r (95% CI)**
Individual ID	0.31 (-0.82, 0.92)	0.37 (-0.77, 0.95)
Bird-year ID	0.36 (-0.85, 0.96)	0.25 (-0.90, 0.97)
Recording ID	**0.44 (0.33, 0.55)**	**0.44 (0.33, 0.55)**

Supplementary Material Table S4

Table S4: Effect size estimates and 95% credible intervals (CI) for predictors affecting mean and residual variance for minimum frequency using the full dataset and the reduced dataset, as well as the correlations between mean and residual variance.

(a) Means	Minimum frequency, full dataset N = 106663	Minimum frequency, reduced dataset N = 21578
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept	-0.05 (-0.10, 0.00)	-0.01 (-0.10, 0.09)
Year 2018	-0.02 (-0.14, 0.05)	-0.04 (-0.17, 0.09)
Year 2019	-0.01 (-0.19, 0.06)	-0.04 (-0.19, 0.10)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	**0.21 (0.08, 0.30)**	**0.23 (0.07, 0.32)**
Bird-year ID	**0.10 (0.00, 0.24)**	**0.11 (0.01, 0.27)**
Recording ID	**0.44 (0.41, 0.49)**	**0.47 (0.42, 0.51)**
Plot-year ID	0.06 (0.00, 0.13)	0.05 (0.00, 0.13)
(b) Residual variances		
Fixed Effects	**φ (95% CI)**	**φ (95% CI)**
Intercept[b]	-0.37 (-0.42, -0.32)	-0.43 (-0.52, -0.35)
Year 2018	0.02 (-0.09, 0.14)	0.03 (-0.08, 0.14)
Year 2019	0.05 (-0.08, 0.16)	0.05 (-0.07, 0.17)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	**0.14 (0.01, 0.26)**	**0.14 (0.01, 0.25)**
Bird-year ID	**0.17 (0.03, 0.25)**	**0.17 (0.02, 0.27)**
Recording ID	**0.37 (0.34, 0.41)**	**0.38 (0.34, 0.42)**
Plot-year ID	0.07 (0.01, 0.14)	0.05 (0.00, 0.13)
(c) Correlations (a x b)	**r (95% CI)**	**r (95% CI)**
Individual ID	0.41 (-0.66, 0.96)	0.37 (-0.77, 0.95)
Bird-year ID	0.21 (-0.89, 0.96)	0.25 (-0.90, 0.97)
Recording ID	**0.47 (0.37, 0.57)**	**0.44 (0.33, 0.55)**

Supplementary Material Text S1

For minimum frequency, the base model was run with three chains, each with 6000 iterations and a warmup/burn in of 1000. From the remaining 5000 iterations, per chain, every third was kept (thin = 3) in order to reduced memory and post-chain processing time pre-emptively. However, as the memory usage and the times required for post-chain processing including diagnostics turned out to be unproblematic and thinning has been shown to be inefficient (Link & Eaton, 2012), we decided to keep every iteration while reducing the overall number of iterations. We adjusted the exact number of chains and iterations for each model following the initial set of analysis depending on the convergence and model fit (see Supplementary Material Table S2 for details on model parameters) to optimize processing times. Model fit and convergence were evaluated using Rhat estimates, as well as by graphical inspection of the trace plots. Furthermore, all models were specified with an adapt-delta of 0.99 and a maximal tree depth of 15 to ensure convergence.

Supplementary Material Table S5

Model	Chains	Iterations/chain	Warmup/burn in	Max tree depth	Adapt delta	Thin
Minimum Frequency:						
Initial model	3	6000	1000	15	0.99	3
Full dataset model	3	6000	1000	15	0.99	3
Base model	3	6000	1000	15	0.99	3
Environmental model	5	2500	500	15	0.99	1
Song type model	5	3000	500	15	0.99	1
Phrase Length						
Base model	3	6000	1000	15	0.99	3
Environmental model	3	3000	1000	15	0.99	1
Song type model	3	3000	500	15	0.99	1

Supplementary Material Table S6

Table S5 **Environmental models**: Effect size estimates and 95% credible intervals (CI) for predictors affecting mean and residual variance for minimum frequency per element and phrase length, including nest stage, test sequence and breeding density as environmental effects, as well as the correlations between mean and residual variance.

(a) Means	Minimum frequency, N = 21239	Phrase length N = 20600
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept	0.11 (-0.05, 0.27)	0.46 (0.44, 0.49)
Year 2018	-0.15 (-0.35, 0.05)	-0.00 (-0.03, 0.03)
Year 2019	-0.14 (-0.34, 0.06)	0.00 (-0.02, 0.04)
Nest stage	-0.08 (-0.17, 0.01)	0.00 (-0.01, 0.01)
Test sequence	-0.02 (-0.10, 0.07)	-0.01 (-0.03, -0.00)
Food treatment	-0.12 (-0.31, 0.07)	-0.02 (-0.04, 0.01)
Year 2018 * food treatment	0.22 (-0.04, 0.48)	0.02 (-0.02, 0.06)
Year 2019 * food treatment	0.19 (-0.11, 0.48)	0.03 (-0.02, 0.07)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0.22 (0.03, 0.33)	0.02 (0.00, 0.04)
Bird-year ID	0.13 (0.01, 0.30)	0.01 (0.00, 0.03)
Recording ID	0.46 (0.42, 0.51)	0.08 (0.07, 0.08)
Plot-year ID	0.05 (0.00, 0.14)	0.01 (0.00, 0.02)
(b) Residual variances		
Fixed Effects	**φ (95% CI)**	**φ (95% CI)**
Intercept[b]	-0.27 (-0.40, -0.13)	-3.70 (-3.92, -3.49)
Year 2018	-0.09 (-0.25, 0.08)	-0.09 (-0.36, 0.18)
Year 2019	-0.08 (-0.25, 0.09)	-0.17 (-0.43, 0.11)
Nest stage	**-0.08 (-0.16, -0.01)**	**-0.15 (-0.28, -0.01)**
Test sequence	-0.03 (-0.11, 0.04)	0.01 (-0.13, 0.13)
Food treatment	-0.17 (-0.34, -0.01)	-0.11 (-0.36, 0.15)
Year 2018 * food treatment	0.23 (0.01, 0.46)	0.29 (-0.08, 0.64)
Year 2019 * food treatment	0.25 (0.00, 0.49)	0.15 (-0.24, 0.54)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0.15 (0.01, 0.26)	0.09 (0.00, 0.23)
Bird-year ID	0.15 (0.01, 0.26)	0.12 (0.01, 0.29)
Recording ID	0.38 (0.34, 0.42)	0.76 (0.71, 0.82)
Plot-year ID	0.05 (0.00, 0.12)	0.08 (0.00, 0.20)
(c) Correlations (a x b)	**r (95% CI)**	**r (95% CI)**
Individual ID	0.31 (-0.78, 0.95)	0.18 (-0.90, 0.96)
Bird-year ID	0.15 (-0.91, 0.95)	0.07 (-0.93, 0.95)
Recording ID	0.45 (0.34, 0.56)	0.23 (0.13, 0.32)

Supplementary Material Table S7

Table S6 **Song type models**: Effect size estimates and 95% credible intervals (CI) for predictors affecting mean and residual variance for minimum frequency per element and phrase length, including nest stage, test sequence and breeding density as environmental effects, song types as random effects, as well as the correlations between mean and residual variance.

(a) Means	Minimum frequency, N = 21239	Phrase length N = 20600
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept	0.24 (0.02, 0.46)	0.50 (0.46, 0.54)
Year 2018	-0.14 (-0.34, 0.06)	0.01 (-0.03, 0.04)
Year 2019	-0.16 (-0.36, 0.04)	-0.00 (-0.04, 0.03)
Nest stage	-0.08 (-0.16, 0.00)	0.00 (-0.01, 0.02)
Test sequence	-0.00 (-0.08, 0.07)	-0.01 (-0.02, 0.01)
Food treatment	-0.10 (-0.29, 0.09)	-0.01 (-0.04, 0.02)
Year 2018 * food treatment	0.17 (-0.09, 0.44)	0.01 (-0.03, 0.06)
Year 2019 * food treatment	0.21 (-0.08, 0.49)	0.03 (-0.02, 0.08)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0.21 (0.07, 0.30)	0.04 (0.02, 0.05)
Bird-year ID	0.10 (0.00, 0.24)	0.01 (0.00, 0.03)
Recording ID	0.40 (0.36, 0.45)	0.07 (0.07, 0.08)
Plot-year ID	0.07 (0.00, 0.10)	0.01 (0.00, 0.02)
Song Type	**0.49 (0.36, 0.66)**	**0.09 (0.07, 0.11)**
(b) Residual variances		
Fixed Effects	**φ (95% CI)**	**φ (95% CI)**
Intercept[b]	-0.23 (-0.42, -0.04)	-3.56 (-3.85, -3.28)
Year 2018	-0.11 (-0.27, 0.05)	-0.20 (-0.49, 0.08)
Year 2019	-0.16 (-0.32, 0.00)	-0.15 (-0.43, 0.13)
Nest stage	-0.09 (-0.16, -0.02)	-0.09 (-0.23, 0.04)
Test sequence	-0.01 (-0.08, 0.05)	0.02 (-0.11, 0.15)
Food treatment	-0.19 (-0.34, -0.04)	-0.08 (-0.34, 0.19)
Year 2018 * food treatment	0.27 (0.05, 0.48)	0.40 (0.03, 0.78)
Year 2019 * food treatment	0.33 (0.10, 0.56)	0.06 (-0.33, 0.46)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0.15 (0.02, 0.25)	0.12 (0.00, 0.23)
Bird-year ID	0.15 (0.01, 0.25)	0.15 (0.01, 0.32)
Recording ID	0.34 (0.30, 0.38)	0.74 (0.68, 0.80)
Plot-year ID	0.04 (0.00, 0.10)	0.08 (0.00, 0.20)
Song Type	**0.43 (0.30, 0.59)**	**0.56 (0.41, 0.75)**
(c) Correlations (a x b)	**r (95% CI)**	**r (95% CI)**
Individual ID	0.44 (-0.61, 0.96)	0.25 (-0.80, 0.95)
Bird-year ID	0.14 (-0.93, 0.95)	0.10 (-0.93, 0.96)
Recording ID	0.48 (0.36, 0.60)	0.17 (0.06, 0.28)
Song Type	**0.72 (0.46, 0.88)**	**0.59 (0.27, 0.81)**

References

Alward, B. A., De Bournonville, C., Chan, T. T., Balthazart, J., Cornil, C. A., & Ball, G. F. (2016). Aromatase inhibition rapidly affects in a reversible manner distinct features of birdsong. *Scientific Reports, 6*(1), 1–9. https://doi.org/10.1038/srep32344

Bartsch, C., Hultsch, H., Scharff, C., & Kipper, S. (2016). What is the whistle all about? A study on whistle songs, related male characteristics, and female song preferences in common nightingales. *Journal of Ornithology*, *157*(1), 49–60. https://doi.org/10.1007/s10336-015-1245-y

Botero, C. A., Rossman, R. J., Caro, L. M., Stenzler, L. M., Lovette, I. J., de Kort, S. R., & Vehrencamp, S. L. (2009). Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. *Animal Behaviour*, *77*(3), 701–706. https://doi.org/10.1016/j.anbehav.2008.11.020

Byers, B. E. (2007). Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behavioral Ecology*, *18*(1), 130–136. https://doi.org/10.1093/beheco/arl058

Byers, B. E., Akresh, M. E., & King, D. I. (2015). A proxy of social mate choice in prairie warblers is correlated with consistent, rapid, low-pitched singing. *Behavioral Ecology and Sociobiology*, *69*(8), 1275–1286. https://doi.org/10.1007/s00265-015-1940-y

Byers, B. E., Akresh, M. E., & King, D. I. (2016). Song and Male Quality in Prairie Warblers. *Ethology, 122*(8), 660–670. https://doi.org/10.1111/eth.12513

Cramer, E. R. A. (2013a). Physically Challenging Song Traits, Male Quality, and Reproductive Success in House Wrens. *PLoS ONE*, *8*(3), e59208. https://doi.org/10.1371/journal.pone.0059208

Cramer, E. R. A. (2013b). Measuring consistency: Spectrogram cross-correlation versus targeted acoustic parameters. *Bioacoustics*, *22*(3), 247–257. https://doi.org/10.1080/09524622.2013.793616

Cramer, E. R. A., Hall, M. L., de Kort, S. R., Lovette, I. J., & Vehrencamp, S. L. (2011). Infrequent Extra-Pair Paternity in the Banded Wren, a Synchronously Breeding Tropical Passerine. *The Condor*, *113*(3), 637–645. https://doi.org/10.1525/cond.2011.100233

De Kort, S. R., Eldermire, E. R. B., Valderrama, S., Botero, C. A., & Vehrencamp, S. L. (2009). Trill consistency is an age-related assessment signal in banded wrens. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1665), 2315–2321. https://doi.org/10.1098/rspb.2009.0127

Faltynek, P. (2017). The effects of corticosterone and social isolation on song stereotypy and neural plasticity in zebra finches. *Electronic Thesis and Dissertation Repository*. https://ir.lib.uwo.ca/etd/4847

Ferreira, A. C., Atwell, J. W., Whittaker, D. J., Ketterson, E. D., & Cardoso, G. C. (2016). Communication value of mistakes in dark-eyed Junco song. *American Naturalist*, *188*(3), 289–305. https://doi.org/10.1086/687520

Geberzahn, N., & Aubin, T. (2014). How a songbird with a continuous singing style modulates its song when territorially challenged. *Behavioral Ecology and Sociobiology*, *68*(1), 1–12. https://doi.org/10.1007/s00265-013-1616-4

Gilman, S., Blumstein, D. T., & Foufopoulos, J. (2007). The Effect of Hemosporidian Infections on White-Crowned Sparrow Singing Behavior. *Ethology*, *113*(5), 437–445. https://doi.org/10.1111/j.1439-0310.2006.01341.x

Grava, T., Grava, A., & Otter, K. A. (2012). Vocal performance varies with habitat quality in black-capped chickadees (Poecile atricapillus). *Behaviour*, *149*(1), 35–50. https://doi.org/10.1163/156853912X625854

Holveck, M.-J., Vieira de Castro, A. C., Lacklan, R. F., ten Cate, C., & Riebel, K. (2008). Accuracy of song syntax learning and singing consistency signal early condition in zebra finches. *Behavioral Ecology*, *19*(6), 1267–1281.

