Evolutionary Interactions Between Sexual Selection, Song Elaboration, and Song Learning: A Combined Empirical and Theoretical Approach

By

Cristina M. Robinson

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in
Biological Sciences

October 31, 2019

Nashville, Tennessee

Approved:
Nicole Creanza, Ph.D.
Kathy Friedman, Ph.D.
Tony Capra, Ph.D.
Ann Tate, Ph.D.
Reyna Gordon, Ph.D.

ACKNOWLEDGMENTS

Thank you to all my lab mates — current and former — for discussions about everything and nothing and for acting as a willing audience for my crafting parties. Nicole Diggins, Begüm Erdoğan, and Léolène Jean Carrington got me through the first half of my dissertation. They were there for editing papers, scientific discussion, and moral support when things got rough. Extra thanks to Nicole, who helped me "accidently" adopt my kitten-y excuse to go home at night, Dusky Ghost. When my first advisor passed away, Nicole Creanza was willing to take a gamble on me as an older graduate student, who was entirely switching fields and would not be eligible for many grants. I am immensely grateful for this and hope I have repaid her in publications and by acting as a resource for the younger students in the lab. Thank you, Abigail Searfoss, for your figure-focused eyes and mind (even if you did sniff my lunches to figure out what they were and punched me regularly). Thank you, Parker Rundstrom, for website coding assistance (as well as snacks and the out of control jungle in the lab). Thank you, Luísa Jabbur, for your extravagant, emotional support cakes (even if they were technically for PI Nicole's Birthday). Finally, thank you to my poor fiancé, Leon, who stuck by me through all the bad days (no matter how much I yelled at the other drivers).

TABLE OF CONTENTS

Page

Chapter

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Song elaboration and sexual selection

Song is a learned vocalization in oscine songbirds and a key aspect of their behavior[1]. At its simplest, song is used for species recognition[2,3], as the songs of different species are distinct from one another; however, it is also used for communication between and within sexes[4] (**Figure 1.1A**). For example, males use song to announce and defend territory boundaries to other males[5,6]. In some species, males engage in song-matching competitions to show aggression or establish dominance[4]. Males also use song to attract females, and males that do not sing a species-typical song are unlikely to find a mate[1]. While this dissertation focuses on male singing behavior, females duet with males in some species[7], and the field has become increasingly aware that in many species where it was assumed that only males sing, females also sing, if only rarely[7,8]. Overall, song is critical for bird communication and strongly linked to reproductive success. Therefore, studying song may provide key insights into songbird ecology, learned behaviors, sexual selection, and evolutionary processes more broadly.

What benefit do females gain from listening to male song? Song has been proposed to act as an honest signal of male fitness[10]. Indeed, reduced developmental stress[11,12], reduced parasite loads[13,14], and greater age[15–17] have been linked to higher song quality. Thus, by choosing males with superior song, females can select for healthier, more experienced males. Indeed, it has been shown that females prefer males that display higher quality song[18,19]. Song quality is often broken into two categories: Song performance includes traits such as song duration, trill rate, and maximum frequency. Song complexity or elaboration focuses on song or syllable repertoire size, where syllables are distinct sound units within songs.

Different species have markedly different syllable repertoire sizes, where syllables are unique sound units in birdsong (**Figure 1.1A-C**). For example, chipping sparrows (*Spizella passerina*)

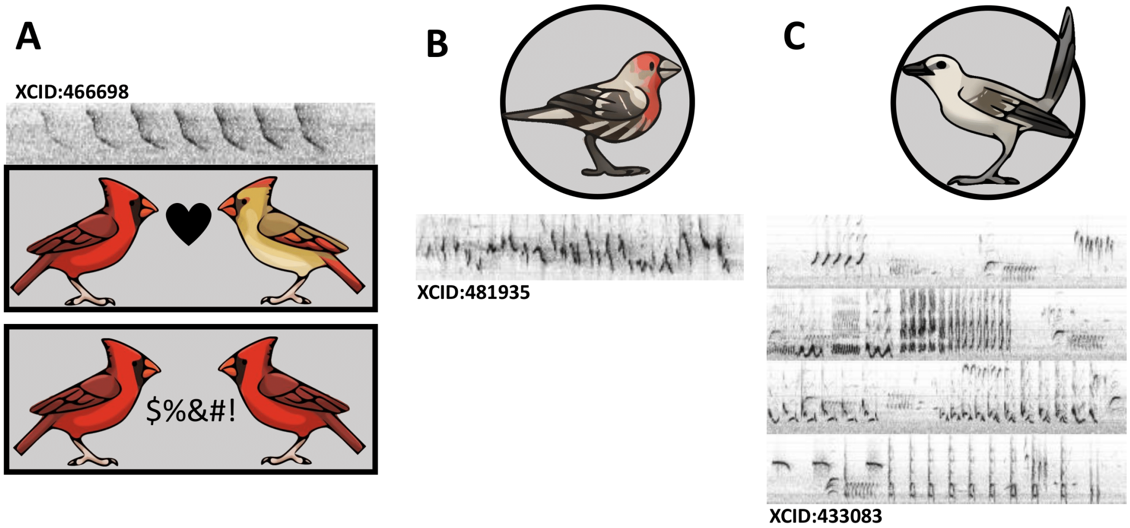


Figure 1.1: Purposes of song and between species differences. (A) This sonogram shows an example song of the northern cardinal (*Cardinalis cardinalis*, recorded by Brian Hendrix, accessible at https://www.xeno-canto.org/466698). Each dark trace in the sonogram is a visually representation of a syllable — a distinct sound unit in the song. Songs are used to attract females and engaged in competitions with other males. (B) Example song of the house finch (*Haemorhous mexicanus*, recorded by Paul Marvin, accessible at https://www.xeno-canto.org/481935). (C) Example song of the northern mockingbird (*Mimus polyglottos*, recorded by Phil Brown, accessible at https://www.xeno-canto.org/433083). (A-C) show the remarkable variability in syllable repertoire size between species, where cardinals only know a handful of syllables, house finches know on average 40 syllables and mockingbirds know hundreds of syllables[9].

have a repertoire of one syllable[20], while nightingales *Luscinia megarhynchos* have a repertoire of over one thousand syllables[21]. If the one-syllable chipping sparrow song is sufficient for species recognition, territory defense, and mate attraction, why do most songbirds have substantially larger syllable repertoires[22–24]? Song has long been considered a sexually selected trait[23,25,26]. If females prefer more elaborate song, and if song elaboration is heritable, then the population should shift to larger syllable and/or song repertoire distributions over evolutionary time. As stated above, female preferences for larger repertoires have been demonstrated in some species, but song is a cultural trait that is vertically inherited from a father to his chicks in only some species[27,28]. In most species, chicks learn obliquely from unrelated males[29]. This would seem to disrupt the heritability of high-quality song. However, females can develop song preferences based on their father's song that they later use to inform mate choice[30,31]. Males may also develop song preferences from their father that affect tutor choice, although this possibility has not yet has not been tested. Furthermore, previous research has shown that young males attend to social interactions between adults and are more likely to learn from males that are successful in these interactions[32–34]. Thus, even if chicks learn only obliquely, superior song is transmitted across generations, because chicks learn from socially successful males.

Although it seems intuitive that sexual selection for larger repertoire sizes would explain elaborate repertoires, the evidence for sexual selection on song elaboration has been mixed. Many laboratory studies have shown that females display more copulation responses when presented with more elaborate songs — even in species with very small repertoires — but these preferences are not always acted upon in the wild[18]. Indeed, it seems likely that song performance[35–38], plumage[39–41], lekking[42,43], and dance displays[44] may be more salient cues for male fitness in some species. Thus, it is not clear which species show strong correlations between individual male song elaboration and reproductive success. It is also unknown whether birds that do show a strong correlation share life history traits with one another that other songbird species lack.

1.2 Song learning

One might hypothesize that a behavior as important for reproduction as song would be innate. Indeed, the song of suboscines — the outgroup clade to the vocal-learning songbirds — is largely or entirely innate and does not require social learning to produce[4,45,46]. There are some putative, innate characteristics of oscine (songbird) songs[47–49]; however, these species must also listen to adult males (tutors) to develop a species-typical song[50,51]. Interestingly, no known lineages of songbirds have lost the ability to learn song. Even though all songbirds learn their song, different species learn for varying lengths of time. The length of the song-learning window varies continuously, but birds are typically broken into two groups[4,52,53]. One group, "age-limited" or "closed-ended" learners, have a set developmental window within which they must hear, practice, and crystallize their adult song. Crystallization occurs during or before the onset of sexual maturity and is correlated with plumage maturation in many species[54]. The second group, "open-ended" learners, can continually or seasonally add, subtract, and/or alter syllables or songs in their repertoire past sexual maturity.

Although these definitions are straightforward, assessing whether a species exhibits adult learning is far more complex. For example, in the field, if a male sings a syllable in his second year that he did not sing in his first year, it cannot be determined whether this "new" syllable was the result of adult learning, or if the male learned it previously as a chick or fledgling and is only now expressing that syllable[55,56]. In the lab, an experimenter can control what sounds a male hears for his entire life, but assessing the length of the song-learning window would require years-long experiments[57–60]. Furthermore, birds may behave differently in the lab than they do in the wild[61,62]. A reasonable proxy for whether birds learn in adulthood that is more easily studied in the field is song stability — whether individuals of a species are observed to modify their syllable repertoires as adults. Open-ended learners are likely to modify their adult syllable repertoires over time, while closed-ended learners would likely have adult songs that are stable from year to year. Therefore, I define species that cease learning before the end of the first breeding season as "song-stable," while species that learn after this time period are deemed "song-plastic."

It is widely believed that longer learning would provide a reproductive benefit, because males with inferior songs as young adults could learn superior songs as they age[1,4]. While little research has sought to uncover whether the evolution of the song-learning window is correlated with the evolution of song traits, it was shown using a small subset of songbird species that longer learning is associated with larger species-level average syllable repertoires[16,17,63]. However, only a limited number of oscine species exhibit song-plasticity in adulthood. To explain this discrepancy between the apparent utility of longer learning and its scarcity in real species, it has been postulated that maintaining the neural plasticity for song learning in adulthood may be metabolically costly and/or that spending longer periods of time learning and practicing song as adults would reduce the time

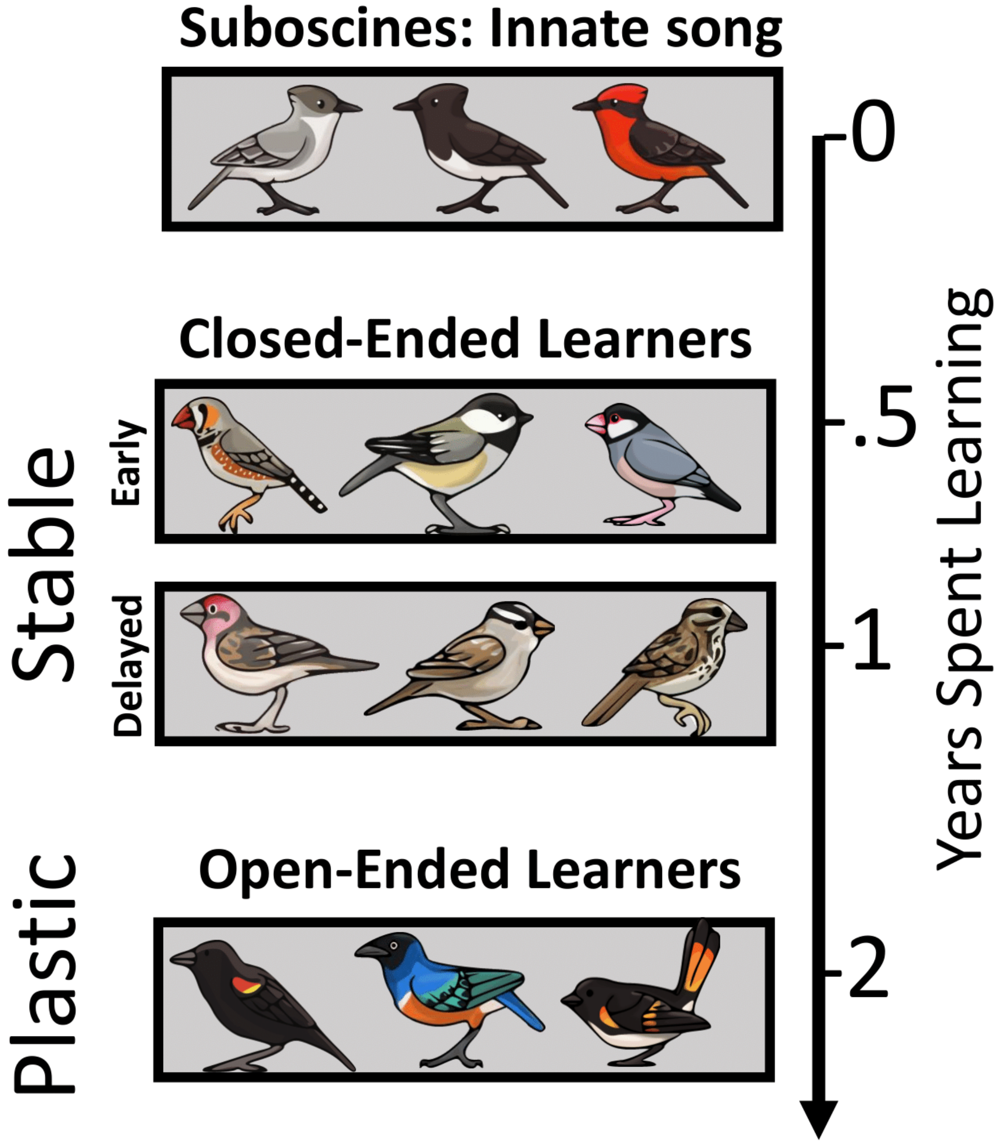


Figure 1.2: Song learning in passerine species. Suboscine species hatch knowing what adult song should sound like and can produce a species typical song without listening to adult male tutors. In contrast, oscine songbird species must listen to adult male tutors to produce a species typical song. Songbirds that learn for a short period of time and then crystallize their adult song and never change it are called “closed-ended” learners. I divide closed-ended learners into two groups: 1) early closed-ended learners cease learning before their first breeding season, whereas 2) delayed closed ended learners learn through their first breeding season. There are also the “open-ended” learners, which are song birds that can learn and modify their songs as sexually mature adults. Because the length of the song-learning window is difficult to quantify, I used measurements of song stability or plasticity as a proxy for the length of the song learning window. Species whose adult songs consist of the same syllables from year to year have song stability, and are presumed to be closed-ended learners. Species have adult song plasticity when their adult syllable repertoire can vary from year to year. These species are presumed to be open-ended learners.

remaining for males to engage in other tasks, like foraging[4,64]. Previous research using computational models has examined the conditions under which learning arises in species that initially have innate songs (analogous to the divergence of oscines from suboscines)[65] or how song learning affects speciation[66–68] and between-species interactions[69]. However, little has been done to examine what selective pressures modulate the length of the song-learning window.

1.3 Purpose of this dissertation

The overall goal of this dissertation is to provide new understanding about the evolutionary interactions between female preferences, song elaboration, and song-learning behavior. In **Chapter 2**, I provide a new solution to the long-standing controversy over whether song elaboration is under sexual selection in songbirds. I examined whether species-level average syllable repertoire or the length of the song-learning window could be used to predict the strength of the correlation between individual male song elaboration and reproductive success. I found that species with larger average syllable repertoires showed stronger correlations between individual male song elaboration and reproductive success. Despite the expectation that song-plastic species (presumed open-ended learners) would show stronger correlations than song-stable species (presumed closed-ended learners), I did not find strong evidence for this possibility. Overall, my results suggest that when females prefer more elaborate songs, this creates a selection pressure that leads to the evolution of larger average syllable repertoire sizes. Furthermore, this pressure persists even after a species has evolved a large average syllable repertoire. This chapter is adapted from my article, *Species-level repertoire size predicts a correlation between individual song elaboration and reproductive success*, which was published in *Ecology and Evolution*[9].

In **Chapter 3**, I tested whether there was correlated evolution between the length of the song-learning window and seven song traits (syllable repertoire size, song repertoire size, syllables per song, song duration, intersong interval, song rate, and song continuity). If such correlated evolution exists, this would strengthen the hypothesis that sexual selection on song features can indirectly drive the evolution of the song-learning window. Song-plastic species had larger syllable and song repertoires than song-stable species, but they were not significantly different for the other five song traits. Nonetheless, intersong interval, song duration, and song rate evolved faster in song-plastic species. I initially saw no difference in the rate of evolution for song or syllable repertoire size between song-stable and song-plastic species. However, when I divided the song-stable group into early song-stable (song crystallization before the first breeding season) and delayed song-stable (song crystallization after the onset of the first breeding season and prior to the end of that breeding season), I found that delayed song-stable species and song-plastic species evolved their song repertoire sizes faster than early song-stable species. Together, this suggests that longer learning allows song elaboration to evolve faster, but it only evolves directionally to larger repertoire sizes when birds continue to learn beyond their first breeding season. While other song traits evolve more quickly in song-plastic species, they do not evolve directionally. Potentially, these features evolve faster in song-plastic species because song structure or performance must change to allow for larger syllable and song repertoire sizes, but there are multiple ways to achieve this. Finally, I found that larger repertoires in song-stable species and smaller repertoires in song-plastic species were evolutionarily unstable; species in these states were predicted to quickly change either their song stability state or average species repertoire size. This chapter is adapted from my co-authored article with Kate T. Snyder, *Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning*, which was published in *eLife*[70].

In **Chapter 4**, I introduce an agent-based model that I designed to study the evolution of syllable repertoire size and the duration of song learning in the presence of female preferences for either 1) larger syllable repertoires, 2) songs that match a female template, or 3) a mix of both. I also allow males to use one of three primary learning strategies: 1) Add, where males attempt to learn all syllables they hear from a tutor. 2) Add/Forget, where males attempt to learn all syllables they hear from a tutor but also attempt to forget any syllables they know but did not hear from a tutor. 3) Conformity, where a male samples multiple tutors and attempts to learn new syllables based on conformity bias and attempts to forget any syllables that no tutor sang. I found that the Add strategy was the most valuable learning strategy (males evolved the longest learning windows) when females preferred larger repertoires. The Conformity strategy was the most valuable when females preferred songs that matched a template (males evolved the highest template-matching accuracy). Interestingly, the Conformity strategy also seemed to be the best at balancing mixed female preferences, because it was the only learning strategy that allowed males to evolve somewhat larger syllable repertoires without losing the main syllables that the females preferred. I also examined whether shorter song-learning windows could evolve in the absence of a fitness cost on longer learning. I was surprised to discover that the hypothesized survival cost on extended leaning was not necessary; apparently, certain combinations of female preferences and learning strategies led to a reduction in song quality the longer males learned, so extended learning was lost from the population. While it still seems likely that extended learning imposes some metabolic and/or resource cost, I have shown that such a cost is not required to evolve shorter learning windows, because there are cases where longer learning is itself detrimental. This chapter is adapted from my article *Modeling the evolutionary interactions of sexual selection and learning strategies on the duration of song-learning in birds*, which is in revision at *PLOS Computational Biology*[71].

In **Chapters 5** and **6**, I briefly explore additional unpublished results I generated using my computational model for two additional lines of inquiry. First, I examine the factors that allow the formation and maintenance of regional song dialects in birds (**Chapter 5**). I found that socially informed tutor choice was critical for dialect formation and that dialects can be formed and maintained via two mechanisms. Second, I examine how the interactions between a wider range of female preferences and male song learning strategies affect song evolution (**Chapter 6**). I found that direct selection on song was required to evolve large repertoire sizes; I never saw cultural hitchhiking occur when females copied one another's mate choices. Furthermore, some learning strategies always led to smaller repertoires regardless of female preferences. Finally, in the absence of vertical learning, conformity learning allowed males to maintain common initial syllables in the population to the end of the simulation, even when females preferred larger repertoires.

In **Chapter 7**, I discuss the overall limitations and caveats of my body of work and synthesize all of my empirical and theoretical findings to provide a coherent set of new hypotheses that can be tested in the field. Finally, I outline a series of necessary ecology research programs to move my work forward, and suggest features that could be added to my computational model to allow it to examine additional questions.

Chapter 2

Species-level repertoire size predicts a correlation between individual song elaboration and reproductive success [1]

2.1 Introduction

High-quality song has been associated with genetic, environmental, and cultural factors[11–13,72–74]. While song is culturally transmitted in songbirds, a high-quality song could be associated with a better genetic background, increased learning ability, a higher quality tutor, a less stressful environment, or some combination, all of which could correlate with fitness. Thus, individual song elaboration, an aspect of song quality, has long been hypothesized to be important in sexual selection[15,23,26,75–83]. Song elaboration could be indicative of males that can achieve greater reproductive success (1) by having greater reproductive output and leaving more offspring, and/or (2) by having increased mating success if females prefer males with larger syllable or song repertoires. For example, repertoire size has been linked to paternal effort in feeding nestlings[84] and the number of eggs laid by a female[83], both of which would potentially increase reproductive output. In addition, laboratory studies in numerous species have suggested that female songbirds tend to prefer more elaborate songs, which would link song and syllable repertoire size to mate choice[80,81,85–89].

However, cross-species analysis of the literature seeking to correlate individual song elaboration with reproductive success has not revealed a strong relationship between the two. One review[18] found that while females show increased copulation responses to recordings of larger repertoires in many laboratory experiments (~80% of studies), field studies were much less likely to find a relationship between song elaboration and mate choice (~35% of studies). A subsequent quantitative meta-analysis consisting of only field studies[19] found a significant effect of song elaboration on reproductive success — as measured by both mate choice and reproductive output.

---

[1] This chapter is adapted from my article *Species-level repertoire size predicts a correlation between individual song elaboration and reproductive success*, which was published in *Ecology and Evolution*[9].

However, the association was weak, with the average effect size (r) ranging between 0.1 and 0.3, depending on how the data were incorporated into the meta-analysis and whether publication bias was controlled for in the final dataset. Thus, there was a marked difference between the results of these analyses[18,19] and the longstanding expectation that song elaboration plays a prominent role in sexual selection[15,23,26,75–83].

To explain this discrepancy, it has been proposed that other song traits, such as performance and stereotypy, or other factors, including territory quality and plumage, may be more influential in female choice than song elaboration in some bird species[35,90–96]. Indeed, both of the aforementioned analyses[18,19] also proposed that song elaboration may not be under universal selective pressure across species. If song elaboration is not under universal selection pressure, it raises a question: is there a subset of bird species for which song elaboration correlates with reproductive success, and do those species have anything in common?

Here, we propose two species-level traits that potentially help predict the strength of the correlation between individual song elaboration and reproductive success. First, we examine an aspect of song — species-level average syllable repertoire size — which may be larger in species undergoing sexual selection. This would be comparable to tail length in birds, which represents a particularly well-studied case of sexual selection. Numerous studies have linked within-species tail length variation to fitness: individuals with longer tails generally have greater reproductive success. However, there is a between-species component to these observations; tail length has primarily been correlated with mating success in species that have elongated tails, whereas tail length has not been correlated to reproductive success in species with shorter tails[97–102]. Thus, it is likely that a species with sexual selection for tail length would, on average, have longer tails than a species that does not (**Figure 2.1**). This may also be the case for syllable repertoire size; if males with greater individual song elaboration (within-species variation) are more reproductively successful and these males tend to produce offspring with larger-than-average repertoires, then the average syllable repertoire size of the entire species could gradually increase under this sexual selection pressure (between-species variation) (**Figure 2.1**). In turn, larger average species syllable repertoires would be a signature of this form of sexual selection. An alternative explanation that relies primarily on reproductive output could hinge on the difficulty of learning the species syllable repertoire. It may be that fairly small average species syllable repertoires can be learned even by less fit males, so individual song elaboration would only correlate with reproductive output in species where the average syllable repertoire size is large enough that less fit males cannot learn the full repertoire.

Alternatively, birds with small average species syllable repertoires might attend more closely to song elaboration, because it is easier to discern which males have more elaborate song when each male only produces a handful of syllables. In birds with larger average species syllable repertoires, it would take more listening time and be more difficult for a female to discern which potential mate shows greater song elaboration[5,103]. Thus, small increases in individual song elaboration may be more meaningful in birds with smaller average species syllable repertoires, where such differences can be quickly perceived. If this correlation is not driven by female preferences, it is more difficult to postulate a reason why a link between individual song elaboration and reproductive output would exist only in species with smaller average species syllable repertoires.

We also propose a second hypothesis: that the association between individual song elaboration and reproductive success could differ between species based on the length of the song-learning window. The length of the song-learning window varies greatly between species, but most can be roughly divided into two categories: (1) “age-limited” or “closed-ended” learners, which learn their song within a developmental window, and (2) “open-ended” learners, which can modify their songs past sexual maturity[4,52,53]. Due to the difficulty in measuring the true length of the song-learning window, we use the stability of a species’ adult songs over time as a proxy for it. Species that modify their adult songs overtime (song-plastic species) are likely open-ended learners, while those that do not (song-stable species) are likely closed-ended learners. Of note, some song-plastic species are known to increase the overall size of their repertoires as they age[15]. This means that in open-ended species, song elaboration can potentially indicate the age of a male in addition to his song-learning capacity (**Figure 2.1**). Furthermore, in some species, older males are preferred

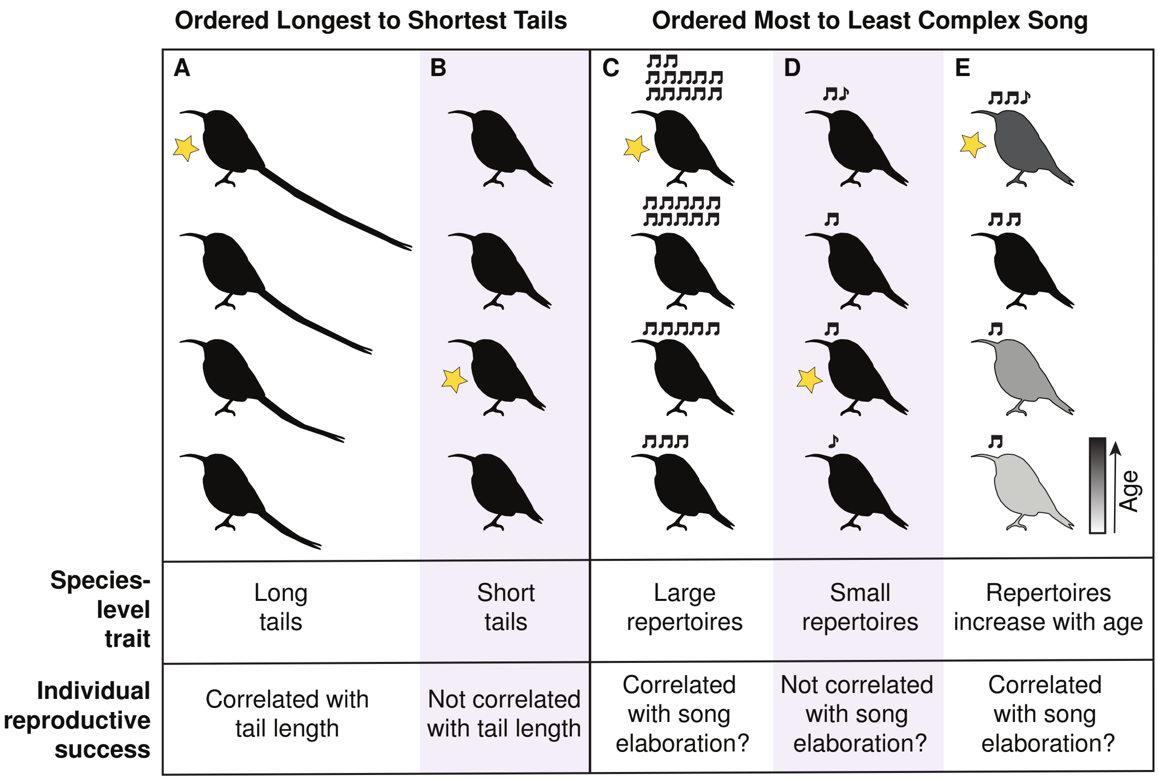


Figure 2.1: Schematic of traits that may predict the presence of sexual selection. (A-B) Males are ranked by tail length (longest tails at the top). (C-E) Males are ranked by song elaboration (most elaborate songs at the top). In all cases, the male with the highest reproductive success is marked by a star. If males with more exaggerated tails are more reproductively successful, due to female preference or genetic superiority, and if tail length is heritable, it is expected that the species distribution will shift toward more exaggerated phenotypes. In this case, one would predict that sexual selection for exaggerated tails is more likely occurring in species with longer average tail lengths (A) than in species with shorter average tail lengths (B). In this same line of thinking, if males with more elaborate songs are more reproductively successful and song elaboration is heritable, then one would predict that sexual selection for more elaborate song is more likely occurring in species with a large average syllable repertoire (C) than in species with a small average syllable repertoire (D). Alternatively, if males can learn more syllables as they age (open-ended learners/song-plastic species), syllable repertoire size could act as a signal for male age (E). Females may prefer older males, because they have more breeding experience than their younger counterparts and have proven their survival capability. Thus, females would prefer mates with larger repertoires if repertoire size correlates with age. In contrast, in closed-ended learners (song-stable species), where song cannot signal age, females would not prefer more elaborate songs.

over younger males[104,105], and, more generally, longevity might be a signal of high genetic quality or greater foraging experience[106–109] (however, see[110]). Therefore, we hypothesize that females from species with plastic adult songs would prefer greater song elaboration more than song-stable species, in which song elaboration cannot signal age.

In this study, we attempt to determine whether species average syllable repertoire size or song stability can predict the strength of the correlation between individual song elaboration and reproductive success. Using a Bayesian multi-level phylogenetic meta-analysis of available field data, we observed that larger species average syllable repertoires predict for a stronger correlation between individual song elaboration and reproductive success, whereas song stability versus plasticity did not predict the strength of this correlation.

2.2 Methods

2.2.1 Data and code availability

https://github.com/CreanzaLab/RepertoireSizeReproductiveSuccess

2.2.2 Data collection

We compiled the list of references in three stages (Workflow in **Figure 2.2**): Stage 1) we obtained field studies that examined the link between individual song elaboration (number of songs or syllables) and reproductive success (reproductive output or mating success) from the references included in the reviews by Byers and Kroodsma[18] and Soma and Garamszegi[19]. Additionally, we searched for relevant studies published since these reviews using the terms “bird” and “song complexity,” “song versatility”, or “repertoire” in combination with “mating success”, “reproductive success,” or “mate choice” in Google Scholar, Web of Science, and ProQuest Dissertation and Theses Global database, which yielded 11 more field studies and 1 thesis with data that was unpublished at the time of data acquisition. This led to a total of 57 studies and 1 thesis. However, 10 studies were discarded, because they correlated reproductive success with aspects of song other than elaboration as we defined it. In the studies that remained, individual song elaboration was measured by either song repertoire size (unique number of songs per individual) or syllable repertoire size (unique number of syllables per individual). We included studies that measured the association between reproductive success and either of these song elaboration metrics, because syllable repertoire size and song repertoire size are correlated between species (**Figure 2.3**) and are likely also correlated within species. For information regarding which studies correlated what form of song elaboration with reproductive success in which species, see columns 1-3 of **STable 1: All Measurements** (available on GitHub).

Stage 2) We performed a literature search using Pubmed, Web of Science, and Google Scholar to gather information on the average syllable repertoire size for each species identified in Stage 1. Species syllable repertoire was defined as the average number of distinct syllables produced across individuals[24]. Although many of the studies gathered in Stage 1 examined the correlation between syllable repertoire size and reproductive success, they often did not report the average species syllable repertoire size, so we searched for other sources (source per species noted in brackets in **AnalysisData.csv** (available on GitHib); bracketed references align with **Appendix A.1**). Studies with average species syllable repertoires were found using the following search terms: Passeriformes or [species name] in combination with "song syllables", "song complexity", and "syllable repertoire." For four species in the full dataset, information on the average species syllable repertoire size could not be found or was ambiguous, so we manually counted the unique syllables sung by individual birds using sonograms of song recordings downloaded from xeno-canto.org[111]. We were able to calculate the average number of syllables across individuals for two of these species (see **Appendix E**).

Stage 3) we performed another literature search to gather information on the length of the song-learning window, using song stability over time as a proxy. Species that sang new syllables after sexual maturity — either by exchanging an old syllable for a novel one or by incorporating a new addition — were considered to have plastic songs. Studies with information about song stability were found using the following search terms: [species name] or [common name] in combination with "open-ended", "close-ended", "closed-ended", "age-limited", "crystal*", "adult learning",

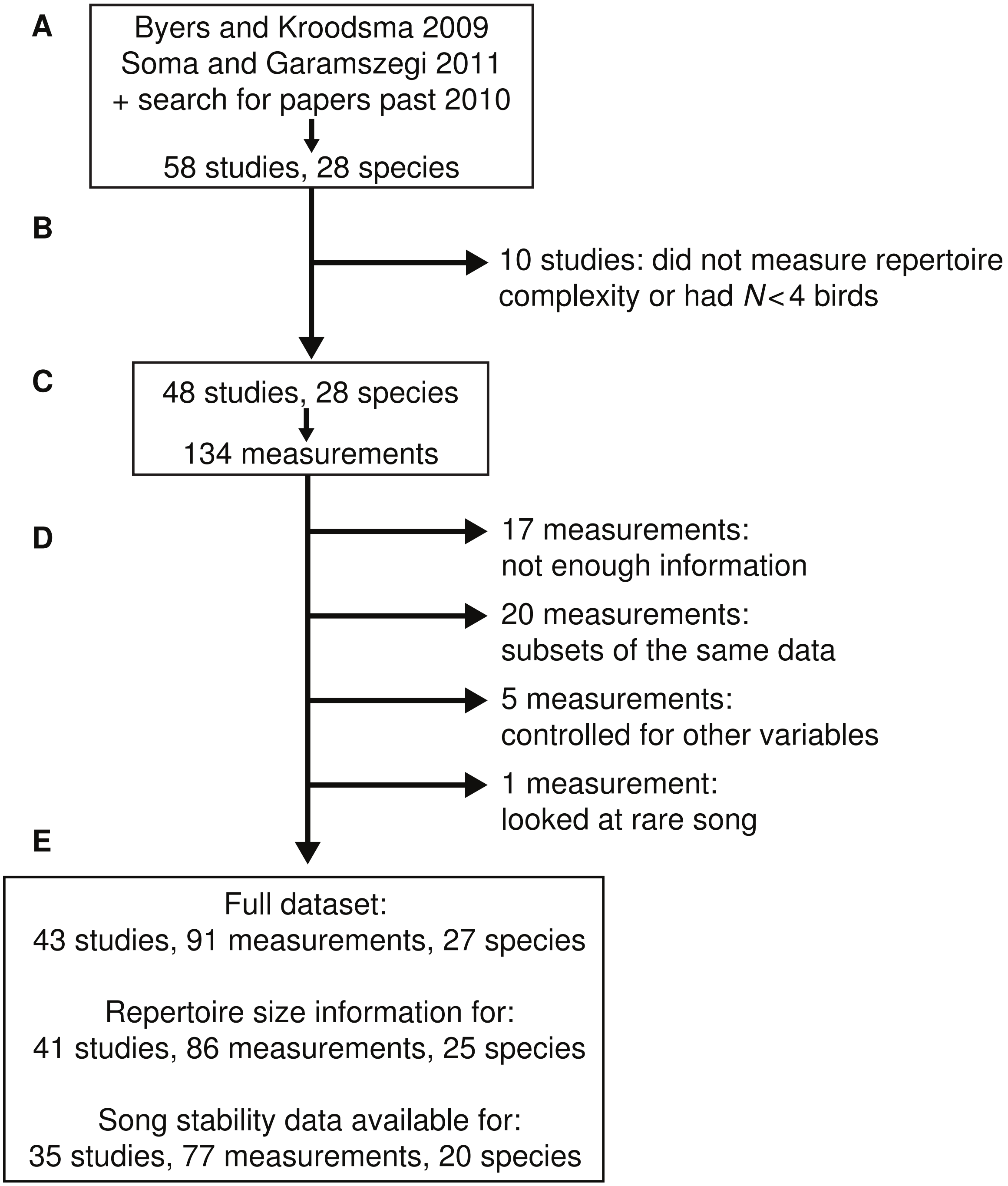


Figure 2.2: Schematic of dataset formation. (A) 57 studies and 1 thesis were compiled from the references in Byers and Kroodsma[18], Soma and Garamszegi[19], and a search for studies and theses published after these analyses. (B) we discarded 10 studies, because they either did not measure song elaboration or studied fewer than 4 individuals. (C) We derived 134 measurements from the remaining 48 studies. (D) We discarded 43 measurements for the reasons listed. (E) The full dataset contained 43 studies (42 papers and 1 thesis), 91 measurements, and 27 species.

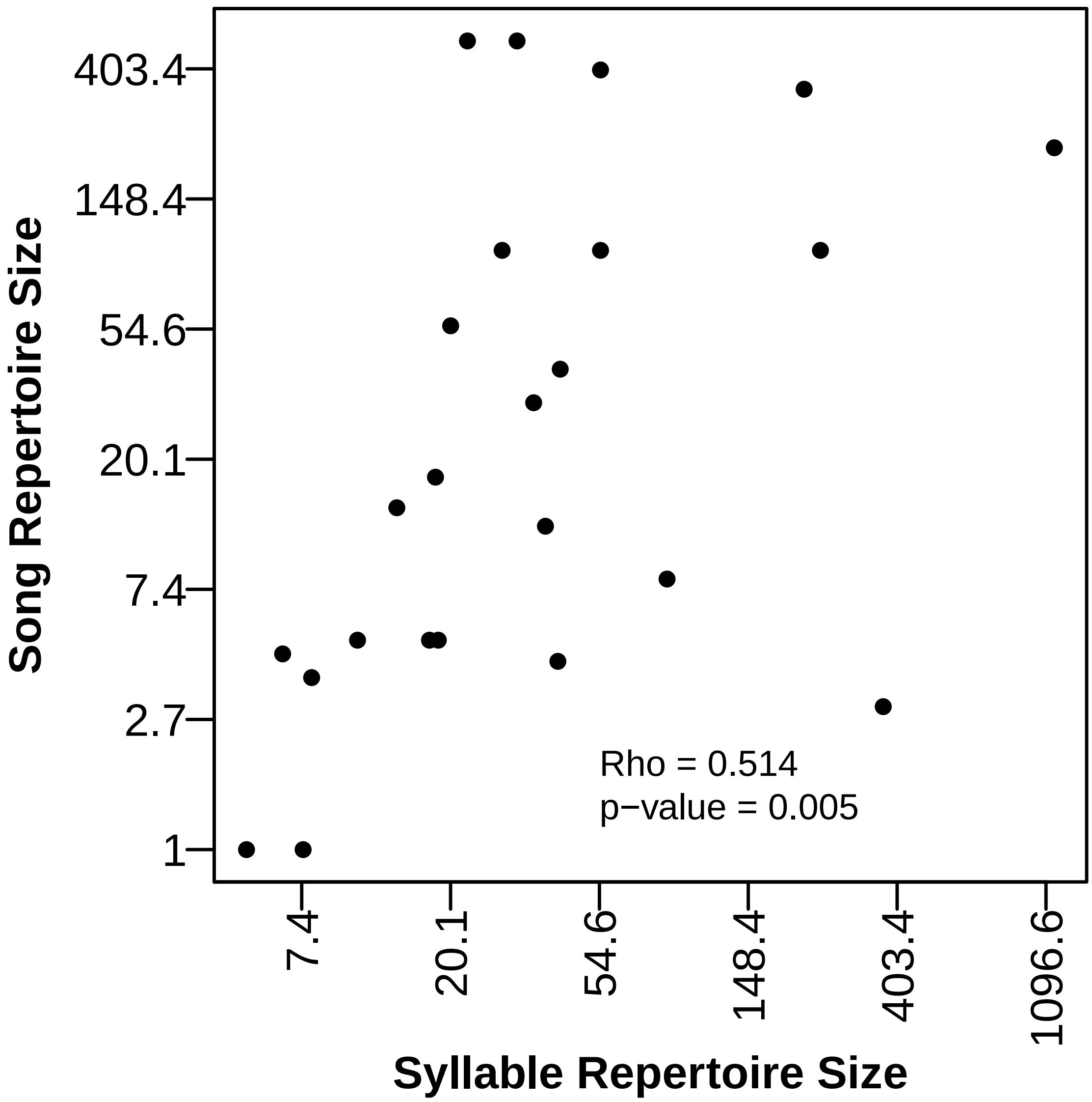


Figure 2.3: Song and syllable repertoire size are correlated between species. Each dot represents a species from this study. All species for which we had a measurement for the average syllable repertoire were included except *Phylloscopus fuscatus*, for which we did not have a measurement of the average song repertoire size. Spearman's ρ and associated p-value included on plot. Song repertoire data obtained from Snyder and Creanza 2019[24]. Specific references for each species are found in **AnalysisData.csv** (available on GitHub), where square-bracketed numbering matches the references in **Appendix A.1**

and “song changes.” Information on the song stability of several species was not available (see **Table 2.1**). Information on the species average syllable repertoire size existed for all species for which we found information regarding song stability.