Iserbyt, A., Eens, M., Baetens, W., Vermeulen, A., & Müller, W. (2017). Within- and between-individual (co)variance partitioning reveals limited pleiotropic effects of testosterone on immune function, sexual signaling, and parental investment. *Behavioral Ecology and Sociobiology*, *71*(4), 1–12. https://doi.org/10.1007/s00265-017-2308-2

James, L. S., & Sakata, J. T. (2019). Developmental modulation and predictability of age-dependent vocal plasticity in adult zebra finches. *Brain Research*, *1721*, 146336. https://doi.org/10.1016/j.brainres.2019.146336

Kriengwatana, B., Wada, H., Schmidt, K. L., Taves, M. D., Soma, K. K., & MacDougall-Shackleton, S. A. (2014). Effects of nutritional stress during different developmental periods on song and the hypothalamic-pituitary-adrenal axis in zebra finches. *Hormones and Behavior*, *65*(3), 285–293. https://doi.org/10.1016/j.yhbeh.2013.12.013

Kubli, S. P., & MacDougall-Shackleton, E. A. (2014). Developmental timing of signals affects information content: song complexity but not consistency reflects innate immune strategy in male song sparrows. *The American Naturalist*, *183*(5), 660–670. https://doi.org/10.1086/675757

Leblois, A., Wendel, B. J., & Perkel, D. J. (2010). Striatal dopamine modulates basal ganglia output and regulates social context-dependent behavioral variability through D1 receptors. *Journal of Neuroscience*, *30*(16), 5730–5743. https://doi.org/10.1523/JNEUROSCI.5974-09.2010

Meitzen, J., Thompson, C. K., Choi, H., Perkel, D. J., & Brenowitz, E. A. (2009). Time course of changes in Gambel's white-crowned sparrow song behavior following transitions in breeding condition. *Hormones and Behavior*, *55*(1), 217–227.

https://doi.org/10.1016/j.yhbeh.2008.10.006

Müller, W., Heylen, D., Eens, M., Rivera-Gutierrez, H. F., & Groothuis, T. G. G. (2013). An experimental study on the causal relationships between (ecto-)parasites, testosterone and sexual signalling. *Behavioral Ecology and Sociobiology*, *67*(11), 1791–1798. https://doi.org/10.1007/s00265-013-1586-6

O'Loghlen, A. L., & Rothstein, S. I. (1993). An extreme example of delayed vocal development: Song learning in a population of wild brown-headed cowbirds. *Animal Behaviour*, *46*(2), 293–304. https://doi.org/10.1006/anbe.1993.1190

Olveczky, B. P., Andalman, A. S., & Fee, M. S. (2005). Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLoS Biology*, *3*(5), e153.

Pytte, C. L., Gerson, M., Miller, J., & Kirn, J. R. (2007). Increasing stereotypy in adult zebra finch song correlates with a declining rate of adult neurogenesis. *Developmental Neurobiology*, *67*(13), 1699–1720. https://doi.org/10.1002/dneu.20520

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

Sakata, J. T., Hampton, C. M., & Brainard, M. S. (2008). Social Modulation of Sequence and Syllable Variability in Adult Birdsong. *Journal of Neurophysiology*, *99*(4), 1700–1711. https://doi.org/10.1152/jn.01296.2007

Schmidt, K. L., Moore, S. D., MacDougall-Shackleton, E. A., & MacDougall-Shackleton, S. A. (2013). Early-life stress affects song complexity, song learning and volume of the brain nucleus RA in adult male song sparrows. *Animal Behaviour*, *86*(1), 25–35. https://doi.org/10.1016/j.anbehav.2013.03.036

Souriau, A., Geberzahn, N., Ivanitskii, V. V., Marova, I. M., Vokurková, J., Reifová, R., Reif, J., & Petrusková, T. (2019). Singing behind the stage: thrush nightingales produce more variable songs on their wintering grounds. *Behavioral Ecology and Sociobiology*, *73*(11), 1–14. https://doi.org/10.1007/s00265-019-2765-x

Sprau, P., Roth, T., Amrhein, V., & Naguib, M. (2013). The predictive value of trill performance in a large repertoire songbird, the nightingale *Luscinia megarhynchos*. *Journal of Avian Biology*, *44*(6), 567–574. https://doi.org/10.1111/j.1600-048X.2013.00113.x

Taff, C. C., Steinberger, D., Clark, C., Belinsky, K., Sacks, H., Freeman-Gallant, C. R., Dunn, P. O., & Whittingham, L. A. (2012). Multimodal sexual selection in a warbler: Plumage and song are related to different fitness components. *Animal Behaviour*, *84*(4), 813–821. https://doi.org/10.1016/j.anbehav.2012.07.002

Teramitsu, I., & White, S. A. (2006). FoxP2 regulation during undirected singing in adult songbirds. *Journal of Neuroscience*, *26*(28), 7390–7394.

https://doi.org/10.1523/JNEUROSCI.1662-06.2006

Troy Smith, G., Brenowitz, E. A., Wingfield, J. C., & Baptista, L. F. (1995). Seasonal changes in song nuclei and song behavior in Gambel’s white-crowned sparrows. *Journal of Neurobiology*, *28*(1), 114–125. https://doi.org/10.1002/neu.480280110

Vehrencamp, S. L., Yantachka, J., Hall, M. L., & de Kort, S. R. (2013). Trill performance components vary with age, season, and motivation in the banded wren. *Behavioral Ecology and Sociobiology*, *67*(3), 409–419. https://doi.org/10.1007/s00265-012-1461-x

Wegrzyn, E., Leniowski, K., & Osiejuk, T. S. (2010). Whistle duration and consistency reflect philopatry and harem size in great reed warblers. *Animal Behaviour*, *79*(6), 1363–1372. https://doi.org/10.1016/j.anbehav.2010.03.012

Manuscript II: Song stability neither predicts within-pair reproductive success nor extra-pair paternity gain

Alexander Hutfluss, Hans Slabbekoorn, Kees van Oers, Niels J. Dingemanse

Unpublished manuscript

Abstract

Individuals consistently differ in behaviour in terms of their average behaviour (“personality”), their degree of plasticity (“individual plasticity), and their “residual variance” (“behavioural stability”) remaining after controlling for personality and plasticity. A major outstanding question is whether behavioural stability is under selection. Bird song represents an ideal candidate behavior to address this question. The stable production of attractive signals is assumed to be costly, suggesting that song stability may be a quality-indicator trait, i.e., a trait under directional selection. Assuming that stable singing signals male quality and “good genes”, we expected stable singers to attract partners that laid bigger clutches, that they gained more extra-pair paternity and lost less paternity at home. Assuming stable songs are more important for males that sing with low frequencies, we also expected selection on stable singing to be more pronounced for males that sing with lower frequencies (i.e., a form of correlational selection). Further, assuming that selection pressures are higher under less preferable conditions, we expected selection on stability to be stronger in years of high competition (i.e., a form of heterogeneous selection). To test these predictions, we monitored 12 nest box populations of great tits (*Parus major*) for three consecutive breeding seasons. We recorded each breeding male’s song up to four times per year, during simulated territory intrusions. At the same time, we measured each male’s reproductive parameters, namely, lay date, clutch size, probability to produce fledglings, and the total number of fledglings. We also assessed paternity, and quantified within-pair paternity loss and extra-pair paternity gain. Surprisingly, we found no support for any link between song stability and metrics of reproductive success, whether within- or outside the social pair. Overall, these findings indicate that the stable production of minimum frequencies is not under selection during the reproductive phase. In other words, we found no evidence for song stability to be used as a fitness-indicator trait by females, or to signal male quality in the first place. Thus, selection on song stability, if existing, would require other mechanisms not considered in this study.

Introduction

Animal behaviour varies at multiple, hierarchically structured levels, all of which contemporary behavioural ecology seeks to understand from an adaptive point of view. Historically, much research has focused on within-individual variation through studies of reversible plasticity or behavioural flexibility (Piersma and Drent 2003). Contemporary research, by contrast, increasingly studies multiple levels of variation in conjunction, for example, the integration between among-individual variation ("animal personality") and reversible plasticity, through studies of personality-related variation in behavioural plasticity (Dingemanse et al. 2010; Mathot et al. 2012). Only recently, attention is focusing on a more cryptic form of variation: the amount of residual within-individual variation (RWV, also intra-individual variation) that remains after controlling for (personality-related) plasticity (Stamps et al. 2012; Biro and Adriaenssens 2013; Westneat et al. 2015). Biologically, RWV represents an individual's level of behavioural "instability" or "unpredictability", attributable to within-individual variation that cannot be explained by the statistical model.

Importantly, behaving (un)predictably can be highly advantageous, but the specific benefits of behavioural stability are often context-specific, making stability itself, but also the ability to adjust it to the environment ("plasticity in stability"), potentially adaptive. For example, behaving less predictable might increase an individual's survival chances during predatory attacks but can also lead to mismatches between optimal and expressed phenotypes when predators are absent (Domenici et al. 2008; Jones et al. 2011; Briffa 2013). Likewise, instability during early life stages, or in novel environmental conditions, may facilitate learning which behaviour to best adopt (Brembs 2011). High levels of stability, on the other hand, can also be favoured by selection, for example, when communicating with conspecifics, stable signals, when costly to produce, can signal an individual's quality, both in sexual and natural selection contexts (Botero et al. 2009; Bartsch et al. 2016). In all of these cases, selection might favour particular optimal levels of stability, as well as the ability to adjust stability plastically. Studies directly estimating the fitness consequences of behavioural stability are, however, largely missing from the literature (Byers 2007).

There is growing evidence that individuals are repeatable in the level of behavioural stability that they express, implying that natural selection should be able act on behavioural stability in natural populations. Briefly, the behaviour of certain individuals is more stable across a range of contexts compared to the behaviour of others (Jolles et al. 2019; Hertel et al. 2021; Mitchell et al. 2021). Unfortunately, there is little formal theory on the proximate or ultimate causes of individuality in behavioural stability, but there are various conceptual ideas. A prominent explanation is that the existence of trade-offs between mean and variance expressed in a single trait or across different traits leads to a flat fitness surface mitigating the erosion of genetic variation despite selection. For example, individuals might experience different ecological conditions that favour different ways to resolve trade-offs. For example, prey species whose main defence is flight behaviour have to balance speed and manoeuvrability (representing stability of direction) to increase survival probability (Jindrich and Qiao 2009; Clemente and Wilson 2015). However, the fastest individuals should focus on speed, while manoeuvrable ones should focus on unpredictability, leading to higher survival for both. Likewise, open habitats might favour speed, whereas dense vegetation might require higher flexibility. Similar trade-offs can be seen for competitive situations that can require individuals to choose between “vigour” and “precision” (Briffa and Lane 2017). Examples include the trade-off between number vs the target accuracy of punches during boxing matches, or the trade-off between the number vs temporal consistency of songs during territorial disputes in birds (Lane and Briffa 2020). Finally, individual differences in stability might arise from individual differences in (genetic) quality. When the stable expression of certain behaviours is associated with costs, individuals might differ in the ability to pay these costs due to those differences in genetic make-up or early-life experiences (Sakata et al. 2008; Auld et al. 2010; Schuett et al. 2010). In such cases, stability is a fitness indicator trait which should be under directional selection acting to erode among-individual variation. However, indirect genetic effects, specifically when they are negatively correlated with direct genetic effects, or mutation-selection balance could heavily constrain selection or even lead to evolutionary stasis (Desai and Fisher 2007; Wilson et al. 2009; Santostefano et al. 2016).