2.2.3 Dataset formation

From the 48 field studies that remained, we compiled 134 relevant measurements (see **STable 1: All Measurements** (available on GitHub)). These commonly used measurements of reproductive success were categorized as follows:

Number of females: how many social mates a male attracts. Males that attract more females are assumed to be more successful.

Latency to pairing date or laying date: these two measures are traditionally used as a measure of reproductive success, because attractive males should pair first. Additionally, birds that produce offspring sooner have a better chance of parenting a larger brood[112] or multiple broods in a single breeding season[113–117]. Chicks born earlier in the season also tend to be more viable[1,117].

Extra-pair paternity: this is often considered a metric of reproductive success because males that sire offspring in extra-pair matings are assumed to be more attractive to females. However, it has been suggested to be an unreliable metric; see[19,118,119] and **Table 2.2** and results below.

Clutch size or number of offspring/recruits: these three measures are affected by both male and female genetic quality; however, it has been shown that females exposed to more elaborate songs can respond by producing larger clutches[83], so male song quality can also potentially affect this metric. The number of offspring or number of recruits (offspring that return to the parental territory) are related to the genetic fitness of males and females, but also to parental investment.

Measurements of the correlation between individual song elaboration and reproductive success (as defined above) that were not Pearson’s correlation coefficient values (r) were converted into r values using standard methods[120–122], see also **()-to-r Converters.R** (available on GitHub) for the exact calculations. Because negative correlations indicate a stronger relationship between individual song elaboration and reproductive success in latency to reproduction measurements, all latency measurements were multiplied by 1 (as in[19]). We converted r values into Fisher's Z values via Fisher's r-to-Z transformation, because Fisher's Z values have normally distributed variance — a criterion for variance estimation in the meta-analysis. This transformation leads to a slight positive bias, so we applied the recommended transformation prior to conversion to correct for that bias[123]. In total, 43 measurements were removed from the analysis for the reasons covered in **Figure 2.2** and are labeled by their reason for exclusion in **STable 1: All Measurements** (available on GitHub). We created three primary datasets:

Full Dataset: includes all species for which we found measurements that correlated individual song elaboration with reproductive success, even if we could not obtain information on the species average syllable repertoire size or song stability state (91 correlation measurements, 27 species).

Species Average Syllable Repertoire Dataset: subset of the species for which we found or counted species average syllable repertoire sizes (86 correlation measurements, 25 species).

Song Stability Dataset: subset of the species for which we found information on song stability (77 correlation measurements, 20 species).

See **Table 2.1** for the list of species in each dataset. We tested for funnel plot asymmetry in all three datasets using Egger's regression test (`regtest`; **R** package `metafor`)[124] and the ranked correlation test (`ranktest`; **R** package `metafor`)[125]. Some studies included measurements of the correlation between individual song elaboration and reproductive success before and after controlling for territory quality or other factors. In these cases, we used the measurements that did not control for other factors in the primary datasets. We created and tested secondary datasets which included the territory-controlled values in place of the non-controlled values.

Table 2.1: Species present in each dataset.

Full	Repertoire	Stability	No Offspring	No Offspring or EPP
Acrocephalus arundinaceus	*Acrocephalus arundinaceus*	*Acrocephalus arundinaceus*	*Acrocephalus arundinaceus*	*Acrocephalus arundinaceus*
Acrocephalus bistrigiceps	*Acrocephalus bistrigiceps*		*Acrocephalus bistrigiceps*	*Acrocephalus bistrigiceps*
Acrocephalus palustris	*Acrocephalus palustris*	*Acrocephalus palustris*	*Acrocephalus palustris*	*Acrocephalus palustris*
Acrocephalus schoenobaenus	*Acrocephalus schoenobaenus*	*Acrocephalus schoenobaenus*	*Acrocephalus schoenobaenus*	*Acrocephalus schoenobaenus*
Agelaius phoeniceus	*Agelaius phoeniceus*	*Agelaius phoeniceus*	*Agelaius phoeniceus*	*Agelaius phoeniceus*
Anthus spinoletta	*Anthus spinoletta*		*Anthus spinoletta*	*Anthus spinoletta*
Cardinalis cardinalis	*Cardinalis cardinalis*	*Cardinalis cardinalis*		
Carpodacus mexicanus	*Carpodacus mexicanus*		*Carpodacus mexicanus*	*Carpodacus mexicanus*
Dendroica pensylvanica	*Dendroica pensylvanica*	*Dendroica pensylvanica*	*Dendroica pensylvanica*	
Emberiza schoeniclus	*Emberiza schoeniclus*	*Emberiza schoeniclus*	*Emberiza schoeniclus*	
Ficedula albicollis	*Ficedula albicollis*	*Ficedula albicollis*	*Ficedula albicollis*	*Ficedula albicollis*
Ficedula hypoleuca	*Ficedula hypoleuca*	*Ficedula hypoleuca*	*Ficedula hypoleuca*	*Ficedula hypoleuca*
Hirundo rustica	*Hirundo rustica*	*Hirundo rustica*	*Hirundo rustica*	*Hirundo rustica*
Luscinia megarhynchos	*Luscinia megarhynchos*	*Luscinia megarhynchos*	*Luscinia megarhynchos*	
Melospiza melodia	*Melospiza melodia*	*Melospiza melodia*	*Melospiza melodia*	*Melospiza melodia*
Mimus polyglottos	*Mimus polyglottos*	*Mimus polyglottos*	*Mimus polyglottos*	*Mimus polyglottos*
Parus caeruleus	*Parus caeruleus*	*Parus caeruleus*	*Parus caeruleus*	*Parus caeruleus*
Parus major	*Parus major*	*Parus major*	*Parus major*	*Parus major*
Phylloscopus fuscatus	*Phylloscopus fuscatus*	*Phylloscopus fuscatus*	*Phylloscopus fuscatus*	
Phylloscopus trochiloides				
Phylloscopus trochilus	*Phylloscopus trochilus*	*Phylloscopus trochilus*	*Phylloscopus trochilus*	*Phylloscopus trochilus*
Plectrophenax nivalis	*Plectrophenax nivalis*		*Plectrophenax nivalis*	*Plectrophenax nivalis*
Saxicola caprata				
Sturnella neglecta	*Sturnella neglecta*	*Sturnella neglecta*	*Sturnella neglecta*	*Sturnella neglecta*
Sturnus vulgaris	*Sturnus vulgaris*	*Sturnus vulgaris*	*Sturnus vulgaris*	*Sturnus vulgaris*
Sylvia communis	*Sylvia communis*	*Sylvia communis*	*Sylvia communis*	*Sylvia communis*
Wilsonia canadensis	*Wilsonia canadensis*			

2.2.4 Random effects meta-analysis

For the initial meta-analytic assessment of the data, we performed a series of random effects meta-analyses, as has been done in this field in the past[19]. We report these analyses in **Appendix B** for full disclosure of all statistical tests used. However, the major caveats of this analysis were that (1) it required discarding a significant portion of the data, (2) the remaining data needed to be analyzed separately by measurement type, and (3) this style of meta-analysis does not control for factors such as phylogenetic relatedness or non-independence of multiple measures from the same species.

2.2.5 Bayesian multi-level phylogenetic meta-analysis

We next analyzed the data as a multi-level phylogenetic meta-analysis using the **R** package `MCMCglmm`[126]. In this package, group differences are assessed by adding them as fixed effects to meta-analytic models. Random effects can be added to account for phylogenetic relatedness, heterogeneity, and non-independence of data as sources of variance. These complex models can be assembled because the program uses a Markov chain Monte Carlo (MCMC) simulation to estimate the amount of variance each of these parameters explain. We chose to include four random effects in the meta-analytic model based on the structure of the data and the best practices suggested in the literature[127–129]:

MType: effects due to the metrics of reproductive success (measurement type) used in a given study

Phylo: effects due to the phylogenetic differences of species studied

Species: effects due to non-phylogenetic differences between the species studied

Study: effects due to non-independence of measurements coming from a given study

Because no study investigated more than one population, differences caused by assessing different populations of the same species would be captured in the Study random effect. We also included a term for the standard error, which was calculated using the standard equation for Fisher's

Z. In this model, the user must hypothesize the amount of variance that each random effect accounts for (prior) and assign a confidence (nu) to this hypothesis before running the simulation. The priors for MType, Species, and Study were set to:

$$\frac{Variance(\text{correlation between individual song elaboration and reproductive success})}{(\text{number of random effects } + 1)} \tag{2.1}$$

We used a low confidence nu set to one (see, e.g., Supplement File 5 of[128]. The prior for standard error was fixed at the values we calculated from the Fisher's Zs. The prior for the Phylo was set by passing a species relatedness matrix — which was calculated as described in the next section — to the `ginverse` argument of `MCMCglmm`. To examine the amount of variance each random effect accounted for, we tested a series of models wherein each random effect was included alone or in combination with the others for each of these fixed effects and calculated the heterogeneity as described previously[127]. Ultimately, we included all variance terms, because inclusion of all terms led to a markedly lower Deviance Information Criterion (DIC) (**Table 2.3**, also see **Appendix 1** of Robinson and Creanza$_{\text{A}}$[9]).

We set the fixed effect in the models as either the full population (i.e. Fisher's Z ~ 1), the natural log of species average syllable repertoire size (as a continuous variable), or song stability (subpopulations that were song-stable or song-plastic). All models were run for 200,000 iterations, with a burn-in of 30,000 iterations and a thinning interval of 10 iterations. All tested models appeared to reach convergence, because they were well mixed with peaks separated from zero (see **Appendix 3** from Robinson and Creanza$_{\text{A}}$[9]), and values for Gelman's $\hat{R}$ were less than 1.1 (`gelman.diag` from **R** package `coda`)[130–132]. The p_{MCMC} is a measure of the fraction of runs that estimated a posterior mean greater than zero; we defined significant evidence of an effect size greater than zero in these models as <5% of the MCMC runs estimating a posterior mean below zero ($p_{MCMC}<0.05$). We performed an additional posterior predictive test on the continuous species average repertoire model to determine whether the model accurately predicted the correlation between individual song elaboration and reproductive success for the species tested.

For direct comparison to the dichotomized song stability results, we also tested the average species syllable repertoire size dichotomized (subpopulations with relatively smaller or larger average species repertoires) and 2) average species syllable repertoire (as subpopulations) (see **Appendix A.2**). To dichotomize species average syllable repertoire size, we tested all possible thresholds that resulted in at least two species in each group. We also tested for an interaction between song stability and dichotomized average species syllable repertoire size (see **Appendix A.2**). Finally, in an attempt to disentangle mating success from reproductive output, we tested models where only metrics of mate choice (genetic or social) or only metrics of social mate choice were used **Appendix A.2**). These models were designed the same way as described for models in the main analysis.

2.2.6 BEST analysis

We tested for between-group differences using "Bayesian Estimation Supersedes the t-Test" (BEST) from the **R** package `BEST`[133]. This test returns the likelihood that the difference between the true means of two groups is greater than zero and gives a 95% credibility interval for the magnitude of this difference. We considered groups to be different if there was <5% chance that there was no real difference, or this difference was in the opposite direction of what the meta-analysis suggested.

2.2.7 Controlling for phylogenetic relationships

To control for phylogenetic effects, we performed these Bayesian meta-analyses with a phylogeny that we generated from publicly available data[134]. Using a list of all species in this study, we extracted a set of 1000 trees via the phylogeny subsets tool on birdtree.org[134]. We included *Sayornis phoebe* as an outgroup to root the tree. We created a consensus tree in **R** using the mean edge length method via the `consensus.edges` function (**R** package `phytools`[135]) and converted it into a relatedness matrix for use in the Bayesian meta-analyses with the `inverseA` function (**R** package `MCMCglmm`[126]. Species not present in a dataset were dropped from the relatedness matrix before being passed to the `MCMCglmm` function. Because this method uses a consensus tree, it removes any phylogenetic uncertainty. To account for phylogenetic uncertainty, we used the R package `mulTree` to generate a series of MCMCglmm models based on 100 randomly chosen trees (two chains per tree) from the original set of 1000 trees for the following four models: 1) the entire population model in the full dataset, 2) the song stability model in the song stability dataset, and 3) the continuous and 4) discrete species average syllable repertoire size in the species average syllable repertoire dataset. Models were otherwise identical to those above, except that they were run for 400,000 iterations.

2.3 Results

2.3.1 Testing for publication bias and examining heterogeneity

To assess the effect of publication bias on these datasets, we tested for funnel plot asymmetry using Egger's regression test and the ranked correlation test[125] in all three datasets. These tests revealed no significant asymmetry (**Figure 2.4**), suggesting that publication bias did not significantly affect these data. We also found no evidence for publication bias when we used territory-controlled values in place of non-controlled values (**Figure A.3.1**). We examined the amount of variance that each random effect term accounted for in the full population models and found that the way that reproductive success was measured (MType) accounted for the most variance, while phylogeny and other species differences accounted for little variance (**Table 2.3**). Because MType accounted for the most variance, we created an additional model with it as a fixed effect. Most measurement types show positive correlations between individual song elaboration and reproductive success, however extra-pair paternity did not (**Table 2.2**). This is in line with what has been shown previously[19,118,119].

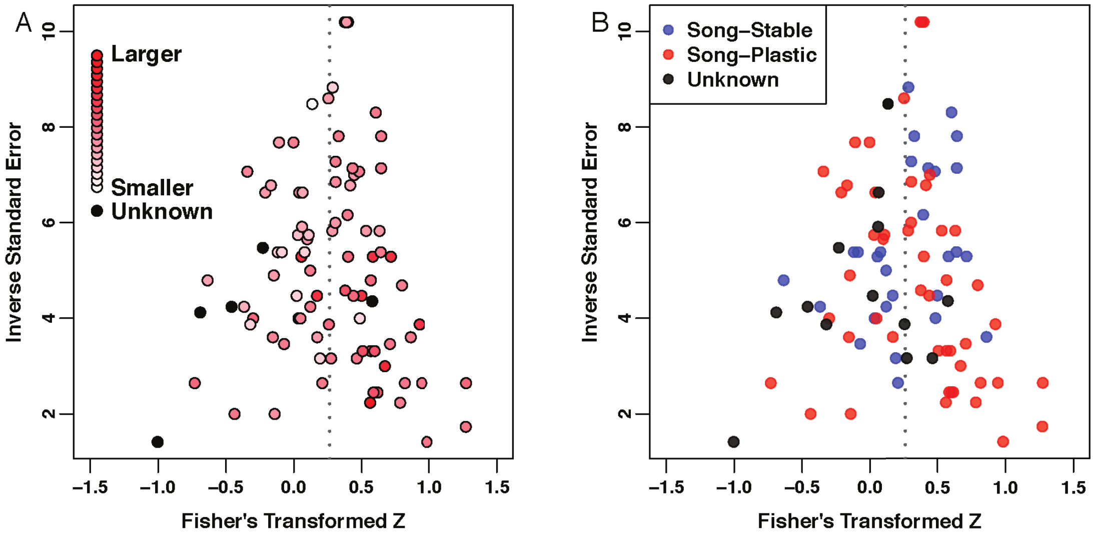


Figure 2.4: Funnel plot asymmetry. Funnel plots show the 91 measurements of the correlation between individual song elaboration and reproductive success from the full dataset. The grey dotted line represents the mean Fisher's transformed Z. (A) Circle color becomes more red as the average syllable repertoire size of the species increases. Black circles show measurements from species for which the syllable repertoire size is unknown. (B) Blue circles indicate measures from song-stable species, while red circles indicate measurements from song-plastic species. Black circles denote species for which no song stability information was available. Egger's regression testing on the full dataset (z=0.9109, p=0.3624), species average syllable repertoire dataset (z=1.5555, p=0.1198), or song stability dataset (z=1.4782, p=0.1394) revealed no significant funnel plot asymmetry. Ranked correlation testing on the full dataset (τ=0.0227, p=0.7523), species average syllable repertoire dataset (τ=0.0539, p=0.4667), or song stability dataset (τ=0.0414, p=0.5974) also revealed no significant funnel plot asymmetry.

Table 2.2: Different metrics of reproductive success have different meta-analytic means. Model tested in the song stability dataset. Females refers to measurements that counted the number of social mates obtained. Fledge refers to measurements of the number of fledglings produced. Clutch refers to measurements of clutch size. Laying refers to measurements of the latency to laying date or latency to hatching date. These two measurements were combined, because there was only one measurement of latency to hatching date. Recruits refers to measurements of the number of recruits a male gained. EPP refers to measurements of extra-pair paternity.

Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Females	5	10	0.404	[0.016;0.802]	0.045
Fledge	12	20	0.345	[−0.035;0.717]	0.059
Pairing	14	25	0.243	[−0.142;0.614]	0.147
Clutch	6	8	0.254	[−0.138;0.648]	0.151
Laying	7	11	0.219	[−0.161;0.633]	0.211
Recruits	3	6	0.224	[−0.187;0.627]	0.212
EPP	9	11	−0.182	[−0.596;0.228]	0.316

Table 2.3: Population variance in the song stability dataset. Different sources of variance and non-independence in the data were added to the model as random effects terms alone and in combination with the others. MType encodes the variance due to the metric of reproductive success used to generate each measurement. Study indicates variance accounted for by studies that reported multiple measurements. Phylo accounts for the effects of phylogeny, while Species encompasses all remaining species-related effects. DIC stands for deviance information criterion.

Random	I^2	DIC
Species	53.52%	17.03
MType	39.35%	33.82
Study	62.89%	−7.63
Species	11.78%	12.05
Phylo	20.28%	“ ”
MType	21.24%	−27
Study	43.5%	“ ”
MType	33.17%	−9.74
Species	6.19%	“ ”
Phylo	12.16%	“ ”
Study	42%	−24.5
Species	4.78%	“ ”
Phylo	8.39%	“ ”
MType	22.7%	−44.37
Species	4.05%	“ ”
Phylo	7.61%	“ ”
Study	19.32%	“ ”

2.3.2 The relationship between song elaboration and reproductive success across all studies

I first estimated the meta-analytic mean for the correlation between individual song elaboration and reproductive success for the entire population (all species) using all three datasets (full, species average syllable repertoire, and song stability dataset). The posterior means of the models using the full dataset and the syllable repertoire dataset were not significantly separated from 0, whereas the model using the song stability dataset was weakly significantly separated from 0 (**Full Dataset:** Posterior Mean=0.213, 95% CredInt=[-0.163;0.607], p_{MCMC}=0.193; **Species Average Syllable Repertoire Dataset:** Posterior Mean=0.242, 95% CredInt=[-0.021;0.521], p_{MCMC}=0.067; **Song Stability Dataset:** Posterior Mean=0.264, 95% CredInt=[0.021;0.53], p_{MCMC}=0.042). We obtained similar results when we used the datasets that included territory-controlled values in place of non-controlled values (**Table A.3.1**) or when we accounted for phylogenetic uncertainty (**Table A.3.2**). While these results across all studies were largely not significant, we note that the magnitudes for the estimated posterior mean of the effect sizes were similar to what was reported previously by Soma and Garamszegi[19].

2.3.3 Testing the effect of species average repertoire size on the strength of the correlation between individual song elaboration and reproductive success

I examined whether there was a linear relationship between average species repertoire size and the correlation between individual song elaboration and reproductive success using the natural log of the species average syllable repertoire size as a continuous variable. The slope of this relationship was significantly greater than 0 (**Intercept:** Posterior Mean=-0.356, 95% CredInt=[-0.753;0.056], p_{MCMC}=0.081; **Slope:** Posterior Mean=0.167, 95% CredInt=[0.071;0.262], p_{MCMC}=0.001) (**Figures 2.5** and **2.6**). Thus, for the 25 species studied here, this model predicts that the strength of the correlation between individual song elaboration and reproductive success grows modestly with every natural-log increase in species average syllable repertoire size. The average syllable repertoires of the bird species studied here range from 5.1 ($ln(5.1) = 1.63$) to 1160 ($ln(1160) = 7.06$) syllables, so the model predicts that, when all other things are equal, the species with the smallest average syllable repertoires will show a very weak correlation, while species with very large average syllable repertoires will show moderate-to-strong correlations. However, we caution readers that this linear model also requires information for the random effects we included (e.g. phylogeny and the metric of reproductive success used) to predict values for new species; the slope and intercept presented here should not be used in isolation to make predictions about other species (see **Figure 2.5** for the distribution of effect sizes when not accounting for the random effects). We next tested how well the data fit the model using a posterior predictive check. This analysis has the model predict the real correlations between individual song elaboration and reproductive success using all the predictor variables (random and fixed effects) for each measurement. We found that the model predicted the real correlations accurately, and no species or measurement appeared to deviate significantly from the model predictions (**Figure 2.7**).

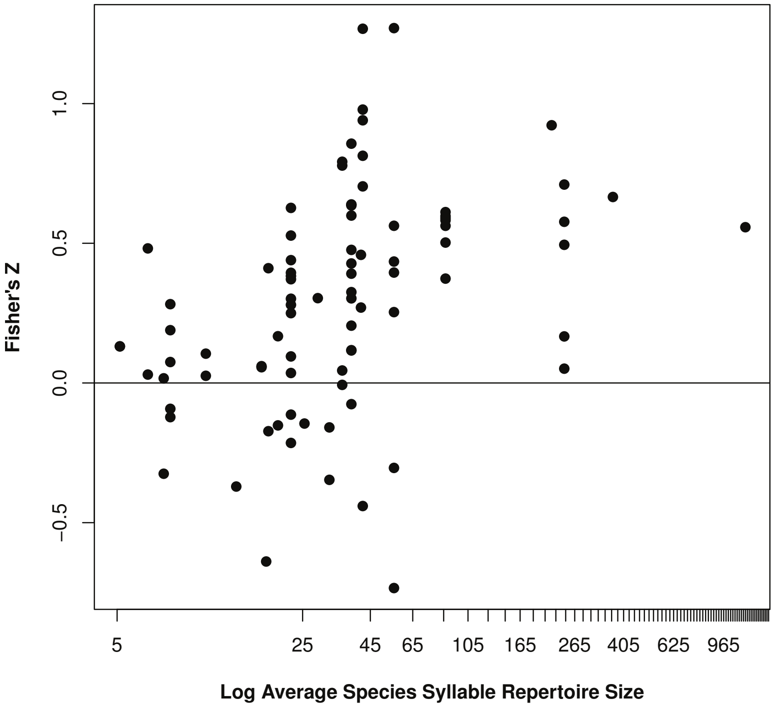


Figure 2.5: Effect sizes as repertoire size increases. The effect sizes for studies used in these meta-analysis plotted against the average syllable repertoire size of the species studied. For species with large repertoires, the effect sizes were generally positive, indicating a positive relationship between individual song elaboration and reproductive success; however, for species with small- to medium-sized repertoires, studies found a wide range of effects, both positive and negative.

Species (Rep) [citation]	**nBirds**	**Fisher's Z**	**[95% CI]**
Anthus spinoletta (5.1) [14]	75	0.135	[−0.096 ; 0.366]
Anthus spinoletta (5.1) [14]	75	0.135	[−0.096 ; 0.366]
Parus caeruleus (6.5) [31]	19	0.034	[−0.456 ; 0.524]
Parus caeruleus (6.5) [31]	19	0.486	[−0.004 ; 0.975]
Plectrophenax nivalis (7.46) [40]	18	−0.321	[−0.827 ; 0.185]
Plectrophenax nivalis (7.46) [40]	23	0.02	[−0.418 ; 0.459]
Parus major (7.9) [33]	32	−0.119	[−0.482 ; 0.245]
Parus major (7.9) [33]	13	0.193	[−0.427 ; 0.813]
Parus major (7.9) [33]	32	−0.089	[−0.453 ; 0.275]
Parus major (7.9) [33]	32	0.079	[−0.285 ; 0.443]
Parus major (7.9) [34]	81	0.286	[0.064 ; 0.508]
Dendroica pensylvanica (10.75) [17]	36	0.109	[−0.232 ; 0.45]
Dendroica pensylvanica (10.75) [17]	36	0.03	[−0.312 ; 0.371]
Cardinalis cardinalis (14) [15]	21	−0.367	[−0.829 ; 0.095]
Wilsonia canadensis (17.43) [48]	38	0.06	[−0.271 ; 0.391]
Wilsonia canadensis (17.43) [48]	47	0.064	[−0.232 ; 0.359]
Emberiza schoeniclus (18.15) [18]	26	−0.635	[−1.044 ; −0.227]
Agelaius phoeniceus (18.5) [13]	49	−0.169	[−0.458 ; 0.12]
Agelaius phoeniceus (18.5) [13]	49	0.415	[0.126 ; 0.704]
Hirundo rustica (20.1) [22]	16	0.171	[−0.373 ; 0.715]
Hirundo rustica (20.1) [20]	27	−0.148	[−0.548 ; 0.252]
Acrocephalus arundinaceus (22.5) [1]	107	0.375	[0.183 ; 0.567]
Acrocephalus arundinaceus (22.5) [1]	107	0.398	[0.206 ; 0.59]
Acrocephalus arundinaceus (22.5) [1]	107	0.387	[0.194 ; 0.579]
Acrocephalus arundinaceus (22.5) [2]	37	0.284	[−0.053 ; 0.62]
Acrocephalus arundinaceus (22.5) [2]	37	0.531	[0.195 ; 0.868]
Acrocephalus arundinaceus (22.5) [2]	37	0.63	[0.294 ; 0.967]
Acrocephalus arundinaceus (22.5) [3]	35	0.099	[−0.248 ; 0.445]
Acrocephalus arundinaceus (22.5) [3]	62	−0.11	[−0.365 ; 0.146]
Acrocephalus arundinaceus (22.5) [3]	52	0.443	[0.163 ; 0.723]
Acrocephalus arundinaceus (22.5) [3]	47	0.04	[−0.256 ; 0.335]
Acrocephalus arundinaceus (22.5) [3]	39	0.305	[−0.021 ; 0.632]
Acrocephalus arundinaceus (22.5) [3]	47	−0.211	[−0.506 ; 0.085]
Acrocephalus arundinaceus (22.5) [5]	77	0.254	[0.026 ; 0.482]
Phylloscopus fuscatus (25.3) [35]	7	−0.141	[−1.121 ; 0.839]
Ficedula hypoleuca (28.4) [21]	50	0.307	[0.022 ; 0.593]
Ficedula albicollis (31.4) [19]	53	−0.343	[−0.62 ; −0.066]
Ficedula albicollis (31.4) [20]	16	−0.155	[−0.699 ; 0.388]
Phylloscopus trochilus (35.1) [38]	62	−0.003	[−0.258 ; 0.252]
Phylloscopus trochilus (35.1) [38]	25	0.796	[0.378 ; 1.214]
Phylloscopus trochilus (35.1) [38]	19	0.048	[−0.442 ; 0.538]
Phylloscopus trochilus (35.1) [39]	8	0.782	[−0.095 ; 1.659]
Melospiza melodia (38) [24]	72	0.603	[0.367 ; 0.839]
Melospiza melodia (38) [24]	41	0.395	[0.077 ; 0.713]
Melospiza melodia (38) [25]	21	0.121	[−0.341 ; 0.583]
Melospiza melodia (38) [25]	15	−0.072	[−0.638 ; 0.494]
Melospiza melodia (38) [25]	10	0.209	[−0.532 ; 0.95]
Melospiza melodia (38) [26]	16	0.86	[0.317 ; 1.404]
Melospiza melodia (38) [22]	54	0.642	[0.368 ; 0.916]
Melospiza melodia (38) [22]	53	0.48	[0.203 ; 0.757]
Melospiza melodia (38) [22]	54	0.432	[0.157 ; 0.706]
Melospiza melodia (38) [22]	32	0.639	[0.275 ; 1.003]
Melospiza melodia (38) [28,20]	28	0.12	[−0.272 ; 0.512]
Melospiza melodia (38) [29]	64	0.643	[0.392 ; 0.894]
Melospiza melodia (38) [29]	56	0.307	[0.037 ; 0.576]
Melospiza melodia (38) [29]	64	0.329	[0.078 ; 0.58]
Carpodacus mexicanus (41.3) [16]	13	0.462	[−0.157 ; 1.082]
Carpodacus mexicanus (41.3) [16]	13	0.274	[−0.346 ; 0.894]
Sturnus vulgaris (41.95) [45]	15	0.708	[0.142 ; 1.273]
Sturnus vulgaris (41.95) [45]	7	−0.437	[−1.417 ; 0.543]
Sturnus vulgaris (41.95) [46]	6	1.272	[0.14 ; 2.403]
Sturnus vulgaris (41.95) [46]	5	0.982	[−0.404 ; 2.368]
Sturnus vulgaris (41.95) [46]	10	0.944	[0.203 ; 1.685]
Sturnus vulgaris (41.95) [46]	10	0.817	[0.076 ; 1.558]
Acrocephalus bistrigiceps (55) [6]	18	0.257	[−0.249 ; 0.764]
Acrocephalus schoenobaenus (55) [9]	26	0.566	[0.158 ; 0.975]
Acrocephalus schoenobaenus (55) [9]	31	0.399	[0.029 ; 0.769]
Acrocephalus schoenobaenus (55) [9]	23	0.438	[0 ; 0.877]
Acrocephalus schoenobaenus (55) [10]	10	1.274	[0.533 ; 2.015]
Acrocephalus schoenobaenus (55) [11]	19	−0.3	[−0.79 ; 0.19]
Acrocephalus schoenobaenus (55) [12]	10	−0.73	[−1.471 ; 0.011]
Sturnella neglecta (86) [43]	9	0.601	[−0.199 ; 1.401]
Sturnella neglecta (86) [43]	14	0.566	[−0.025 ; 1.157]
Sturnella neglecta (86) [43]	9	0.615	[−0.185 ; 1.416]
Sturnella neglecta (86) [43]	14	0.594	[0.003 ; 1.185]
Sturnella neglecta (86) [43]	9	0.586	[−0.214 ; 1.386]
Sturnella neglecta (86) [43]	14	0.506	[−0.085 ; 1.097]
Sturnella neglecta (86) [44]	24	0.377	[−0.05 ; 0.805]
Mimus polyglottos (216) [30]	18	0.926	[0.42 ; 1.432]
Acrocephalus palustris (241) [7]	23	0.17	[−0.268 ; 0.609]
Acrocephalus palustris (241) [7]	23	0.498	[0.06 ; 0.937]
Acrocephalus palustris (241) [8]	31	0.581	[0.21 ; 0.951]
Acrocephalus palustris (241) [8]	31	0.714	[0.344 ; 1.084]
Acrocephalus palustris (241) [8]	31	0.055	[−0.315 ; 0.425]
Sylvia communis (367.5) [47]	12	0.67	[0.016 ; 1.323]
Luscinia megarhynchos (1160) [23]	8	0.561	[−0.315 ; 1.438]

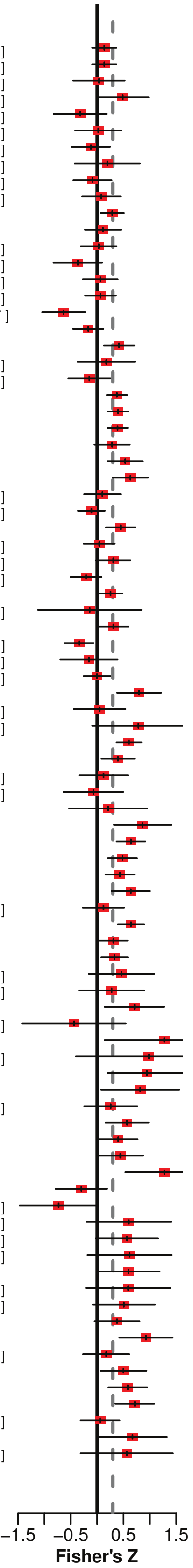


Figure 2.6: Forest plot of the species average syllable repertoire dataset. Columns show the study reference, species studied, the number of birds used to generate a measurement, the Fisher's Z form of the estimate, and its 95% confidence intervals. Ticks in the boxes mark the Fisher's Z and black horizontal lines show the confidence interval. The grey, dashed vertical line shows the population mean. When the same study is listed in more than one row on the plot, multiple different metrics of reproductive success were obtained from that study.

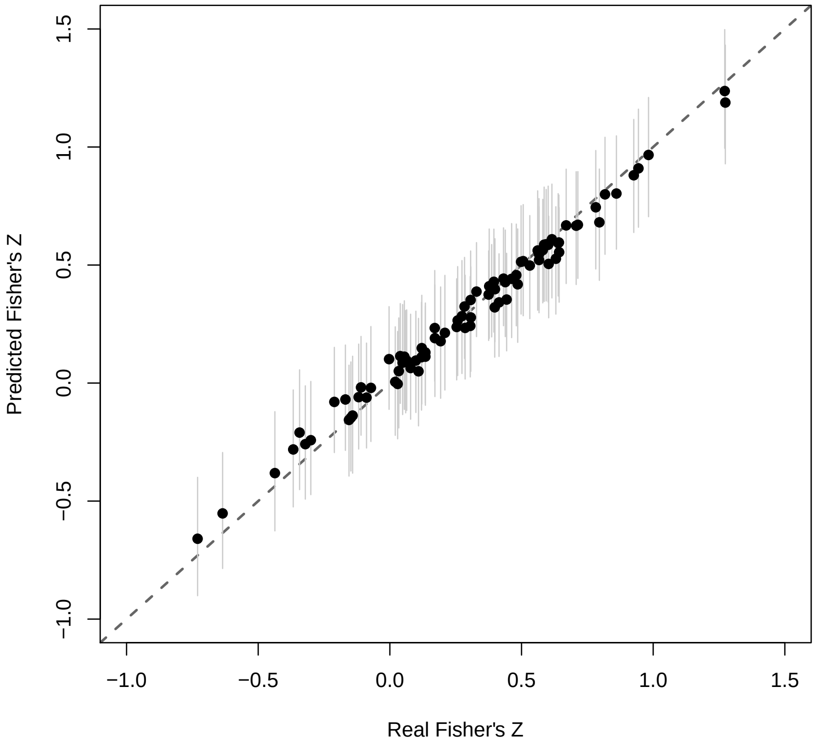


Figure 2.7: Posterior predictive check of the continuous syllable repertoire size model. Each black dot compares the actual correlation in Fisher's Z between individual song elaboration and reproductive success measured in the field (x-axis) to the value predicted for that correlation (y-axis) when the model was given the repertoire size of the studied species and information about the random effects (i.e. MType, Study, Phylo, and Species). The dark-grey, dashed line shows the line of unity, where the predicted correlation is identical to the real correlation. Light grey solid lines show the 95% confidence interval around each predicted correlation. All dots are very close to or overlapping with the line of unity, showing that model-predicted correlations were very similar to the real correlations.

To test the robustness of these results, we performed a jackknife analysis where each species was removed in turn. This did not significantly affect these results (**Table A.3.3**). Furthermore, we tested whether these results were driven by the species with the smallest or largest average repertoires by excluding those species. The three species with the largest average syllable repertoires could be excluded and the model still predicted a significant relationship between species average syllable repertoire size and the correlation between individual song elaboration and reproductive success (**Table A.3.4**). In addition, up to nine species with the smallest average syllable repertoires could be excluded, and the model still predicted a significant relationship (**Table A.3.5**). For many species, there were multiple measures of species average syllable repertoire size reported in the literature, and we used the median value for the main analysis. Using the maximum or minimum literature-reported values instead of the median values did not significantly affect the results (**Table A.3.6**), nor did using territory-controlled values in place of non-controlled values (**Table A.3.7** and **Figure A.3.2**), or accounting for phylogenetic uncertainty (**Table A.3.2**). This finding was also robust to methodological changes in which repertoire size was broken into discrete groups (**Appendix A.2**).

2.3.4 Probing the differences between species with stable or plastic songs

To examine whether song stability could predict the strength of the correlation between individual song elaboration and reproductive success in a given species, we tested song stability as a fixed effect using the song stability dataset (**Figure 2.8**). The effect size for song-stable species was not significantly separated from zero (Posterior Mean=0.149, 95% CredInt=[-0.226;0.511], p_{MCMC}=0.39). The effect size for song-plastic species was predicted to be positive (Posterior Mean=0.31, 95% CredInt=[0.034;0.594], p_{MCMC}=0.028), but song stability did not appear to be a reliable predictor of the correlation between individual song elaboration and reproductive success. First, the song-plasticity estimate was not strongly significant (p_{MCMC}=0.028), particularly when it is considered that we tested two independent hypotheses in this dataset. Second, the song-plastic group's posterior mean and 95% credibility interval were qualitatively similar to those seen for the entire population, and its 95% credibility interval overlapped substantially with that for song-stable species. Finally, we examined whether there was a difference between the song-stable and song-plastic groups using “Bayesian Estimation Supersedes the t-Test” (BEST) analysis, and we did not find evidence for a significant difference between song-stable and song-plastic species (BEST%>0=21.7%, Mean Difference=0.075, 95% CredInt=[-0.109,0.259]). In other words, song-stable and song-plastic species did not show a significant difference in their distribution of effect sizes. Thus, these results suggest that song stability may not be a species trait that can reliably predict the strength of the correlation between individual song elaboration and reproductive success. However, we have few song-stable species in this dataset (6), so this hypothesis should be re-evaluated as more data becomes available. This discrete analysis of adult song stability should not be directly compared to the continuous analysis of species average syllable repertoire, so we also provided a discrete analysis of repertoire size, which is concordant with the results from the continuous analysis (**Appendix A.2**).

We classified a species as song-plastic if individual birds changed their repertoires over time. This included both species that 1) increase their repertoire size with age and have the potential to signal their age via their repertoire size, as well as 2) species that replace old syllables with new ones and maintain a constant repertoire size overtime. To address this, we repeated this analysis, reclassifying all studied species as those which do increase their repertoire size with age and those that do not. The reclassification scheme did not significantly affect these results (**Tables A.3.8** and **A.3.9**). Field studies examining song stability often examine a small number of birds, so it is possible that one small-scale study might conclude that a species does change its repertoire when another study might conclude that the species does not[136]. Therefore, we re-assigned each species in turn to the opposite song-stability classification. This did not significantly affect the results (**Table A.3.10**). Using territory-controlled measurements in place of non-controlled measurements also did not significantly affect the results (**Tables A.3.11** and **A.3.12** and **Figure A.3.3**). Finally, accounting for phylogenetic uncertainty also did not significantly affect the results (**Table A.3.2**).

Taken together, these results do not support the hypothesis that song stability can be used to predict the strength of the correlation between individual song elaboration and reproductive success.

Species (Rep) [citation]	nBirds	Fisher's Z	[95% CI]
Acrocephalus palustris (241) [7]	23	0.17	[−0.268 ; 0.609]
Acrocephalus palustris (241) [7]	23	0.498	[0.06 ; 0.937]
Acrocephalus palustris (241) [8]	31	0.581	[0.21 ; 0.951]
Acrocephalus palustris (241) [8]	31	0.714	[0.344 ; 1.084]
Acrocephalus palustris (241) [8]	31	0.055	[−0.315 ; 0.425]
Cardinalis cardinalis (14) [15]	21	−0.367	[−0.829 ; 0.095]
Emberiza schoeniclus (18.15) [18]	26	−0.635	[−1.044 ; −0.227]
Melospiza melodia (38) [24]	72	0.603	[0.367 ; 0.839]
Melospiza melodia (38) [24]	41	0.395	[0.077 ; 0.713]
Melospiza melodia (38) [25]	21	0.121	[−0.341 ; 0.583]
Melospiza melodia (38) [25]	15	−0.072	[−0.638 ; 0.494]
Melospiza melodia (38) [25]	10	0.209	[−0.532 ; 0.95]
Melospiza melodia (38) [26]	16	0.86	[0.317 ; 1.404]
Melospiza melodia (38) [22]	54	0.642	[0.368 ; 0.916]
Melospiza melodia (38) [22]	53	0.48	[0.203 ; 0.757]
Melospiza melodia (38) [22]	54	0.432	[0.157 ; 0.706]
Melospiza melodia (38) [22]	32	0.639	[0.275 ; 1.003]
Melospiza melodia (38) [28,20]	28	0.12	[−0.272 ; 0.512]
Melospiza melodia (38) [29]	64	0.643	[0.392 ; 0.894]
Melospiza melodia (38) [29]	56	0.307	[0.037 ; 0.576]
Melospiza melodia (38) [29]	64	0.329	[0.078 ; 0.58]
Parus caeruleus (6.5) [31]	19	0.034	[−0.456 ; 0.524]
Parus caeruleus (6.5) [31]	19	0.486	[−0.004 ; 0.975]
Parus major (7.9) [33]	32	−0.119	[−0.482 ; 0.245]
Parus major (7.9) [33]	13	0.193	[−0.427 ; 0.813]
Parus major (7.9) [33]	32	−0.089	[−0.453 ; 0.275]
Parus major (7.9) [33]	32	0.079	[−0.285 ; 0.443]
Parus major (7.9) [34]	81	0.286	[0.064 ; 0.508]
Acrocephalus arundinaceus (22.5) [1]	107	0.375	[0.183 ; 0.567]
Acrocephalus arundinaceus (22.5) [1]	107	0.398	[0.206 ; 0.59]
Acrocephalus arundinaceus (22.5) [1]	107	0.387	[0.194 ; 0.579]
Acrocephalus arundinaceus (22.5) [2]	37	0.284	[−0.053 ; 0.62]
Acrocephalus arundinaceus (22.5) [2]	37	0.531	[0.195 ; 0.868]
Acrocephalus arundinaceus (22.5) [2]	37	0.63	[0.294 ; 0.967]
Acrocephalus arundinaceus (22.5) [3]	35	0.099	[−0.248 ; 0.445]
Acrocephalus arundinaceus (22.5) [3]	62	−0.11	[−0.365 ; 0.146]
Acrocephalus arundinaceus (22.5) [3]	52	0.443	[0.163 ; 0.723]
Acrocephalus arundinaceus (22.5) [3]	47	0.04	[−0.256 ; 0.335]
Acrocephalus arundinaceus (22.5) [3]	39	0.305	[−0.021 ; 0.632]
Acrocephalus arundinaceus (22.5) [3]	47	−0.211	[−0.506 ; 0.085]
Acrocephalus arundinaceus (22.5) [5]	77	0.254	[0.026 ; 0.482]
Acrocephalus schoenobaenus (55) [9]	26	0.566	[0.158 ; 0.975]
Acrocephalus schoenobaenus (55) [9]	31	0.399	[0.029 ; 0.769]
Acrocephalus schoenobaenus (55) [9]	23	0.438	[0 ; 0.877]
Acrocephalus schoenobaenus (55) [10]	10	1.274	[0.533 ; 2.015]
Acrocephalus schoenobaenus (55) [11]	19	−0.3	[−0.79 ; 0.19]
Acrocephalus schoenobaenus (55) [12]	10	−0.73	[−1.471 ; 0.011]
Agelaius phoeniceus (18.5) [13]	49	−0.169	[−0.458 ; 0.12]
Agelaius phoeniceus (18.5) [13]	49	0.415	[0.126 ; 0.704]
Dendroica pensylvanica (10.75) [17]	36	0.109	[−0.232 ; 0.45]
Dendroica pensylvanica (10.75) [17]	36	0.03	[−0.312 ; 0.371]
Ficedula albicollis (31.4) [19]	53	−0.343	[−0.62 ; −0.066]
Ficedula albicollis (31.4) [20]	16	−0.155	[−0.699 ; 0.388]
Ficedula hypoleuca (28.4) [21]	50	0.307	[0.022 ; 0.593]
Hirundo rustica (20.1) [22]	16	0.171	[−0.373 ; 0.715]
Hirundo rustica (20.1) [20]	27	−0.148	[−0.548 ; 0.252]
Luscinia megarhynchos (1160) [23]	8	0.561	[−0.315 ; 1.438]
Mimus polyglottos (216) [30]	18	0.926	[0.42 ; 1.432]
Phylloscopus fuscatus (25.3) [35]	7	−0.141	[−1.121 ; 0.839]
Phylloscopus trochilus (35.1) [38]	62	−0.003	[−0.258 ; 0.252]
Phylloscopus trochilus (35.1) [38]	25	0.796	[0.378 ; 1.214]
Phylloscopus trochilus (35.1) [38]	19	0.048	[−0.442 ; 0.538]
Phylloscopus trochilus (35.1) [39]	8	0.782	[−0.095 ; 1.659]
Sturnella neglecta (86) [43]	9	0.601	[−0.199 ; 1.401]
Sturnella neglecta (86) [43]	14	0.566	[−0.025 ; 1.157]
Sturnella neglecta (86) [43]	9	0.615	[−0.185 ; 1.416]
Sturnella neglecta (86) [43]	14	0.594	[0.003 ; 1.185]
Sturnella neglecta (86) [43]	9	0.586	[−0.214 ; 1.386]
Sturnella neglecta (86) [43]	14	0.506	[−0.085 ; 1.097]
Sturnella neglecta (86) [44]	24	0.377	[−0.05 ; 0.805]
Sturnus vulgaris (41.95) [45]	15	0.708	[0.142 ; 1.273]
Sturnus vulgaris (41.95) [45]	7	−0.437	[−1.417 ; 0.543]
Sturnus vulgaris (41.95) [46]	6	1.272	[0.14 ; 2.403]
Sturnus vulgaris (41.95) [46]	5	0.982	[−0.404 ; 2.368]
Sturnus vulgaris (41.95) [46]	10	0.944	[0.203 ; 1.685]
Sturnus vulgaris (41.95) [46]	10	0.817	[0.076 ; 1.558]
Sylvia communis (367.5) [47]	12	0.67	[0.016 ; 1.323]

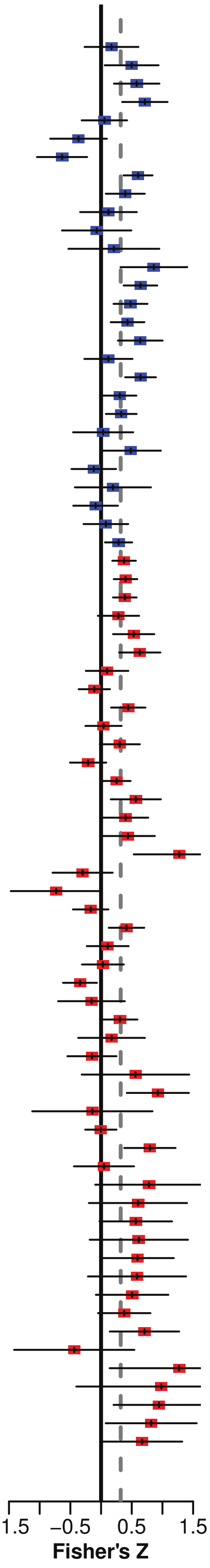


Figure 2.8: Forest plot of the song stability dataset. Columns shows the study reference, species studied, the number of birds used to generate a measurement, the Fisher's Z form of the estimate, and its 95% confidence intervals. Blue boxes mark measurements in the song-stable group, while red boxes mark measurements in the song-plastic group. Ticks in the boxes mark the Fisher's Z, and black horizontal lines show the confidence interval for each box. The grey, dashed vertical line shows the population mean. When the same study is listed in more than one row on the plot, multiple different metrics of reproductive success were obtained from that study.

2.4 Discussion

The relationship between reproductive success and song elaboration has long been proposed to exist[15,23,26,75–83], and, in laboratory settings, females have been observed to prefer larger repertoires in mate-choice tests[18,19,80,81,85–89] and to lay more eggs in response to playbacks of larger repertoires[83]. However, analysis of available field data has not provided significant evidence for a strong correlation between reproductive success and song elaboration in nature[18,19]. Thus, there has been a longstanding controversy over the putative link between individual song elaboration and reproductive success, with some research claiming "elaborate songs... are the acoustic equivalent of the peacock's tail"[26] and other research stating "it is unlikely that sexual selection for more elaborate songs is widespread among songbirds"[137]. Here, we reconcile these contradictory interpretations by showing that the correlation between individual song elaboration and reproductive success is stronger in species with elaborate syllable repertoires than in species with simple repertoires. Unexpectedly, song stability did not provide predictive power regarding the correlation between individual song elaboration and reproductive success.

These results and the above prediction stand in contrast to the past hypothesis that species with smaller average repertoires might be more likely exhibit a correlation between individual song elaboration and reproductive success, because it would be easier to discern which males have the most elaborate songs when each male of that species only produce a few syllables[5,103]. Why would these species not exhibit this preference in nature when they do in laboratory studies[80,81,85–89]?

Small average species repertoires are suggested to be more advantageous in contexts where dialects or song-matching are an important facet of male-male competition[138–142]. If males with larger repertoires tend to lose these competitions and are thus low in dominance hierarchies or cannot hold a territory, this would counteract female preferences for larger repertoires (see, e.g.[143]). These findings would not rule out a model wherein each bird species balances its average syllable repertoire size between more elaborate songs optimized for attracting females and less elaborate songs optimized for other functions, such as individual recognition, territory defense, or other aggressive interactions, but further investigation would be required. Alternatively, if females have a stronger preference for traits reflecting song learning accuracy[93,144], song performance (e.g. trill length or note frequency)[35–38] , or other exaggerated physical or behavioral traits (e.g. ornate plumage[39–41], lekking[42,43], or dance[44]), these characteristics may be more indicative of male quality, and would thus have greater influence on mate choice in nature[145].

Alternatively, the measured association between individual song elaboration and reproductive success could hinge on the ability of repertoire size to act as an honest signal of male fitness[146]. Therefore, the species average syllable repertoire would need to be large enough that inferior males cannot learn all of it. Large repertoires have been proposed to be costly to learn[10] due to the metabolic costs of the neural underpinnings of song learning[82,147,148] and because of the time and energy that must be dedicated to learning, practicing, and displaying large repertoires[77]. Theoretically, small species average syllable repertoires would not lead to large resource requirements and would be less costly to learn. Thus, inferior males would be able to produce all species-typical syllables. In this case, performance characteristics may be more indicative of male quality and eventual reproductive output than repertoire size in species with small average syllable repertoires, as performance would likely still be affected by male quality[146,149].

We hypothesized that open-ended learners would be more likely to show a correlation between individual song elaboration and reproductive success than closed-ended learners, because open-ended learners could potentially signal their age with their song. Extending the song learning window is expected to be metabolically costly[4,150], so longer learning windows should be present only in species where there is selection for song traits that could benefit from extended learning windows. Indeed, it has been suggested that adult song-learning is associated with the evolution of larger repertoires[150]. However, we found that the strength of this correlation in song-plastic species was not significantly different from song-stable species. It may be that open-ended learning is beneficial in multiple contexts; in some species, males may increase their repertoire size over time to signal their age, whereas, in other species, song stability may assist in song-matching and counter singing if male-male interactions are critical to reproductive success. Thus, song stability overall would not be predictive of a correlation between individual song elaboration and reproductive success. Narrowing the definition of song-plastic species to those that increase their repertoire size with age did not yield significant results (**Tables A.3.8** and **A.3.9**). However, the analysis of the interaction between species average syllable repertoire size and song stability allows us to cautiously propose that there may be an interaction between these traits (**Table A.2.13**). Further research will be required to conclude whether adult song stability can predict the strength of the correlation between individual song elaboration and reproductive success.

While these findings take an important step in elucidating the link between song elaboration and reproductive success, we note that this meta-analysis was done with the goal of generating testable predictions for future field studies, which is by definition limited by the number of existing studies. This meta-analysis was performed on the relatively small number of species for which the correlation between individual song elaboration and reproductive success were measured in the field. It remains to be seen whether these results will apply across all bird species, and we caution against making songbird-wide generalizations from a meta-analysis of relatively few species[151]. As more data are collected, it will also be important to investigate other factors that have been proposed to affect the strength of sexual selection in a species, such as polygyny, extra-pair paternity, breeding synchrony, and migration behaviors[1,19,22,24,152–159].

To date, most research on the relationship between individual song elaboration and reproductive success has been conducted in species with small to moderate-sized syllable repertoires (see **Figure 2.5**). Currently, we have data from four species with average repertoires larger than 100 syllables; the observed, significant relationship between species average syllable repertoire size and the correlation between individual song elaboration and reproductive success persists if three of these four species are removed, but not if all four are removed **Table A.3.4**. With more species surveyed at the higher end of species average syllable repertoire size, the field could better evaluate the observed trend. Thus, these results suggest that the field would particularly benefit from surveying more species with very large average syllable repertoires to assess the relationship between individual song elaboration and reproductive success.

Here, we re-evaluated the link between individual song elaboration and reproductive success with a Bayesian meta-analysis of decades of field studies that integrated additional between-species variables that may interact with sexual selection on individual song elaboration. This meta-analysis brings the results of these studies into sharper focus and proposes new hypotheses for future research to explore the origins and long-term effects of sexual selection on elaboration in learned mating signals. We find that individual male song elaboration appears to be most correlated to reproductive success in species that have evolved unusually large syllable repertoires, potentially implying both past and ongoing sexual selection for larger individual song elaboration in these species. If so, it will be important to consider the factors that initially drive the evolution of this elaboration and whether this trend is driven by mate choice and/or reproductive output. These factors could include 1) the species-specific importance of male traits for which song acts as an honest signal (e.g. health, developmental stresses, and song-learning capacity), 2) species lifestyle and ecological niche traits (e.g. mating system, migratory status, and breeding synchrony), and 3) tension between the importance of different uses of song. As more relevant variables are revealed, it will be possible to build better models to explain the different forces influencing sexual selection in song. Such models would be powerful tools not only for understanding bird species, but also for gaining insight into the behavioral and ecological forces that mediate the expression of sexually selected traits in different species.

Chapter 3

Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning [1]

3.1 Introduction

The duration of song-learning has been studied for decades[1,160,161]. In some species, learning is restricted to a short sensitive period early in life, also called a “critical period” (for example, ~day 25-90 in zebra finches (*Taeniopygia guttata*)), after which no new songs are acquired[162–164]. Other species appear to delay song crystallization until some time in adulthood[109,165,166]. For example, chipping sparrows *Spizella passerina* have been observed to have a second sensitive period immediately after their first migration, following which the song is crystallized[20,167]. Together, these species are labeled as “closed-ended” learners. Open-ended learners can continue to acquire new syllables or songs throughout their lives[17,168–172]. Previous research has focused on factors that shape the duration song learning, such as environmental variation and breeding season length[173–175], but it is still largely unknown how variation in the duration of song-learning interacts with the evolution of song itself. However, evidence from a small-scale comparative analysis suggests that a longer learning window may be associated with a larger average syllable repertoires[150].

Evolutionary pressures can act on both culturally and genetically inherited features of birdsong. Two key modes of selection might act in conjunction: on one hand, female choice can favor certain song traits, such as superior repertoire size, learning quality, or song performance[4,10,15,15,176], all of which could be improved upon by longer learning windows. On the other hand, the inherent metabolic cost of neuroplasticity should theoretically favor a shorter song-learning window and therefore reduce the opportunity for a bird to alter its song in adulthood[173,177,178]. Thus, while

---

[1] This chapter is adapted from my co-authored article with Kate T. Snyder *Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning*, which was published in *eLife*[70].

learning in adulthood or elongated sensitive periods have not been shown to be directly under positive selection (sexual or natural) or to play an explicit role in female preferences, sexual selection acting on certain song traits could indirectly favor longer or shorter song-learning windows. However, this theorized connection between sexual selection and the duration of song learning hinges on establishing a relationship between learning and the evolution of song traits, which has not yet been done.

Furthermore, sexual selection is hypothesized to be amplified in species with polygynous social mating systems and those with high rates of extra-pair paternity (EPP). A recent large-scale study found that polygyny drives faster, but non-directional, evolution of syllable repertoire size, and that syllable repertoire size is negatively correlated with the rate of EPP[24]. Thus, it may be higher rates of EPP and polygyny are associated with longer learning windows.

Here, we take a comparative, computational approach to the evolutionary history of open- and closed-ended song learning. We mined the literature for longitudinal field and laboratory observations to classify species as exhibiting “adult song stability” (likely closed-ended learners) or “adult song plasticity” (likely open-ended learners). Ultimately, we classified the song stability state of 67 species. For these species, we also compiled a database of seven species-level song traits that can represent either song elaboration (syllable repertoire, syllables per song, and song repertoire) or song performance (song duration, inter-song interval, song rate, and song continuity). We then perform phylogenetically-controlled analyses to evaluate the evolution of song and mating behavior alongside the relative plasticity of song over time. We find that adult song plasticity has evolved numerous times in bird species. We also find evidence of correlated evolution between the adult song stability and social mating system, with shifts in social mating system occurring more rapidly in lineages with adult song plasticity. In addition, we find a significant evolutionary pattern: in general, species with plastic songs have larger repertoires than species with stable songs. Specifically, the evolution of larger syllable and song repertoires appears to drive an evolutionary transition toward open-ended learning.

3.2 Methods

3.2.1 Data and code availability

https://github.com/CreanzaLab/SongLearningEvolution

3.2.2 Data collection

We performed a literature search to gather information on the song stability state of males in oscine species. Studies with information about learning style were found via Google Scholar using the following search terms: [species name] or [common name] in combination with "open-ended," "close-ended," "closed-ended," "age-limited," "crystal*," "adult learning," and "song changes." We used three strategies to assign the song stability state of species. We first defined a species as having adult song stability ('song-stable species') if the literature indicated that males did not modify their songs after the first breeding season. Species in which males modified their syllable repertoires after their first breeding season were classified as having adult song plasticity ('song-plastic species'). This strategy was meant to mimic the dichotomy of open- and closed-ended learners often used in the field. We made a two exceptions to this categorization for the species *Cacicus cela* and *Phoenicurus ochruros*, in which males do not gain their mature plumage until their second breeding season and may therefore be delayed in reaching sexual maturity relative to other bird species[179,180]. Because these birds cease modifying their repertoires before reaching their second breeding season with mature plumage, they were considered song-stable (Dataset S1). Additionally, past research defined *Melospiza lincolnii* as an "open-ended improviser." It is unclear whether improvisation throughout the lifespan is equivalent to learning throughout the lifespan, however it does fit our definition for adult song plasticity. For the main analysis, we considered this species to have adult song plasticity, but we repeated the main analyses with this species reclassified as having adult song stability (**Tables C.1.1** and **C.1.2**).

For our second strategy, we separated song-stable species into two sub-groups: early song-stable (species that cease modifying their songs before the first breeding season) and delayed song-stable (species that modify their songs during their first breeding season but not after). There was not enough information to make this determination for some species, so our dataset was reduced to 59 species. We used these same 59 species to generate a continuous measure of song stability for our third strategy. There is little information about the prevalence of song changes beyond the second breeding season for most of the species in our dataset; therefore, this continuous measure only ranges from 0 to 2 years, and all song-plastic species were assigned a value of 2. Furthermore, it was not clear exactly when most of the delayed song-stable species stopped learning, so they were given a value of 1.33, at which point the first breeding season should have ended. The two species mentioned above that display delayed song and plumage maturation were assigned a value of 2.

To gather data on the song traits, we performed a literature search via Google Scholar and Web of Science using the search terms "Passeriformes" or [species name] in combination with "syllable repertoire" or "song repertoire." This yielded a mix of primary sources and studies that had previously aggregated repertoire size data. We also gathered data from the curated field guides Birds of North America[181] and Handbook of Birds of the World[182]. We did not perform explicit searches for any of the other included song traits, but we collected this data whenever we encountered it. song trait nomenclature is variable across studies, so, when possible, we read the methods of the primary sources to ensure the authors used the same definitions for song traits that we did.

We utilized the following definitions for song traits:

Syllable Repertoire: the mean number of unique syllables produced per individual.

Song Repertoire: the mean number of unique songs produced per individual.

Syllables per Song: the mean number of unique syllables per song.

Song Duration: the mean length of a song, delineated by pauses or the beginning of a repeated motif (seconds).

Intersong Interval: the mean pause length between songs (seconds).

Song Rate: the mean number of songs produced per minute.

Calculated value: $Song\ Rate = \frac{60}{(Duration + Interval))}$

Continuity: the fraction of singing time spent actively producing song

Calculated value: $Continuity = \frac{Duration}{(Duration + Interval))}$

For three species with song-learning window data, syllable repertoire size estimates were not available in the literature, so we estimated these repertoire sizes from published sources or song recordings (**Appendix E**). When the song repertoire for a species equalled one, we assumed that its species' syllable repertoire was equal to its number of syllables per song. In many cases, there were multiple studies that gave different estimates for a given song variable in one species. To handle these discrepancies, we created three datasets: 1) the main dataset, wherein we used the median value across studies. 2) The minimum values reported in the literature. 3) The maximum values reported in the literature. We log normalized all song trait data.

We also catalogued species-level mating behavior data. In particular, we assembled binary classifications of social mating system (monogamy vs polygyny) and extra-pair paternity (low EPP vs high EPP). We considered a species to be monogamous or polygynous when a source unambiguously categorized that species' social mating system; we did not assign a social mating system to species labeled "probably," "usually," "mostly," "normally," "typically," and "generally monogamous/polygynous," etc. unless quantitative measurements were also provided. When quantitative data were available, species were defined as polygynous when at least 5% of males have more than one social mate, as in[24]. A review of extra-pair paternity studies estimated an average of ~11% of offspring per nest were attributable to extra-pair mates across species[183]. In accordance with this estimate and with previous studies that used a binary classification of EPP ([19,24], we used a 10% threshold for either extra-pair young or nests containing at least one extra-pair chick to estimate the frequency of extra-pair paternity in that species ($<$10% = low EPP; $\geq$10% = high EPP).

3.2.3 Assessing the evolutionary history of adult song stability

To predict the rate of transition between adult song stability and adult song plasticity, we used the `ace` function from the **R** package `Phytools`[135] and a publicly available phylogeny[134]. We tested whether an all-rates different (ARD) model fit the data significantly better than the equal rates (ER) model using an ANOVA. We then used the better-fit equal rates model to generate 10,000 trees with `make.simmap` (textbfR package `Phytools`). This function uses the rate from `ace` and a phylogenetic tree with annotated tips to create stochastic simulation maps for the potential evolutionary transitions between the song-stable and song-plastic states. We found the predicted ancestral state for each of these 10,000 simulations and used `countSimmap` (textbfR package `Phytools`) to count the total number of transitions that occurred in each map. The minimum number of predicted evolutionary transitions across these 10,000 simulations was considered to be the most parsimonious; we also compared this to a manual count of evolutionary transitions starting from either ancestral state, and found the same result.

3.2.4 Correlated evolution of song traits and song stability

To test whether there were any significant differences between song-stable species and song-plastic species for the song traits, we performed a phylogenetically-controlled ANOVA (`phylANOVA`, `Phytools`) for each song trait. We repeated this analysis with the subset of species we classified into early song-stable, delayed song-stable, and song-plastic. Because there was a limited number of early song-stable species in this dataset, we only performed this re-analysis for song traits that had data on at least 9 early song-stable species (**syllable repertoire size:** 9 species with early song stability; **song repertoire size:** 9 species; **syllables per song:** 10 species). We visualized the predicted ancestral traits on trees with color and pie graphs here, however, the raw values are available in (**Figure 2 - source data 1** from Robinson, Snyder, and Creanza 2019[70]. To test for correlations between song traits and the continuous values for song stability, we performed a phylogenetic generalized least squares (PGLS) analysis. We used the function `gls` (**R** package: `nlme`), with the "correlation" parameter lambda (λ) computed using the function `corPagel` (**R** package: `ape`).

To test whether adult song stability state affected the rate of evolution for the song traits, we used the function `brownie.lite` (**R** package: `phytools`). This function first calculates a one-rate model of evolution for a song trait using a phylogenetic tree and the current states of the tips for that song trait. This one-rate model assumes that change in the value of the song trait is random across evolutionary time and can be estimated via Brownian motion. Next, a model is generated wherein two different rates are calculated; this two-rate model assumes that the evolution of the song trait has one rate in the song-stable state and a different rate in the song-plastic state. This model requires estimations for the ancestral states of song stability for each branch of the phylogeny. To create these estimates, we used the function `ace` (**R** package: `phytools`)to calculate the rate of transition between the song-stable and song-plastic states for the full dataset. We then used these transition rates to generate 1,300 different stochastic simulation maps (`make.simmap`) (**R** package: `phytools`) for the subset of species that had data for each song trait. For the Brownie analysis, we tested whether the two-rate model fit the data significantly better than the one-rate model by performing a chi-square test on the mean log likelihoods of the two models. We repeated this analysis for the set of species we classified as early song-stable, delayed song-stable, and song-plastic for traits for which we had data on at least 9 early song-stable species. We compared the three-rate model to the one-rate model. We also reran the two-rate model in this reduced dataset by combining the early and delayed song-stable groups and testing whether the three-rate model was better than the two-rate model. Because the delayed song-stable trace peaked at a similar position to the song-plastic trace for syllable and song repertoire size (Figure 4D-E), we also compared the three-rate model to another version of the two-rate model, in which one group was early song-stable (shorter learning), and the other was delayed song-stable and song-plastic combined (longer learning).

We used `BayesTraits` [184,185] to test for correlated evolution between song stability and song traits, or, in other words, whether the rate and direction of evolutionary transitions of one trait are dependent on the state of another trait, and whether an order of transition events can be inferred. Specifically, we tested the hypothesis that an evolutionary change in song stability increases the likelihood of an evolutionary change in certain song variables or mating behaviors, or vice versa. `BayesTraits` compares two models of discrete trait evolution for a pair of binary traits and a given phylogenetic tree: 1) an independent model (i.e. the evolution of one trait does not depend on the other trait) and 2) a dependent model (i.e. the evolutionary transitions of each trait depend on the state of the other trait, suggesting correlated evolution). `BayesTraits` reports marginal likelihoods for the computed dependent and independent models (function `Discrete` in **R** package `btw`), allowing us to determine whether the dependent model describes the data significantly better than the independent model. We used function `LRtest` (**R** package: `lmtest`) to perform the likelihood ratio test. Since this model requires both input traits to be binary, we classified the continuous song traits as binary groups ("low" or "high") based on a delineating threshold. Instead of choosing the threshold arbitrarily, we used each unique value of the song trait data as the threshold and repeated the analysis 100 times at each threshold. This method of using a spectrum of thresholds to delineate the low and high value categories resulted in transition rates that varied dramatically depending on where the threshold was placed. In essence, when the threshold is set as a value in the bottom third of the unique trait values present in the data, the analysis evaluated the rate transition from low to moderate and larger values for a song trait and vice versa. When the threshold is set as a value in the upper range of the unique trait values present in the data, the analysis calculates the rates of transition from higher song trait values to medium and lower values. To account for this nuance, we binned the threshold data into two to five bins, with three bins as the default: low (bottom third of unique trait values), medium (middle third) and high (top third). We then calculated the mean of each state transition rate in each bin. In addition to the song traits, we also analyzed song stability versus social mating system (i.e. social monogamy or polygyny) and rate of EPP. These analyses were performed for 1,000 runs each.

3.2.5 Jackknife analysis

Some families of birds were well represented in our sample, while others were only represented by one or two species. To test whether any well-represented family significantly skewed our results, we removed each family that was represented by four or more species in the full dataset in turn, and repeated the `phylANOVA`, `brownie.lite`, and `BayesTraits` analyses. Jackknife analyses were only performed when significant results were obtained in the main analysis. Thus, all song traits except continuity were tested in the `phylANOVA` and `brownie.lite` jackknife analysis, while only syllable repertoire, song repertoire, and syllables per song were tested in the `BayesTraits` jackknife analysis. Each brownie analysis was run on 1,300 unique stochastic character maps, and each BayesTraits analysis was repeated 20 times. We determined the family of each species based on its classification in the 2017 version of the eBird Clements Integrated Checklist[186]. The family *Locustella* was combined with *Acrocephalidae*, as *Acrocephalidae* was paraphyletic when *Locustella* was considered to be a separate family. The *Mimidae* family alone had a large effect on the syllables per song metric, so we performed another jackknife analysis with `phylANOVA` and `brownie.lite` wherein we removed each mimid species in turn.

3.2.6 Correction for multiple testing

We used a Holm-Bonferroni correction to control for testing multiple hypotheses with the same data[187]. This correction is appropriate for data wherein the outcome of one hypothesis is likely to be related to the outcome of another hypothesis, as we believed would be the case for song traits.

3.3 Results

3.3.1 Evolution of song stability across clades

We were interested in examining the rate of evolution of adult song stability versus adult song plasticity and whether we could predict the ancestral state of clades on a phylogenetic tree[134]. As with any reconstruction of evolutionary history, these simulations cannot exactly predict the ancestral states but aim to approximate them. Furthermore, only a subset of oscine families could be included in our analysis. Ultimately, we could not make a conclusion about ancestral state for the last common ancestor of all the included species, but our results hint that there might have been several early transitions in the oscine lineage, leading to clades that predominantly have adult song stability or plasticity, coupled with a number of more recent transitions (see pie charts in **Figure 3.1 A** for the predicted likelihood of each state at each node). We found that a model allowing the transition rate from song stability to plasticity to be different from the transition rate from plasticity to stability (all-rates-different model — ARD) did not fit the data significantly better than a simpler model allowing for only one rate of transition back and forth between song stability and plasticity (equal rates model — ER) ($\text{LogLikelihood}_{\text{ER}}=-38.22$, $\text{LogLikelihood}_{\text{ARD}}=-38.21$, $p=0.87$). At least 14 transitions were required to explain the current binary song-stability states of our subset of bird species. Explaining the distribution of song plasticity in our subset of species most parsimoniously requires at least 9 transitions to adult song plasticity if the last common ancestor was song-stable and 7 transitions to song stability if the common ancestor was song-plastic (**Figure 3.2**).

Figure 3.1: Syllable repertoire size is larger in species with adult song plasticity even when controlling for phylogeny. These phylogenies show the calculated evolution of natural-log transformed syllable repertoire size and either (A) stable and plastic song stability states or (B) early song-stable, delayed song-stable, and song-plastic states. Dots at the tips of branches represent the current song-stability state. Pie charts represent the likelihood that the common ancestor at that node was in each song-stability state. Dark purple colors represent small syllable repertoires while white represents large repertoires. For the sake of visualization, the color range was truncated based on the distribution of the data, such that the lowest value was the 25th percentile minus the range of the 25th to 50th percentile and the highest value was the 75th percentile plus the range of the 50th to 75th percentile. See **Table 3.1** for `PhylANOVA` results.

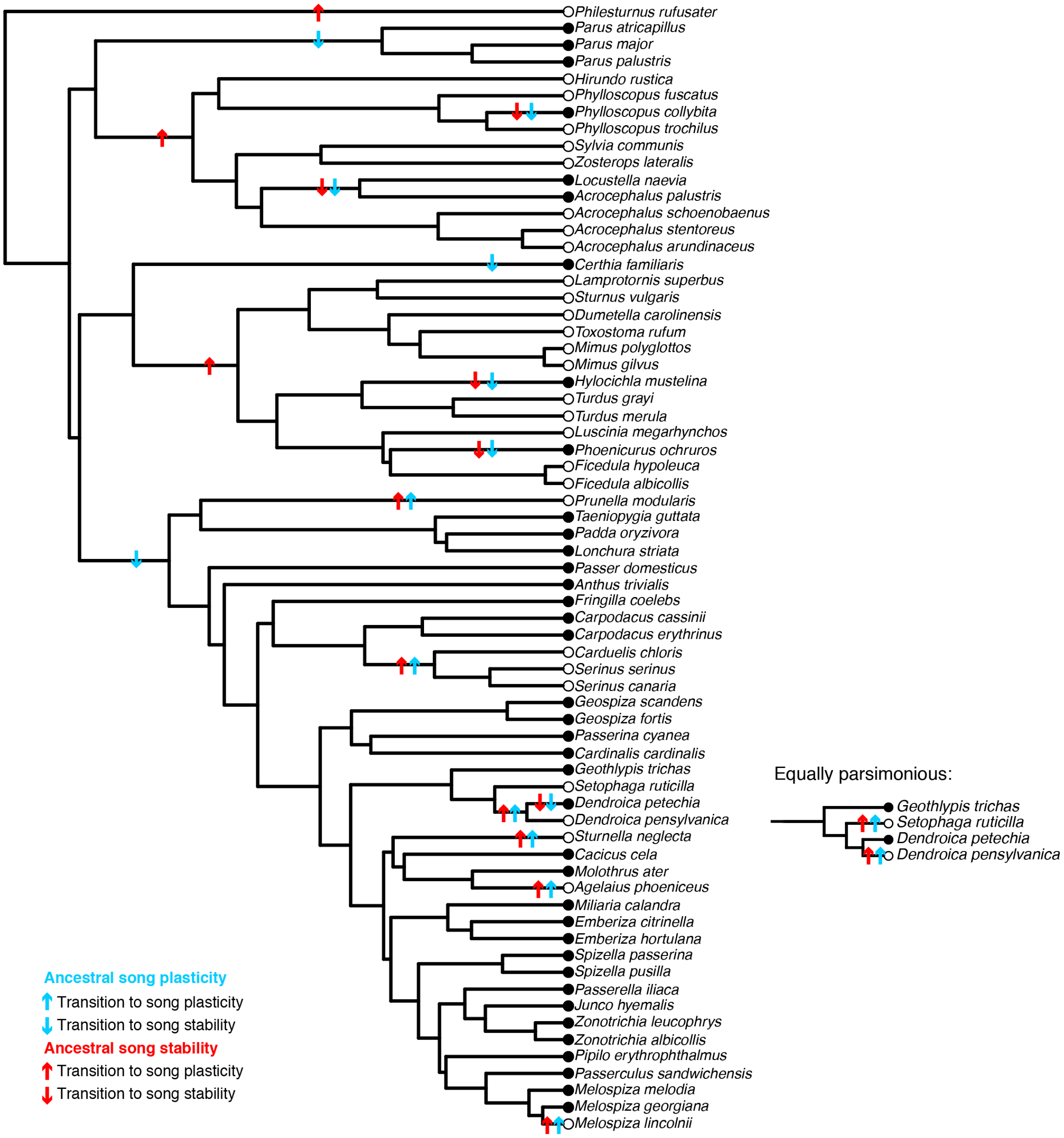


Figure 3.2: The minimum number of evolutionary transitions required to recapitulate the current song stability states of birds in this study. White dots at the tips show song-plastic species, while black dots show song-stable species. Inset shows an equally parsimonious set of transitions for the labeled species. At least 7 transitions to the song-stable state are required to generate the current distribution of adult song plasticity from a song-plastic last common ancestor. At least 9 transitions to the song-plastic state are required if the last common ancestor was a song-stable species, and 14 transitions between adult song stability and adult song plasticity are required, regardless of whether we assume a song-plastic or song-stable ancestral state.

3.3.2 Song trait differences based on song stability state

We tested whether song traits were affected by the length of the song-learning window on an evolutionary scale. Intuitively, it makes sense that a species that has a longer time-window to learn might be able to accumulate a larger repertoire. Indeed, this relationship is suggested by the pattern of song stability and repertoire size in *Phylloscopus* species (**Figure 3.3**); however, many individual species do not follow this prediction: for example, *Acrocephalus palustris* appears to learn a large repertoire in a single year[165], and *Philesturnus rufusater* modifies its song for multiple years but maintains a small repertoire[188]. Further, numerous species with adult song plasticity do not significantly increase their repertoire sizes over time[189–192]. Thus, it is important to test the potential evolutionary link between larger repertoire sizes and adult song plasticity. Using a phylogenetically-controlled ANOVA[135,193], we found that species with adult song plasticity did possess significantly larger syllable repertoires than species with adult song stability (**Figure 3.1A**, **Figure 3.4A**, and **Table 3.1**). This concurs with a previous analysis using a smaller dataset[150]. Similarly, we found that song-plastic species had significantly larger song repertoires than song-stable species (**Figure 3.4B**, **Figure 3.5**, and **Table 3.1**).