Bird song is a prime example of a trait where stability both varies among individuals and is assumed to be costly to produce (Suthers and Zollinger 2004; Sakata and Vehrencamp 2012; Kubli and MacDougall-Shackleton 2014; Byers et al. 2015; Iserbyt et al. 2017). For this reason, stability in bird song has repeatedly been predicted to be under directional selection (Chapter 3)(Lambrechts and Dhondt 1988; Suthers and Zollinger 2004; Byers 2007). Studies on the adaptive potential of song stability are, however, mostly limited to those demonstrating indirect links with other fitness indicators, e.g. dominance, age or size (Botero et al. 2009; De Kort et al. 2009; Bartsch et al. 2016). Importantly, such traits might not reliably signal fitness in all species, or their significance might be mitigated by other factors. Increasing reproductive success with age, for example, might only be present in fast pace-of-life individuals (Dingemanse et al. 2020) or partially explained by selective disappearance masking senescence (Bouwhuis et al. 2009). Therefore, what we require is insight into the direct routes by which directional selection might act on stability (Byers 2007; Cramer et al. 2011).

From previous studies we know that reproductive success can be achieved via three different pathways: female clutch size (and subsequent within-pair success), within-pair paternity loss, and extra-pair paternity gain (Webster et al. 1995; Griffith et al. 2008; Araya-Ajoy, Kuhn, et al. 2016; Araya-Ajoy, Dingemanse, et al. 2016) (Figure 1). Links between singing and specific pathways have been reported for a number of song traits (McGregor et al. 1981; Forstmeier et al. 2002; Hardouin et al. 2009; Nemeth et al. 2012; but see Garamszegi and Møller 2004 and Soma and Garamszegi 2011). Consequently, stable singers can be expected to also exhibit increased fitness via some or all three pathways (Figure 1). First, stable singers should be high-quality birds that attract high-quality mates, therefore occupying better territories and having mates producing larger clutches (Araya-Ajoy, Kuhn, et al. 2016). Ultimately, this should lead to more fledglings, in potentially better condition, for males that sing more stable songs. Second, if stable singing indicates a male’s genetic quality (good genes hypothesis), social mates of stable singers might be more faithful or refrain from seeking alternative sires for their offspring, therefore reducing the paternity loss of stable singers. Finally, stable singers can be expected to be chosen more often as an extra-pair mates, therefore enjoying increased reproductive success outside their social bond.

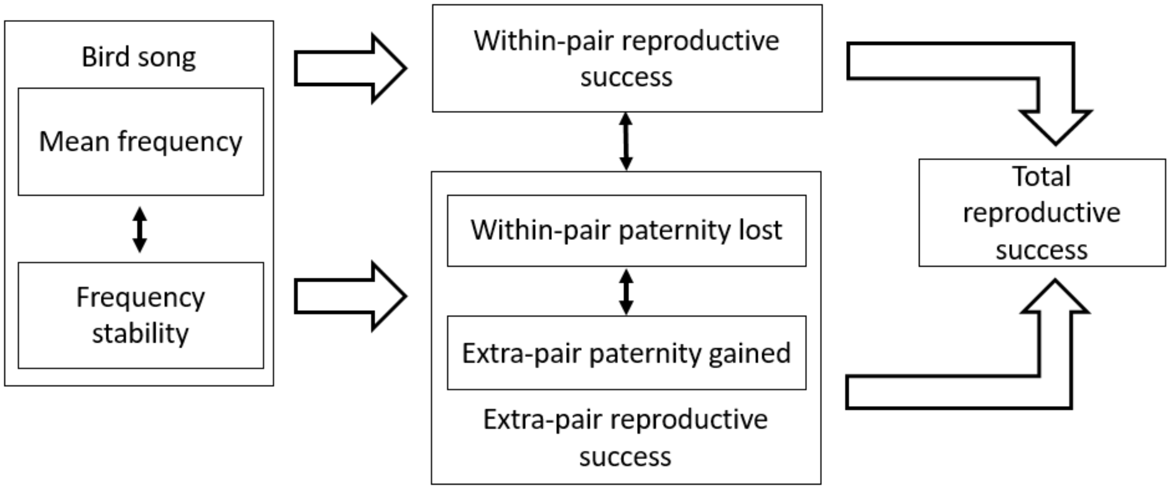


Figure 1: Illustration of the three pathways towards male reproductive success. The path diagram illustrates how mean frequency, the stability of frequency and the interaction between those two song traits affect each. Double-sided arrows indicate potential trade-offs/interactions.

In great tits and other song birds, minimum frequency represents a key song parameter for which both mean and stability are repeatable (Chapter 3)(Halfwerk et al. 2011; Zollinger et al. 2017). In many species minimum frequency is under selection, thus suggesting that the stable production of low frequencies might be advantageous (ten Cate et al. 2002; Gil and Gahr 2002; Hardouin et al. 2009). However, evidence for selection on stability is limited (Byers 2007; Cramer et al. 2011). In this study, we therefore examined the links between song stability in minimum frequency and within- and extra-pair reproductive success. We monitored great tit males over the course of three years, repeatedly recording their song in response to simulated territory intrusions and measuring different fitness parameters including paternity. We expected the following. First, we expected stable singers to have social partners that started breeding earlier and laid more eggs, and themselves to be more likely to produce at least one fledgling and to ultimately produce more fledglings. Second, we expected females mated to stable singers to be more faithful, leading to fewer offspring in the focal male's nest sired by extra-pair males. And we expected stable singers to gain more extra-pair fertilizations, thus gaining paternity in other nests. In addition, we expected stronger selection on the stable production of lower frequencies, i.e., a form of correlational selection, considering the strong preference for lower frequencies in many species (Byers 2007; Halfwerk et al. 2011; Huet des Aunay et al. 2014; Zollinger et al. 2017). Finally, we expected stronger selection in years of high competition, i.e. a form of heterogeneous selection, arguing that selection pressures are stronger under less preferable conditions.

Methods

Data were collected during three consecutive breeding seasons (2017–2019) in 12 great tit nest box populations (established in 2009) located between Starnberg and Herrsching (Bavaria, Germany). We monitored all nest boxes from the beginning of April following a standardized protocol (see Nicolaus et al., 2015 for details). We assessed key reproductive parameters, including lay date, clutch size, hatch date, brood size, and fledgling number. Ten days after hatching, we caught both adult breeders in the nest box using trapdoors. We measured weight, tarsus and wing length, ringed birds if not banded previously, and collected 10µl blood samples. For nestlings the same procedure was followed on day 14. We determined the number of fledglings by checking the nests for dead nestlings 19-21 days after hatching and comparing it to the number of nestlings measured on day 14.

Simulated territory intrusions

To acquire repeated song recordings from the same set of males, we subjected each breeding male to four simulated territorial intrusions. We chose this approach rather than recording songs during the dawn chorus or unprovoked singing bouts as many other studies (Halfwerk et al. 2011; Byers et al. 2015; Zollinger et al. 2017), because it allowed us to sample songs in an efficient and unbiased way for all individuals equally. One could argue that simulated territory intrusions represent male-male competition rather than female choice. However, at least for blue and great tits singing behaviour during the dawn chorus predicts singing behaviour during territory intrusions (Poesel et al. 2004; Snijders et al. 2015) indicating that responses are consistent across contexts. Furthermore, the main assumption behind this study was that song stability functions as a quality-indicator trait. Unless males use different song traits to assess an opponent’s quality compared to females, we would not expect substantial differences in the singing between the two contexts.

The simulated territory intrusions were conducted during first breeding attempts (defined as clutches started within 30 days after the first egg in all populations was found); two tests during egg-laying (1 and 3 days after the first egg was observed) and two tests during incubation (1 and 3 days after clutch incubation was confirmed). Tests were conducted between 7h00AM and 12h00PM and postponed if weather conditions were unfavourable. We used a taxidermic great tit model as visual and a playback of a great tit song as acoustic stimuli. The model and speaker (Foxpro Shockwave) were placed 1m in front of the subject's nest box. The model was placed on a 1.2m wooden pole for consistent visibility, the speaker on the ground next to the pole. Playbacks were chosen randomly from a set of 23 models and 13 (2017) or 174 (2018 and 2019) playbacks (for further details see Chapter 3).

We recorded the focal male's behaviour for three minutes after the male entered a 15 m radius around the nest box. The observer, positioned in front of the box at the edge of the radius, recorded the song of the focal male for the full test period using a directional microphone (Sennheiser ME66/K6) and a recorder (TASCAM DR-05, recording 44.1 kHz, 16-bit WAV files). In addition, observers estimated the minimum approach distance to the taxidermic model as measure of aggressiveness. Subjects not arriving within 15 minutes were scored as "non-responsive".

Acoustic measurements

We extracted acoustic parameters from the recordings using Avisoft SAS Lab Pro (Avisoft Bioacoustics, Specht 2002) following a semi-automatic protocol detailed in Hutfluss et al (in submission). After applying "Time Domain FIR Filters" (between 2.0 and 8.0 kHz) and normalizing all files to 75%, we produced separate spectrograms (FFT length: 1024, Overlap: 87.5%, Window: Hamming) for each strophe of the focal bird. We then extracted spectral and temporal features of each element using the automatic measurements function (three thresholds: -24 dB, start: -16 dB, end: -12 dB, hold time: 10 ms, minimum length: 30 ms). For each element, we measured start and end time, and peak frequency at the start and end. The lower of those two frequencies was then taken forward as minimum frequency.

Extra-pair paternity (EPP)

Immediately after the collection, we transferred each blood sample into Eppendorf tubes containing Queen's Lysis buffer (Seutin et al. 1991) and later extracted DNA using a NucleoSpin Blood QuickPure extraction kit (Macherey-Nagel, Düren, Germany).

To determine paternity of the nestlings we used five highly variable microsatellite markers PmaC25, PmaD105, PmaGAn27, PmaTAGAn71 and PmaTGAn33 (Saladin et al. 2003). PCR products were prepared and transferred on ABI plates together with a molecular size standard

(GeneScan LIZ, Applied Biosystems, Foster City, CA, USA). Sequencing was carried out by BaseClear BV (BaseClear BV, Leiden, The Netherlands). We sequenced 145 of 220 broods for the breeding season 2017 and 155 of 197 broods for 2018. We were able to determine paternity for 295 of 300 available broods (145 in 2017 and 150 in 2018). Five broods from 2018 containing 32 offspring had to be excluded, because of a mix-up of samples. We categorized nestlings as within-pair offspring if all of their loci matched those of the social mother and father (no trio-mismatches). We also categorized nestlings with a maximum of one mismatch with the social father in one locus, if the social father was still the most likely father. We did this To do this, we tested all nestlings against their supposed fathers using the Windows-based program CERVUS version 5.0 (Marshall et al. 1998; Kalinowski et al. 2007). None of the loci deviated significantly from Hardy Weinberg equilibrium when the genotypes of all individuals in the analysis were included. We calculated critical LOD values using the following parameters in CERVUS: 10,000 cycles, 98% of loci typed, error rate 0.01% and two candidate parents). We then compared the genotypes of nestlings categorized as extra-pair with the genotypes of all males in the population in the respective year in CERVUS (‘open analysis’), using the mother as ‘known parent’. We calculated critical values using the following parameters in CERVUS: 10,000 cycles, 98% of loci typed, error rate 0.01%. We assigned paternity to a candidate male if its genotype matched the offspring’s genotype perfectly and if this match was more likely (higher LOD) than the next best matching male. Candidate males with one mismatch with the nestling were also assigned as sire, if they had a perfect match with another nestling in the same nest (parsimony). Only males breeding in the same plot were assigned as biological fathers of extra pair offspring, as great tits show high side-fidelity (personal observation), extra-pair fathers typically inhabit territories that are closer than 200 m from the extra-pair nest (unp. data) and plots are separated by at least 1.5 kilometres. For extra-pair nestlings that had no likely candidate father within the same plot, we assumed an unsampled male to be the biological father.

Statistical analysis

The main aim of this study was to examine whether song stability predicted reproductive success. Initially, we used multivariate double hierarchical mixed models (DHGLM) to directly model links between the stable production of minimum frequencies and different reproductive parameters. DHGLMs represent an expansion of regular mixed effect models as they are specifically designed to model the distribution of residual variance across random effect levels or along environmental gradients or other covariates (Lee and Nelder 2006; Rönnegård et al. 2010). Within the multivariate DHGLM we combined a DHGLM estimating residual within-individual variance in song with regular linear mixed effect-models estimating reproductive parameters. However, these state-of-the-art methods require large amounts of data; despite our substantial dataset with hundreds of individuals and tens of thousands of acoustic measurements, these models did not converge.

Instead of this multivariate DHGLM, we, therefore, used a two-step approach running the DHGLM and the linear mixed models successively. We started with a univariate DHGLM to estimate individual differences in mean and stability of minimum frequency, using this trait as response variable. In previous work, we have shown considerable among-individual variation in this trait (Chapter 3). We added random intercepts for individual identity, recording identity and bird-year (i.e. the unique combination of bird and year) to the mean and residual part of the DHGLM. To account for variance at spatio-temporal scales (e.g. breeding density), we added plot-year as random effect. In addition, to account for variation among breeding seasons, we included year as a fixed effect in the mean and residual part. From this DHGLM, we extracted the Best Linear Unbiased Predictors (BLUPs) for each individual's mean ("minimum frequency") and stability in minimum frequency ("stability frequency").