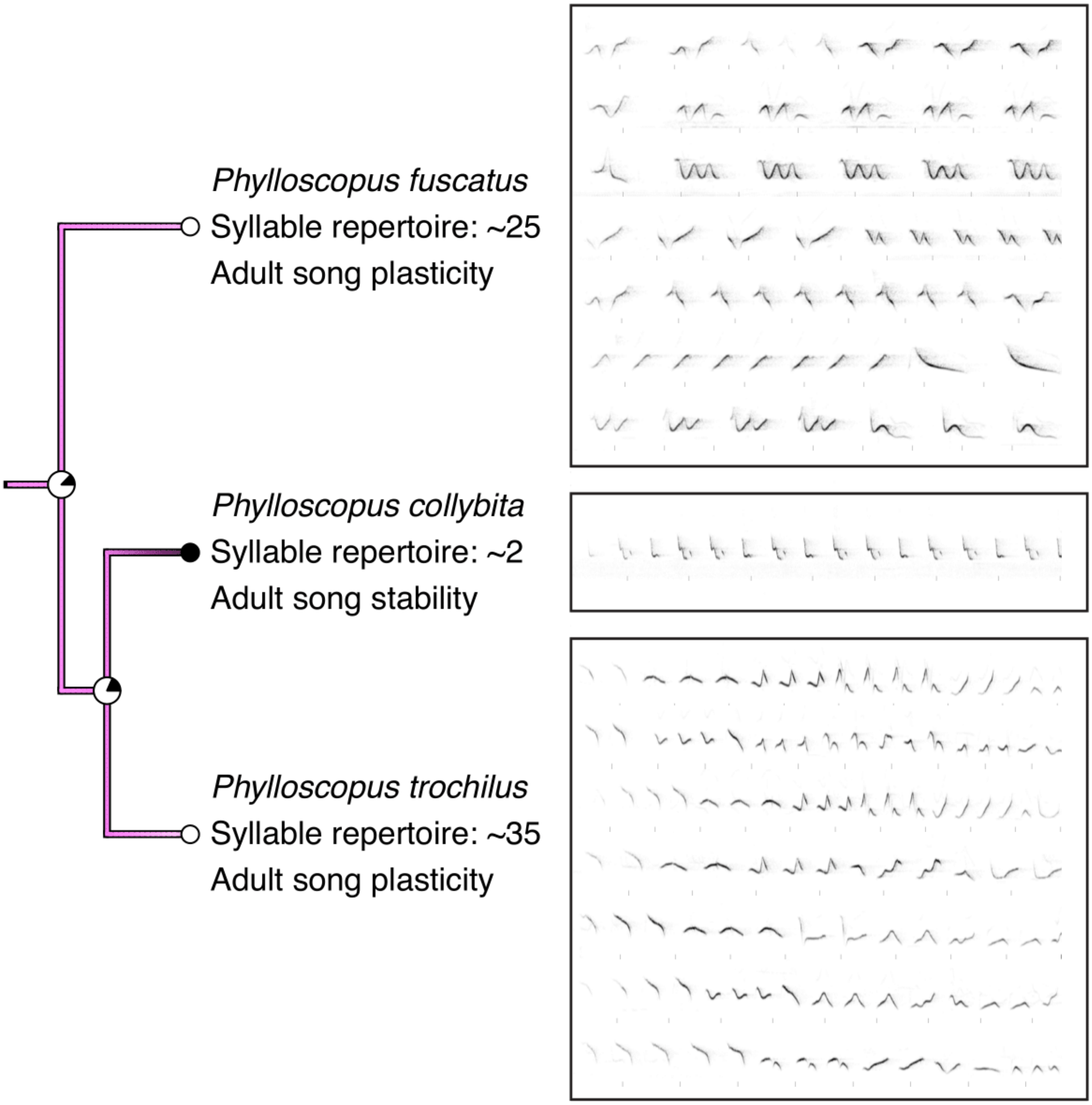


Figure 3.3: Recent transitions in song stability state and syllable repertoire size in *Phylloscopus* species. *P. collybita*, a species with adult song stability, has a smaller syllable repertoire size than *P. trochilus* and *P. fuscastus*, two species with adult song plasticity. Colors of branches and nodes correspond with Figure 2. Sonograms were generated from recordings obtained from xeno-canto.org: XC340281 recorded by Tom Wulf (*P. fuscatus*, accessible at: www.xeno-canto.org/XC340281), XC414221 recorded by Frank Lambert (*P. collybita*, accessible at: www.xeno-canto.org/XC414221), and XC402265 recorded by Hans Matheve (*P. trochilus*, accessible at: www.xeno-canto.org/XC402265). Sonograms are used only to demonstrate comparative repertoire size from one individual for each species and were stretched horizontally to fit the allotted space.

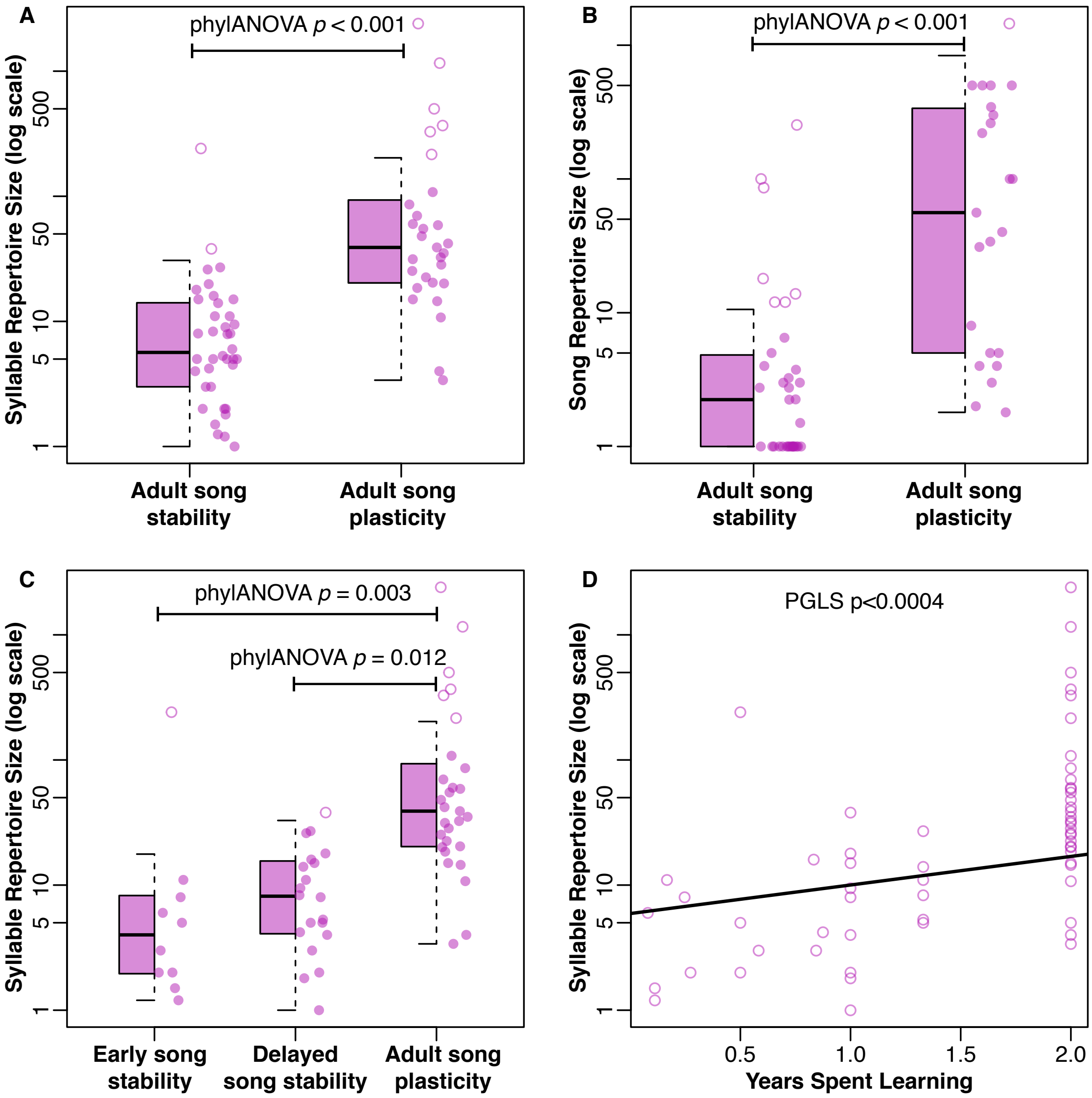


Figure 3.4: Distribution of repertoire sizes in species with different song stability states. (A) shows the distribution of syllable repertoires and (B) shows the distribution of song repertoires when species are broken into two groups based on song stability. (C) shows the distribution of syllable repertoires when species are broken into three groups based on song stability. Boxes indicate the 25th, 50th, and 75th percentile. The lower whisker is either the minimum value or the 25th percentile minus 1.5 times the interquartile range, whichever was larger. The upper whisker is either the maximum value or the 75th percentile plus 1.5 times the interquartile range, whichever was smaller. Dots are the raw values as a scatter plot. Solid dots are within the range of the box and whiskers, while open dots are outliers. See **Table 3.1** for full `PhylANOVA` results for (A-C). (D) shows the continuous relationship between syllable repertoire size and song stability when song stability is truncated at 2 years due to lack of data in subsequent years.

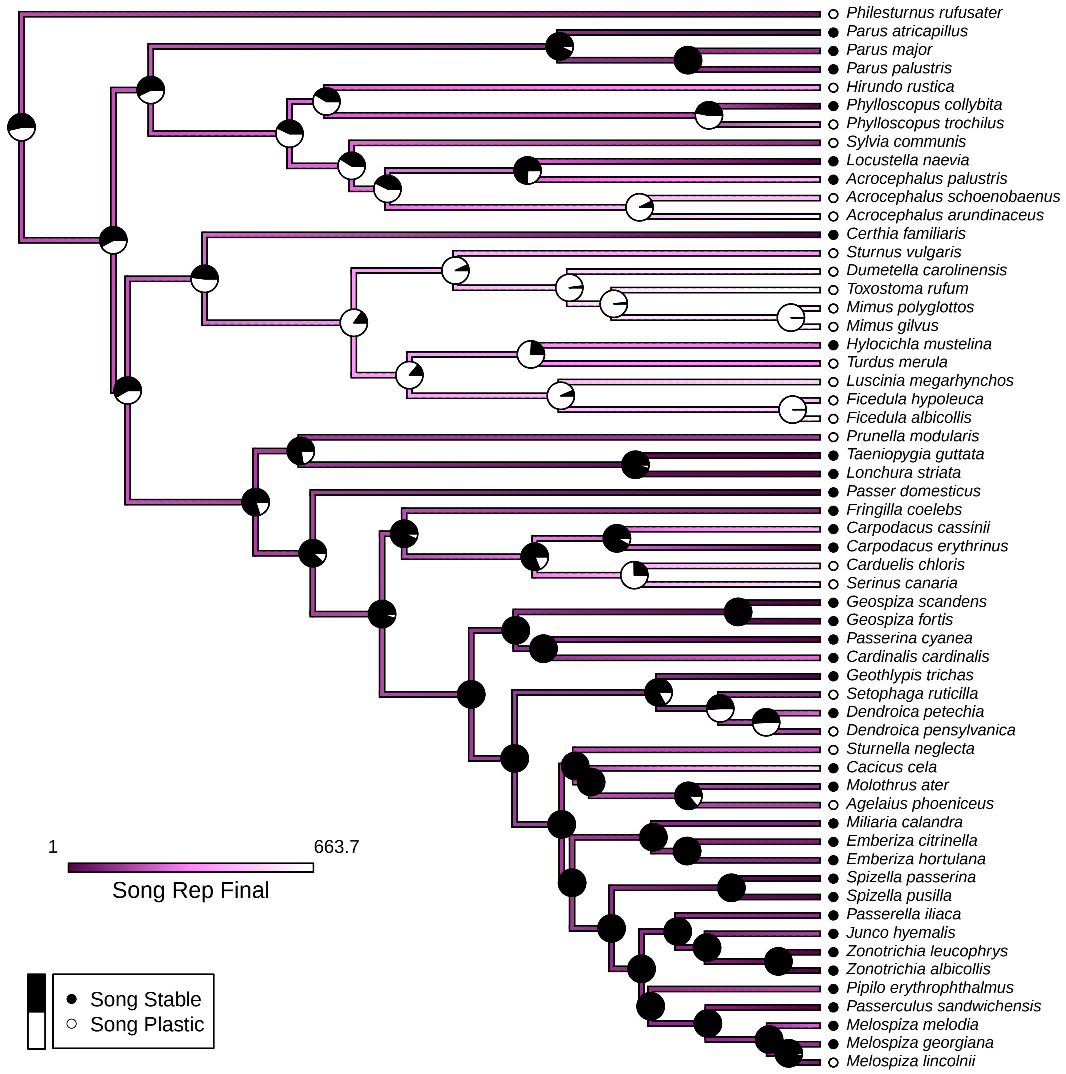


Figure 3.5: There was a relationship between song repertoire and song stability when controlling for phylogeny. The estimated ancestral character states are mapped on the tree for both adult song stability state and for log-transformed song repertoire size. Black and white dots represent a species that is currently in a stable or plastic state respectively. Black and white in the pie charts at each node represent the likelihood that the common ancestor was in the stable or plastic state. Dark purple colors represent small repertoires while white represents large repertoires. For the sake of visualization, the color range was truncated based on the distribution of the data, such that the lowest value was the 25th percentile minus the range of the 25th to 50th percentile and the highest value was the 75th percentile plus the range of the 50th to 75th percentile. `PhylANOVA` results for **Figures 3.5** to **3.10** are available in **Table 3.1**.

Table 3.1: `PhylANOVA` results for all song traits when birds are divided into species with adult song stability or adult song plasticity. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote significant groups.

Song Trait	Song-Stable	Song-Plastic	F-value	Corrected α	*p*-value
Syllable Rep	1.8807	3.946	41.5064	0.0071	<0.001*
Song Rep	1.1055	3.8688	33.8334	0.0083	<0.001*
Syll\Song	1.2556	2.2962	9.2658	0.01	0.094
Duration	0.7736	1.2927	2.0783	0.0125	0.42
Continuity	−1.3453	−1.0286	2.1537	0.0167	0.474
Interval	1.6075	1.218	1.3879	0.025	0.567
Song Rate	1.8969	2.0971	0.6079	0.05	0.713

There were no significant differences between song-plastic and song-stable species for the other song traits that we tested: syllables per song, inter-song interval, song duration, song continuity, or song rate (**Table 3.1** and **Figures 3.6** to **3.10**). When we used the classification scheme with three states, we could only test for differences in syllable repertoire, song repertoire, and syllables per song between groups, as there were very few early song-stable species for which we had data on the other song traits. We found no significant differences between early song-stable and delayed song-stable species for any tested traits, but both of these groups had significantly smaller syllable and song repertoires compared to song-plastic species (**Figure 3.1B**, **Figure 3.4C**, and **Tables 3.2** and **3.3**). When performing a PGLS analysis using the continuous values for song-stability, we found similar results; both syllable repertoire and song repertoire were correlated with the song plasticity duration, such that repertoire size increased as the duration of song plasticity increased (**Figure 3.4D** and **Table 3.4**).

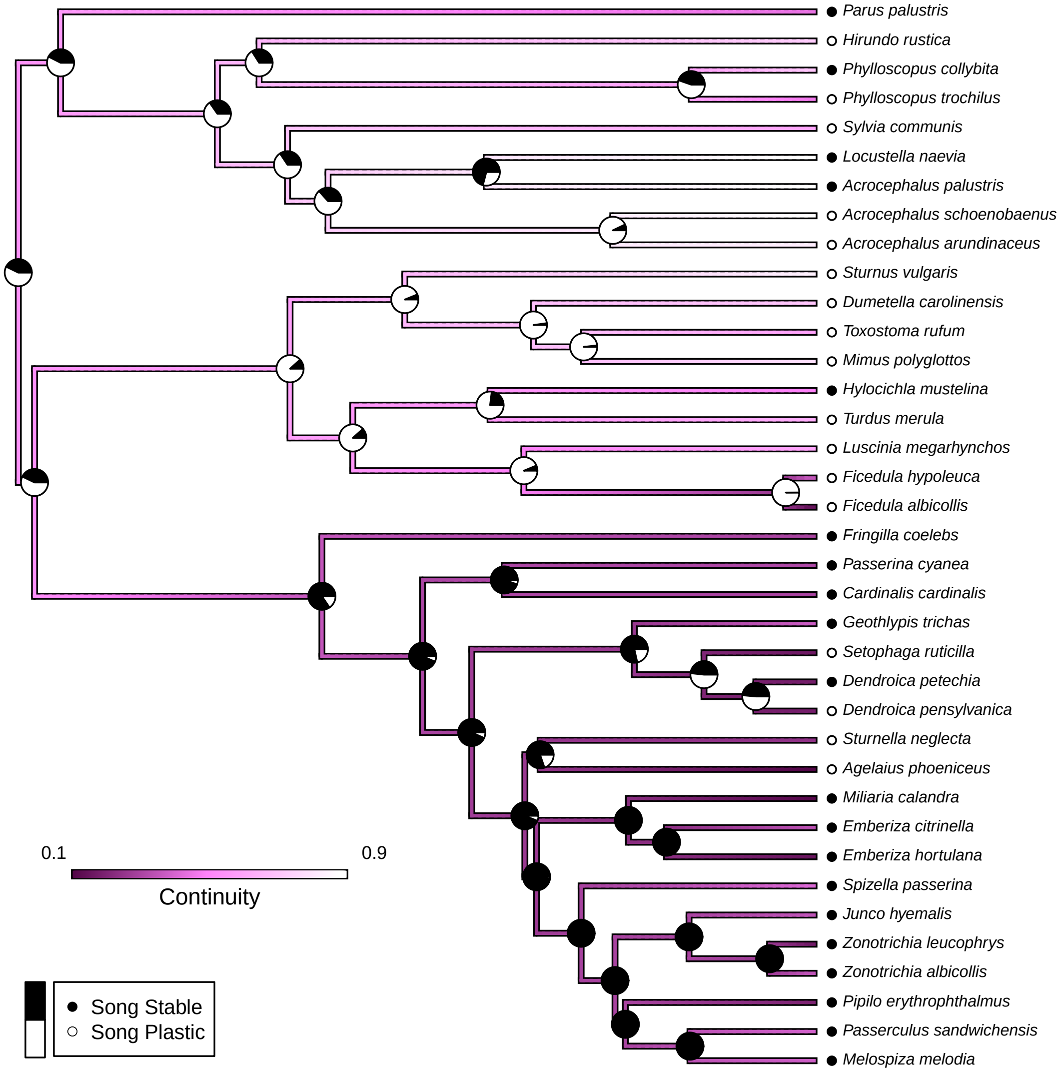


Figure 3.6: There was no relationship between song continuity and song stability when controlling for phylogeny. The estimated ancestral character states are mapped on the tree for both adult song stability versus plasticity (black versus white) and for log-transformed song continuity (purple). Labeling is the same as in **Figure 3.5**.

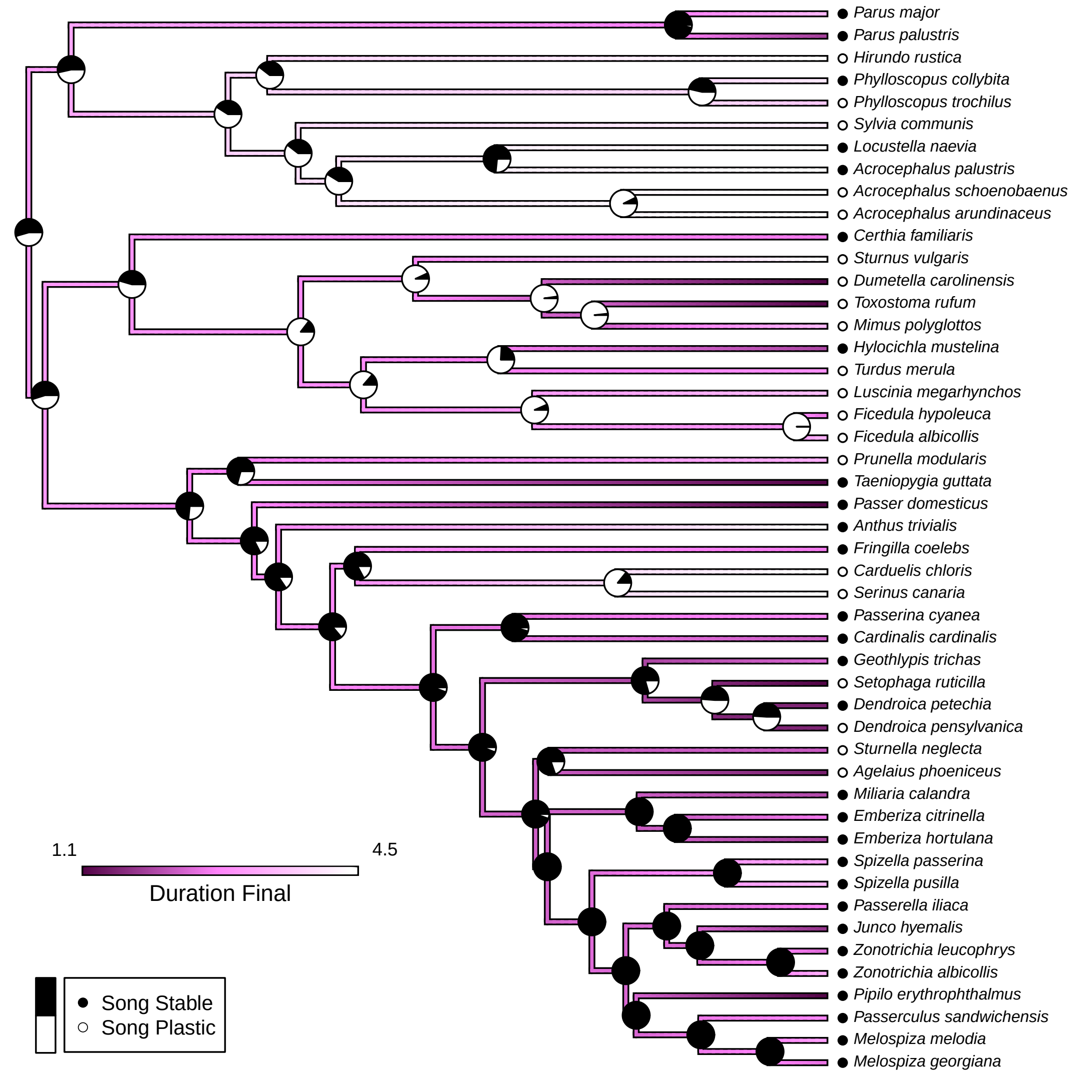


Figure 3.7: There was no relationship between song duration and song stability when controlling for phylogeny. The estimated ancestral character states are mapped on the tree for both adult song stability versus plasticity (black versus white) and for log-transformed song duration (purple). Labeling is the same as in **Figure 3.5**.

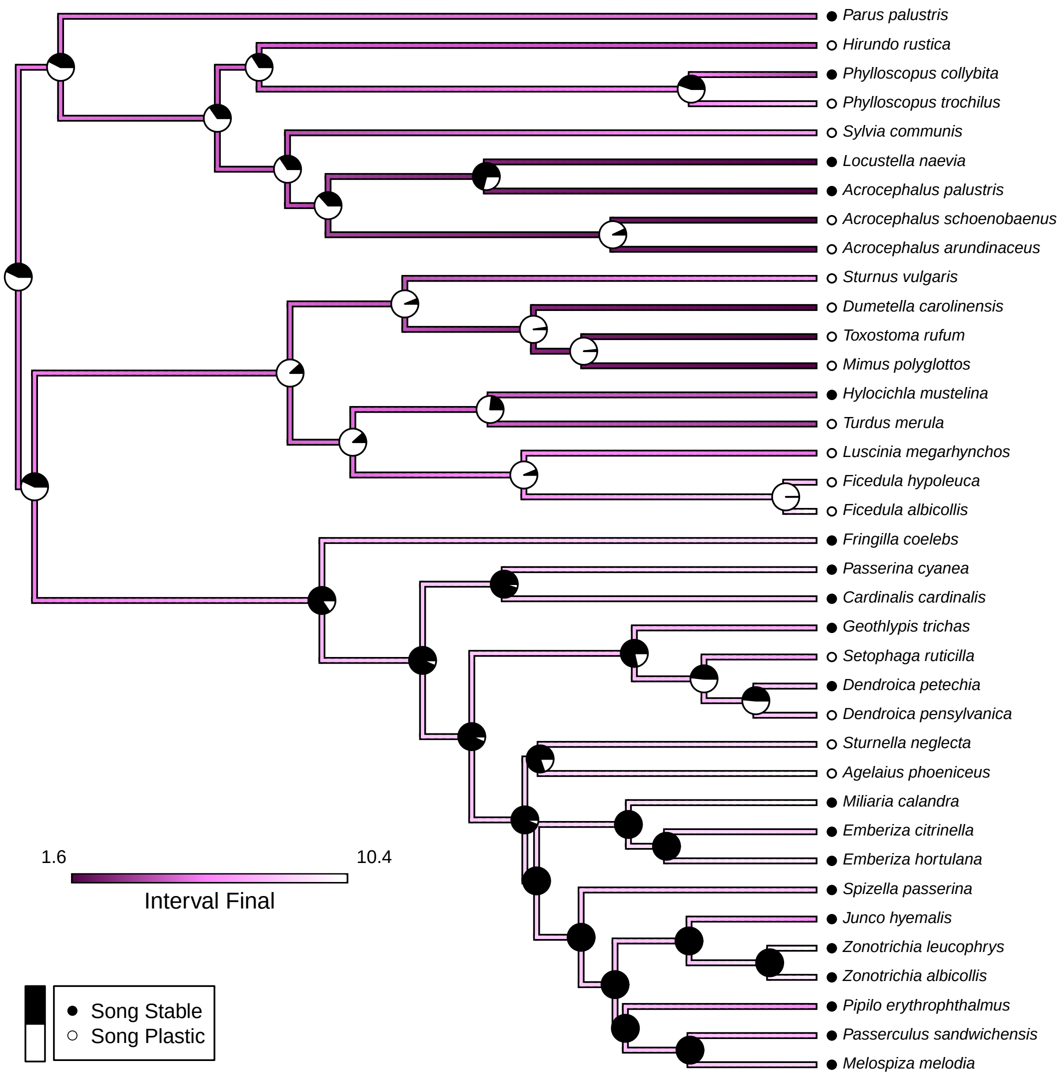


Figure 3.8: There was no relationship between intersong interval and song stability when controlling for phylogeny. The estimated ancestral character states are mapped on the tree for both adult song stability versus plasticity (black versus white) and for log-transformed song interval (purple). Labeling is the same as in **Figure 3.5**.

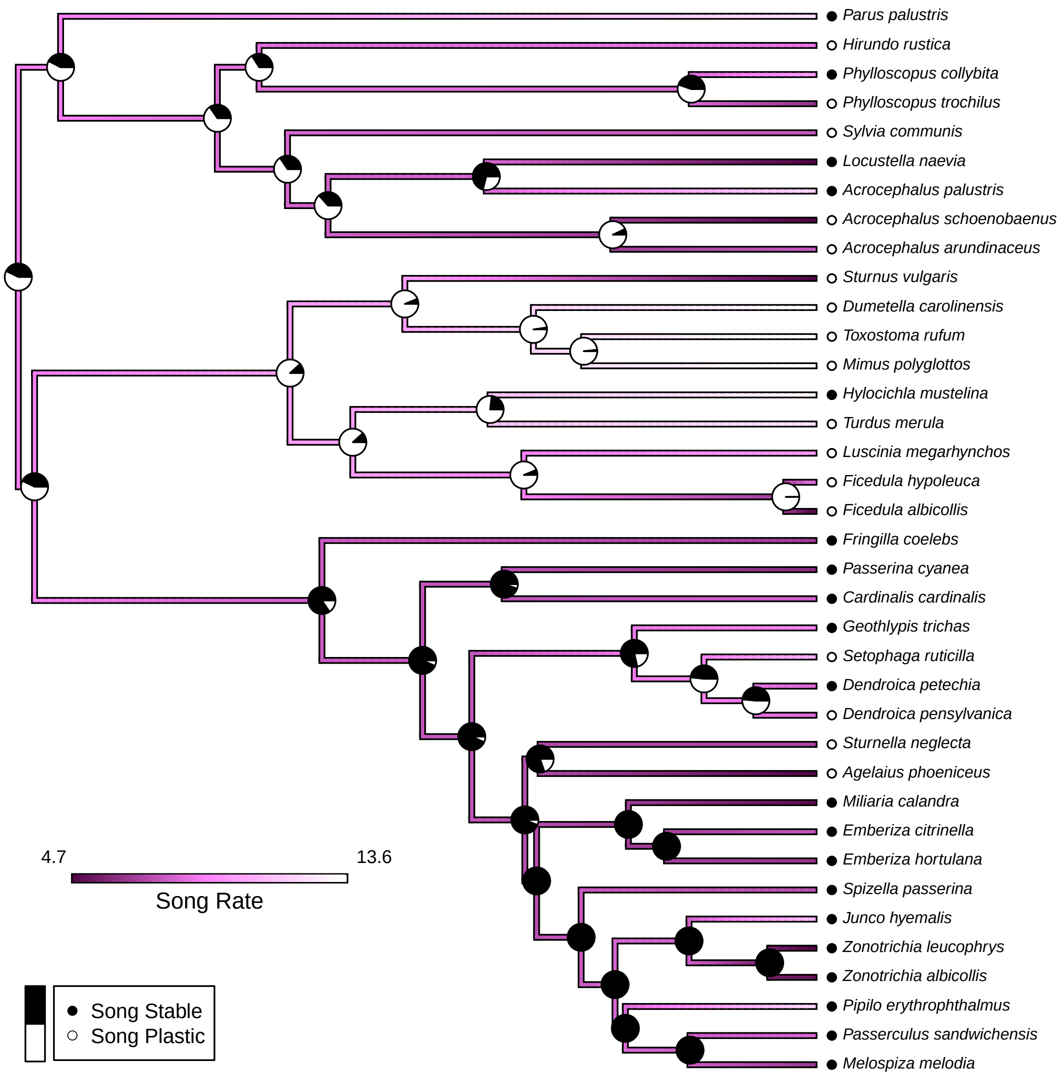


Figure 3.9: There was no relationship between song rate and song stability when controlling for phylogeny. The estimated ancestral character states are mapped on the tree for both adult song stability versus plasticity (black versus white) and for log-transformed song rate (purple). Labeling is the same as in **Figure 3.5**.

Figure 3.10: There was no relationship between syllables per song and song stability when controlling for phylogeny. The estimated ancestral character states are mapped on the tree for both adult song stability versus plasticity (black versus white) and for log-transformed syllables per song (purple). Labeling is the same as in **Figure 3.5**.

Table 3.2: `PhylANOVA` results for all song traits when birds are divided into early song stability, delayed song stability, and song plasticity. Song traits are sorted from most to least significant. Early, Delayed, and Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote significant groups.

Song Trait	Early	Delayed	Plastic	F-value	Corrected α	p-value
Syllable Rep	1.6436	2.0062	3.946	17.1099	0.0071	0.003*
Song Rep	0.6788	1.4819	3.8688	12.88	0.0083	0.011*
Syll\Song	1.2852	1.2467	2.2962	3.6877	0.01	0.252

Table 3.3: Post-hoc pairwise `phylANOVA` tests for significant song traits when birds are divided into early song stability, delayed song stability, and song plasticity. Asterisks (*) denote significant groups.

Song Trait	State 1	State 2	T-value	*p*-value
Syllable Rep	Plastic	Delayed	4.8995	0.012*
Syllable Rep	Early	Plastic	4.6091	0.003*
Syllable Rep	Early	Delayed	0.6872	0.659
Song Rep	Plastic	Delayed	4.0268	0.044*
Song Rep	Early	Plastic	4.3074	0.015*
Song Rep	Early	Delayed	1.0444	0.55

Table 3.4: Results of PGLS analysis between song traits and continuous song stability. Test performed on the natural log scale values of song traits. λ is the value by which off-diagonal elements in the Brownian motion model are multiplied to make the correlation structure. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Song traits are sorted from most to least significant. Asterisks (*) denote significant slopes.

Song Trait	Slope	Std Error	λ	T-value	Corrected α	*p*-value
Syllable Rep	0.9067	0.2449	0.8913	3.7021	0.0071	<0.001*
Song Rep	1.1013	0.3123	0.8316	3.5263	0.0083	<0.001*
Syll\Song	0.3701	0.2224	0.4699	1.6642	0.01	0.1029
Interval	0.4221	0.2646	0.8823	1.5953	0.0125	0.1215
Continuity	−0.2135	0.1439	0.8832	−1.4838	0.0167	0.1486
Duration	0.3702	0.2569	1.0163	1.441	0.025	0.1578
Song Rate	−0.2113	0.25	0.7307	−0.8453	0.05	0.4048

3.3.3 The effect of song stability on the rate of song trait evolution

Our result that species with adult song plasticity had significantly larger syllable and song repertoires raised the question of whether song stability versus plasticity also affected the rate of evolution for any of the song traits. To examine this possibility, we used the `Brownie` algorithm[194], which tests whether a model with two rates of evolution for each song traitone rate for ancestral periods of song stability and another rate for song plasticityfits the data significantly better than a model that allows for only one rate of evolution of each song trait regardless of the ancestral states of song stability. Each calculation of the two-rate model is based on one stochastic projection of the ancestral traits across the phylogenetic tree, so we generated 1,300 different stochastic simulation maps to use with Brownie. We plotted the distribution of potential rates and compared the average log likelihood of the two-rate models to the log likelihood of the one-rate model.

We found that allowing for two different rates of song trait evolution depending on song stability or plasticity did not lead to a significantly better fit model than using one Brownian-motion rate for either syllable repertoire size or song repertoire size, even though syllable repertoires and song repertoires were both significantly larger in species with adult song plasticity (**Figure 3.11A** and **B** and **Table 3.5**). In contrast, the two-rate model led to a significantly better fit for syllables per song, song rate, inter-song interval, and song duration (**Figure 3.11C**, **Figure 3.12**, and **Table 3.5**), indicating that evolution of these song traits was faster in song-plastic lineages (**Figure 3.11C** and **Figure 3.12**, red traces).

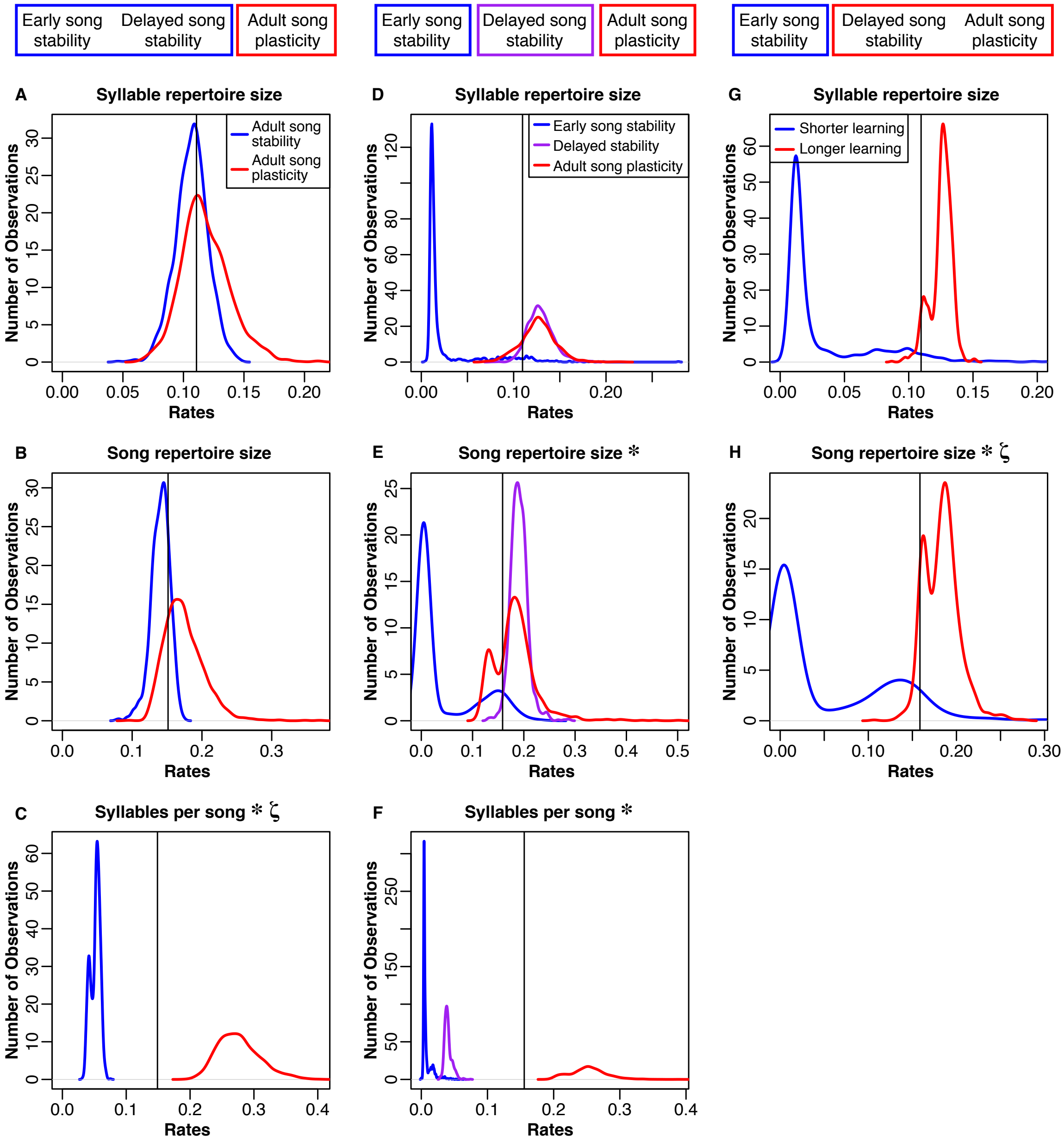


Figure 3.11: Distributions of rates for natural-log transformed song traits related to elaboration. The boxes at the top illustrate how we grouped the species for each model. **Column 1 (A-C):** Blue traces are song-stable, while red traces are song-plastic. **Column 2 (D-F):** Blue traces are early song-stable, purple traces are delayed song stable, and red traces are song-plastic. **Column 3 (G-H):** Blue traces are early song-stable, while red traces are delayed song-stable and song-plastic combined. The black line shows the rate value for the one-rate model in all columns. Asterisks indicate that the rate of evolution of that song trait significantly differed between groups. Lowercase zeta (ζ) indicates the multi-rate model that best fit the data while using the fewest number of rates. In the case of syllable repertoire, the multi-rate models were not significantly better than the one-rate model). The three-rate models (D-F) were not significantly better than the two-rate models (G,H, and C). See **Tables 3.5** to **3.9** for chi-square test results.

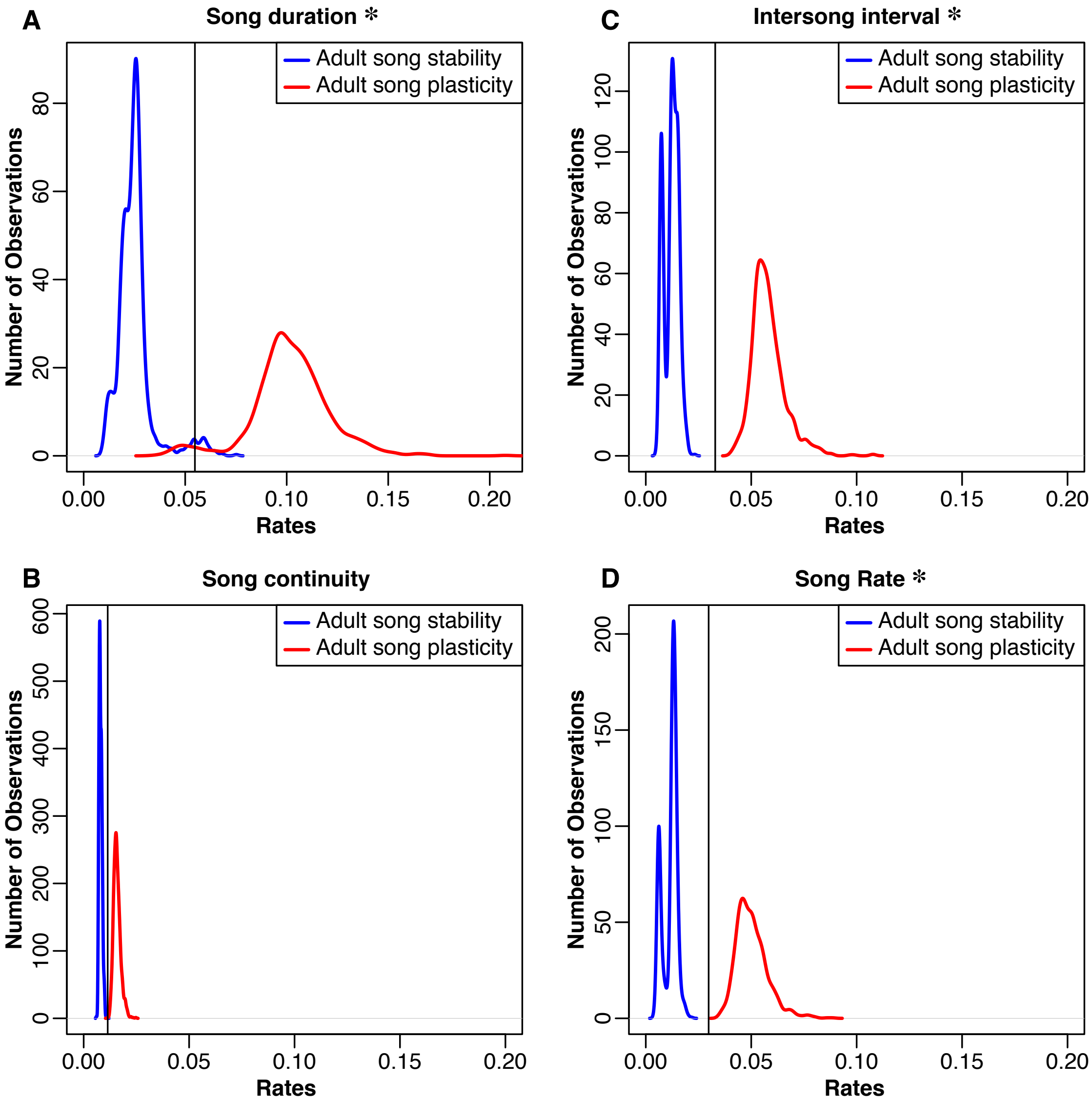


Figure 3.12: Distributions of rates for natural-log transformed song traits related to performance. Blue traces song-stable, while red traces are song-plastic. The black line shows the rate for the one-rate model. Asterisks indicate that the rate of evolution of that song trait significantly differed between song-stable and song-plastic lineages. See **Table 3.5** for chi-square test results.

Table 3.5: Brownie results for song traits when birds are divided into species with adult song stability or adult song plasticity. For **Tables 3.5** to **3.9**, rate columns show mean log likelihood. Song traits are sorted from most to least significant. Asterisks (*) denote traits where the more complex model fit the data significantly better than the simpler model.

Song Trait	One Rate	Two Rates	*p*-value
Syll\Song	−110.6482	−100.7673	<0.001*
Song Rate	−43.4397	−38.4938	0.002*
Interval	−45.2842	−40.5004	0.002*
Duration	−71.2042	−66.3122	0.002*
Continuity	−25.6471	−24.7285	0.175
Syllable Rep	−120.2983	−120.0695	0.499
Song Rep	−113.5829	−113.3706	0.515

We repeated this analysis with the three-state categorization for syllable repertoire, song repertoire, and syllables per song; for other song traits, we did not have enough species in the early song stability group. We found that the three-rate model was significantly better than the one-rate model for syllables per song and song repertoire, but not for syllable repertoire (**Figure 3.11A-C** and **Table 3.6**). However, the three-rate model was only significantly better than the two-rate model for song repertoire (**Table 3.7**). Thus, the two-rate model sufficiently approximates the evolution of syllables per song. We noticed that for both song repertoire and syllable repertoire, the rate of evolution in delayed song-stable lineages (purple traces in **Figure 3.11 D** and **E**) was very similar to the rate in song-plastic lineages (corresponding red traces). We tested one more set of models where we combined delayed song-stable species with song-plastic species to create a "longer learning" group, while early song-stable species were assigned to a "shorter learning" group. For this comparison of shorter versus longer learning, the two-rate model was significantly better than the one-rate model for song repertoire and trending in that direction for syllable repertoire (**Figure 3.11 G** and **H** and **Table 3.8**). The three-rate model was not significantly better than the longer/shorter-learning two-rate model for either syllable or song repertoire (**Table 3.9**). Taken together with our `phylANOVA` results, this pattern suggests that species with early song stability evolve their song repertoires and potentially their syllable repertoires at a slower rate than delayed song-stable and song-plastic species; however, only song-plastic species directionally evolve towards larger song and syllable repertoires.

Table 3.6: `Brownie` results for song traits when birds are divided into early song stability, delayed song stability, and song plasticity.

Song Trait	One Rate	Three Rates	*p*-value
Syll\Song	−97.8349	−86.3206	<0.001*
Song Rep	−100.812	−97.7647	0.014*
Syllable Rep	−107.3206	−105.5895	0.063

Table 3.7: `Brownie` results for song traits when birds are divided into either song stability (early and delayed) and song plasticity (Two Rates) or early song stability, delayed song stability, and song plasticity (Three Rates).

Song Trait	Two Rates	Three Rates	*p*-value
Song Rep	−100.691	−97.7148	0.015*
Syllable Rep	−107.1332	−105.5532	0.075
Syll\Song	−86.3125	−86.3447	1

Table 3.8: `Brownie` results for song traits when birds are divided into shorter learning (early song stability) and longer learning (delayed song stability and song plasticity).

Song Trait	One Rate	Two Rates	*p*-value
Song Rep	−100.812	−97.9918	0.018*
Syllable Rep	−107.3206	−105.8488	0.086

Table 3.9: `Brownie` results for song traits when birds are divided into either shorter learning (early song stability) and longer learning (delayed song stability and song plasticity) (Two Rates) or early song stability, delayed song stability, and song plasticity (Three Rates).

Song Trait	Two Rates	Three Rates	*p*-value
Syllable Rep	−105.8156	−105.5532	0.469
Song Rep	−97.9372	−97.7148	0.505

3.3.4 Evolutionary interactions between song stability and the evolution of song traits or mating strategies

We used `BayesTraits`[184,185] to test whether the rate and order of evolutionary transitions in one trait is dependent on the state of another trait. Because the song traits were continuous variables, we binarized them by setting a series of threshold values to delineate “low” and “high” categories, using each observed song trait value as a threshold in turn. We then tested whether there was correlated evolution between the binary classifications of adult song stability and each of the seven song traits.

In the lowest third of syllable repertoire thresholds, adult song plasticity with small syllable repertoires was an evolutionarily unstable state, with rapid transitions primarily towards a song-stable state and secondarily towards larger syllable repertoires (82% of runs significant in this range, **Figure 3.13**). In the middle third of syllable repertoire thresholds, song-stability with smaller syllable repertoires is an evolutionarily stable attractor state, with high rates of transition observed from large to small syllable repertoires in song-stable species and from plasticity to stability with a small syllable repertoire. These rate differences are highly significant (100% of runs significant in this range). In the highest third of syllable repertoire thresholds, adult song stability with a large syllable repertoire is an evolutionarily unstable state, transitioning primarily towards adult song plasticity (86% of runs significant in this range, **Figure 3.13**). We found similar trends when using two, four, or five bins with subtle differences. When using four or five bins, we still observe that song stability with larger syllable repertoires is an unstable combination; however, for the highest bin of threshold values, the transition rates are faster when changing to song plasticity, whereas for the second-highest bin, we observe faster transition rates toward repertoire size increases (see **Figure 6 - figure supplement 1** from Robinson, Snyder, and Creanza 2019[70]).

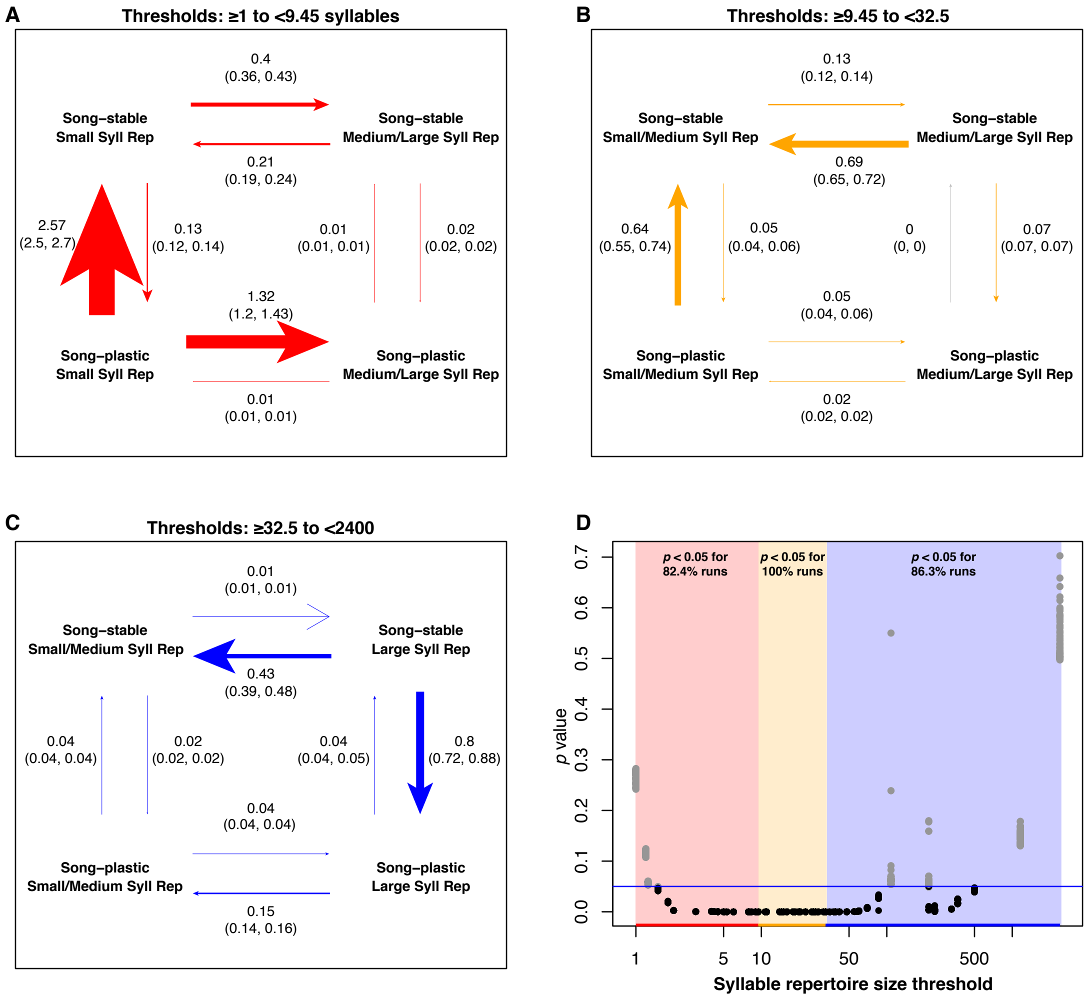


Figure 3.13: Analysis of correlated evolution between adult song stability and syllable repertoire size. We repeated the BayesTraits test using each value of the continuous song trait as the threshold delineating the larger and smaller syllable repertoire groups. We performed a total of 100 runs per threshold. We pooled the results of all of the runs into three groups based on whether the threshold was in the lowest, middle, or highest third of the unique trait values. Within these groups, we computed the mean percentage of runs that were significant at $p<0.05$ at each threshold. (A-C) Rate of transition plots when the lowest (red), middle (yellow), and highest (blue) thirds of the unique syllable repertoire values in the dataset were used as the threshold. Rates are the average across all runs when the threshold denoting small/large repertoire sizes was defined as each value within each segment. Arrows are labeled with the mean rate and the 95% confidence interval. Arrow weights are scaled to the mean rate values. (D) p-values from the 100 runs per threshold, plotted against threshold. Colored bars denote low, middle, and high threshold segments. Blue line denotes p=0.05.

At low song repertoire threshold values, song plasticity with a small song repertoire is an evolutionarily unstable state; there is rapid transition away from this combination, predominantly trending towards song stability but also transitioning secondarily to a larger song repertoire, with very high significance (100% of runs significant in this range). At moderate song repertoire thresholds, the highest rate is observed for species with small repertoires transitioning from song plasticity to song stability, also with very high significance (100% of runs significant in this range). At high song repertoire thresholds, the primary shift is from song stability to song plasticity in species with large song repertoires (89% of runs significant in this range) (**Figure 3.14**). When analyzing the results using five bins, transitions in the upper range of song repertoire values becomes more nuanced; in the highest bin, song stability with a larger song repertoire is very unstable, but is relatively stable in the second-highest bin. In the lower four bins, the dominant transition is from plastic to stable song with smaller song repertoires (see **Figure 6 - figure supplement 1** from Robinson, Snyder, and Creanza 2019[70]). The results for syllables per song show some general trends that are complicated by the strong effect of the mimid species (**Figure 3.15**, also see results of jackknife analysis below). We did not find evidence for correlated evolution between adult song stability or plasticity state and any of the other song traits (**Figures 3.16** to **3.19**, also see **Figure 6 - figure supplement 1** from Robinson, Snyder, and Creanza 2019[70]).

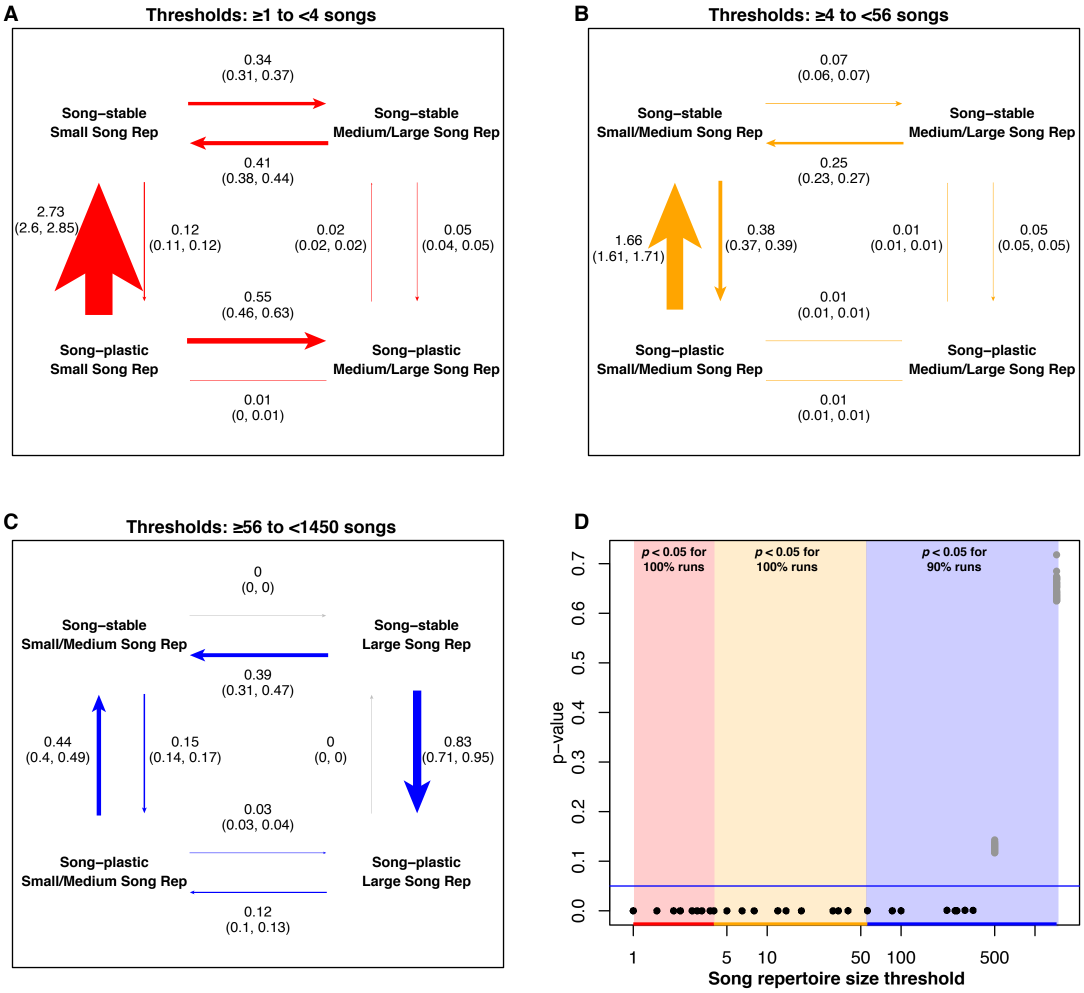


Figure 3.14: Analysis of correlated evolution between adult song stability and song repertoire size. Labeling is the same as in **Figure 3.13**.

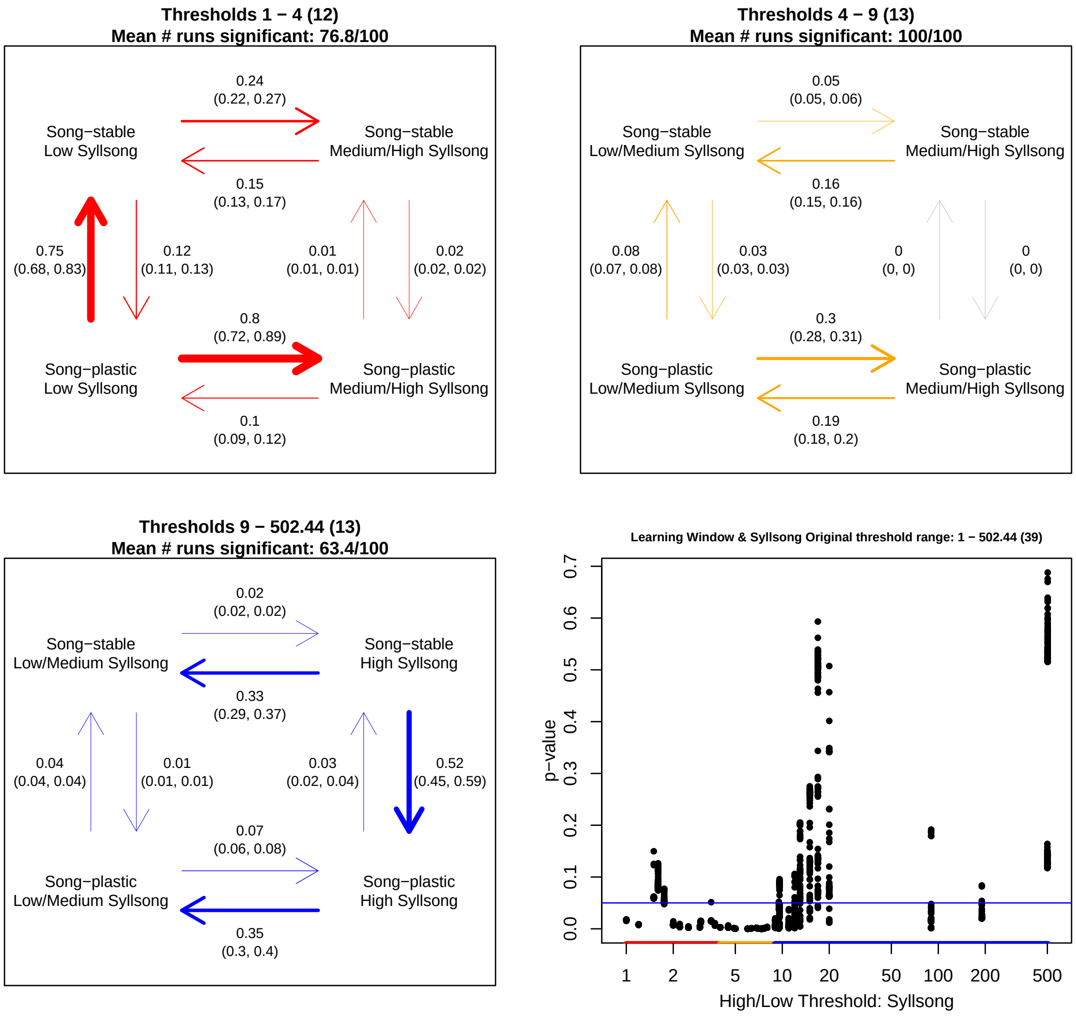


Figure 3.15: Analysis of correlated evolution between adult song stability and syllables per song. Labeling the same as in **Figure 3.13**.

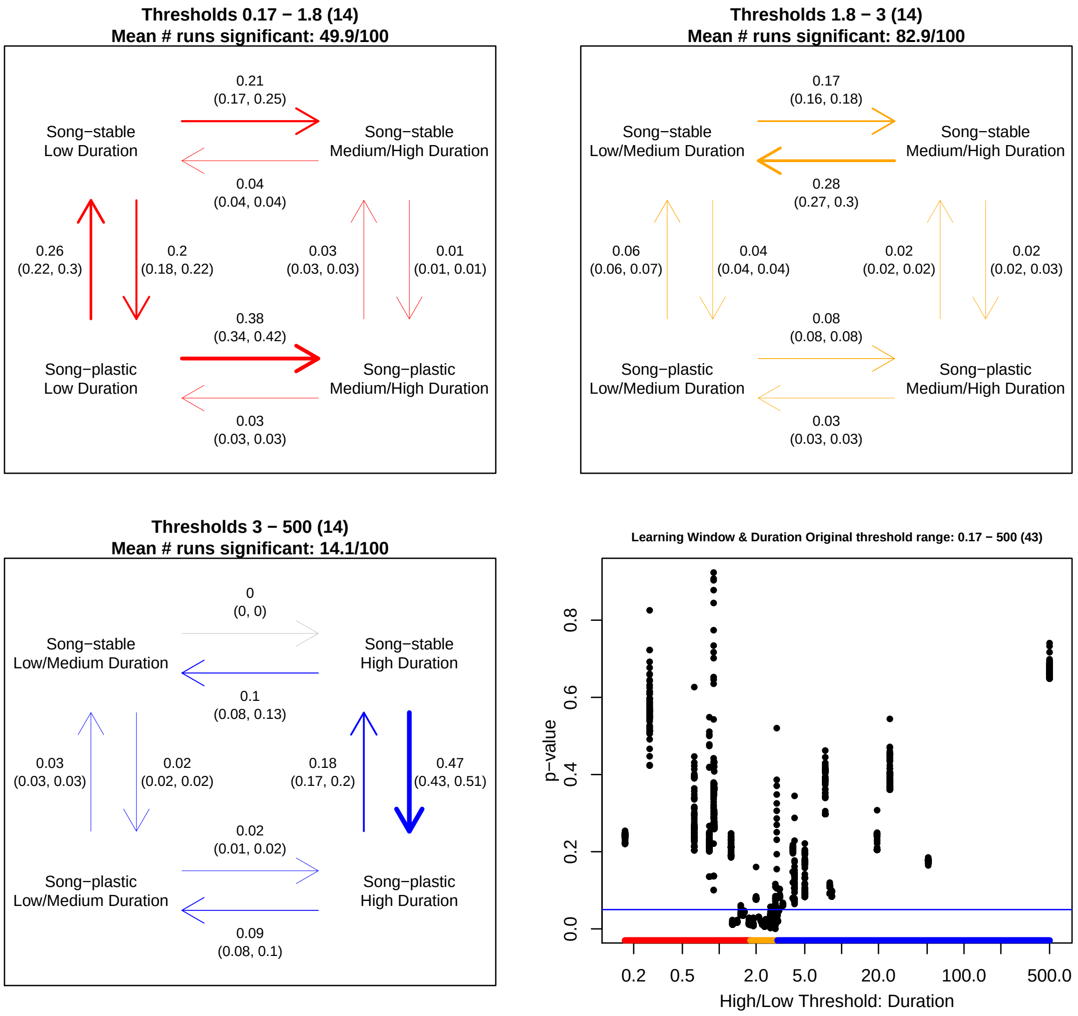


Figure 3.16: Analysis of correlated evolution between adult song stability and song duration. Labeling the same as in **Figure 3.13**.

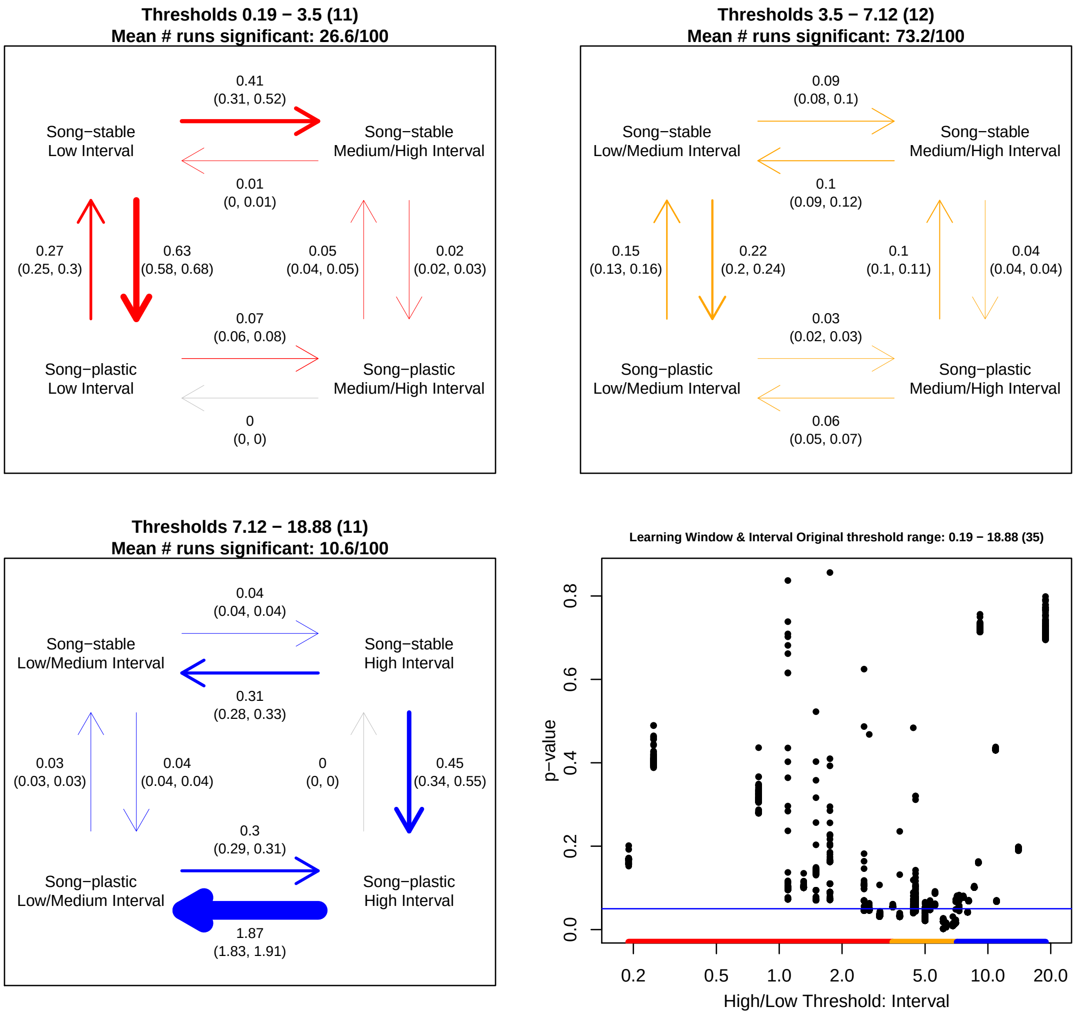


Figure 3.17: Analysis of correlated evolution between adult song stability and intersong interval. Labeling the same as in **Figure 3.13**.

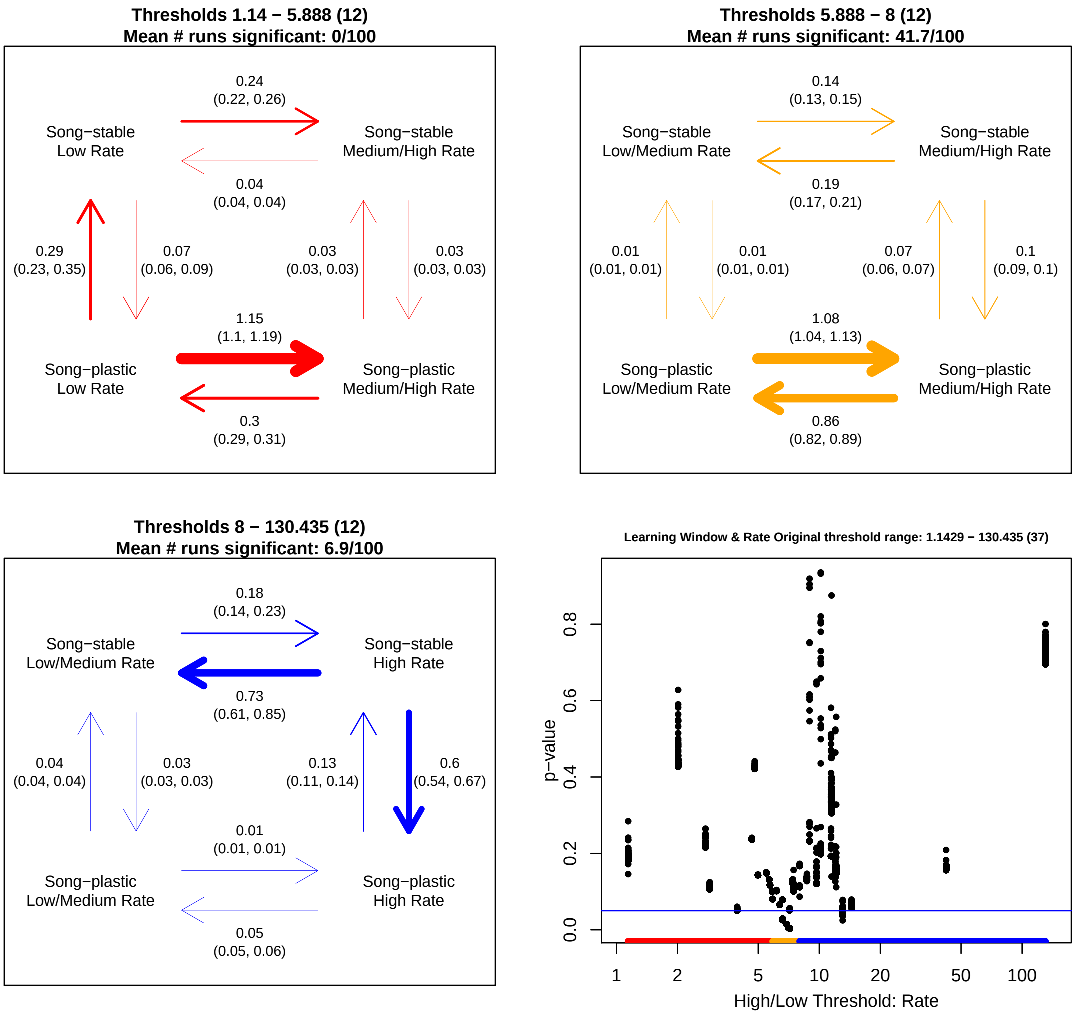


Figure 3.18: Analysis of correlated evolution between adult song stability and song rate. Labeling the same as in **Figure 3.13**.

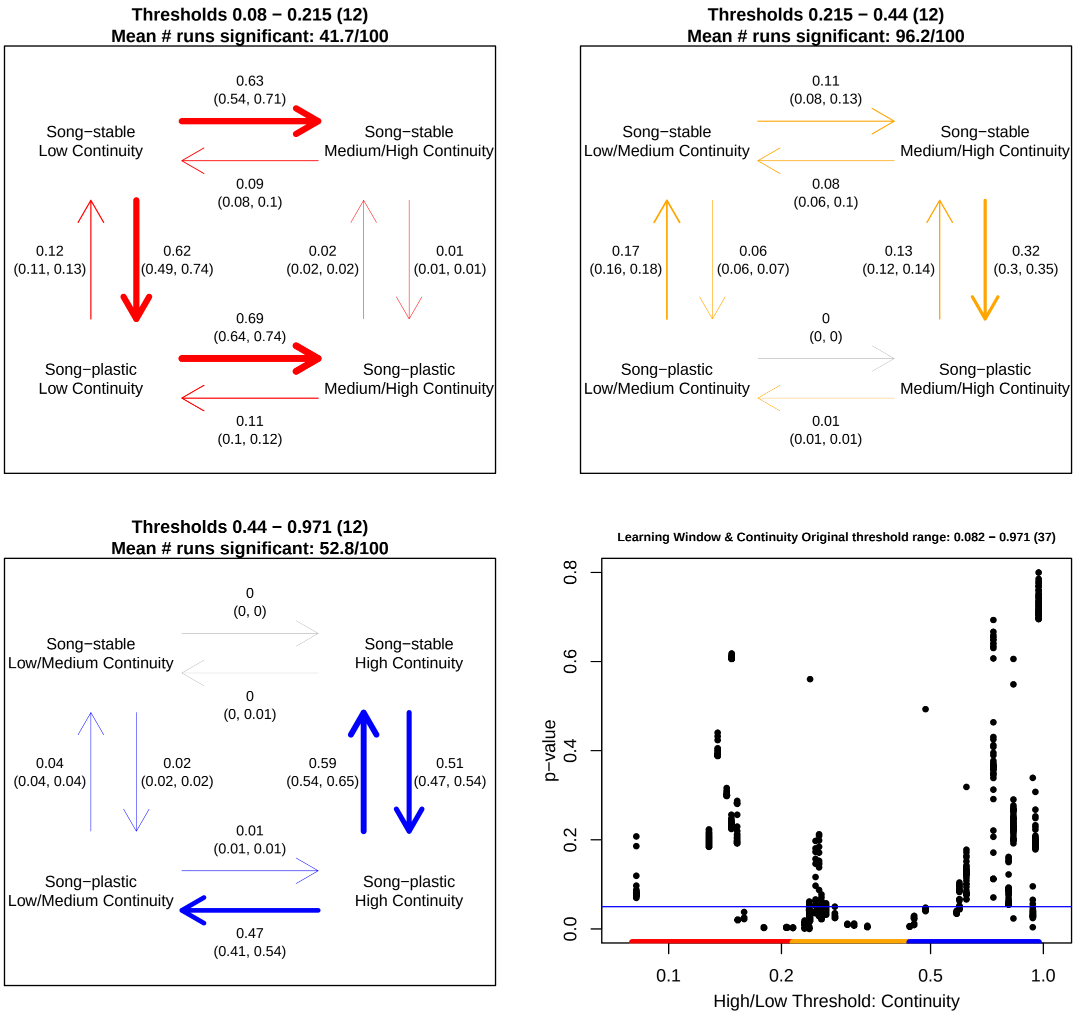


Figure 3.19: Analysis of correlated evolution between adult song stability and continuity. Labeling the same as in **Figure 3.13**.

In addition, it has been proposed that polygyny and extra-pair paternity (EPP) may increase sexual selection pressures on sexually selected traits, including song[152,195], and increased selection pressure due to polygyny was theorized to accelerate the evolution of song learning in a mathematical model[65]. There was evidence for correlated evolution between polygynous/monogamous mating systems and song plasticity (100% of runs significant), with elevated rates of transition between polygyny and monogamy in the song-plastic state. We did not find evidence for correlated evolution between EPP and adult song plasticity (**Figure 3.20**).

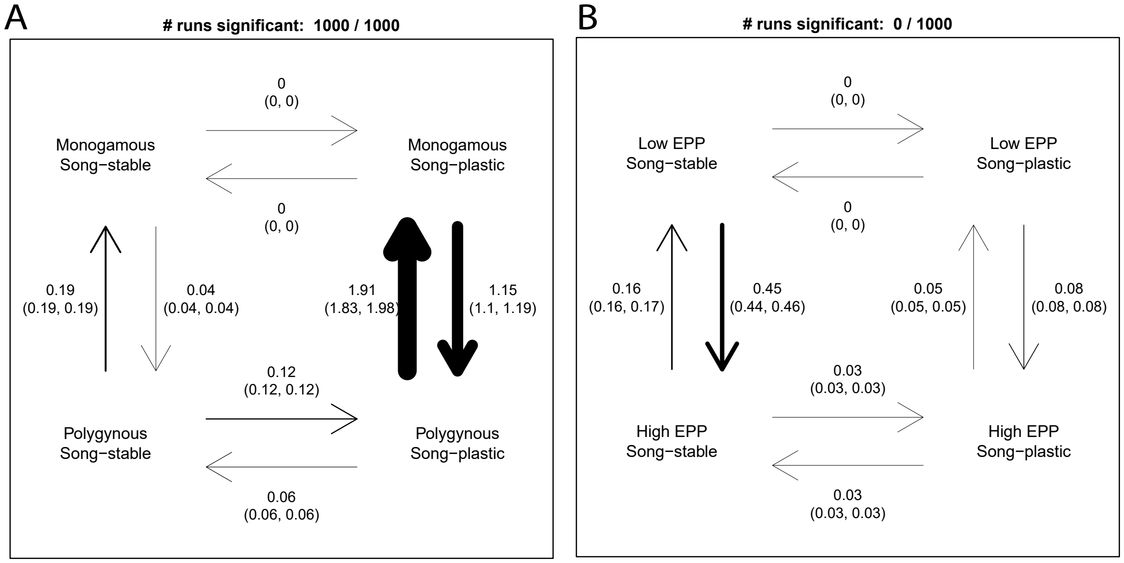


Figure 3.20: Analysis of correlated evolution between song stability and mating behaviors. Performed over 1000 runs. Transition arrows are labelled with the mean rate over all 1000 runs. A) Mating system (polygyny vs. monogamy): 1000 of 1000 runs were significant at $p<0.05$. B) Extra-pair paternity (Low EPP vs. High EPP): No runs were significant at $p<0.05$.