We then took the BLUPs forward and included them as covariates in the linear mixed-effect models to estimate links between stability and reproductive success. We focused on six key breeding parameters: 1) the day the first egg was laid ("lay date") 2) number eggs laid ("clutch size), 3) the probability to produce at least one fledgling ("nest success), 4) the number of offspring in a focal male's nest box for boxes that had at least one fledgling ("number of fledglings"), 5) the probability to sire at least one offspring with an extra-pair mate ("paternity gained"), and 6) the probability to have at least one offspring in the focal male's nest sired by an extra-pair male ("paternity lost"). We pragmatically chose to use these binary categories due to the strong prevalence of birds without any paternity gains or losses (see Results). We ran model 4 with successful nests only because we assumed complete nest failures to probably be caused by extreme events such as cold spells or predation rather than by gradual differences in traits related to song stability (Marques-Santos and Dingemanse 2020). For each response variable, we fitted random intercepts for individual identity, plot, and plot-year, to account for repeated measures and variation at different spatial and spatio-temporal levels. We included year as fixed effect in all three models.

In two subsequent analyses we adjusted the reproductive success models 1 to 6 to examine two specific types of selection: 1) stronger selection on the stable production of lower frequencies (correlational selection) and 2) stronger selection on song stability in years with stronger competition (heterogeneous selection). To address the first prediction, we modelled the interaction between the two BLUPs (mean and stability) as a covariate in all models. To address the second prediction, we added breeding density defined as breeding pairs per hectare per plot as covariate and its two-way interaction with both the BLUPs for mean minimum frequency and stability in minimum frequency as covariates to all models.

Model parametrization

All analyses were performed in R v.4.0.2 (R Core Team 2020). Classic and double hierarchical mixed effect models were implemented using the packages “brms”, “rstan”, and “MCMCglmm” (Bürkner 2017). The DHGLMM was run with the default priors for all model parameters. It was run with three chains, each with 6000 iterations and a warmup of 1000 From the remaining 5000 iterations per chain, every third was kept (thin = 3). Furthermore, all models were specified with an adapt-delta of 0.99 and a maximal treedepth of 15 to ensure convergence. Model fit and convergence were evaluated using Rhat estimates, as well as by graphical inspection of the traceplots. Mean (β) and 95% credible intervals (CIs) for each fixed and random effect parameter are presented. We regarded effects significant in the frequentist’s sense when their 95% CIs did not overlap zero (Nakagawa and Cuthill 2007). Furthermore, we regarded effects as absent when their 95% CIs were equally distributed around zero with an average effect size estimate close to zero. Finally, we regarded effects as neither absent nor significant when their average effect size deviated from zero and their 95% CIs were not equally distributed around zero but still overlapped zero.

The usage of BLUPs in subsequent analysis has been criticized, as this approach does not account for the uncertainty associated with these estimates (Hadfield et al. 2010; Houslay and Wilson 2017). One possible solution is to use the posterior distribution of possible BLUP estimates rather than just the mean estimate. Subsequent analyses are run repeatedly with different BLUP estimates randomly drawn from this posterior distribution. Using simulations a recent study has shown that this method can lead to an underestimation of the mean effect size, thereby resulting in conservative estimates (Dingemanse et al. 2020) . We, therefore, pragmatically decided to focus on the effects of the average BLUPs for mean and stability of minimum frequency throughout the main text.

Results

Over the course of three breeding seasons, we extracted acoustic parameters from 586 recordings from 273 males, for which we then acquired individual estimates for mean and stability of minimum frequency. During the study period we monitored 366 breeding males, which produced 469 broods. In 2017 and 2018 we assigned paternity for 299 broods of 249 distinct males. Of these males, 187 (75.1%) did not gain any extra-pair paternity and 152 (61.0%) did not lose any. For the 62 males that gained paternity, number of offspring sired ranged from one to six (mean = 1.87). For the 97 males that lost paternity, number of offspring sired by different males ranged from one to nine (mean = 2.16) (Figure 2).

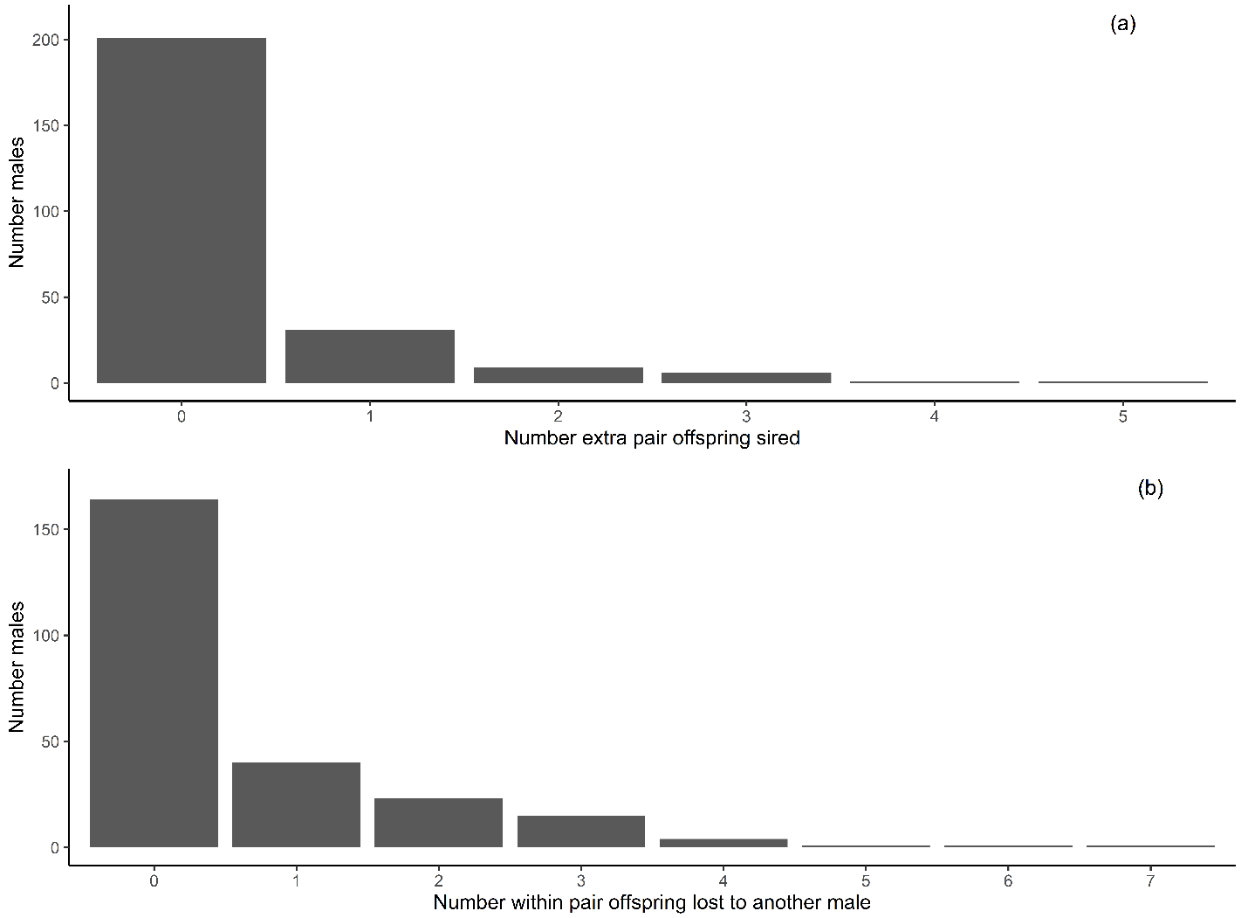


Figure 2: Total number of males for each number of (a) extra pair offspring sired and (b) within pair offspring sired by extra-pair males.

Within-pair reproduction

We found no support for any of the reproductive parameters to be influenced by mean minimum frequency or stability of minimum frequency (Table 1, Figure 3). There was, however, moderate support for a positive link between lay date and residual within-individual variance. This support was indicated by the respective point estimate that had 95% CIs skewed away from zero (towards positive values), indicative of weak evidence for stable singers to be mated to females that initiated clutches earlier. At the same time, our models strongly support the absence of an effect of either mean minimum frequency or its stability on clutch size or the number of fledglings as their point estimates were centred on zero. This indicates that stable singers or males producing songs with lower minimum frequencies were not mated to females laying larger clutches, and neither did they produce more offspring with their social mate. Along the same lines, our analysis supports the absence of an effect of mean minimum frequency on lay date, indicating that males with lower frequencies were not mated to early breeding females. Finally, estimates of mean frequency and stability on nest success were supporting neither the absence nor the existence of links between singing and this reproductive success parameter (Table 1). This was indicated by mean effect sizes deviating from zero and credible intervals not equally distributed around but still overlapping zero. All reproductive parameters differed strongly among years. Compared to 2017, 2018 was characterized by higher nest success and more fledglings, indicating it was a “good” year. 2019, on the other hand, was characterized by smaller clutches, lower nest success and fewer fledglings, thus representing a “worse” year. Finally, birds did show consistent individual differences in all parameters except the nest success, although among-individual variation in lay date and clutch size was likely caused by differences among females consistently pairing with the same male.

There was no strong support for correlational selection, i.e. selection on specific combinations of frequency and stability, e.g. low and stable frequencies (indicated by the interaction effect between the two BLUPs). There was also no strong support for the absence of such selection, as the estimates of the interaction between mean and stability were not centered around zero for any of the reproductive parameters. Only for clutch size, we found a slight tendency for a positive interaction. This means that birds singing higher frequencies with lower stability, i.e. higher RWV) had partners that laid fewer eggs (Supplementary Material Table S1). There was also no strong support for heterogeneous selection, i.e. selection to be more pronounced when competition was higher. This means higher breeding densities did not lead to increased reproductive success for stable singers, or males that produced lower frequencies. Overall, there was also no strong support for the absence of such effects, indicated by the interactions between density and minimum frequency or stability frequency not equally centred on zero. For breeding density itself, there was moderate support for a negative effect on lay date, indicating that broods were started earlier in years with more breeding pairs (Supplementary Material Table S2).

Table 1: Effects of minimum frequency and stability frequency on lay date, clutch size, nest success and the number of fledglings shown as 95% credible interval around the estimated mean parameter estimate.

(a) Means	Lay date N = 362	Clutch size N = 362	Nest success N = 363	Number fledglings N = 321
Fixed effects	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Intercept[a]	7.80 (6.43, 9.22)	8.39 (8.05, 8.72)	2.19 (1.59, 2.80)	5.00 (4.60, 5.39)
Year 2018	13.01 (11.60, 14.46)	0.05 (-0.31, 0.41)	0.92 (-0.10, 1.94)	2.23 (1.76, 2.71)
Year 2019	4.99 (3.46, 6.45)	-0.68 (-1.11, -0.28)	-0.96 (-1.72, -0.20)	-0.76 (-1.32, -0.21)
Minimum Frequency	-0.37 (-5.13, 4.37)	-0.19 (-1.73, 1.28)	-1.20 (-4.08, 1.67)	-0.08 (-1.75, 1.51)
Stability Frequency	9.28 (-0.37, 19.14)	-0.24 (-3.24, 2.79)	2.52 (-3.20, 8.03)	0.25 (-2.91, 3.43)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	3.27 (2.64, 4.05)	0.44 (0.36, 0.53)	0 (0, 0)	0.22 (0.17, 0.27)
Plot-Year	0.43 (0.24, 0.66)	0.003 (0.001, 0.004)	0 (0, 0)	0.11 (0.06, 0.18)
Plot	4.33 (1.72, 8.19)	0.14 (0.05, 0.26)	0 (0, 0)	0.11 (0.04, 0.21)
Residual	23.16 (20.12, 26.63)	1.94 (1.67, 2.24)		2.32 (1.97, 2.70)

[a] Reproductive parameters for birds breeding in 2017 with population level mean and RWV of minimum frequency