3.3.5 Robustness analysis

In many cases, there were multiple studies that gave different estimates for a given song trait in one species, so we used the median values across studies for our main analysis. To test whether our results depended on the particular values we used, we repeated the `PhylANOVA` and `Brownie` analysis using the either the maximum or minimum values reported in the literature. This did not significantly alter our results (**Tables C.1.3** to **C.1.7**). In our dataset, 24 songbird families were represented by 1 to 11 species each. This meant that families had unequal influence on the outcomes of the analyses. We performed a jackknife analysis to examine whether our results were affected by excluding individual families represented by 4 or more species in the full dataset for the `PhylANOVA`, `Brownie`, and `BayesTraits` analysis. Exclusion of individual families did not significantly alter most results significantly (**Tables C.1.8** to **C.1.19**, **Figure 6 - figure supplement 8-9, and Figure 7 - figure supplement 1** from Robinson, Snyder, and Creanza 2019[70]), but there was one notable exception.

Removal of *Mimidae* (4 species) from the `Brownie` analysis of syllables per song drastically changed the results, such that the two-rate model was no longer significantly better fit than the one-rate model (**Table C.1.17**). All of the included mimid species are song-plastic, so these data suggest that mimids may be driving the fit of the two-rate model for syllables per song. Indeed, a secondary jacknife analysis suggested that this trend was driven entirely by the inclusion of *Mimus polyglottos* and *Mimus gilvus* together, and removal of either led to non-significant results ((**Table C.1.20**). In addition, removal of *Mimidae* from the `BayesTraits` analysis of song plasticity and syllables per song altered some of the observed trends (see **Figure 6 - figure supplement 9** from Robinson, Snyder, and Creanza 2019[70]). In combination with the difficulty in quantifying syllables per song in mimid species[196], these results suggest that the evolutionary trends for syllables per song seen in the full dataset do not represent real evolutionary patterns across bird families; however, further study of additional mimid species will be required to examine whether the evolution of syllables per song is markedly different in mimids compared with other songbird families.

3.4 Discussion

Previously, it was unknown whether the song-learning window evolved in concert with song traits associated with sexual selection, as predicted by a computational model of song learning[150]. Here, we performed phylogenetically-controlled analyses to assess the interactions between the length of the song-learning window — using song stability as a proxy — and the evolution of song traits. Interestingly, we noted that several evolutionary events relatively early in passerine evolution accounted for much of the diversity in the song-plasticity period in our sample of species. Our results predict that a species' ability to modify adult song not only affects the characteristics of song in a generation but also interacts with the evolution of larger repertoire sizes, since sexual selection for large repertoires could drive evolution of increased song plasticity.

We found two key trends in the trait correlation (`phylANOVA`) and evolutionary rate (`Brownie`) analyses. First, the direction of evolution in traits that can be considered metrics of song elaboration (syllable and song repertoire size, **Figures 3.1**, **3.4** and **3.5** and **Table 3.1**) was increased in song-plastic species, leading to larger repertoires. Further, species with early song stability evolved their repertoires at a slower rate than species with longer learning (delayed song stability and song plasticity)(**Table 3.5**), but song-plastic species did not evolve their repertoires at a faster rate than species with early or delayed song stability combined (**Tables 3.6** to **3.9**). Thus, while repertoires only evolve directionally in song-plastic species (**Tables 3.2** and **3.3**), our results suggest that extended learning through the first breeding season allows for faster, but not directional, evolution of repertoire size. It is possible, then, that delayed song learning allows individuals to modify their songs after migration and thus adapt their song to their new surroundings once they establish a territory, without necessarily corresponding with sexual selection for larger repertoires. This ability for an individual to adapt to a new local song might be beneficial, particularly when species have local dialect structure; however, it would not lead to directional evolution. Second, song plasticity increased the rate of evolution primarily in traits that can be considered metrics of song performance (song duration, intersong interval, and song rate **Figure 3.12** and **Table 3.5**). While these performance-related song traits evolved faster in song-plastic lineages, this rapid evolution did not lead to significant differences in those song traits compared to song-stable lineages. A possible explanation may be that increases in repertoire size necessitate changes to song structure, but multiple structural aspects of song can be altered to accommodate these changes. Thus, there is no overall pattern of directional evolution in these other song traits. Alternatively, bird species may be required to adapt to increasing repertoire sizes while maintaining species-specific constraints imposed by innate aspects of song structure necessary for species recognition or female preferences for different performance characteristics. In the latter case, if information about innate characteristics and female preferences are known, it may be possible to predict how song traits will change in response to increasing repertoire sizes and greater adult song plasticity.

With our analyses of correlated evolution, we aimed to detect whether the state of the repertoire size or adult song stability versus plasticity consistently changes first in evolutionary history, facilitating a change in the other trait. Overall, our results suggest that there is not a consistent order of evolutionary transitions (**Figures 3.13** and **3.14**). For example, song stability with very large syllable or song repertoires, and song plasticity with very small syllable or song repertoires both formed evolutionarily unstable states, with high evolutionary rates of transition in both repertoire size and song-learning window. However, we do note that the fastest rates of transition in our analyses were those switching between song-stability and song-plasticity to leave those unstable states. This trend suggests that the magnitude of a species' repertoire may be more likely to drive the evolution of learning window than vice versa. This is consistent with the idea that selection acting upon song traits could indirectly place selective pressures on the learning window. We propose several hypotheses that could explain these evolutionary dynamics: 1) it may be disproportionately costly to maintain song plasticity when syllable or song repertoire sizes are very small, perhaps because the benefit of extra time to learn does not outweigh the metabolic cost of maintaining plasticity, or 2) species with small syllable or song repertoires may have highly stereotyped songs which are selected for based on accuracy of learning and consistency of song production, favoring males that only learn from their fathers or early-life neighbors. Alternatively, in species where females prefer larger repertoires, 3) it may simply require more time to learn a large song or syllable repertoire than is available with a short learning window, or 4) learning large repertoires may require too much energy devoted to song practice during the crucial period of development before and during fledging, favoring birds that can learn for longer periods. Further research into the physiological or reproductive costs of song plasticity is needed.

Beginning with Darwin[25], numerous researchers have proposed that some mating behaviors could lead to amplified sexual selection[152,195]. We tested for correlated evolution between song stability and both social mating system (polygyny vs. monogamy) and extra-pair paternity (low vs. high EPP), with the caveat that many species in our dataset lacked mating behavior classifications (57 species with social mating system data, 41 with EPP data). We did not find evidence for correlated evolution between song stability and EPP. In contrast, all of our runs testing for correlated evolution between social mating system and song stability versus plasticity were significant, showing an elevated rate of transition between monogamy and polygyny in the song-plastic state relative to the song-stable state (**Figure 3.20**). This result suggests an interesting hypothesis for further investigation: perhaps having a plastic song-learning program facilitates evolutionary transitions in mating systems.

Our results provide key evidence that sexual selection upon song traits might indirectly act upon the song-learning window. We do not fully understand the mechanisms underlying the maintenance or reopening of the song learning window in adulthood, but genetic, environmental, hormonal, and social factors are likely contributors[28,59,161,197,198]. For example, when zebra finches were reared in isolation, their sensitive periods were lengthened. These isolated birds maintained both gene expression profiles associated with song learning in the song system and high levels of neuronal addition to the HVC (a key region in the song system of the songbird brain) for longer than birds reared with an adult male tutor[199,200], linking the neural underpinnings of song learning to the length of the song-learning window[4,160,201,202]. Furthermore, a positive association between HVC volume and song repertoire size has been demonstrated both intraspecifically[82] and interspecifically[148]. In light of our findings that adult song plasticity correlates with an increase in song repertoire size, there is a logical prediction that extended song learning may be associated with increased HVC volume across species. This is an important avenue for future research.

Although our dataset includes species from 24 different songbird families, many families are not represented due to a lack of data about song stability. It will be important to expand this dataset in future studies. It would also be interesting to explore the evolutionary interactions between adult song plasticity and mimicry of heterospecific sounds, which has been observed in *Mimidae* and numerous other clades across the songbird lineage[203]. With our current dataset, we could not adequately explore the effects of mimicry on the evolution of song learning outside of the mimids, but the repeated evolution of mimicry makes it a particularly interesting topic for follow-up studies on the length of the song-learning window. In addition, different song metrics that are tailored to mimicry would be important in studying the evolution of vocal mimics and the dynamics of their unique vocal patterns. There is also increasing interest in the importance of female song in species, which is more common than previously thought[7,8]. Our dataset includes numerous species wherein females are known to sing at least occasionally[7], but the length of the song-learning window in females has not been assessed in any of these species. There is, however, some evidence that female birds can modify their song preferences in adulthood[204]. Thus, it remains an open question whether song plasticity affects the evolution of female song in the same way it affects male song, and whether species with adult song plasticity in males also have adult song plasticity in females.

Our findings shed new light on the broader subject of song evolution — specifically the evolution of the process of song learning. We hypothesize that sexual selection on certain aspects of song could in turn alter the selection pressures on the length of the song-learning window. Here, we perform phylogenetically-controlled analyses across 67 songbird species to assess the evolutionary interactions between song traits and song plasticity in adulthood. With these analyses, we show the first evidence for this evolutionary relationship. Song stability is evolutionarily dependent upon the properties of the song itself: large syllable and song repertoires appear to drive the evolution of adult song plasticity and thus open-ended song learning. This provides context for the remarkable interspecific variation in song-learning windows across the songbird lineage and suggests an evolutionary mechanism by which sexual selection might have influenced the evolution of songbird brains.

Chapter 4

Modeling the evolutionary interactions of sexual selection and learning strategies on the duration of song-learning in birds [1]

4.1 Introduction

The duration of the song-learning window is highly variable among songbirds. In closed-ended learners, it may only last for several months post-hatch (e.g. house sparrow[205] and zebra finch[206]) or could extend into the second breeding season in species with delayed sexual maturation (e.g. black redstart (*Phoenicurus ochruros*)[179] and yellow-rumped caciques (*Cacicus cela*)[180]). In contrast, open-ended learners exhibit extended learning windows and can modify their song as sexually mature adults[4,52]. Although open-ended learners are typically thought of as life-long learners, we note that it is not clear how long these extended periods of plasticity last, as few studies have examined adult learning beyond the second breeding season[70]. Nonetheless, longer learning has been assumed to be valuable, because it would allow birds to modify their songs if they were unsuccessful at attracting a mate with their current song or, in species that engage in song-matching, continual modification of songs could be advantageous if the males on neighboring territories differ from year to year[1,4]. Because song is so important for mate attraction[207,208], it has long been wondered why all birds do not engage in open-ended learning. To explain this discrepancy, it is hypothesized that maintaining the neural plasticity required for extended periods of song learning could be metabolically costly, or that devoting additional time to song learning and practicing would reduce the time remaining for other critical activities like foraging[4,52]. However, experimentally testing the existence of these proposed costs and quantifying their effect on overall reproductive fitness in live birds would be extraordinarily difficult.

In addition to a variable length of the song-learning window, different bird species appear to use different learning strategies to acquire and modify their adult songs[51,52]. Some species

---

[1] This chapter is adapted from my article *Modeling the evolutionary interactions of sexual selection and learning strategies on the duration of song-learning in birds*, which is in revision at *PLOS Computational Biology*[71]

replicate their father's song with high accuracy (e.g. Darwin's finches[27]), while other species instead conform to a short local dialect (e.g. white-crowned sparrows[209,210]), and still other species can modify their song to better match a neighbor once they reach their breeding ground as adults (e.g. chipping sparrows (*Spizella passerina*)[20] and swamp sparrows (*Melospiza georgiana*)[211]). Species whose song consists of a large repertoire with a high proportion of mimicked sounds (e.g. mimids, starlings, and lyrebirds) can learn new syllables from conspecific and heterospecific individuals, including non-avian species, as well as inanimate objects[170,212–214]. Unsurprisingly, there are marked differences in song structure and syllable number between species that are known to employ different learning strategies. However, in many species, it is not clear which learning strategy males use to create their song or the extent to which song structure is affected by song-learning strategy.

If different learning strategies can be leveraged to produce songs with specific characteristics, it seems likely that some learning strategies are more valuable than others in certain selective contexts. In particular, female preference for larger repertoires may favor indiscriminate addition of novel sounds, while a preference for specific song features would likely favor high-accuracy learning from a bird's father or neighboring conspecifics. However, it is likely that the length of the song-learning window interacts with learning strategy to conform to selection pressures, which complicates these otherwise straightforward hypotheses. Furthermore, analysis of these interactions would require examining a population on an evolutionary timescale. Previous work modeling the interactions between learning and sexual selection have focused on speciation[66–68], between-species interactions[69], or acquisition and maintenance of learned song in place of innate song[65], but none have considered how selection pressures on song interact with the male song-learning strategy to affect the evolution of song and/or the duration of song learning.

Here, we present a model of birdsong evolution to address these relationships. This model simulates a population of birds with evolving syllable repertoires and song-learning traits. We test which learning strategies are selected for when males with either larger repertoires or better song-matching are most likely to reproduce; for the purposes of this manuscript, we frame these selection pressures in terms of female preferences. We also test whether longer learning windows must be costly to allow for the evolution of shorter windows by adding a conditional fitness penalty on longer learning in a subset of simulations and setting this penalty to zero in others. Overall, we found support for the hypothesis that a strong female preference for large repertoires would favor indiscriminate syllable addition and extended learning windows, while either learning purely vertically from a male's father or learning obliquely according to conformity bias with a short learning window were beneficial strategies when females strongly prefer songs that matched a template. Unexpectedly, we found that a fitness cost on longer learning windows was not required to evolve shorter song-learning windows, though it allowed this transition to occur more readily.

4.2 Methods

4.2.1 Data and code availability

C# library, programs, and walk-through:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/Csharp-Library-and-Programs

R package, code manual, and walk-though:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/R-Package

Parameter generators and data visualizing code:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/OverviewPaper

4.2.2 Model design

The song evolution model was coded as a **C#** library that is incorporated into console applications. We used the `BetaScaled.PERT` function from `MathNet.Numerics v4.5.1`; all other scripts were custom. **R** scripts were used to generate the parameter set files and plot the data output from console applications. Furthermore, we developed a functionally identical **R** version of the model, which is available as a package, but was not used to generate any simulation data in this study, because the **C#** version ran substantially faster. For this study, the default values were used for all parameters unless otherwise noted in figure legends or **Table 4.1**. A full explanation of the defaults and components exists in both the **C#** Walk-Through and **R** Walk-Through. Note that the walk-throughs are exhaustive and also contain information about features not implemented in this manuscript. Any feature not covered below was not used.

Table 4.1: All parameters modified in this paper and their defaults. A full list of all parameters and their defaults is available in the walk-through. Song-learning traits can evolve if they are given an inheritance noise (ε) greater than 0 (see methods for equation); if $\varepsilon = 0$, all members of the population are initialized with the same initial value for a trait and chicks inherit this trait perfectly from their fathers. For this paper, learning accuracy and learning threshold can evolve, while chance to invent and forget are static. The average population repertoire size can also evolve, though this is based on song-learning behavior, not inheritance from the father with noise. Additionally, in the full model, female preference for larger repertoires and female preference for better template matches are two separate parameters. Because the two always sum to 100% in this paper, we talk about them as though they were a single parameter below and in the heatmaps.

Variable	**Meaning**	**Default**	**Range**
Initial Repertoire size	number of syllables that an each individual in a population has a 90% chance to know at simulation start. Also affects the female template, which does not change over time for this paper.	5	5, 20
Initial Accuracy	the mode learning accuracy for the population at the start of the simulation	70%	10-90%, 95%
Inherited Accuracy Noise	the boundary for the distribution of learning accuracy values that a male chick can inherit is his father's learning accuracy +/- this value	15%	0-25%
Initial Learning Threshold	mode length of the song-learning window for the population at simulation start	2	.5-3.5, 10
Chance to Invent	chance that a male will invent a new syllable when he attempts and fails to correctly learn a tutor's syllable	10%	5%, 10%, 50%
Chance to Forget	chance that a male will forget syllable he knows, but his tutor did not sing	20%	20%, 50%, 80%
Learning Penalty	fitness cost applied to males that have a learning window longer than 1 time step (analogous to 1 year); is prorated so birds that learn longer are more severely affected	.75	0-2

Learning Strategy	the method males use to learn obliquely. **Add:** learner attempts to add all the syllables that his tutor sang to his repertoire. **Add/Forget:** learner attempts to learn all the syllables that his tutor sang, but he also has a chance to forget syllables he knows, but which the tutor did not sing. **Conformity:** learner listens to multiple tutors and adds or loses syllables based on conformity bias.	NA	Add, Add/Forget, Conformity
Conformity Number of Tutors	number of tutors a male listens to when using the Conformity learning strategy	8	2-12
Repertoire Size Preference	amount that female preference relies on preference for larger repertoires; for this paper, if this value is less than 100%, the remaining percentage is female preference for matching to the female song template	NA	0-100%
Vertical Learning	whether males learn from their fathers	True	True, False
Oblique Learning	whether males learn from from males in the population post-fledging	True	True, False
Listening Threshold	the amount of a tutor's repertoire that a learner hears. It is either a percentage of the repertoire or a discrete number of syllables.	7	7, 0%, 30%, 60%, 100%
Father Listening Threshold	the amount of a father's repertoire that a son hears. It is either a percentage of the repertoire or a discrete number of syllables.	100%	0%, 30%, 60%, 100%
Minimum Learned Syllables	the minimum number of syllables a male learns when the listening threshold is a percentage	7	0, 3, 7, 15

4.2.3 Model operations in brief

See **Figure 4.1** for a visual summary of the model. Given a set of parameters, a population of 400 male birds in a 20x20 matrix is initialized with a syllable repertoire and song-learning traits. The syllable repertoire is in essence an array of length 500. Syllables are coded in binary terms such that males either have a specific syllable or they do not. For this paper, at time step 0, males are initialized with similar syllable repertoires in the same region of the syllable vector (e.g. they can know syllables 1-7, and are most likely to know 1-5). However these repertoires are generated stochastically, so males do not all have the exact same syllable repertoire. Female birds are initialized with a level of preference between 0% and 100% for repertoire size (preference for mates whose song contains more syllables) and template matching (preference for mates whose song more closely matches their template song). Females use this preference to score potential mates, and they are more likely to choose males with higher scores. The levels of these two preferences combined always sum to 100% in this paper. All females in a simulation have identical preferences. If females display any preference for template-matching, one female template song is generated at the start of a simulation, following the same rules and in the same region of the syllable space as the male templates. All females are given a copy of this single template, which does not evolve over the course of the simulation. If the level of template-matching preference is 0%, a song template is not generated, because it is not used by the females in mate choice.

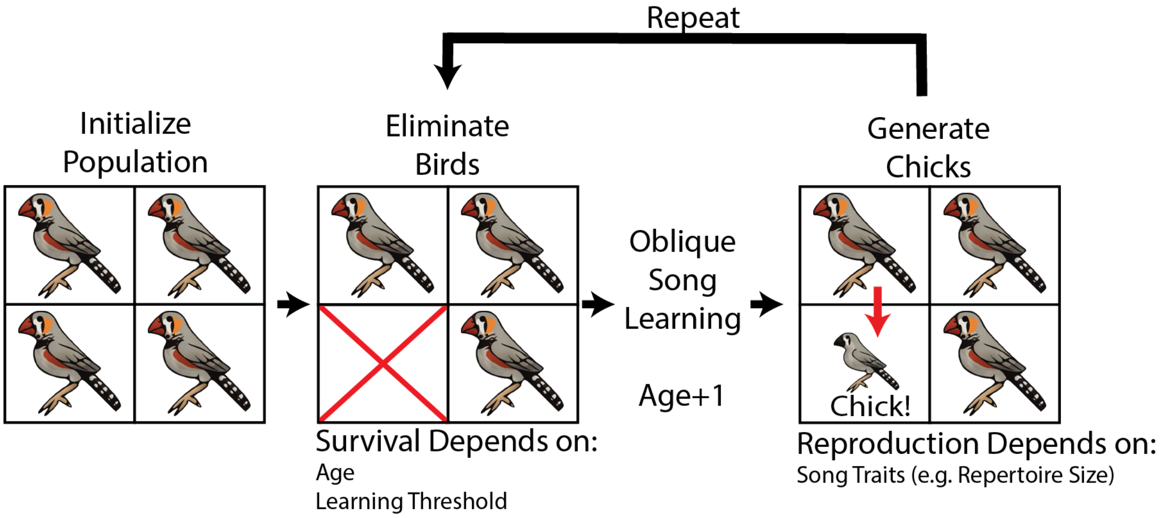


Figure 4.1: Summary of model events. First, a population of birds is initialized. Next, a proportion of birds in of each age group are chosen to be eliminated. If there is a learning penalty, birds with thresholds longer than one year (time step) are more likely to be eliminated. Longer learning windows increase the chance of being chosen. Surviving birds then can learn obliquely if they are younger than their learning threshold and oblique learning is enabled. Birds then age up by one year. Finally, males are chosen to sire chicks to replace eliminated males. Males are chosen to reproduce based on their song traits (i.e. having larger repertoires or better matching the female song template). Chicks inherit song-learning traits from their fathers with some amount of inheritance noise. Unless the inheritance noise is set to zero, chicks are similar to, but not identical to, their fathers. Chicks then vertically learn their song from their fathers if vertical learning is enabled. The cycles returns to the elimination step and continues for a user-specified number of time steps.

4.2.4 Song learning and male traits

The male song-learning traits include:

Length of the song learning window ('learning threshold'): the age at which a male can no longer modify his syllable repertoire. Males must have a learning threshold of at least 0.25 to learn vertically. Longer thresholds give males a chance to learn obliquely as well. The learning threshold is continuous, even though age is always a discrete integer. To account for this, the fractional portion of learning threshold is the chance that a male learns obliquely during the time step when he is only one time step older than his learning threshold; i.e. a bird with a learning threshold of 2.5 will learn vertically once and obliquely once in his first year, once in his second year, and has a 50% chance of learning obliquely again in his third year.

Learning accuracy ('accuracy'): the chance that a male will successfully learn a syllable that he attempts to learn.

Chance to invent: if a learner fails to learn a syllable, the chance that a novel syllable is created instead.

Chance to forget: the chance that learner loses a syllable if he does not hear it sung by tutors.

In this paper, only the learning threshold and learning accuracy can evolve; chance to invent and forget are static values that all males in a simulation share. Each male also has a stored value for his 1) age, which increases by one every time step, 2) the total number of syllables he knows, which is updated after learning, and 3) the match between his song and his resident female's template, if one exists, which is also updated after learning. The match is calculated as the number of shared syllables minus the number of extra syllables a male knew divided by the total syllable repertoire of the female.

Each time step, multiple males are chosen to die. For this paper, death is dependent on the age of the birds. Male chicks have a 30% survival rate[215–218], and adult males have an approximately

83% survival rate at each time step, following a type II survival curve, which is typical of birds[219] (see **walk-through Section: Implementing a Type II Survival Curve**). Thus, approximately 37% of the male population is replaced every time step. Males from each age group are randomly chosen to die with a probability based on their learning threshold. Males with learning thresholds greater than 1 time step are penalized, such that they are more likely to be chosen to die than birds with learning threshold of 1 or less. This penalty is prorated such that birds which learn until the max age of the population are more heavily penalized than birds with learning thresholds that are only slightly above 1. In some cases, the penalty parameter is set to 0, and all males in an age group are equally likely to be chosen to die regardless of the length of the song-learning window.

For most simulations, males younger than or equal to their learning threshold have the opportunity to learn up to seven new syllables per tutor. We also test alternative cases where males learn a discrete **minimum number of syllables** (Minimum Learned Syllables) plus a **percentage** (Listening Threshold) of the tutor's repertoire, where marked. In these cases, the amount a male can learn from his tutor is calculated as follows:

$$\text{Learnable Syllables} = \text{Min Syllables} + (\text{Tutor Syllables} - \text{Min Syllables}) * \text{Percentage} \qquad (4.1)$$

Any fractional portion remaining is the chance for a male to learn one additional syllable. We implement three different oblique learning strategies: 'Add', in which new syllables can be added from a tutor; 'Add/Forget', in which syllables can be both added and lost to better match a tutor's song; and 'Conformity', in which a bird surveys multiple tutors, adding syllables based on conformity bias and attempting to lose syllables the tutors lack. For the Add and Add/Forget learning strategies, only one tutor is chosen per time step, while for the Conformity strategy 8 tutors are chosen per time step, unless otherwise noted. For this paper, tutors are chosen randomly from all males in the population that are alive, know at least one syllable, and are not chicks (age > 0). The age of the birds is increased by one after the learning step.

In each time step, males are randomly chosen to father new male chicks, which will replace the males that have died. The chance of being selected as a father is based on how well each male conforms to female preferences. For this paper, all males in the population that are alive and know at least one syllable can father chicks that will acquire a random territory previously inhabited by an eliminated male.

Male chicks are generated based on their father's song-learning traits; for each trait that evolves, a beta-PERT distribution is created that satisfies the following:

$$\text{Mode} = V \tag{4.2a}$$

$$\text{Max} = min(\text{X},\ V + \varepsilon) \tag{4.2b}$$

$$\text{Min} = max(0,\ V - \varepsilon) \tag{4.2c}$$

$$\text{Where}:\ V = \text{Father's Trait},\ \varepsilon = \text{Parameterized Noise},\ \text{X} = \text{max possible value}$$

X is 20 when calculating a chick's learning threshold, and X is 1 for all other traits. Thus, chicks are more likely to have learning trait values similar to their father's learning trait values, and chosen values are always within the absolute minimum and maximum bounds for a trait. A random value is drawn from this distribution and assigned to the chick. For most simulations, a chick can then learn his father's entire repertoire based on his song-learning traits. However, we test parameter sets where chicks can only learn a percentage of the father's repertoire or cannot learn vertically, where indicated.

4.2.5 Parameter sweeps

To examine how different combinations of parameters interaction with one another, we performed parameter sweeps. Sweeps either entailed changing a few levels for many parameters to look for large scale interactions between multiple parameters or changing many levels for a few parameters to look at interactions in finer detail. Each parameter set was run for 4,000 time steps, and repeated for 50 individual iterations. Box plots show the averages at the final time step, while show the traces for the 50 iterations of each parameter set in the overview sweep (see **Figures S1-S3** and **Table S1** for parameters in each run from (Robinson and Creanza 2019[71]). Traces show that 400 time steps was a sufficient number for most parameter sets to reach an equilibrium state, where the population song traits had stabilized (i.e. traces had a slope of approximately 0). There were several cases where 4,000 time steps was not sufficient to allow all iterations to reach an equilibrium state, so we reran these parameter sets for 20,000 time steps (see **Figures S4** (Robinson and Creanza 2019[71]). Running the simulations longer allowed most if not all iterations to reach the equilibrium state, but did not change the overall results, so we continued running our experiments for 4,000 time steps. See **Table 4.1** and figure legends for information about which parameters were changed for each sweep. Any parameters not mentioned in this paper were set as the model defaults (see **walk-through Section: .SEMP Files**).

4.2.6 Heat maps

For each heat map, 50 iterations of each parameter set were run for 4,000 time steps each. The heat map data shows the average of the 50 final time steps for a parameter set. See text and figures for information about which parameters were changed for each sweep. Any individual parameters not mentioned were set as the defaults for this paper, as explained above.

4.2.7 Learning invasion

For invasion experiments, 1,000 iterations were run for each parameter set. Simulations began without invaders, and then 4 invaders (1% of the population) were introduced at time step 500. Invaders were created by randomly choosing territories and reassigning the learning threshold at that territory to the invading value. The age for invaders was set to 1, so they would have the best chance of surviving several time steps. Invasion simulations ended when all invaders and their progeny were removed from the population, they overtook the population (100% conversion to the invader phenotype), or 400 time steps had passed, whichever occurred first. We set the inheritance noise for the learning-window (ε) to 0, so males within a simulation would be initialized at time step 0 with the same learning window duration, and chicks would inherit this trait perfectly from their fathers. We conducted these simulations with three defined learning-window states: 1) Closed-ended learners had a learning window of 0.25 years, so they would only learn vertically from their fathers. 2) Delayed closed-ended learners had a learning window of 1 year, so birds would learn vertically and would make one attempt to learn obliquely (not penalized for longer learning). This was meant to represent birds that maintain song plasticity until shortly after their first migration. 3) Open-ended learners have a learning window of 2 years, so they learn vertically and attempt to learn obliquely twice (thus, they are penalized for longer learning). We chose 2 years because most research examining the length of the song learning window did not test birds past their second breeding season, so it is unknown how prevalent life-long learning is in birds[70]. For each invasion test, all birds in the model began in one state.

4.2.8 Validation

For the current study, we expected to see the following if our model worked as intended: (1) Increasing the learning penalty should decrease the length of the song-learning window. (2) When females choose mates based on exclusively on how well they match a song template (match preference 100%), the average repertoire size should be approximately the length of the template. (3) When females prefer larger repertoires and a population of birds starts with a small repertoire, the average repertoire size should evolve to larger values during the simulation. All of these expectations were met **Figure 4.2**.

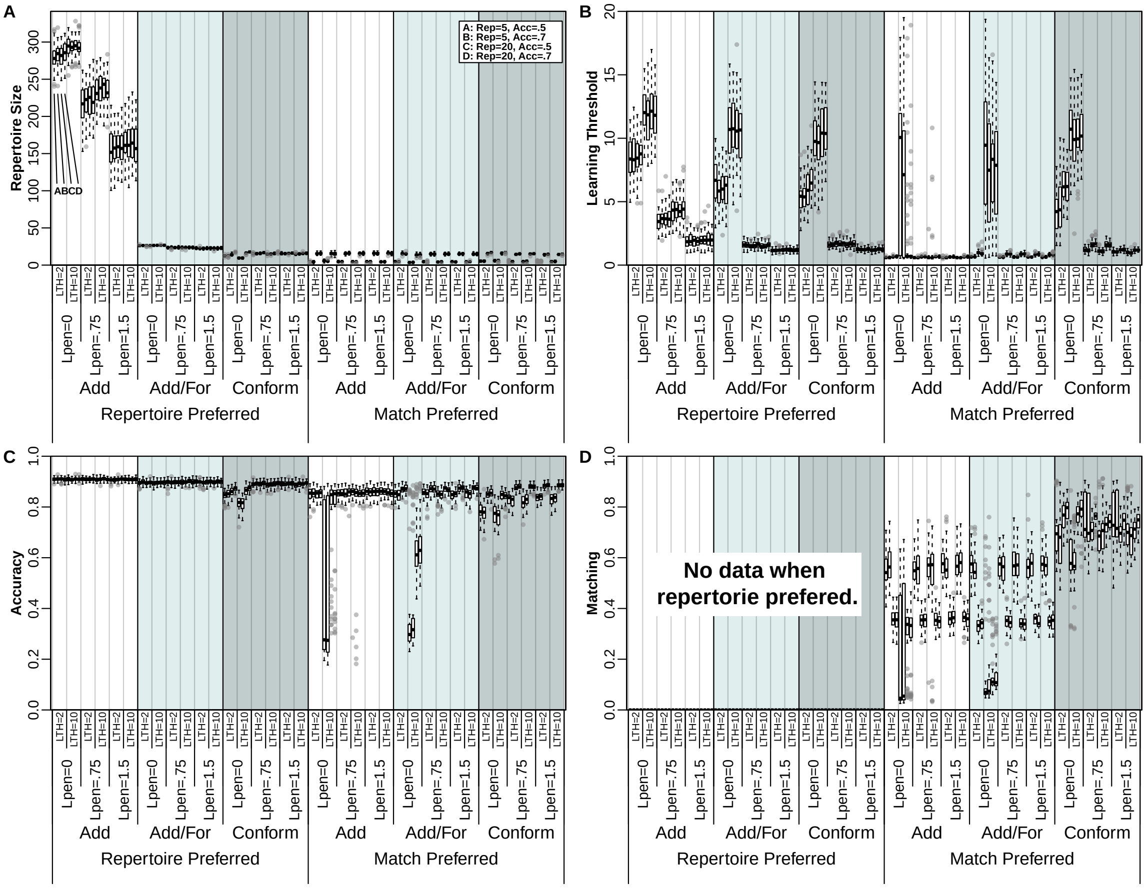


Figure 4.2: Non-aggregated data for the initial parameter sweep. Each box plot represents 50 individual iterations. The same iterations were used in all four plots here and in the corresponding **Figures 4.3** and **4.10**. **Abbreviations:** LTH = Learning Threshold, Lpen = Learning Penalty, Rep = Initial Repertoire Size, Acc = Initial Learning Accuracy.

4.3 Results

4.3.1 Model overview

We chose to model two types of female preferences that have been proposed in the literature as selection pressures: 1) preference for larger repertoires and 2) preference for songs that match a specific song template (analogous to preference for specific song dialects). Selection pressures were implemented by giving the highest opportunity to breed to male birds that best fit these preferences. We assigned males four song-learning traits: 1) learning accuracy 2) length of the song-learning window, 3) chance to invent a new syllable, and 4) chance to forget a syllable. Of note, birds with song-learning windows of less than 0.25 (analogous to 3 months) do not learn.

Males with learning thresholds of 0.25 will learn vertically from their fathers. Longer learning windows give males the chance to learn obliquely from one or more randomly chosen adult males for as many years as the learning threshold allows (e.g. a threshold of 1.5 means a male will learn obliquely once and has a 50% chance to learn obliquely a second time). It is likely that different species use different strategies to learn obliquely, so we implemented three learning strategies. The 'Add' strategy allows the learner to acquire syllables indiscriminately. The 'Add/Forget' strategy allows the learner to both add new syllables and try to forget syllables that he had previously acquired which the tutor did not sing, enabling a closer match to the tutor's song. The 'Conformity' strategy allows the learner to sample from several tutors and learn new syllables based on conformity bias and try to forget known syllables that no tutor sang.

A final component we considered was the debate over the cost of longer song-learning windows. It has been proposed that the ability to learn in adulthood would induce fitness costs, such as a greater metabolic cost[4,52]. While this is a reasonable hypothesis, a cost to longer learning has not been quantified in real birds, so it is not clear if this cost exists and whether it constrains the evolution of the song learning window. To examine whether such a cost is necessary for the evolution of shorter learning windows, we implemented a variable learning penalty on birds with a learning threshold greater than 1 year, which affected their chance to survive to the next time step. We divided most of the results for this paper into trends seen when females exclusively prefer larger repertoires or when females exclusively prefer template matching, but the results from these two sections can also be compared with one another since we conducted our simulations with the same parameter sets, as noted in the figure legends (e.g. **Figure 4.3** and **Figure 4.10** can be compared). If the value for a parameter is not labeled in the figure, we used the default value noted in **Table 4.1** (for parameters we changed in this paper) or in the **walk-through** (for parameters we never changed in this paper).

4.3.2 Female preference for larger repertoires

We conducted a large parameter sweep to assess the effects of different parameters in our model. In the overview sweep, when females preferred larger repertoires, the Add strategy always led to repertoires of over 100 syllables, while the Add/Forget and Conformity strategies led to repertoire of less that 50 syllables (**Figure 4.3A**). While all learning strategies led to the maintenance of adult learning when there was no learning penalty on extended learning windows (average learning window >5 years, **Figure 4.3B**), the length of these windows decreased as the learning penalty was increased. Of note, males using the Add strategy maintained longer learning in the face of the strongest learning penalty (~2 years, meaning that most males would be penalized, **Figure 4.3B**) than males using the Add/Forget and Conformity strategies (~1 year, meaning that males can avoid the learning penalty, **Figure 4.3B**). When the learning penalty was increased, average repertoire sizes decreased for all learning strategies, but this was most noticeable for the Add strategy (**Figure 4.3A**). Learning accuracy at the end of the simulation was similar for all parameter sets (**Figure 4.3C**).

We next examined how changing several other parameters affected song evolution. Increasing the chance to invent a new syllable after inaccurately learning a tutor's syllable led to larger repertoire sizes **Figure 4.4**. Increasing the chance to forget led to smaller repertoire sizes and increased the learning accuracy for simulations using the Conformity strategy **Figure 4.5**. Decreasing the number of tutors sampled for the Conformity strategy led to larger repertoires **Figure 4.6**.

The tutor-choice strategy employed here allows males to learn a significant portion of their repertoire from their fathers. However, extensive or exclusive vertical song-learning in male birds appears to be rare[220–224]. Furthermore, learners could only obtain a maximum of 7 new syllables per oblique tutor, and allowing for a greater amount of learning from oblique learning could affect the evolutionary patterns we found. We tested this possibility by allowing birds to learn more syllables from oblique tutors, and we found that changing these parameters had a large effect on the magnitude of the final repertoire size but did not greatly affect the evolutionary relationships between the different song-learning traits (**Figure 4.7** and **Figure 4.8**).

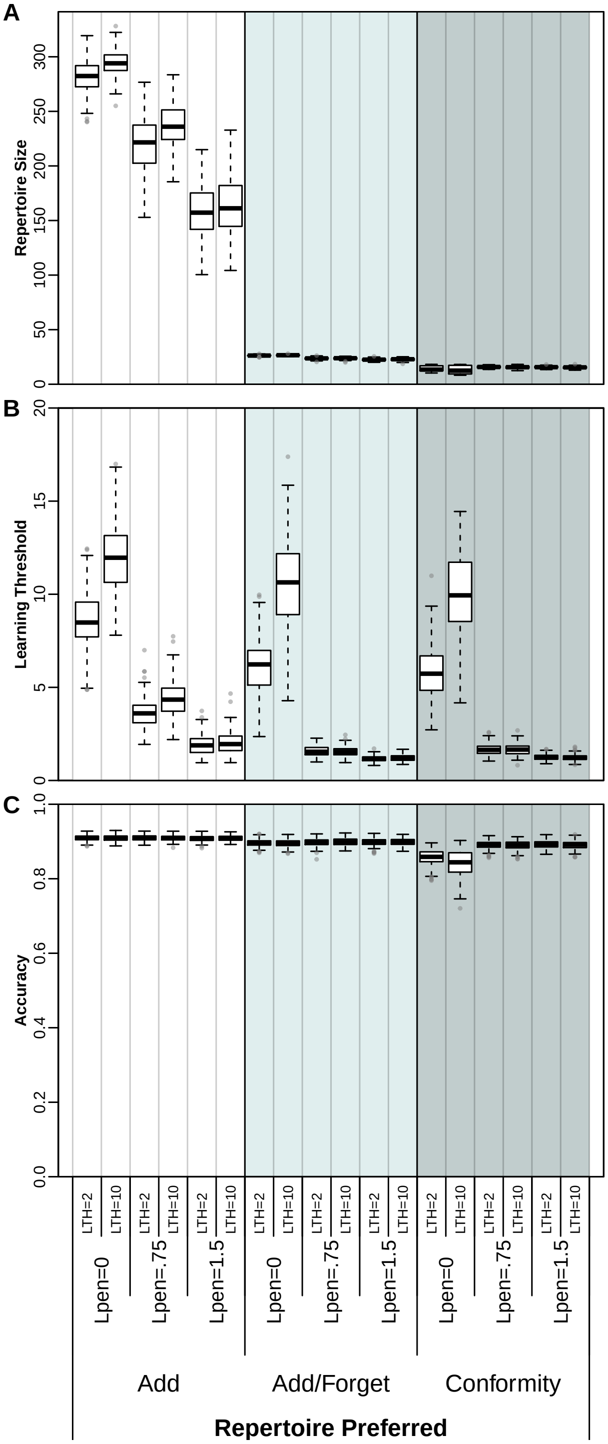


Figure 4.3: Parameter sweep when females exclusively prefer larger repertoires. Each box plot represents 200 individual iterations, because simulations that differed only in their initial learning accuracy and repertoire size were aggregated (see **Figure 4.2** for non-aggregated results). The same iterations were used in all three plots. This plot can be directly compared to **Figure 4.10**. **Abbreviations:** LTH = Learning Threshold, Lpen = Learning Penalty.

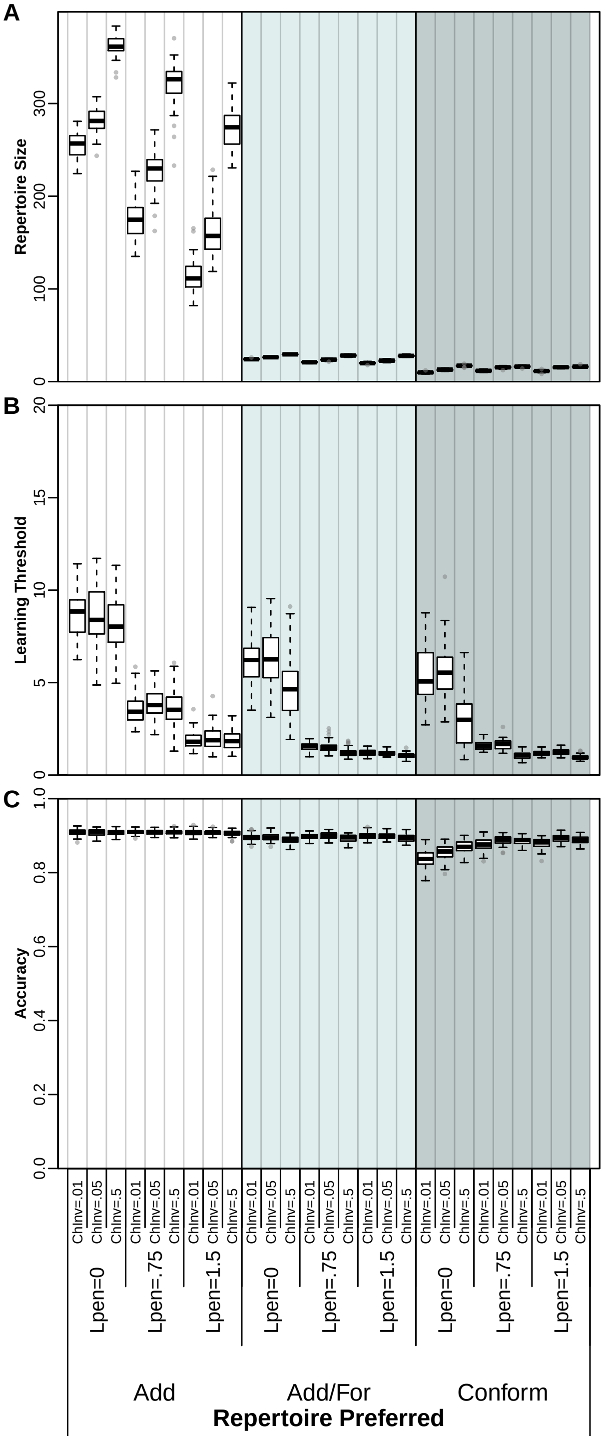


Figure 4.4: Increasing the chance to invent increases repertoire size when females exclusively prefer larger repertoires. Each box plot represents 50 individual iterations. The same iterations were used in all three plots. **Abbreviations:** LTH = Learning Threshold, Lpen = Learning Penalty, ChInv = Chance to Invent.

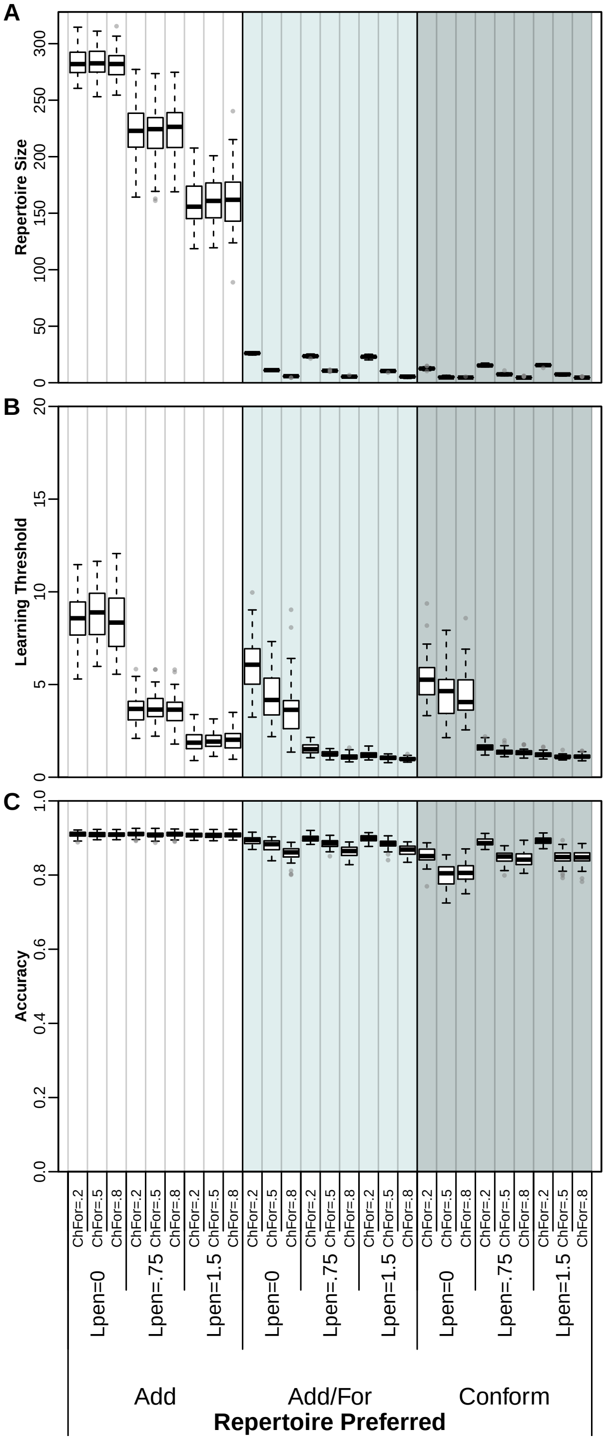


Figure 4.5: Increasing the chance to forget decreases repertoire size when females exclusively prefer larger repertoires. Each box plot represents 50 individual iterations. The same iterations were used in all three plots. **Abbreviations:** LTH = Learning Threshold, Lpen = Learning Penalty, ChFor = Chance to Forget.

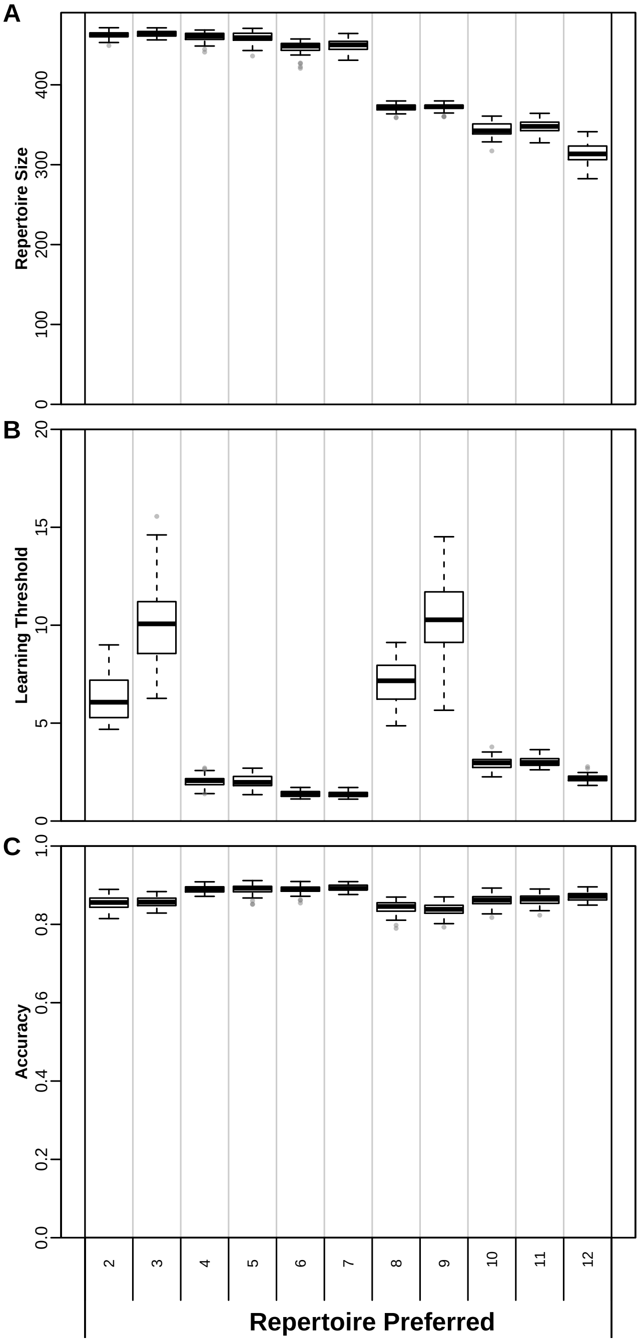


Figure 4.6: Number of conformity strategy tutors has little effect on evolution when females exclusively prefer larger repertoires. However, when the number of tutors was very small (2) the average repertoire size was somewhat larger. Each box plot represents 50 individual iterations. The same iterations were used in all three plots. Numbers on the x-axis show the number of tutors sampled.

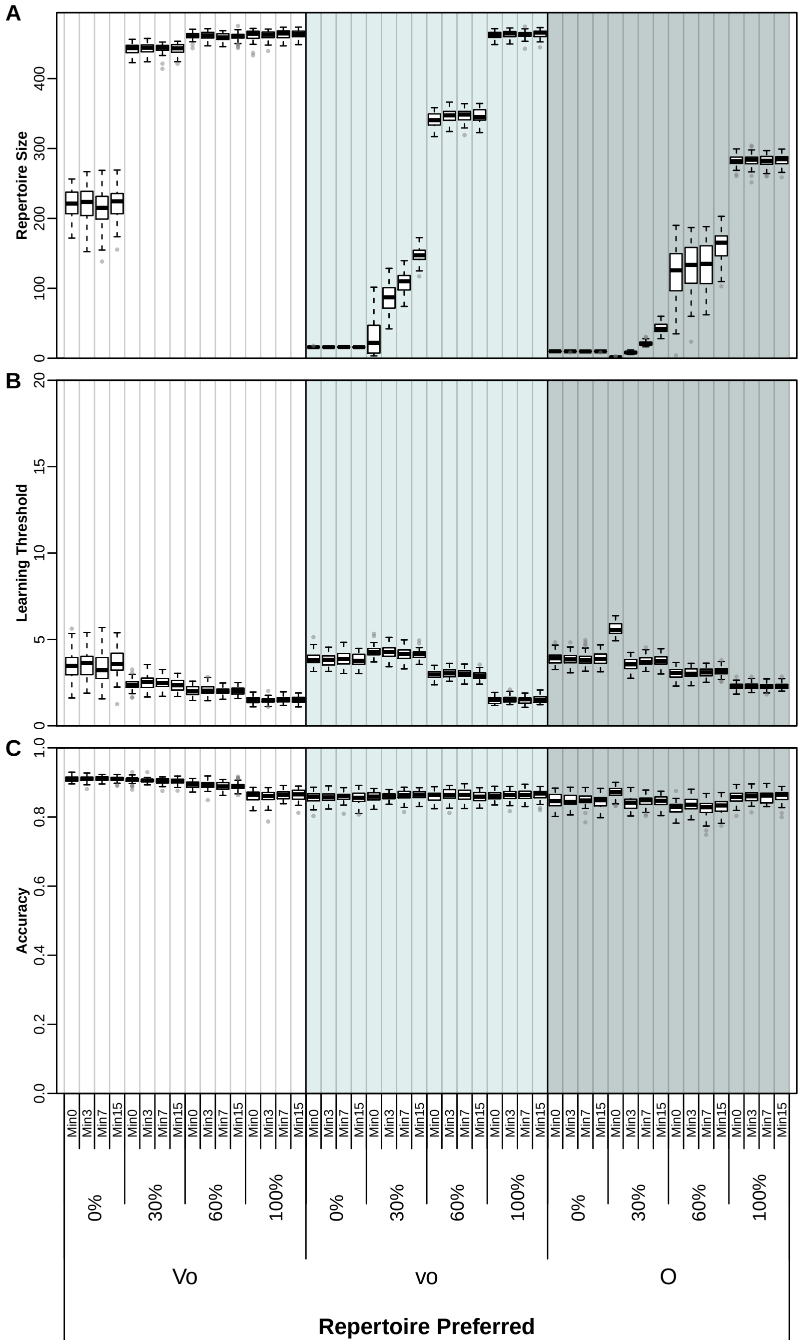


Figure 4.7: Different tutor-choice strategies affected the final repertoire size for the Add learning strategy. The percentage learned from tutors also affected the length of the song-learning window. Each box plot represents 50 individual iterations. The same iterations were used in all three plots. This plot can be directly compared to **Figure 4.10**. Tutor strategies are as follows: 1) Vo-a male hears his father's entire repertoire and none-to-all of his oblique tutor's repertoire, 2) vo-a male learns hears the same amount of his vertical and oblique tutors' repertoire, 3) O-a male only learns from his oblique tutor. **Abbreviations:** Tutor = tutor strategy, % heard = the percent of the repertoire a male heard, Min=minimum number of syllables heard.

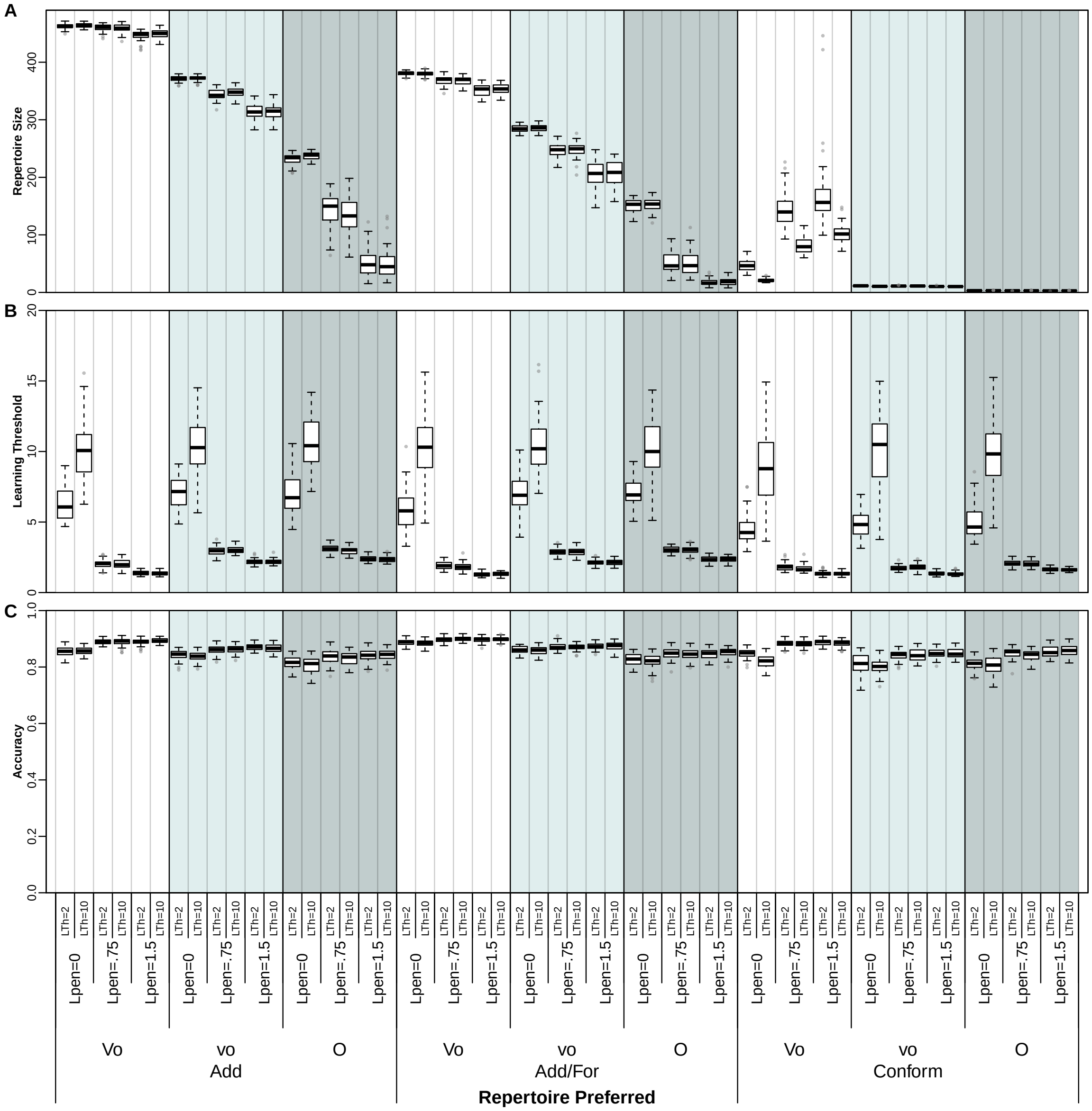


Figure 4.8: Different tutor-choice strategies affected the final repertoire size for all learning strategies. As seen before, increasing the learning penalty decreased the duration of the song-learning window and repertoire size. However, different tutor-choice strategies only interacted with the learning penalty to affect repertoire size. Tutor-choice strategy had little other affect on evolutionary dynamics. Each box plot represents 50 individual iterations. The same iterations were used in all three plots. This plot can be directly compared to **Figure 4.10**. Tutor strategies are as follows: 1) Vo-a male hears his father's entire repertoire and none-to-all of his oblique tutor's repertoire, 2) vo-a male learns hears the same amount of his vertical and oblique tutors' repertoire, 3) O-a male only learns from his oblique tutor. **Abbreviations:** Tutor = tutor strategy, % heard = the percent of the repertoire a male heard, Min=minimum number of syllables heard.

Finally, we conducted an invasion assay to determine the contexts in which open-ended learning might out-compete closed-ended learning and vice versa. For these invasion analyses, we quantified the probability of conversion (all males are progeny of invaders) as the fraction of simulations in which the invading state overtook the initial population state (out of 1000 iterations). When there was no learning penalty, for all learning strategies, there was a high probability of conversion from closed-ended to delayed closed-ended learning and from closed-ended to open-ended learning (**Figure 4.9A**). The probability of conversion from delayed closed-ended to open-ended learning was very small, unless the learning strategy was Add (**Figure 4.9A**), and there was no conversion to shorter learning windows. In contrast, when there was a strong learning penalty, the conversion from closed-ended to open-ended learning was somewhat reduced, and the conversion from delayed closed-ended to open-ended learning was markedly reduced (**Figure 4.9B**). Additionally, there was a small probability of conversion from open-ended to delayed closed-ended for the Add/Forget and Conformity strategies (**Figure 4.9B**).

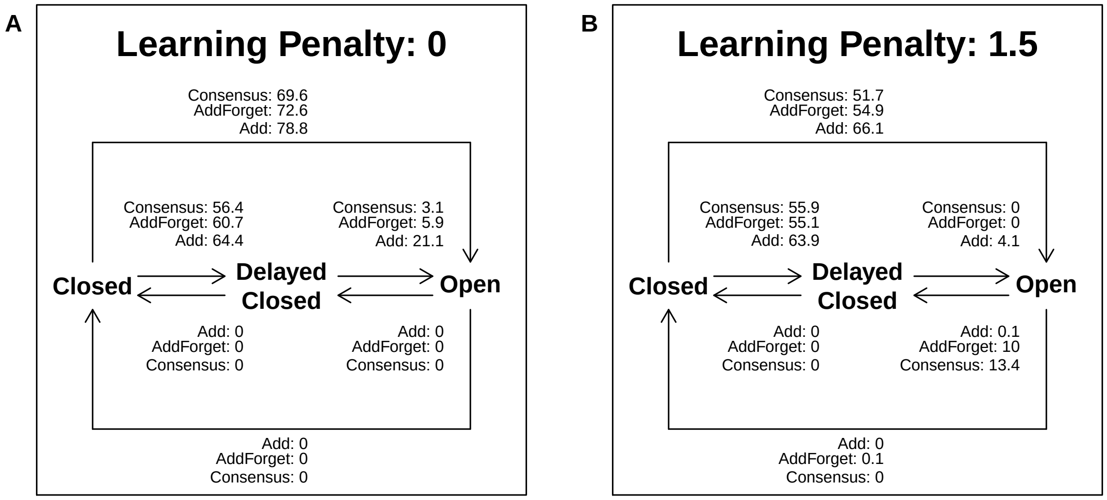


Figure 4.9: Percentage of conversion to invader learning strategies when females exclusively prefer larger repertoires. This plot can be directly compared to **Figure 4.15**.

4.3.3 Female preference for template-matching

When females preferred songs that matched a template, all learning strategies could be used to maintain repertoire sizes similar to the number of syllables in that template (**Figure 4.10A**). When the learning strategy was Add or Add/Forget, populations generally evolved learning windows of ~0.5 years (**Figure 4.10B**), even when extended learning was not costly. The learning window inheritance noise was $\varepsilon = 0.25$ and the lower bound for vertical learning was 0.25 years, meaning birds with these learning strategies tried to optimize maintaining vertical learning while avoiding oblique learning. In other words, our results suggest that oblique learning with the Add and Add/Forget strategies caused detrimental changes to song when females preferred template matching. In contrast, birds using the Conformity learning strategy that experienced a learning penalty typically evolved an average learning window of ~1 year (**Figure 4.10B**), suggesting that birds with this strategy benefited from oblique learning. They also tended to maintain longer learning windows if there was no learning penalty (average learning window >5 years) (**Figure 4.10B**). Together, this suggests that Conformity learning was beneficial to song matching, but even a small fitness penalty was enough to drive birds to evolve windows short enough to avoid the penalty. Learning accuracy was generally high (~85%, **Figure 4.10C**), though it was lower than that seen when females preferred repertoire size (~90% **Figure 4.3C**). This was surprising because we expected accurate learning to be more advantageous when males need to match a specific template rather than simply acquire more syllables. The average match to the song template at the end of the simulation was poor for the Add and Add/Forget strategies (average 50%, **Figure 4.10D**). The Conformity learning strategy did better (~80%, **Figure 4.10D**), but was still lower than some of the values that have been reported in the literature for birds with very simple repertoires, such as zebra finches[225]. Interestingly, there was an unexpected relationship where longer learning windows were associated with lower learning accuracy and worse matching (**Figure 4.10B-D**).

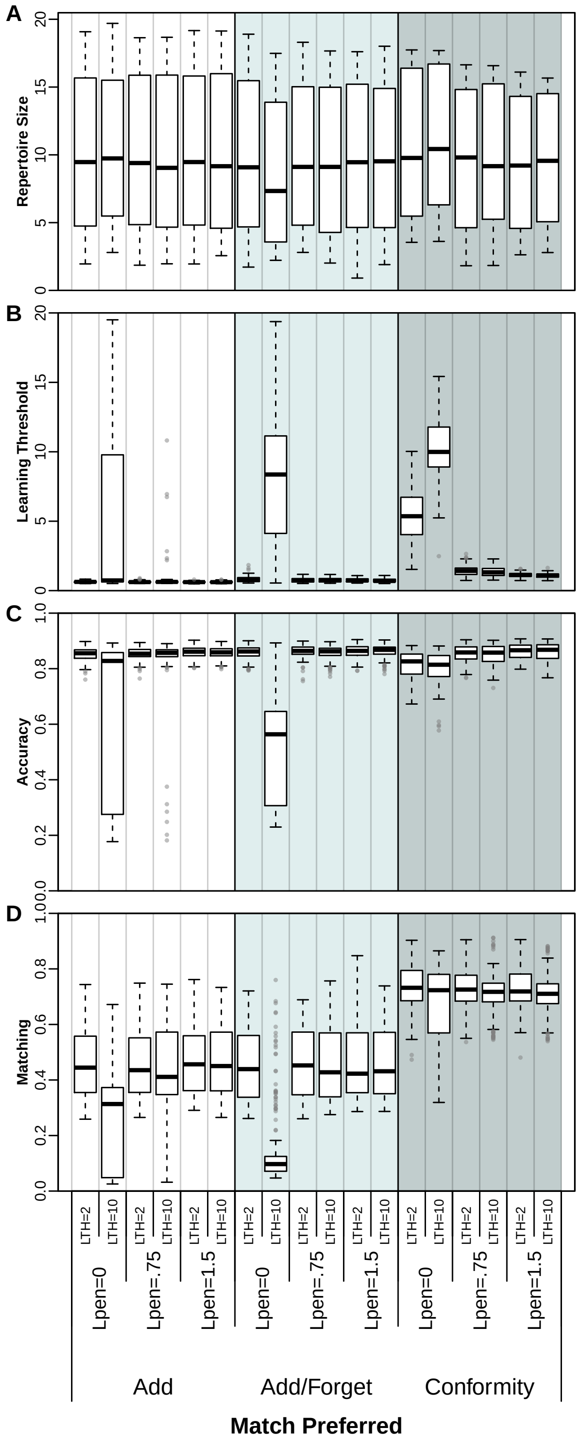


Figure 4.10: Parameter sweep when females exclusively prefer template matching. Each box plot represents 200 individual iterations, because simulations that differed only in their initial learning accuracy and repertoire size were aggregated (see **Figure 4.2** for non-aggregated data). The same iterations were used in all four plots. This plot can be directly compared to **Figure 4.3**. **Abbreviations:** LTH = Learning Threshold, Lpen = Learning Penalty.

One could hypothesize that the best template-matching would occur when birds only learn vertically and do so with high learning accuracy across generations (e.g. Darwin's finches[226,227]), so we wondered whether preventing oblique learning or reducing the inheritance noise ε for learning accuracy would lead to better matching. We found this was the case (**Figure 4.11**). When learning accuracy was inherited faithfully (small ε), birds maintained smaller repertoire sizes, higher learning accuracy, and better matches (**Figure 4.11D**). The best song template matching was achieved when ε was small and birds could only learn vertically. In contrast, Conformity learning was not as strongly affected by the magnitude of the inheritance noise (ε), and this strategy led to the best matches when ε was large (**Figure 4.11D**).

Increasing the chance to invent when syllables were inaccurately learned led to worse matching (**Figure 4.12**), while the effect of chance to forget on matching depended on the learning strategy (**Figure 4.12**). Increasing the number of Conformity tutors increased the length of the song-learning window and led to better matches (**Figure 4.14**).

In invasion assays, when there was no learning penalty, only Conformity learners converted from closed-ended to open-ended learners or from closed-ended to delayed closed-ended learners, but they did so a large proportion of the time (**Figure 4.15A**). Open-ended learners converted to delayed closed-ended learners some of the time if the learning strategy was Add or Add/Forget (**Figure 4.15A**). When there was a large learning penalty, the conversion from closed-ended to open-ended learning dropped by half (**Figure 4.15B**). The conversion from open-ended to delayed closed-ended learning increased somewhat for the Add and Add/Forget strategy, and increased substantially for the Conformity strategy (**Figure 4.15B**).

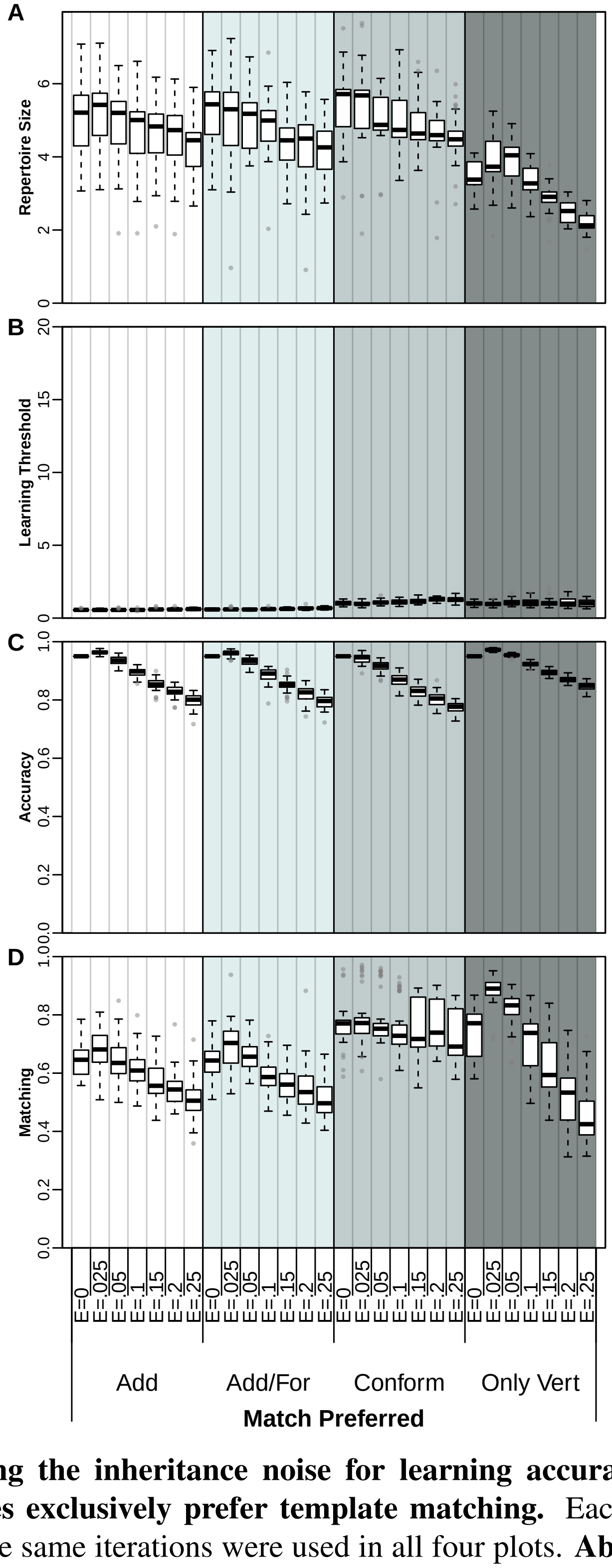


Figure 4.11: Increasing the inheritance noise for learning accuracy decreases template-matching when females exclusively prefer template matching. Each box plot represents 50 individual iterations. The same iterations were used in all four plots. **Abbreviations:** Noise (ε) = inheritance noise for learning accuracy.

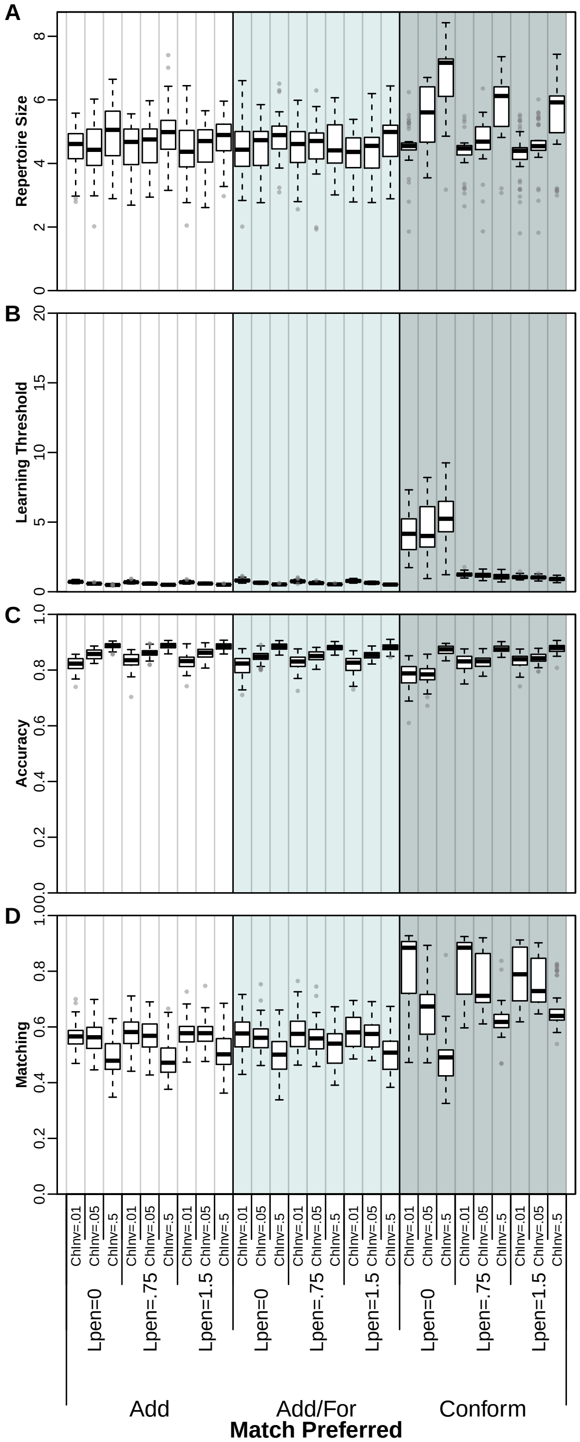


Figure 4.12: Increasing the chance to invent increases repertoire size when females exclusively prefer template matching. Each box plot represents 50 individual iterations. The same iterations were used in all four plots. **Abbreviations:** LTH = Learning Threshold, Lpen = Learning Penalty, ChInv = Chance to Invent.

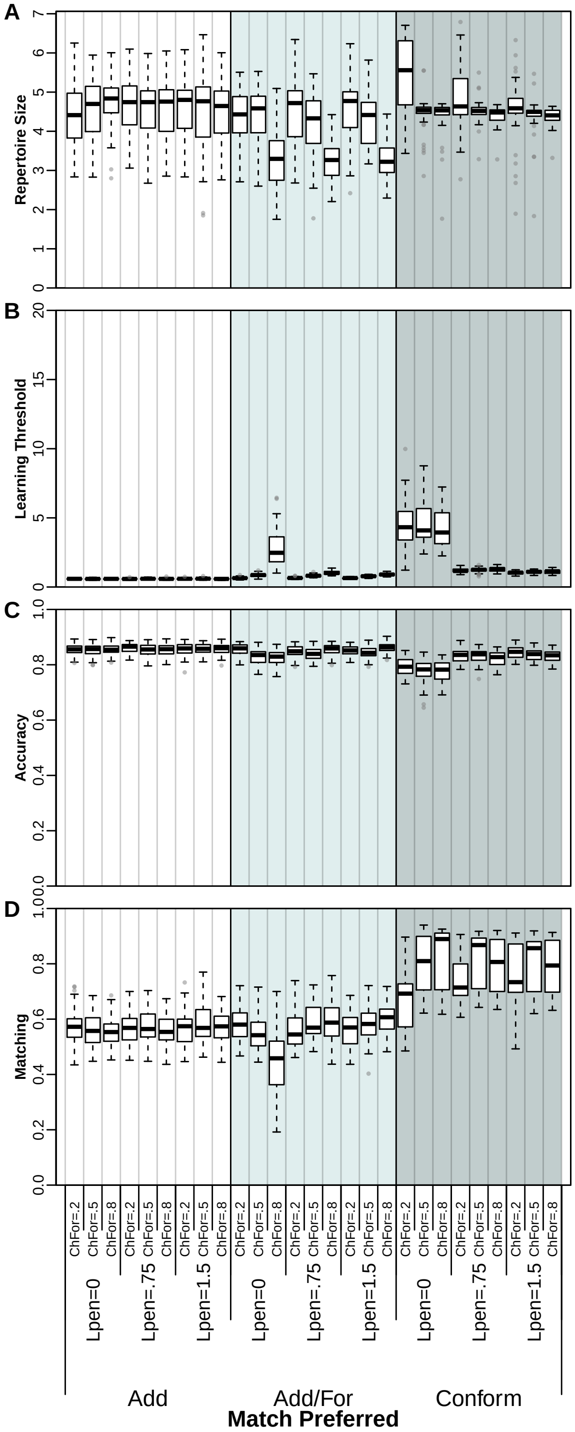


Figure 4.13: Increasing the chance to forget decreases repertoire size when females exclusively prefer template matching. Each box plot represents 50 individual iterations. The same iterations were used in all four plots. **Abbreviations:** LTH = Learning Threshold, Lpen = Learning Penalty, ChFor = Chance to Forget.

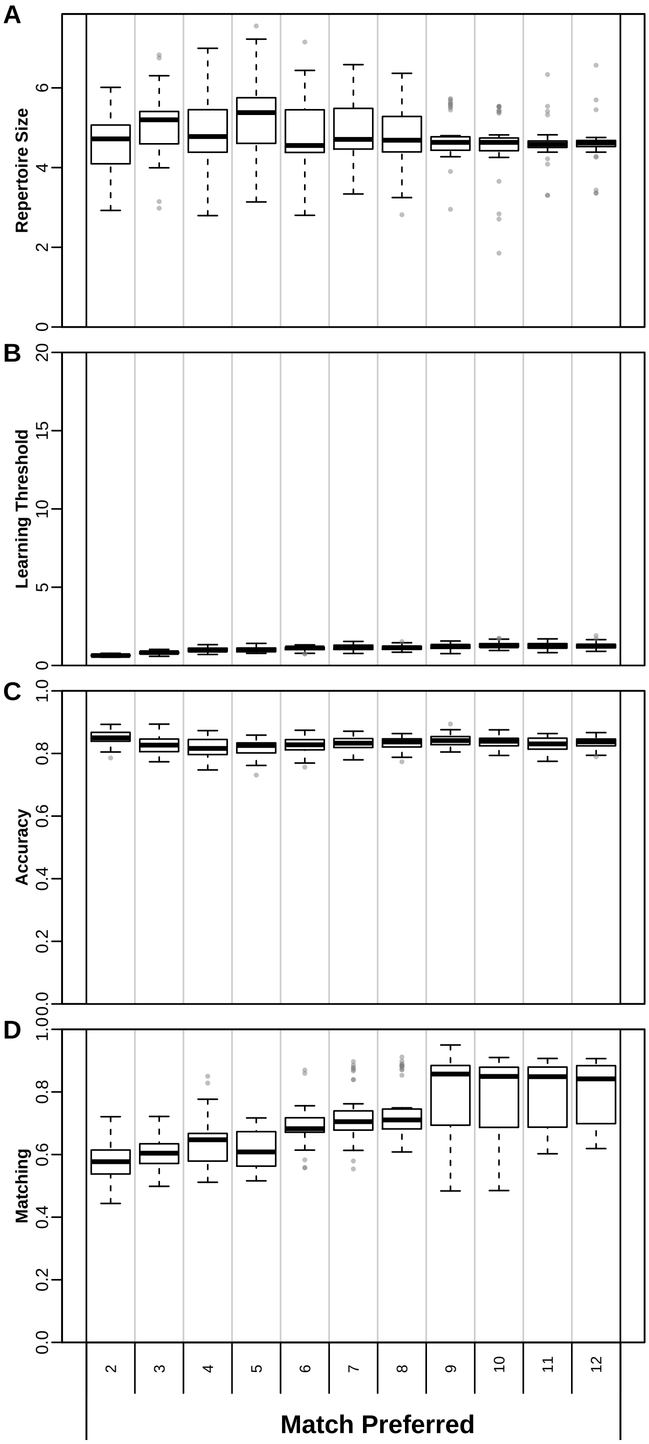


Figure 4.14: Number of Conformity strategy tutors has little effect on evolution when females exclusively prefer template matching. Each box plot represents 50 individual iterations. The same iterations were used in all four plots. Numbers on the x-axis show the number of tutors sampled.