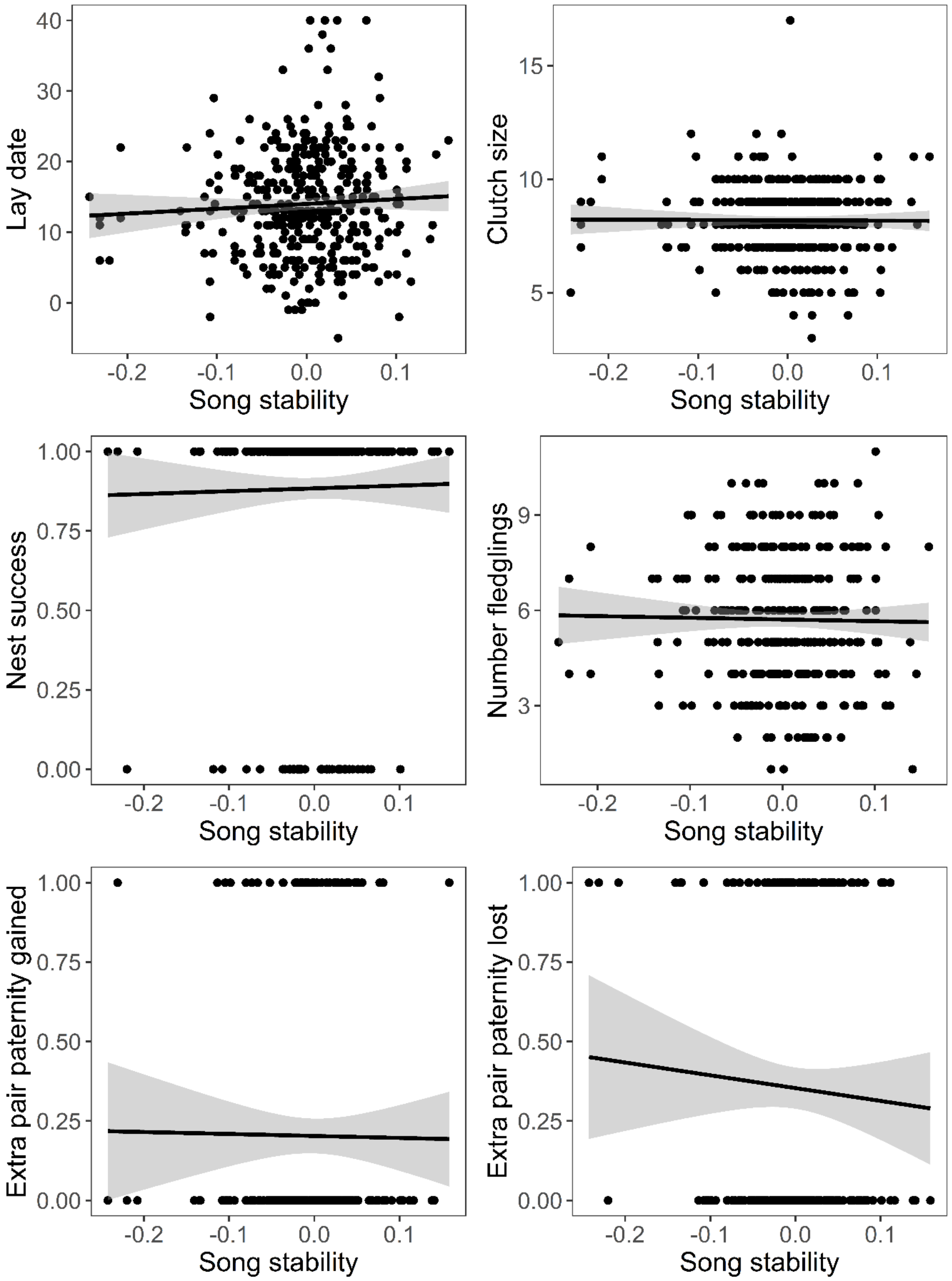


Figure 3: Song stability represented by individual BLUPs for RWV for minimum frequency plotted against all six reproductive success parameters. Black lines represent the regression line, while the grey areas indicate the 95% confidence interval. Graph based on raw data.

Extra-Pair Paternity

Overall, we did not find strong support for an influence of mean minimum frequency or stability on the probabilities to gain or lose paternity for at least one nestling (Table 2, Figure 2). In fact, parameter estimates supported the absence of effects of mean and stability on extra-pair paternity gained, indicating that males were not chosen as extra-pair mates based on their minimum frequencies. On the contrary, our results for extra-pair paternity lost supported neither the existence nor the absence of an effect, due to their credible intervals not centred on zero. Nonetheless, the estimates suggested that stable singer and males producing low frequencies were not better at preventing their female from seeking extra-pair fertilizations. The probability to lose or gain at least one EPP offspring did not differ between the two study years. Neither of the two traits showed consistent individual differences, indicating that males did not differ in the probability to sire EPP offspring or to lose paternity to EPP males. However, this could also be a statistical artefact (see below).

There was no support for correlational selection, i.e. a link between EPP paternity and specific combinations of frequency and stability, e.g. low and stable frequencies (indicated by the interaction effect between the two BLUPs) (Supplementary Material Table S3). Neither was there strong support for the absence of such selection. Interestingly, when accounting for the interaction between mean and stability, consistent individual differences in both, EPP gain and loss were found. This indicates that either the base model was incomplete or that the previously not found individual differences were a statistical artefact. There was no strong support for heterogeneous selection, i.e. an effect of competition on EPP paternity. Neither was there strong support for the absence of such selection. This indicates that birds singing stable or low frequency songs were not favoured when competition was high (Supplementary Material Table S4).There was also no direct effect of breeding density on either of the two paternity measures, meaning that higher numbers of breeding pairs did not affect the probabilities to lose or gain EPP. This means higher breeding densities did not lead to increased probabilities of EPP gained or lost.

Table 2: Effects of minimum frequency and stability frequency on the probabilities to gain or lose at least one extra-pair offspring shown as 95% credible interval around the estimated mean parameter estimate.

(a) Means	Extra-Pair Gain (Y/N) N = 257	Extra-Pair Lost (Y/N) N = 257
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept[a]	-1.47 (-1.93, -1.02)	-0.58 (-0.97, -0.20)
Year 2018	0.20 (-0.45, 0.85)	-0.04 (-0.62, 0.57)
Minimum Frequency	-0.32 (-3.39, 2.81)	0.96 (-1.55, 3.49)
Frequency Stability	-0.11 (-5.72, 5.52)	-2.39 (-7.18, 2.19)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0 (0, 0)	0 (0, 0)
Plot-Year	0 (0, 0)	0 (0, 0)
Plot	0.02 (0.01, 0.04)	0 (0, 0)
Residual	3.29[b]	3.29[b]

[a] Number of fledglings for birds breeding in 2017 with population level mean and RWV of minimum frequency

[b] Residual variance fixed to $\pi^2/3$ for binary models

Discussion

This study examined multiple routes by which selection might act on behavioural stability in the context of bird song. Specifically, we investigated links between song stability in minimum frequency as a key song trait and multiple reproductive success parameters. We considered that reproductive success can be achieved via different pathways by differentiating between within- and extra-pair success. The latter was further divided into within-pair paternity lost to other males and extra-pair paternity gained. Overall, we found no evidence for links between the stable production of low frequency songs and any of the investigated song parameters. There was weak evidence at best for stable singers to be mated to females laying earlier. In addition, stable singers were not more likely to avoid nest failures or the loss of paternity within their own nest. Our models instead gave relatively conclusive support for the *absence* of links between song stability and clutch size, the number of fledglings and the probability to sire at least one EPP offspring: point estimates were largely zero with credible intervals centred on zero. These results imply that stable singers were not mated to females laying larger clutches nor did they produce more offspring or were more likely to sire EPP offspring. In conclusion, specific results did not support our expectations but instead refuted them, indicating that song stability is not affecting reproductive parameters in general. In addition, we found no support for correlational or heterogeneous selection. Overall, this implies that females are unlikely to use it as a signal of male quality or good genes. Considering the existence of consistent individual differences in stability these findings lead to further questions regarding the cause and maintenance of such individuality in stability and the adaptive value of song stability. If selection also does not act via any other mechanism, e.g. survival, song stability would not be a fitness-indicator trait, and mutation-selection balance cannot explain this variation.

Within-pair reproductive success

Song stability can be linked to within-pair reproductive success via two non-independent pathways, both of which were not supported by our findings. Importantly, both ways are based on the assumption that the acoustic trait of interest represents male quality. First, birds are assumed to mate assortatively in regards to quality, either by mutual mate choice or by preference for quality in one sex and competition for quality mates within the other (Holveck and Riebel 2010; Jiang et al. 2013). If stable singing represents an underlying male quality and assortative mating for quality occurs, stable singers should have mated with high-quality females indicated by earlier breeding initiation and larger clutches (Hatchwell 1991; Araya-

Ajoy, Kuhn, et al. 2016). Second, if song stability correlates with quality indicators such as competitiveness leading to the occupation of better territories or parental care, stable singers should be more successful in raising offspring, independent of female quality. The absence of either of these patterns strongly indicates that the individual differences in song stability that we report in Chapter 3 are not used by females as basis of their mate choice, neither are they signalling male quality in the first place.

The available, but limited evidence found in the literature indicates that stable singers mated with early breeding females and had higher reproductive output but were not likelier to find a social mate (Taff et al. 2012; Byers et al. 2015). Consequently, song stability could represent a quality-indicator trait that is preferred by high quality females and ultimately under selection. Furthermore, the notion that unmated males and males mated to fertile females did not substantially differ in stability indicates that biases in our data due to the exclusion of unmated and thus supposedly less stable males seems unlikely (Taff et al. 2012). Our findings, however, contradict this assumption and suggest that song stability in this particular song trait is simply not under selection. This discrepancy might be explained by two linked issues. First, the scarce direct empirical evidence for selection on stability makes it very difficult to draw general conclusions. Second, selection pressures and female preference can be highly species and trait specific (Macdougall-Shackleton 1997; Soma and Garamszegi 2011). In great tits, repertoire sizes rather than other song traits seem to associated with female choice, survival and reproductive success (McGregor et al. 1981; Baker et al. 1986; Rivera-Gutierrez et al. 2010). Song stability, on the other hand, seems to signal age (Rivera-Gutierrez et al. 2010). However, in great tits the links between age and reproductive success are affected by personality and the existence of selective disappearance masking senescence (Bouwhuis et al. 2009; Dingemanse et al. 2020). Finally, the available estimates on within-pair reproductive success in relation to song traits might have been overestimated due to publication biases or confounding factors (Soma and Garamszegi 2011). As such, our results, especially the evidence for the absence of links between stability and reproductive success might not be surprising but simply reflect underrepresented representations of non-existent patterns of selection.

Extra-pair paternity

Being chosen as an extra-pair mate based on acoustic traits represents an alternative pathway by which males can acquire reproductive success. This pathway follows the good-genes hypothesis, which argues that females benefit from extra-pair matings if they choose males with traits that signal high-quality. However, our results do not indicate a link between song stability and extra-pair paternity gains or losses. In fact, our findings strongly suggest that extra-pair paternity is not gained via stable singing in great tits. Consequently, the stable production of this trait is unlikely used by females to choose suitable sires, potentially because it does not convey information on male quality. This would also be in line with the notion that the good-genes hypothesis finds almost no support in great tits (Krokene 1998; Strohbach et al. 1998; van Oers et al. 2008). Specifically, offspring sired by extra-pair males does not show higher growth rates, body mass or survival, thereby not leading to fitness benefits of extra-pair matings for the female. Consequently, even if females use specific traits to choose extra-pair mates, these traits are unlikely reliable fitness indicators, at least in great tits. Alternatively, extra-pair paternity in great tits might be explained by other mechanisms then by active choice, since females are not actively visiting extra-pair mates territories (Bircher et al. 2020; Bircher et al. 2021).

Nonetheless, our findings are opposing previous studies in other species showing stable singers to gain more extra-pair offspring (Byers 2007; Cramer et al. 2011; Taff et al. 2012). One explanation for this difference could be the substantial amount of within-individual variation in song stability (Taff et al. 2012; Cramer 2013, Chapter 3). In studies not accounting for this variation, positive links between song stability and extra-pair paternity gains could be explained by temporal or spatial autocorrelations. Factors leading to increased extra-pair matings might simultaneously lead to increased song stability (Westneat and Stewart 2003 and citations therein). Vegetation density or habitat complexity, for example, might require birds to sing more stable songs to transfer information while simultaneously impeding mate guarding (Sherman and Morton 1988; Brown and Handford 2000). By using individual stability estimates based on long-term data, we could reduce the potential influences of this within-individual variation and forcefully demonstrate that individual differences in stability are not linked to extra-pair paternity gains or losses. In addition, a recent meta-analysis indicated that there is no overall link between acoustic traits such as song complexity or output and extra-pair paternity (Garamszegi 2004). Although this pattern might be different for the stability of song traits, it is in line with our results supporting the absence of such links. Ultimately, females might base their choice of extra-pair mates on other non-acoustic signals, behaviours or genetic similarity, rather than specific song traits (Arct et al. 2015; Araya-Ajoy, Kuhn, et al. 2016; Roeder et al. 2019).

Patterns of selection

Besides a trait- and species-specific signal value of acoustic features, temporal and spatial patterns of selection in relation to environmental factors might offer an alternative explanation for both our findings and diverging results in the literature (Doutrelant et al. 2000; Quinn et al. 2009; Dingemanse and Réale 2013; Mouchet et al. 2021). Specifically, the selective advantage of supposedly high quality signals, e.g. stable production of song traits, might be more pronounced during times of scarcity. When life is good, i.e. plenty of food, nesting sites or low predation, all individuals should be able to mate and reproduce. As a consequence, individual differences in underlying quality will not be reflected by differences in reproductive success. However, when conditions are harsh, e.g. when competition for limited resources is high, only the "best" individuals might be able to properly reproduce leading to visible selection pressures for quality signals (Both 1998; Nicolaus et al. 2016). Such changing patterns of selection could lead to year-specific links between song traits and reproductive success as well as to null results on long-term mean values of reproductive success such as ours. However, testing for year-specific differences in breeding density, and thus competition, did not reveal these types of heterogeneous selection.

Individuality in stability

The absence of selection on song stability during the reproductive phase raises the question why consistent individual differences in stability exist in the first place. This finding obviously contradicts the assumption that stability functions as a quality signal. Alternative explanations for the present individuality in stability could be found in the positive correlation between mean and RWV found in a previous study (Chapter 3). Importantly, the positive link, i.e. higher frequencies are associated with lower stability (higher RWV), refutes a potential trade-off between the two traits. That means that individual differences in stability are not caused by differences in how individuals resolve such trade-offs, e.g. due to different ecological conditions they experience. This was further supported by the lacking support for correlational selection on specific combinations of mean minimum frequency and stability of minimum frequency. Alternatively, the positive link could indicate that certain environmental conditions simultaneously favour certain frequency and stability levels. If these environmental conditions differ among individuals we would expect the observed individual differences in stability. One example could be higher vegetation density, for which lower frequencies and stable singing are required to ensure signal transmission compared to open habitats (Badyaev 1997; Brown and Handford 2000; Derryberry 2009).

Importantly, individual differences in stability could also exist without stability being a biological important trait in itself. Positive correlations between mean and variance are common in nature and can often be attributed to mathematical or methodological artefacts as a by-product of selection on mean minimum frequency (Taylor 1961; Viswanathan 2005; Westneat et al. 2015). Alternatively, high-quality individuals might be able to produce lower frequencies very consistently, even though the stability in itself has no signal value. Together, this could mean that selection on lower frequencies present in many species leads to indirect selection on higher stability (Halfwerk et al. 2011; Zollinger et al. 2017). Individual differences in minimum frequency, e.g. due to size differences, would then cause individual differences in stability as a by-product. Along the same lines, environmental conditions that favour different frequencies could lead to individuality in minimum frequency and indirectly individuality in stability. Noise levels, for example, could affect the transmission of low frequency songs, leading to higher frequencies and consequently lower stability (Halfwerk and Slabbekoorn 2009; Halfwerk et al. 2011). Surprisingly, our results oppose this explanation as there was no evidence for selection on the mean minimum frequency either.

Conclusion

Overall, this study implies that song stability is not under selection during the reproductive phase. We demonstrated that stable singers did not gain increased reproductive success via any of three main fitness pathways: 1, stable singers were not mated to females laying larger clutches or produced more offspring with their social mate, 2, stable singers were not better at preventing their social mate from seeking extra-pair fertilizations, 3, stable singers were not better at acquiring extra-pair matings. These findings, therefore, imply that the stable production of minimum frequency is not used by females to choose mating partners or that it does not function as a quality-indicator trait in the first place. We also found no evidence for correlational or heterogeneous selection. Our results improve our understanding of the adaptive potential of song stability as they represent some of the first estimates of selection on this trait. Ultimately, further studies are needed to reveal the mechanisms that cause and maintain individual differences in song stability.

ACKNOWLEDGEMENTS

A.H. and N.J.D. were supported by the German Science Foundation (grant no. DI 1694/1-1). N.J.D. was supported by the Max Planck Society. H.S. was supported by the Institute of Biology (Leiden University). We thank all past members of the "Evolutionary Ecology of Variation" group at the Max Planck Institute for Ornithology, and the "Behavioural Ecology" group at the LMU, field assistants and students for help in data collection. We thank Martijn van der Sluijs and Christa Mateman for lab work related to microsatellite analyses. All procedures complied with guidelines from the District Government of Upper Bavaria (Regierung von Oberbayern) for Animal Care.

References

Araya-Ajoy YG, Dingemanse NJ, Kempenaers B. 2016. Timing of extrapair fertilizations: within-pair fertilization trade-offs or pair synchrony spillovers? Behav Ecol. 27(2):377–384. doi:10.1093/beheco/arv187.

Araya-Ajoy YG, Kuhn S, Mathot KJ, Mouchet A, Mutzel A, Nicolaus M, Wijmenga JJ, Kempenaers B, Dingemanse NJ. 2016. Sources of (co)variation in alternative siring routes available to male great tits (Parus major). Evolution (N Y). 70(10):2308–2321. doi:10.1111/evo.13024.

Arct A, Drobniak SM, Cichoń M. 2015. Genetic similarity between mates predicts extrapair paternity—a meta-analysis of bird studies. Behav Ecol. 26(4):959–968. doi:10.1093/beheco/arv004.

Auld JR, Agrawal AA, Relyea RA. 2010. Re-evaluating the costs and limits of adaptive phenotypic plasticity. Proc R Soc B Biol Sci. 277(1681):503–511. doi:10.1098/rspb.2009.1355.

Badyaev A V. 1997. Habitat Associations of Song Characteristics in Phylloscopus and Hippolais Warblers. Auk. 114(1):40–46. doi:10.2307/4089063.

Baker MC, Bjerke TK, Lampe H, Espmark Y. 1986. Sexual Response of Female Great Tits to Variation in Size of Males' Song Repertoires. Am Nat. 128(4):491–498. doi:10.1086/284582.

Bartsch C, Hultsch H, Scharff C, Kipper S. 2016. What is the whistle all about? A study on whistle songs, related male characteristics, and female song preferences in common nightingales. J Ornithol. 157(1):49–60. doi:10.1007/s10336-015-1245-y.

Bircher N, van Oers K, Hinde CA, Naguib M. 2020. Extraterritorial forays by great tits are associated with dawn song in unexpected ways. Griffin A, editor. Behav Ecol. 31(4):873–883. doi:10.1093/beheco/araa040.

Bircher N, van Oers K, Naguib M. 2021. Anyone listening? No evidence for eavesdropping on male singing interactions in the great tit, Parus major. Anim Behav. 176:67–76. doi:10.1016/j.anbehav.2021.03.009.

Biro PA, Adriaenssens B. 2013. Predictability as a personality trait: consistent differences in intraindividual behavioral variation. Am Nat. 182(5):621–9. doi:10.1086/673213.

Botero CA, Rossman RJ, Caro LM, Stenzler LM, Lovette IJ, de Kort SR, Vehrencamp SL. 2009. Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. Anim Behav. 77(3):701–706. doi:10.1016/j.anbehav.2008.11.020.

Both C. 1998. Experimental evidence for density dependence of reproduction in great tits. J Anim Ecol. 67(4):667–674. doi:10.1046/j.1365-2656.1998.00228.x.

Bouwhuis S, Sheldon BC, Verhulst S, Charmantier A. 2009. Great tits growing old: Selective disappearance and the partitioning of senescence to stages within the breeding cycle. Proc R Soc B Biol Sci. 276(1668):2769–2777. doi:10.1098/rspb.2009.0457.

Brembs B. 2011. Towards a scientific concept of free will as a biological trait: spontaneous actions and decision-making in invertebrates. Proc R Soc B Biol Sci. 278(1707):930–939. doi:10.1098/rspb.2010.2325.

Briffa M. 2013. Plastic proteans: reduced predictability in the face of predation risk in hermit crabs. Biol Lett. 9(5):20130592. doi:10.1098/rsbl.2013.0592.

Briffa M, Lane SM. 2017. The role of skill in animal contests: A neglected component of fighting ability. Proc R Soc B Biol Sci. 284(1863). doi:10.1098/rspb.2017.1596. [accessed 2021 May 28]. http://dx.doi.org/10.1098/rspb.2017.1596.

Brown TJ, Handford P. 2000. Sound Design for Vocalizations: Quality in the Woods, Consistency in the Fields. Condor. 102(1):81–92. doi:10.1093/condor/102.1.81.

Bürkner P-C. 2017. Advanced Bayesian Multilevel Modeling with the R Package brms. R J. 10(1):395–411.

Byers BE. 2007. Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. Behav Ecol. 18(1):130–136. doi:10.1093/beheco/arl058.

Byers BE, Akresh ME, King DI. 2015. A proxy of social mate choice in prairie warblers is correlated with consistent, rapid, low-pitched singing. Behav Ecol Sociobiol. 69(8):1275–1286. doi:10.1007/s00265-015-1940-y.

ten Cate C, Slabbekoorn H, Ballintijn MR. 2002. Birdsong and male—male competition: Causes and consequences of vocal variability in the collared dove (Streptopelia decaocto). Adv Study Behav. 31(C):31–75. doi:10.1016/S0065-3454(02)80005-5.

Clemente CJ, Wilson RS. 2015. Balancing Biomechanical Constraints: Optimal Escape Speeds When There Is a Trade-off between Speed and Maneuverability. Integr Comp Biol. 55(6):icv103. doi:10.1093/icb/icv103.

Cramer ERA. 2013. Measuring consistency: spectrogram cross-correlation versus targeted acoustic parameters. Bioacoustics. 22(3):247–257. doi:10.1080/09524622.2013.793616.

Cramer ERA, Hall ML, de Kort SR, Lovette IJ, Vehrencamp SL. 2011. Infrequent Extra-Pair Paternity in the Banded Wren, a Synchronously Breeding Tropical Passerine. Condor. 113(3):637–645. doi:10.1525/cond.2011.100233.

Derryberry EP. 2009. Ecology Shapes Birdsong Evolution: Variation in Morphology and Habitat Explains Variation in White-Crowned Sparrow Song. Am Nat. 174(1):24–33. doi:10.1086/599298.

Desai MM, Fisher DS. 2007. Beneficial Mutation–Selection Balance and the Effect of Linkage on Positive Selection. Genetics. 176(3):1759–1798. doi:10.1534/genetics.106.067678.

Dingemanse NJ, Kazem AJN, Réale D, Wright J. 2010. Behavioural reaction norms: animal personality meets individual plasticity. Trends Ecol Evol. 25(2):81–89. doi:10.1016/j.tree.2009.07.013.

Dingemanse NJ, Moiron M, Araya-Ajoy YG, Mouchet A, Abbey-Lee RN. 2020. Individual variation in age-dependent reproduction: Fast explorers live fast but senesce young? J Anim Ecol. 89(2):601–613. doi:10.1111/1365-2656.13122.

Dingemanse NJ, Réale D. 2013. What is the evidence for natural selection maintaining animal personality variation?. In: Animal personalities: Behavior, physiology, and evolution. University of Chicago Press. p. 201–220.

Domenici P, Booth D, Blagburn JM, Bacon JP. 2008. Cockroaches Keep Predators Guessing by Using Preferred Escape Trajectories. Curr Biol. 18(22):1792–1796. doi:10.1016/j.cub.2008.09.062.

Doutrelant C, Blondel J, Perret P, Lambrechts MM. 2000. Blue Tit song repertoire size, male quality and interspecific competition. J Avian Biol. 31(3):360–366. doi:10.1034/j.1600-048X.2000.310312.x. [accessed 2021 Jul 15]. http://doi.wiley.com/10.1034/j.1600-048X.2000.310312.x.

Forstmeier W, Kempenaers B, Meyer A, Leisler B. 2002. A novel song parameter correlates with extra-pair paternity and reflects male longevity. Proc R Soc London Ser B Biol Sci. 269(1499):1479–1485. doi:10.1098/rspb.2002.2039.

Garamszegi LZ. 2004. Extrapair paternity and the evolution of bird song. Behav Ecol. 15(3):508–519. doi:10.1093/beheco/arh041.

Garamszegi LZ, Møller AP. 2004. Extrapair paternity and the evolution of bird song. Behav Ecol. 15(3):508–519. doi:10.1093/beheco/arh041.

Gil D, Gahr M. 2002. The honesty of bird song: multiple constraints for multiple traits. Trends Ecol Evol. 17(3):133–141. doi:10.1016/S0169-5347(02)02410-2.

Griffith SC, Owens IPF, Thuman KA. 2008. Extra pair paternity in birds: a review of interspecific variation and adaptive function. Mol Ecol. 11(11):2195–2212. doi:10.1046/j.1365-294X.2002.01613.x.

Hadfield JD, Wilson AJ, Garant D, Sheldon BC, Kruuk LEB. 2010. The misuse of BLUP in ecology and evolution. Am Nat. 175(1):116–25. doi:10.1086/648604.

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. Proc Natl Acad Sci U S A. 108(35):14549–14554. doi:10.1073/pnas.1109091108.

Halfwerk W, Slabbekoorn H. 2009. A behavioural mechanism explaining noise-dependent frequency use in urban birdsong. Anim Behav. 78(6):1301–1307. doi:10.1016/j.anbehav.2009.09.015.

Hardouin LA, Bretagnolle V, Tabel P, Bavoux C, Burneleau G, Reby D. 2009. Acoustic cues to reproductive success in male owl hoots. Anim Behav. 78(4):907–913.

Hatchwell BJ. 1991. An Experimental Study of the Effects of Timing of Breeding on the Reproductive Success of Common Guillemots (Uria aalge). J Anim Ecol. 60(3):721. doi:10.2307/5410.

Hertel AG, Royauté R, Zedrosser A, Mueller T. 2021. Biologging reveals individual variation in behavioural predictability in the wild. Bijleveld A, editor. J Anim Ecol. 90(3):723–

737. doi:10.1111/1365-2656.13406.

Holveck MJ, Riebel K. 2010. Low-quality females prefer low-quality males when choosing a mate. Proc Biol Sci. 277(1678):153–160. doi:10.1098/rspb.2009.1222. [accessed 2021 Jun 17]. https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2009.1222.

Houslay TM, Wilson AJ. 2017. Avoiding the misuse of BLUP in behavioural ecology. Behav Ecol. 28(4):948–952. doi:10.1093/beheco/arx023.

Huet des Aunay G, Slabbekoorn H, Nagle L, Passas F, Nicolas P, Draganoiu TI. 2014. Urban noise undermines female sexual preferences for low-frequency songs in domestic canaries. Anim Behav. 87(C):67–75. doi:10.1016/j.anbehav.2013.10.010.

Iserbyt A, Eens M, Baetens W, Vermeulen A, Müller W. 2017. Within- and between-individual (co)variance partitioning reveals limited pleiotropic effects of testosterone on immune function, sexual signaling, and parental investment. Behav Ecol Sociobiol. 71(4):1–12. doi:10.1007/s00265-017-2308-2.

Jiang Y, Bolnick DI, Kirkpatrick M. 2013. Assortative mating in animals. Am Nat. 181(6). doi:10.1086/670160.

Jindrich DL, Qiao M. 2009. Maneuvers during legged locomotion. Chaos. 19(2):026105. doi:10.1063/1.3143031.

Jolles JW, Briggs HD, Araya-Ajoy YG, Boogert NJ. 2019. Personality, plasticity and predictability in sticklebacks: bold fish are less plastic and more predictable than shy fish. Anim Behav. 154:193–202. doi:10.1016/j.anbehav.2019.06.022.

Jones KA, Jackson AL, Ruxton GD. 2011. Prey jitters; protean behaviour in grouped prey. Behav Ecol. 22(4):831–836. doi:10.1093/beheco/arr062.

Kalinowski ST, Taper ML, Marshall TC. 2007. Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. Mol Ecol. 16(5):1099–1106. doi:10.1111/j.1365-294X.2007.03089.x.

De Kort SR, Eldermire ERB, Valderrama S, Botero CA, Vehrencamp SL. 2009. Trill consistency is an age-related assessment signal in banded wrens. Proc R Soc B Biol Sci. 276(1665):2315–2321. doi:10.1098/rspb.2009.0127.

Krokene C. 1998. The function of extrapair paternity in blue tits and great tits: good genes or fertility insurance? Behav Ecol. 9(6):649–656. doi:10.1093/beheco/9.6.649.

Kubli SP, MacDougall-Shackleton EA. 2014. Developmental timing of signals affects information content: song complexity but not consistency reflects innate immune strategy in male song sparrows. Am Nat. 183(5):660–70. doi:10.1086/675757.

Lambrechts MM, Dhondt AA. 1988. The anti-exhaustion hypothesis: a new hypothesis to explain song performance and song switching in the great tit. Anim Behav. 36:327–334.

Lane SM, Briffa M. 2020. Perceived and actual fighting ability: determinants of success by decision, knockout or submission in human combat sports. Biol Lett. 16(10):20200443. doi:10.1098/rsbl.2020.0443.

Lee Y, Nelder JA. 2006. Double hierarchical generalized linear models. J R Stat Soc Ser C Appl Stat. 55(2):139–185. doi:10.1111/j.1467-9876.2006.00538.x.

Macdougall-Shackleton SA. 1997. Sexual Selection and the Evolution of Song Repertoires. In: Current Ornithology. Springer US. p. 81–124.

Marques-Santos F, Dingemanse NJ. 2020. Weather effects on nestling survival of great tits vary according to the developmental stage. J Avian Biol. 51(10):jav.02421. doi:10.1111/jav.02421.

Marshall TC, Slate J, Kruuk LEB, Pemberton JM. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol. 7(5):639–655. doi:10.1046/j.1365-294x.1998.00374.x.

Mathot KJ, Wright J, Kempenaers B, Dingemanse NJ. 2012. Adaptive strategies for managing uncertainty may explain personality-related differences in behavioural plasticity. Oikos. 121(7):1009–1020. doi:10.1111/j.1600-0706.2012.20339.x.

McGregor PK, Krebs JR, Perrins CM. 1981. Song Repertoires and Lifetime Reproductive Success in the Great Tit (Parus major). Am Nat. 118(2):149–159. doi:10.1086/283811.

Mitchell DJ, Beckmann C, Biro PA. 2021. Understanding the unexplained: The magnitude and correlates of individual differences in residual variance. Ecol Evol.:ece3.7603. doi:10.1002/ece3.7603.

Mouchet A, Cole EF, Matthysen E, Nicolaus M, Quinn JL, Roth AM, Tinbergen JM, van Oers K, van Overveld T, Dingemanse NJ. 2021. Heterogeneous selection on exploration behavior within and among West European populations of a passerine bird. Proc Natl Acad Sci U S A. 118(28). doi:10.1073/pnas.2024994118.

Nakagawa S, Cuthill IC. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. Biol Rev. 82(4):591–605. doi:10.1111/j.1469-185X.2007.00027.x.

Nemeth E, Kempenaers B, Matessi G, Brumm H. 2012. Rock Sparrow Song Reflects Male Age and Reproductive Success. Bartell PA, editor. PLoS One. 7(8):e43259. doi:10.1371/journal.pone.0043259.

Nicolaus M, Mathot KJ, Araya-Ajoy YG, Mutzel A, Wijmenga JJ, Kempenaers B, Dingemanse NJ. 2015. Does coping style predict optimization? An experimental test in a wild passerine bird. Proc R Soc B Biol Sci. 282(1799):20142405–20142405. doi:10.1098/rspb.2014.2405.

Nicolaus M, Tinbergen JM, Ubels R, Both C, Dingemanse NJ. 2016. Density fluctuations represent a key process maintaining personality variation in a wild passerine bird. Grether G, editor. Ecol Lett. 19(4):478–486. doi:10.1111/ele.12584.

van Oers K, Drent PJ, Dingemanse NJ, Kempenaers B. 2008. Personality is associated with extrapair paternity in great tits, Parus major. Anim Behav. 76(3):555–563. doi:10.1016/j.anbehav.2008.03.011.

Piersma T, Drent J. 2003. Phenotypic flexibility and the evolution of organismal design.

Trends Ecol Evol. 18(5):228–233. doi:10.1016/S0169-5347(03)00036-3.

Poesel A, Dabelsteen T, Pedersen SB. 2004. Dawn song of male blue tits as a predictor of competitiveness in midmorning singing interactions. Acta Ethol. 6(2):65–71. doi:10.1007/s10211-004-0086-0.

Quinn JL, Patrick SC, Bouwhuis S, Wilkin TA, Sheldon BC. 2009. Heterogeneous selection on a heritable temperament trait in a variable environment. J Anim Ecol. 78(6):1203–1215. doi:10.1111/j.1365-2656.2009.01585.x.

Rivera-Gutierrez HF, Pinxten R, Eens M. 2010. Multiple signals for multiple messages: Great tit, Parus major, song signals age and survival. Anim Behav. 80(3):451–459. doi:10.1016/j.anbehav.2010.06.002.

Roeder D V., Husak MS, Murphy MT, Patten MA. 2019. Size, ornamentation, and flight feather morphology promote within-pair paternity in a sexually dimorphic passerine. Behav Ecol Sociobiol. 73(7):1–12. doi:10.1007/s00265-019-2704-x.

Rönnegård L, Felleki M, Fikse F, Mulder HA, Strandberg E. 2010. Genetic heterogeneity of residual variance - Estimation of variance components using double hierarchical generalized linear models. Genet Sel Evol. 42(1):1–10. doi:10.1186/1297-9686-42-8.

Sakata JT, Hampton CM, Brainard MS. 2008. Social Modulation of Sequence and Syllable Variability in Adult Birdsong. J Neurophysiol. 99(4):1700–1711. doi:10.1152/jn.01296.2007.

Sakata JT, Vehrencamp SL. 2012. Integrating perspectives on vocal performance and consistency. J Exp Biol. 215(2):201–209. doi:10.1242/jeb.056911.

Saladin V, Bonfils D, Binz T, Richner H. 2003. Isolation and characterization of 16 microsatellite loci in the Great Tit Parus major. Mol Ecol Notes. 3(4):520–522. doi:10.1046/j.1471-8286.2003.00498.x.

Santostefano F, Wilson AJ, Araya-Ajoy YG, Dingemanse NJ. 2016. Interacting with the enemy: indirect effects of personality on conspecific aggression in crickets. Behav Ecol. 27(4):1235–1246. doi:10.1093/beheco/arw037.

Schuett W, Tregenza T, Dall SRX. 2010. Sexual selection and animal personality. Biol Rev. 85(2):217–246. doi:10.1111/j.1469-185X.2009.00101.x.

Seutin G, White BN, Boag PT. 1991. Preservation of avian blood and tissue samples for DNA analyses. Can J Zool. 69(1):82–90. doi:10.1139/z91-013.

Sherman PW, Morton ML. 1988. Extra-pair fertilizations in mountain white-crowned sparrows. Behav Ecol Sociobiol. 22(6):413–420. doi:10.1007/BF00294979.

Snijders L, van Rooij EP, Henskens MFA, van Oers K, Naguib M. 2015. Dawn song predicts behaviour during territory conflicts in personality-typed great tits. Anim Behav. 109:45–52. doi:10.1016/j.anbehav.2015.07.037.

Soma M, Garamszegi LZ. 2011. Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. Behav Ecol. 22(2):363–371. doi:10.1093/beheco/arq219.

Stamps JA, Briffa M, Biro PA. 2012. Unpredictable animals: individual differences in intraindividual variability (IIV). Anim Behav. 83(6):1325–1334. doi:10.1016/J.ANBEHAV.2012.02.017.

Strohbach S, Curio E, Bathen A, Epplen J, Lubjuhn T. 1998. Extrapair paternity in the great tit (*Parus major*): a test of the “good genes” hypothesis. Behav Ecol. 9(4):388–396. doi:10.1093/beheco/9.4.388.

Suthers RA, Zollinger SA. 2004. Producing Song: The Vocal Apparatus. Ann N Y Acad Sci. 1016(1):109–129. doi:10.1196/annals.1298.041.

Taff CC, Steinberger D, Clark C, Belinsky K, Sacks H, Freeman-Gallant CR, Dunn PO, Whittingham LA. 2012. Multimodal sexual selection in a warbler: Plumage and song are related to different fitness components. Anim Behav. 84(4):813–821. doi:10.1016/j.anbehav.2012.07.002.

Taylor LR. 1961. Aggregation, variance and the mean. Nature. 189(4766):732–735. doi:10.1038/189732a0.

Viswanathan M. 2005. Measurement error and research design. Sage.

Webster MS, Pruett-Jones S, Westneat DF, Arnold SJ. 1995. Measuring the effects of pairing success, extra-pair copulations and mate quality on the opportunity for sexual selection. Evolution (N Y). 49(6):1147–1157. doi:10.1111/j.1558-5646.1995.tb04441.x.

Westneat DF, Stewart IRK. 2003. Extra-Pair Paternity in Birds: Causes, Correlates, and Conflict. Annu Rev Ecol Evol Syst. 34:365–396. doi:10.1146/annurev.ecolsys.34.011802.132439.

Westneat DF, Wright J, Dingemanse NJ. 2015. The biology hidden inside residual within-individual phenotypic variation. Biol Rev. 90(3):729–743. doi:10.1111/brv.12131.

Wilson AJ, Gelin U, Perron M-C, Réale D. 2009. Indirect genetic effects and the evolution of aggression in a vertebrate system. Proc R Soc B Biol Sci. 276(1656):533–541. doi:10.1098/rspb.2008.1193.

Zollinger SA, Slater PJB, Nemeth E, Brumm H. 2017. Higher songs of city birds may not be an individual response to noise. Proc R Soc B Biol Sci. 284(1860):20170602. doi:10.1098/rspb.2017.0602.

Electronic Supplementary Material of Chapter 4

Supplementary Material Table S1

Effects of minimum frequency, stability frequency and their interaction on lay date, clutch size, nest success and the number of fledglings shown as 95% credible interval around the estimated mean parameter estimate.

(a) Means	Lay Date N = 362	Clutch Size N = 362	Nest Success N = 363	Number Fledglings N = 321
Fixed effects	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Intercept[a]	7.80 (6.38, 9.21)	8.34 (7.98, 8.69)	2.18 (1.56, 2.79)	4.95 (4.56, 5.36)
Year 2018	13.02 (11.66, 14.41)	0.05 (-0.33, 0.42)	0.95(-0.05, 1.93)	2.23 (1.74, 2.71)
Year 2019	5.02 (3.48, 6.54)	-0.68 (-1.10, -0.29)	-0.96 (-1.73, -0.21)	-0.76 (-1.31, -0.21)
Minimum Frequency	9.21 (-0.64, 18.74)	-0.52 (-3.55, 2.62)	2.60 (-3.19, 7.96)	-0.11 (-3.16, 3.02)
Stability Frequency	-0.56 (-5.78, 4.61)	-0.62 (-2.27, 0.99)	-1.31 (-4.27, 2.06)	-0.46 (-2.27, 1.35)
Mean x Stability	7.12 (-67.64, 80.28)	18.09 (-3.00, 40.45)	2.98 (-43.75, 47.52)	15.95 (-7.28, 39.58)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	3.34 (2.71, 4.08)	0.44 (0.36, 0.53)	0 (0, 0)	0.21 (0.17, 0.26)
Plot-Year	0.44 (0.25, 0.69)	0.0098 (0.0055, 0.015)	0 (0, 0)	0.12 (0.07, 0.19)
Plot	4.36 (1.69, 8.30)	0.13 (0.05, 0.25)	0 (0, 0)	0.10 (0.04, 0.21)
Residual	23.12 (20.02, 26.85)	1.92 (1.66, 2.23)		2.30 (1.99, 2.69)

[a] Lay date, clutch size, nest success and number fledglings for birds breeding in 2017 with population level mean and RWV of minimum frequency

Supplementary Material Table S2

Effects of minimum frequency, stability frequency, breeding density and the two-way interactions between density and the two song measurements on lay date, clutch size, nest success and the number of fledglings shown as 95% credible interval around the estimated mean parameter estimate.

	Lay date	Clutch size	Nest success	Number fledglings
(a) Means	N = 362	N = 362	N = 363	N = 321
Fixed effects	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**	**β (95% CI)**
Intercept[a]	7.34 (5.99, 8.68)	8.33 (7.95, 8.70)	2.32 (1.63, 2.99)	4.93 (4.52, 5.38)
Year 2018	13.06 (11.59, 14.54)	0.07 (-0.31, 0.47)	0.94 (-0.07, 1.96)	2.24 (1.73, 2.75)
Year 2019	6.12 (4.15, 8.10)	-0.56 (-1.07, -0.03)	-1.27 (-2.24, -0.34)	-0.61 (-1.32, 0.08)
Breeding Density	-1.13 (-2.32, 0.02)	-0.12 (-0.44, 0.18)	0.26 (-0.17, 0.70)	-0.14 (-0.48, 0.21)
Minimum Frequency	-0.16 (-5.16, 4.79)	-0.18 (-1.76, 1.36)	-1.10 (-4.20, 1.82)	-0.33 (-1.95, 1.28)
Stability Frequency	9.10 (-1.17, 18.98)	-0.28 (-3.44, 2.82)	2.93 (-3.09, 8.75)	0.03 (-3.17, 3.41)
Mean x Density	8.39 (-3.09, 19.89)	-0.25 (-1.90, 1.40)	-0.94 (-4.29, 2.44)	-1.44 (-3.31, 0.47)
Stability x Density	-0.88 (-6.34, 4.53)	0.58 (-2.99, 4.10)	-0.91 (-7.65, 5.63)	-0.95 (-4.45, 2.63)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual	3.45 (2.75, 4.28)	0.46 (0.37, 0.56)	0 (0, 0)	0.17 (0.13, 0.21)
Plot-Year	0.67 (0.38, 1.04)	0.009 (0.005, 0.01)	0.09 (0.06, 0.15)	0.13 (0.08, 0.21)
Plot	2.35 (0.88, 4.48)	0.14 (0.05, 0.27)	0 (0, 0)	0.12 (0.04, 0.22)
Residual	22.97 (19.85, 26.57)	1.92 (1.64, 2.23)	3.29[b]	2.34 (2.00, 2.71)

[a] Lay date, clutch size, nest success and number fledglings for birds breeding in 2017 with population level mean and RWV of minimum frequency

[b] Residual variance fixed to $\pi^2/3$ for binary models

Supplementary Material Table S3

Effects of minimum frequency, stability frequency and their interaction on the probabilities to gain or lose at least one extra-pair offspring shown as 95% credible interval around the estimated mean parameter estimate.

(a) Means	Extra-Pair Gain (Y/N) N = 257	Extra-Pair Lost (Y/N) N = 257
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept[a]	-1.60 (-2.12, -1.11)	-0.61 (-1.04, -0.20)
Year 2018	0.44 (-0.19, 1.05)	0.05 (-0.47, 0.55)
Minimum Frequency	-1.14 (-5.99, 3.93)	-2.47 (-7.08, 1.98)
Stability Frequency	0.67 (-2.04, 3.55)	0.15 (-2.39, 2.65)
Mean x Stability	18.35 (-19.46, 54.38)	-1.83 (-36.56, 33.03)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0.20 (0.17, 0.25)	0.23 (0.19, 0.28)
Plot-Year	0 (0, 0)	0 (0, 0)
Plot	0 (0, 0)	0 (0, 0)
Residual	3.29[b]	3.29[b]

[a] Probability to sire at least one extra-pair offspring or lose at least one offspring for birds breeding in 2017 with population level mean and RWV of minimum frequency

[b] Residual variance fixed to $\pi^2/3$ for binary models

Supplementary Material Table S4

Effects of minimum frequency, stability frequency, breeding density and the two-way interactions between density and the two song measurements on the probabilities to gain or lose at least one extra-pair offspring shown as 95% credible interval around the estimated mean parameter estimate.

(a) Means	Extra-Pair Gain (Y/N) N = 257	Extra-Pair Lost (Y/N) N = 257
Fixed Effects	**β (95% CI)**	**β (95% CI)**
Intercept[a]	-1.33 (-1.84, -0.83)	-0.48 (-0.91, -0.07)
Year 2018	0.15 (-0.55, 0.85)	-0.13 (-0.71, 0.45)
Breeding Density	0.38 (-0.14, 0.90)	0.24 (-0.19, 0.68)
Minimum Frequency	-0.86 (-4.31, 2.65)	1.83 (-1.25, 4.90)
Stability Frequency	0.51 (-5.94, 6.68)	-3.28 (-9.10, 2.28)
Mean x Density	-2.36 (-7.31, 2.70)	3.28 (-1.10, 7.60)
Stability x Density	2.25 (-7.06, 12.04)	-4.36 (-12.50, 4.58)
Random Effects	**σ^2 (95% CI)**	**σ^2 (95% CI)**
Individual ID	0 (0, 0)	0 (0, 0)
Plot-Year	0 (0, 0)	0 (0, 0)
Plot	0 (0, 0)	0 (0, 0)
Residual	3.29[b]	3.29[b]

[a] Probability to sire at least one extra-pair offspring or lose at least one offspring for birds breeding in 2017 with population level mean and RWV of minimum frequency

[b] Residual variance fixed to $\pi^2/3$ for binary models

Acknowledgments

Throughout the over four years of my PhD, I have had the pleasure to meet and work with so many wonderful people. Some might not have contributed to this thesis scientifically, but all of them have supported me in their own ways, even if was only with a little piece of advice or a friendly conversation at a conference. In the end, this PhD has allowed me to meet all of them and they have made this period of my life enjoyable and a lot of fun.

Most and foremost, I need to thank my supervisor Niels Dingemanse. Niels, thank you for offering me this PhD position, for the trust you put in me, the scientific and personal freedom to solve tasks and take decisions independently, but also the constructive feedback and guidance you gave when necessary. I am well aware that it is not natural to have a supervisor that is always available when needed, even if it is only for a small chat during coffee. Thank you again for the patience, the support and the help in growing as a scientist and as a person.

I also want to thank Hans Slabbekoorn and Eira Bermúdez-Cuamatzin, for sharing your knowledge of bird song and how to measure it and providing repeated feedback and input. Without you, I would not have been able to produce this thesis. I really enjoyed our exchanges, my visits to the Netherlands and the things I learned from you. Especially your different perspectives on aspects like scientific approaches and writing gave me new experiences and expanded my skill set and knowledge beyond my working group.

Speaking of the working group, I need to thank all past and current members of the Behavioural Ecology group at the LMU Munich. It was an exceptional experience to work with such excellent scientists and to pick their brains for conceptual, statistical and personal advice when needed; advice they were always happy to give. Beyond the work, all members of this group have created the best working atmosphere one can hope for. I enjoyed every lunch break, every social event, and every conference we visited together.

A special thank you needs to go to Alexia Mouchet for being the best co-worker I could have asked for. Always motivated, always engaged, always organized, you made this project work. You were the not-so-secret leader of the project, managing the fieldwork, the students, and the data collection; basically everything. Your hard work and experience always guaranteed the successful and smooth execution of our field work and data collection, making my PhD life so much easier. And even beyond the breeding season, you were always there to provide help and advice when needed.

I also need to thank all the students who have contributed to this thesis, directly or indirectly. I have had the pleasure to meet and supervise many enthusiastic and excellent young researchers. Some of you joined the field seasons, always leading to great teams, some of you extracted the acoustic measurements; a tedious task so essential to this thesis and one you all performed so reliably. Some of you even became co-authors in the chapters of this thesis. And some of you have already started your own scientific careers. I like to imagine that I contributed at least to a tiny degree to your personal growth and success and hope you enjoyed the shared time as much as I did.

Finally, I need to thank my family and my girlfriend for supporting me without questions and without conditions. To my parents, who never doubted my decisions and the path I have chosen. Thank you for all the support and unconditional love. Without this, I would have never been able to successfully finish this PhD. And to Valérie, whom I met in the last and certainly hardest year of this thesis. You shared the hardship of these exceptional years, 2020 and 2021, with me and made it less unbearable.

Author contributions

Publication 1 Great tits responding to territorial intrusions sing less but alarm more on colder days

AFTS, AH and NJD conceived and designed the study, AFTS and AH collected and analysed the data, AFTS and AH wrote the paper with input from NJD

Publication 2 Does song overlap signal aggressiveness? An experimental study with repeated measures in free-ranging great tits

AH, VAR, EBC, HS and NJD conceived and designed the study, AH, VAR, SS and LS collected and analysed the data, AH wrote the paper with input from EBC, HS and NJD.

Manuscript 1 Within-individual stability in bird song shows cross-year repeatability and is affected by food supplementation in a wild passerine bird

AH, HS and NJD conceived and designed the study. EBC, MB, HS and NJD gave conceptual and methodological advice and help on the collection, extraction and analysis of the data. AH and AM collected the data, AH analysed the data. AH wrote the manuscript with input from all co-authors.

Manuscript 2 Song stability neither predicts within-pair reproductive success nor extra-pair paternity gain

AH and NJD conceived and designed the study. HS, KVO and NJD gave conceptual and methodological advice on the collection, extraction and analysis of the data. AH collected the data in the field and took acoustic measurements, KVO analysed paternity data. AH performed the statistical analysis. AH wrote the manuscript with input from all co-authors.

Niels J. Dingemanse
Doktorvater

Alexander Hutfluss
Doktorant

Addresses of co-authors

Eira Bermúdez-Cuamatzin
Behavioural Biology, Institute of Biology
Leiden University,
Leiden, The Netherlands
Email: eira.bc@gmail.com

Mark Briffa
School of Biological and Marine Sciences, Animal Behaviour Research Group
University of Plymouth,
Plymouth, United Kingdom
Email: mark.briffa@plymouth.ac.uk

Niels Dingemanse
Behavioural Ecology, Department Biology II
Ludwig-Maximilians University of Munich,
Planegg-Martinsried, Germany
Email: dingemanse@biologie.uni-muenchen.de

Alexia Mouchet
Behavioural Ecology, Department Biology II
Ludwig-Maximilians University of Munich,
Planegg-Martinsried, Germany
Email: mouchet@biologie.uni-muenchen.de

Kees van Oers
Department of Animal Ecology,
Netherlands Institute of Ecology (NIOO-KNAW),
Wageningen, The Netherlands
Email: k.vanoers@nioo.knaw.nl

Veronika Rohr
Forschungsgruppe Küpper,
Max-Placnk Institute for Ornithology,
Seewiesen, Germany
Email: Vrohr@orn.mpg.de

Saray Scheidt
Behavioural Ecology, Department Biology II
Ludwig-Maximilians University of Munich,
Planegg-Martinsried, Germany
Email: saray.scheidt@hotmail.com

Hans Slabbekoorn
Behavioural Biology, Institute of Biology
Leiden University,
Leiden, The Netherlands
Email: h.w.slabbekoorn@biology.leidenuniv.nl

Linda Steinbichl
Behavioural Ecology, Department Biology II
Ludwig-Maximilians University of Munich,
Planegg-Martinsried, Germany
Email: linda.steinbichl@kabelmail.de

Aurelia Strauß
Faculty of Science and Engineering,
GELIFES, Groningen Institute for Evolutionary Life Sciences
Groningen, The Netherlands
Email: a.strauss@rug.nl