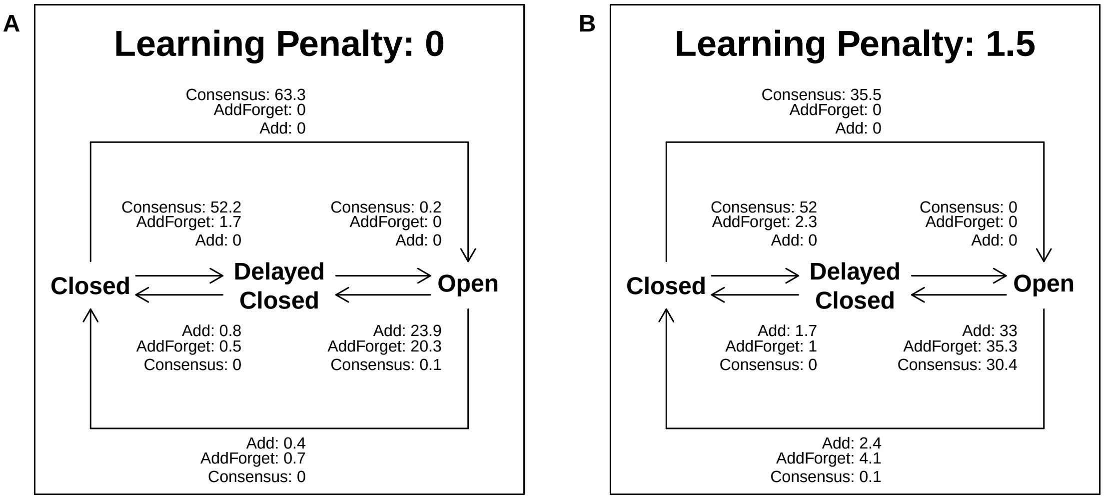


Figure 4.15: Percentage of conversion to invader learning strategies when females exclusively prefer template matching. This plot can be directly compared to **Figure 4.9**.

4.3.4 Mixed female preferences

The previous experiments examined a scenario in which females exclusively prefer either larger repertoires or songs that match a song template. However, it seems likely that many species would have a combination of these preferences, such as preferring larger repertoires but only if they also included species-typical syllables or syntax. We examined how mixed preferences interact with learning strategy and the presence or absence of a learning penalty by making heat maps showing the value of a given song-learning trait when we varied the female preference from 100% repertoire-size preference (0% template-matching preference) to 100% template-matching preference (0% repertoire-size preference) and also varied the learning penalty (**Figure 4.16**). We found that accurate matching was only maintained when the female preference for template-matching was very high and the learning strategy was Add or Add/Forget (>80%, **Figure 4.16A-B, D-E**). However, when the learning strategy was Conformity, matches were maintained fairly well until the preference for larger repertoires was greater than 50% (**Figure 4.16C, F**). While there was little compromise between matching and repertoire size for Add and Add/Forget strategies, when males used a Conformity strategy and the female preference for matching was between 40-80%, matches became somewhat worse and repertoires became somewhat larger. As we had seen previously, increasing the learning penalty decreased repertoire size and the length of the learning window (**Figure 4.16G-I**). We did not see strong evidence for an interaction between mixed female preferences and the learning penalty for any learning strategy for any song-learning trait (**Figure 4.16**); however, there is a weak trend for longer learning windows in males using the Conformity learning strategy in the face of weak learning penalties that is dependent on the female preference for larger repertoires (**Figure 4.16**I). We tested whether there was an interaction between female preferences and the initial length of the learning window and found that this was the case for all song-learning traits if the learning strategy was Add or Add/Forget (**Figure 4.17A-B, D-E, G-H**). We did not see this relationship when the learning strategy was Conformity (**Figure 4.17C, F, I**). We also tested whether there was an interaction between female preferences and initial learning accuracy, and found no strong interaction for any song-learning trait given any song-learning strategy (**Figure 4.18**).

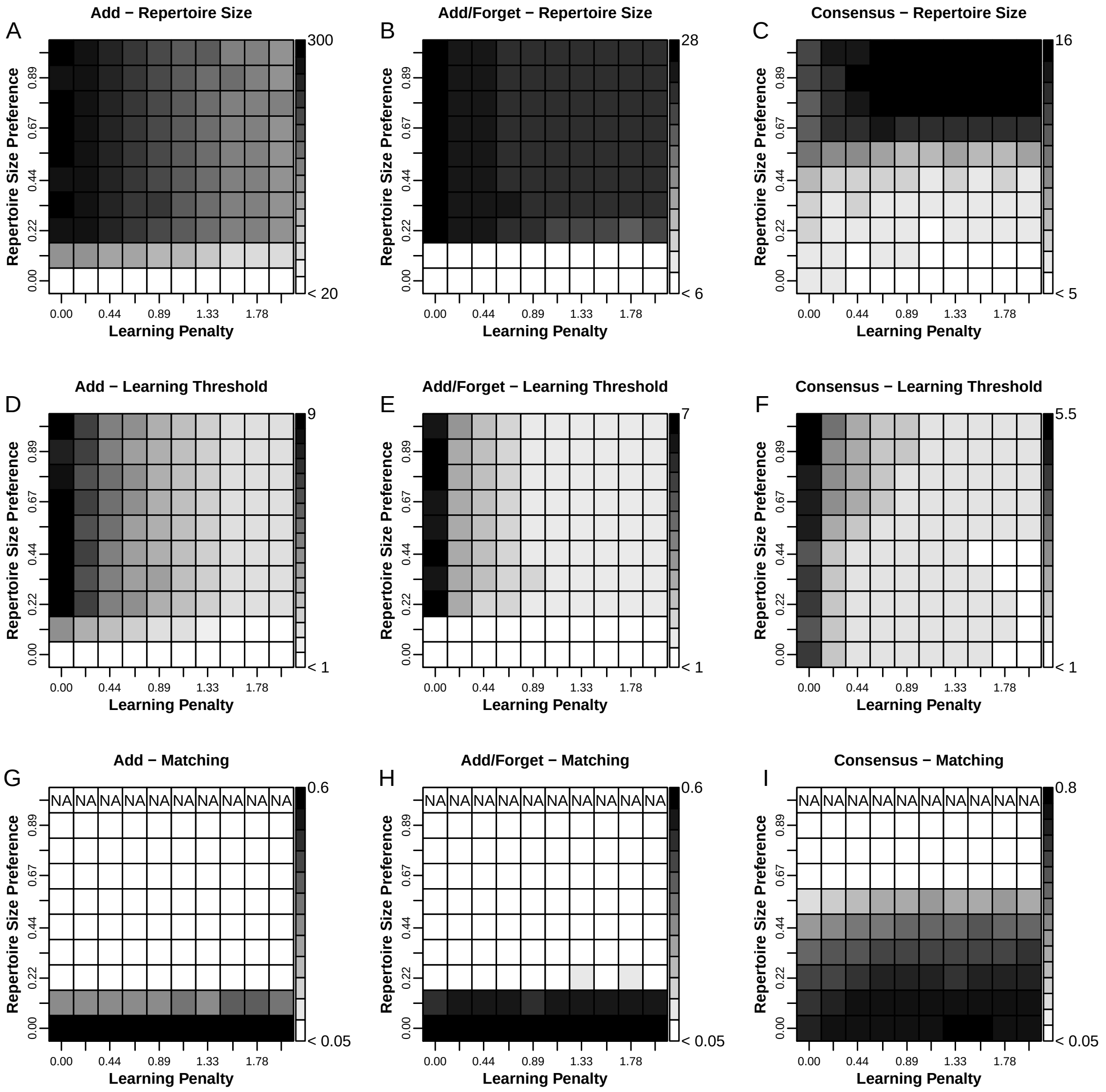


Figure 4.16: Although the magnitude of the learning penalty and female preferences individually affect song evolution, they largely do not interact; the patterns are driven by one parameter or the other. The only notable exception to this occurs with the Conformity learning strategy (e.g. panels C, F); the length of the learning threshold was less strongly affected by the learning penalty when the preference for larger syllable repertoires was greater. Darker colors indicate larger values. The color scale for each graph is unique and marked on the right of the graph. The female preferences for larger repertoires and matching a song template always add to 1, so a smaller preference for larger repertoires means a larger preference for template matching. NAs note that data were not collected for matching, because females templates are not generated when they exclusively prefer larger syllable repertoires.

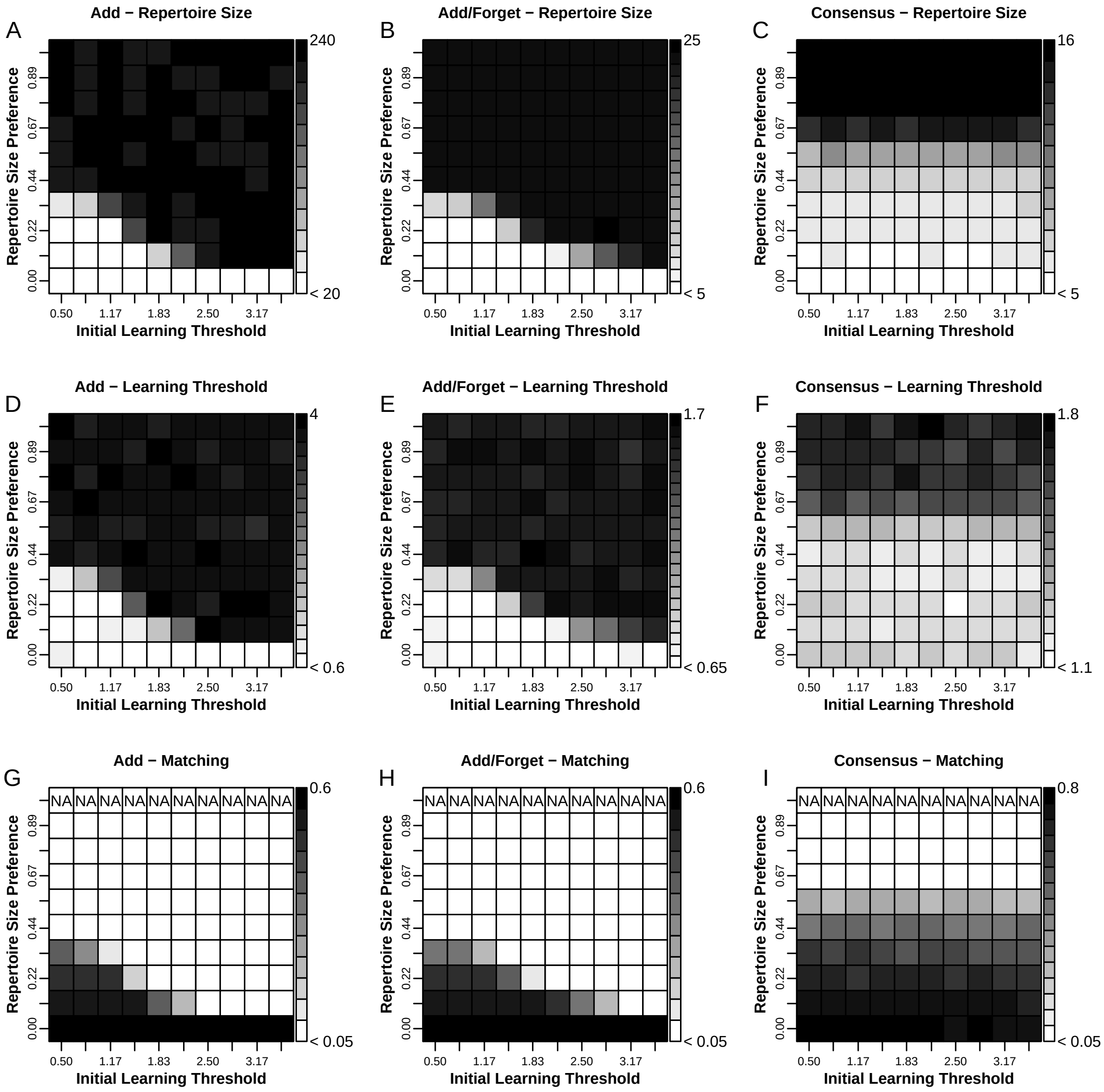


Figure 4.17: Initial learning threshold interacts with female preferences to affect song evolution. For the Add and Add/Forget strategies, the diagonal patterning shows that repertoire size and the learning threshold grow larger as the initial learning threshold increases, but stronger preference for matching can overcome this. Darker colors mean larger values. The color scale for each graph is unique and marked on the right of the graph. The female preferences for larger repertoires and matching a song template always add to 1, so a smaller preference for larger repertoires means a larger preference for template matching. NAs note that data were not collected for matching, because females templates are not generated when they exclusively prefer larger syllable repertoires.

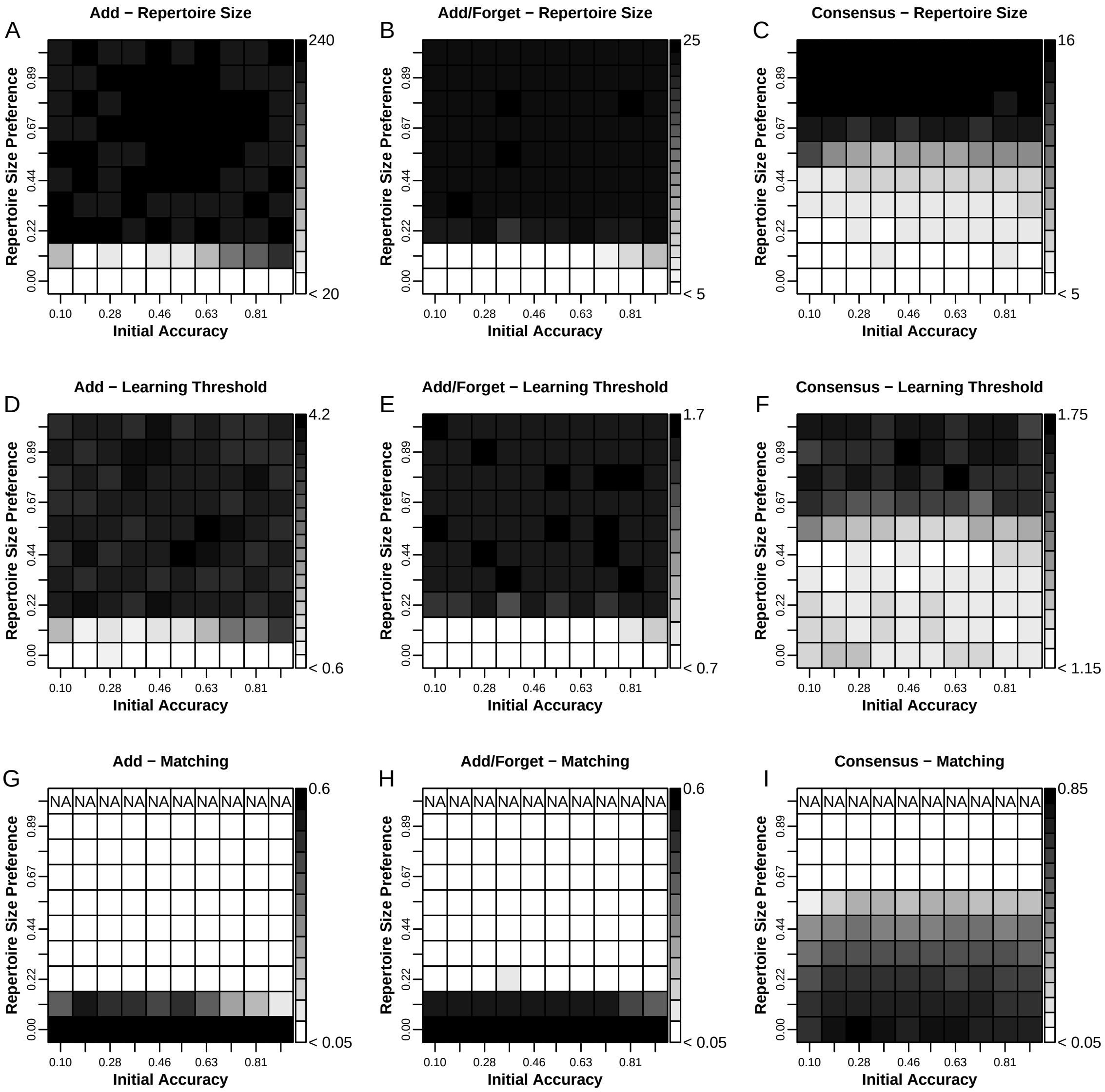


Figure 4.18: Initial learning accuracy has little affect on song evolution.Darker colors mean larger values. The color scale for each graph is unique and marked on the right of the graph. The female preferences for larger repertoires and matching a song template always add to 1, so a smaller preference for larger repertoires means a larger preference for template matching. NAs note that data were not collected for matching, because females templates are not generated when they exclusively prefer larger syllable repertoires.

4.4 Discussion

There is evidence from field and laboratory studies that female birds prefer different aspects of song, that different birds species employ different strategies to learn song, and that the length of the song-learning window varies between species. Because all these variables could affect birdsong evolution, it seems likely that they would interact, but examining these interactions in live birds would be difficult or impossible. Here, we used a computational model that simulates how different female preferences and male learning strategies affect the length of the song-learning window, syllable repertoire, and other song-learning traits.

4.4.1 Which learning strategies were best suited to different female preferences?

When females preferred larger repertoires exclusively, the Add strategy, where males indiscriminately learn new syllables, led to the largest syllable repertoires. Furthermore, males using this strategy maintained longer learning windows (>2 years) even when extended learning resulted in a large fitness cost. This makes intuitive sense; if females prefer elaboration and not specific features of song, then learning to produce every possible sound is in a male's best interest. While it is unlikely that females of many bird species prefer males based only on elaboration with no regard to syntax or species-specific features, some species show evidence for this pattern: mimids, starlings, and lyrebirds add sounds to their repertoires that were originally acquired from other species, or even from abiotic sources, and have very large repertoires[170,212,214]. The Add/Forget and Conformity strategy did not lead to such large repertoires or long learning windows, but simulations using these strategies maintained oblique learning, suggesting that the oblique learning after vertical learning did confer a benefit over pure vertical learning when females prefer larger repertoires. We also note that even when the female preference for repertoire size was reduced (to 20% for the Add strategy and 30% for the Add/Forget strategy), males still maintained oblique learning and increased their repertoire sizes. Together, these results suggest that longer learning windows could be generally beneficial for males in the face of moderate-to-strong female preference for larger repertoires. Indeed, we showed previously in a cross-species meta-analysis of actual field data that longer learning windows were associated with larger repertoire sizes **Chapter 3**[70].

When females preferred males with songs that match song templates, the Add and Add/Forget strategies for oblique learning were detrimental to template matching, and males would evolve shorter learning windows that minimized oblique learning. This was not surprising for the Add strategy, but was unexpected for the Add/Forget strategy. We thought that the Add/Forget strategy would allow a learner to lose inappropriate syllables that he had learned from his father, which would make extended learning valuable, as has been suggested for selection-based learning strategies[221,228]. However, in the invasion assays, where males were in non-evolvable states of closed-ended (learning window duration 0.25 years), delayed closed-ended (1 year), and open-ended (2 years), there was a very low probability of conversion from open- or delayed closed-ended to closed-ended learning for models using these learning strategies. This suggests that learning obliquely is more beneficial than exclusive vertical learning when the entire population has to pick one strategy or the other. These two results together imply that even with the strategies that greatly reduced the length of the song-learning window, having some percentage of oblique learning in a population can be valuable when females prefer songs that match a template, though most males will avoid engaging in oblique learning. This may help explain cases such as Great Tit (*Parus major*), where one paper showed that individuals of this species can change their syllable repertoire across years[229], while a larger-scale study in a different population of this species showed that the syllable repertoire was consistent across years[136]. It may be that a only small percentage individuals in this species is capable of or chooses to modify their syllable repertoires, while the vast majority do not. Alternatively, it has been demonstrated in a number of species that the length of the song-learning window is subject to social and environmental experiences and hormonal cues[28,59,161,197,198], and this flexibility is believed to be beneficial for birds raised in less than ideal circumstances. Although we did not include a mechanism to allow birds with poor songs to choose whether to learn or not, our results provide further support for this hypothesis that flexible timing for the closing of the song-learning window is valuable.

On the other hand, simulations using the Conformity strategy did maintain at least one year of oblique learning when females preferred songs that matched a template. This strategy was fairly effective in maintaining matches in our main parameter sweep, and modifying other parameters (the inheritance error of learning accuracy, the number of tutors sampled, the chance to forget, or the chance to invent) was sufficient to allow for average population matches of 90% or greater. Together, these results suggest that oblique learning via conformity bias is effective when females prefer template-matching. Indeed, conformity bias has also been suggested to allow for extremely stable song traditions in swamp sparrows[211]. However, simulations where only vertical learning was possible (similar to Darwin's finches[226]) and learning accuracy was highly heritable led to the tightest distribution of accurate matches (greater than 90%). This leaves open the question of what benefit Conformity learning offers over just vertical learning. When females showed mixed preferences for larger repertoire sizes and template matching, we found that the Conformity strategy allowed the balancing of these two preferences, and repertoire sizes would grow somewhat larger while also maintaining some ability to match a template. Perhaps, the evolutionary value of Conformity learning could be in balancing strong-to-moderate preferences for matching with a weak-to-moderate preference for larger repertoires. Alternatively, in species where song-matching between territorial neighbors is critical, such as in swamp sparrows, the father's song template may not be appropriate for the final breeding territory of his offspring, so extended or delayed learning may be necessary even if it is more costly.

4.4.2 Must extended learning be costly to prevent the evolution of universally long learning windows?

Although it is likely that there is some cost to maintaining the ability to learn song in adulthood, we found that longer learning does not need to incur a cost for shorter learning windows to be advantageous. When females preferred songs that matched a template, males with the Add or Add/Forget strategies tended to evolve shorter song-learning windows (~0.5 years, **Figures 4.2** and **4.10**) even when longer learning had no fitness cost. Furthermore, the invasion assays showed that open-ended learners using the Add or Add/Forget strategies could be converted into delayed closed-ended learners even when there was no learning penalty. Overall, these results imply that oblique learning can potentially produce song that fits female preferences less well than it would have been without this extra learning; this effect is sufficiently detrimental to mate choice that birds can evolve shorter song-learning windows in the absence of a specific metabolic cost.

That said, we found that the evolution of shorter learning windows occurred more readily if a learning penalty existed. Models with the Add or Add/Forget strategy universally evolved short learning windows in the presence of a learning penalty, even if the initial learning window was 10 years, and increased conversion from open-ended to delayed closed-ended learning in the invasion assay. Importantly, application of a fitness cost on longer learning did not prevent models from evolving longer learning windows (greater than 1 year) when learning was beneficial, though the learning windows evolved to be shorter as the penalty became greater. Together, these results show that an extended learning window does not not need to be metabolically costly to evolve shorter learning windows, though such costs do make this transition occur more readily. Of greater importance, our results call attention to an idea not often considered in birdsong: adult learning does not always lead to better song, and can in fact lead to worse song. To understand why some species engage in adult learning, while others do not, assessing the positive and negative outcomes of different learning strategies given different song constraints and selection pressures will be as important as identifying the metabolic and time commitment costs of this learning.

4.5 Limitations and future directions

Although we frame our interpretation in the lens of female preference, our results are not limited to questions of mate choice; as implemented in this paper, the model chooses which males will breed based on their repertoire size and/or their match to a single, template. Thus, these results can be generalized to any selective pressure that would give males with certain song traits a reproductive advantage, such as male-male song competitions over territory boundaries or status in dominance hierarchies. However, this model also omits several interesting facets of song learning.

It has been shown that the social context in which a learner hears song is important[32–34]. Juvenile birds may be more likely to attend to and learn from interactions between multiple birds, where they are perhaps judging whether to learn from a tutor based on whether he successfully attracts a mate or fends off another male. In this model, the only fitness information that a juvenile bird knows about other males is that its father reproduced successfully, so he would make a good tutor. In contrast, learners randomly choose oblique tutors, regardless of the tutor's previous success in attracting mates. If we recoded the model so that learners could discriminate between successful and unsuccessful males and only learn from successful males, then oblique learning would likely be more valuable. However, it is unclear whether adding this feature would simply increase the speed at which song evolves to fit selective pressures or whether it would allow song to better adapt to those pressures leading to higher accuracy matching, larger repertoires, and a better balancing between preferences in mixed preference contexts. We also conducted most of our analysis with one tutor-choice strategy, where males learn most of their father's song and some additional syllables from oblique tutors, even though real birds exhibit numerous patterns of song learning[220–224]. However, allowing for other tutor-choice strategies, greater oblique learning, and reduced vertical learning did not greatly affect relationships between song-learning traits when females preferred larger repertoires. Because the song-matching templates in the majority of experiments were small (5 syllables +/- 2), and repertoire sizes did not grow significantly, implementing different tutor-choice strategies would not greatly affect the outcomes of simulations where females preferred males whose songs matched a template.

Furthermore, our model only examined the utility of learning when selective forces remain static. If female preferences are evolving, then some learning strategies may be better suited to adapt to the changing pressure. Based on the results shown in our heat maps for the learning strategies we tested **Figure 4.16**, we would hypothesize that the Conformity learning strategy would be best suited to changing female preferences, since there was a range over which it allowed males to optimize between repertoire size and template matching instead of picking one strategy over the other. Additionally, our model does not allow for males to evolve different learning strategies. It is unknown what biological and environmental conditions lead to the evolution of new learning strategies; however, it is likely that, in some species, males might be able pick one of several learning strategies that persist in the population. Indeed, it has been shown that males adopt different mate guarding strategies within a single population, and these strategies exist in an equilibrium with one another[230].

Finally, other female preferences have been proposed which we did not examine here. In canaries (*Serinus canaria*), it has been shown that females respond more strongly to "sexy syllables" within song[231,232]. However, it is not clear from these studies whether males are able to figure out which syllables females prefer. In our model, syllables either all have the same value when larger repertoires are preferred, or they are "good" (within a song template) or "bad" (not in the song template) when females prefer template matching. The model is constructed such that implementing feature-weighted syllables to examine how sexy syllables affect song evolution would be possible in the future. Our model also does not explicitly consider syntax or how well males perform complex syllables. Because "syllables" in our model are just the presence or absence of a value in a male's repertoire, what we define as a syllable could be re-framed as a syntax element that is performed correctly or not. However, it may not be possible to examine continuous traits like preference for better performance in our model as currently implemented.

In the future, our model is poised to analyze many other important questions regarding learning and song evolution. Currently, it allows for the establishment of dialects within a population, evolution of female song template preferences, the prevention of vertical or oblique learning, and for mate and tutor choices to be made globally as we did in this study or from a pool of nearby neighbors. By providing the open-source code, we hope to encourage other researchers to use and expand on the model to generate testable predictions for other questions regarding the evolution of song-learning. Here, we provide evidence that short learning windows can evolve in the absence of an inherent metabolic cost, because not all learning strategies are beneficial in all sexual selection contexts. Furthermore, our results suggest that even in populations that do not typically learn as adults, it can be beneficial to the population for some individuals to learn as adults. This population-level variability could potential allow a species to more quickly adapt to changing female preferences.

Chapter 5

Mechanisms for the establishment and maintenance of songbird dialects

5.1 Introduction

Regional song dialects occur when songbirds of one species sing one song type in one region and members of the same species in another location sing a different song[1,209,233]. In species with song dialects, males typically only know one dialect, and researchers rarely find mixed-dialect singers that know multiple dialects or males that sing a song that is a hybrid between two dialects[234–236]. Furthermore, dialects appear to be fairly stable. For example, in a population of corn buntings (*Emberiza calandra*), where dialects have been particularly well studied, both the syllable composition of individual dialects and the boundaries between dialect regions remained fairly constant from 1977 to 1985[234,237]. This stability implies that dialects are important in species that have them, however it is still not clear how dialects are formed, maintained, or what purpose they serve.

It could be that dialects are formed simply as a byproduct of fledglings learning from neighbors at their breeding location, and no other specific behavior is necessary. One computational model attempted to study whether dialects could form based solely on error-prone learning from neighbors[238]. In this model, song sharing between more than a few neighbors could only be formed if the error-rate on learning and mortality rate were unnaturally low. Therefore, the authors suggested that their learning-only model was more suitable to examining the formation of small song neighborhoods than large dialects, and that dialect formation relies on other factors.

Dialects may be the target of sexual selection if females prefer specific dialects. Perhaps dialects are used to advertise kinship to promote or prevent assortative mating[91]. Indeed, it has been shown in laboratory tests that female white-crowned sparrows (*Zonotrichia leucophrys*) gather more suitable nesting material when exposed to their natal dialect rather than a foreign one[239], whereas brown-headed cowbirds (*Molothrus ater*) respond with more copulation displays to foreign dialects[240]. For this reason, it has been suggested that dialect formation relies on genetic differentiation between populations[233,241]. However, multiple dialect regions can be adjacent to one another and fledglings do not always learn the same dialect as their father[242]. Indeed, analysis of white-crowned sparrows showed that isolation by distance explained genetic differences between populations as well as isolation by dialects[243] (also see[244]; however[245]), and analysis of rufous-collared sparrows (*Zonotrichia capensis*) showed little genetic differentiation between birds on either side of a dialect boundary[246] (see also[247]). Overall, this evidence suggests that dialects can be a cultural phenomenon that do not always reflect genetic divisions between different populations.

Alternatively, females may prefer common songs. Indeed, indigo buntings (*Passerina cyanea*) that share songs with their neighbors are more likely to earn mates[248] and song sparrow (*Melospiza melodia*) females attend more closely to songs that are similar to their mate's[249]. Consistent female preference for common songs could drive the formation of local dialects, as males with rare songs would be less fit and thus less likely to pass on rare songs onto the next generation. It could also be that male-male interactions like song-matching are critical for territorial interactions, and males that do not sing the correct dialect cannot hold territories[250].

Here, I used my agent-based computational model described in **Chapter 4** to examine the contexts that allow dialects to establish and be maintained in a population. For these simulations, females preferred songs that were more like their song-template, and males could evolve across generations to maintain dialects or evolve songs that matched the female song-template with high template-matching accuracy. I found that sexual selection on dialects was required for regional dialects to be maintained. I also found that males needed to learn song preferentially from reproductively successful tutors to establish dialects. Furthermore, learners were required to learn the entirety of a song from one male either vertically or obliquely to maintain or establish dialects, though higher accuracy template-matching was achieved if males subsequently listened to other tutors and modified this initial song according to conformity bias. Together, my model predicts a set of behaviors that are theoretically necessary for dialect formation and maintenance, which creates new hypotheses for field researchers to test in real bird species.

5.2 Methods

5.2.1 Data and code availability

C# library, programs, and walk-through:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/Csharp-Library-and-Programs

R package, code manual, and walk-though:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/R-Package

Parameter generators and data visualizing code:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/DialectsPaper

5.2.2 Model features

I used the **C#** version of the model for all experiments, though R was used to generate the parameter sets and plot the data. The model operates as described in **Chapter 4**, though I employed additional functions here. The parameters that were modified for this chapter and their possible values are listed in **Table 5.1**. The default learning accuracy inheritance noise (ε) was reduced to 0.1 for all simulations based on results from **Chapter 4**. All other unmentioned parameters were left as the model defaults (see the **Walk-through**). For these simulations, when females have a 0% preference for template-matching, this means that females have no preferences for any song feature. This differs from **Chapter 4**, wherein females that had no preference for template-matching had a preference for larger repertoires. Here, when females have a 0% template-matching preference, the only selection pressures on song evolution is that males that sing zero syllables cannot breed.

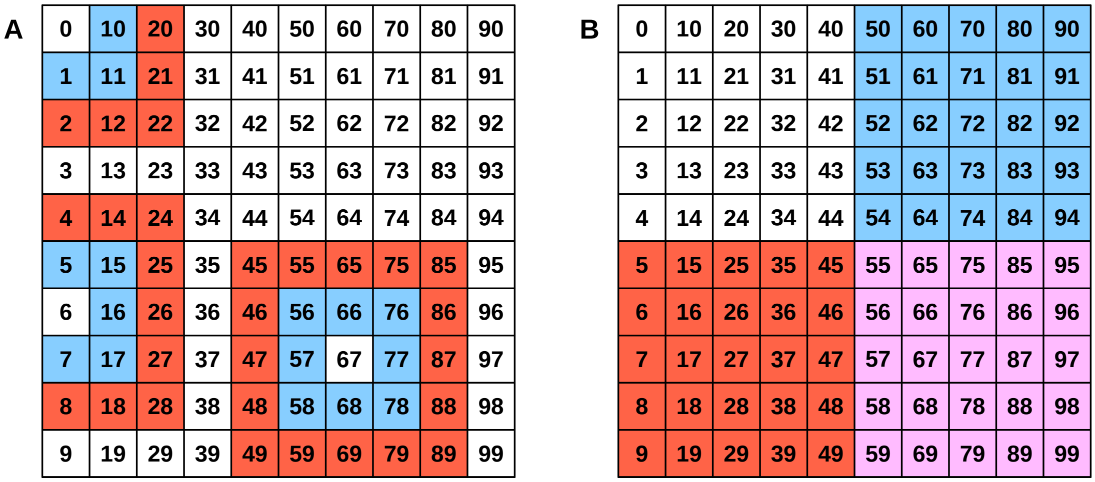
Table 5.1: All parameters modified in this chapter and their defaults. A full list of all parameters and their defaults is available in the walk-through.

Variable	**Meaning**	**Default**	**Range**
Breed	Whether chicks are generated from fathers local to a vacant territory (nest-site fidelity), or if any male can father a chick for any territory (no nest-site fidelity).	NA	True, False
Tutor	Whether learners choose local males as tutors, or if any male can tutor a learner from any territory.	NA	True, False
Vertical	Whether vertical learning is enabled.	True	True, False
Social	Whether learners are more likely to pick tutors that bred in the last time step.	False	True, False
Learning Strategy	The method birds use to learn obliquely. **Add:** the learner attempts to add all the syllables that his tutor sang to his repertoire. **Conformity:** the learner listens to multiple tutors and adds syllables based on conformity bias, and can lose syllables that no tutor sang.	Conformity	Conformity, Add
Female Template-Matching Preference	The amount of female preference that relies on a male matching her song template.	100%	100%, 0%
Dialects	The number of regional dialects in the matrix.	NA	1, 2, 4, 8, 16
Uniform Templates	Whether female templates are uniform in a dialect region or if each female has her own variant of the regional dialect	NA	True, False
Male Dialect	How the initial male dialects are established. **None:** males have uniquely generated songs composed of syllables from regional dialect 1. **Similar:** each male has a uniquely generated song composed of syllables from his regional dialect. **Same:** each male has a song that is identical to his resident female's song template.	NA	None, Same, Similar
Initial Syllable Repertoire size	the number of syllables birds have a 90% chance to know at the start of a simulation	5	5, 1

Syllable Overhang	This values times the syllable repertoire size is the number of syllables birds have a 10% and 1% chance to know. For most simulations birds have a 90% chance to know each of syllables 1-5 (Initial Syllable Repertoire Size = 5), a 10% chance to know syllable 6 and a 1% chance to know syllable 7.	20%	20%, 0%
Rows	The number of rows in the matrix.	20	4, 5, 10, 20
Cols	The number of columns in the matrix.	20	100, 8, 40, 20

In my previous work, any bird in the matrix could be a potential tutor for all other birds (global tutor choice), and when chicks were generated, they could populate any territory in the matrix (no nest-site fidelity). However, the formation of regional dialects almost certainly depends on local interactions between birds, so I also implemented local restrictions on these steps. In other words, for local tutor choice, males could only pick their immediate neighbors as tutors and, during the breeding step, empty territories were repopulated with a chick fathered by an immediate neighbor to the empty territory (high nest-site fidelity). The immediate neighbors for three example territories are shown by the blue squares in **Figure 5.1A**. For these local mechanisms, if there are not enough immediate neighbors at a specific territory for tutor choice during learning (i.e. neighbors are dead, chicks, or have a zero-syllable song) or father choice during breeding (i.e. neighbors are dead or have a zero-syllable song), the scope for local birds around that territory temporarily expands by one step (red squares in **Figure 5.1A**). The scope for local birds will continue to expand until there are enough local birds to complete the learning or breeding process.

Regional dialects are created by shifting the syllable space to the right. In other words, if regional dialect 1 can be composed of syllables 1-7, then regional dialect 2 can be composed of syllables 8-14, and regional dialect 3 can be composed of syllables 15-24. This means that initially, there is no overlap in the syllable repertoires of males in different dialect regions; however, syllable repertoire overlap between dialects can form as the simulation progresses if some learners acquire syllables from males in dialect regions that the learners do not reside in. Each bird is assigned a regional dialect based on its position in the matrix. The matrix is divided into regions such that each region is as square as possible and the same number of birds belong to each region (see **Figure 5.1B** for illustration).



A

0	10	20	30	40	50	60	70	80	90
1	11	21	31	41	51	61	71	81	91
2	12	22	32	42	52	62	72	82	92
3	13	23	33	43	53	63	73	83	93
4	14	24	34	44	54	64	74	84	94
5	15	25	35	45	55	65	75	85	95
6	16	26	36	46	56	66	76	86	96
7	17	27	37	47	57	67	77	87	97
8	18	28	38	48	58	68	78	88	98
9	19	29	39	49	59	69	79	89	99

B

0	10	20	30	40	50	60	70	80	90
1	11	21	31	41	51	61	71	81	91
2	12	22	32	42	52	62	72	82	92
3	13	23	33	43	53	63	73	83	93
4	14	24	34	44	54	64	74	84	94
5	15	25	35	45	55	65	75	85	95
6	16	26	36	46	56	66	76	86	96
7	17	27	37	47	57	67	77	87	97
8	18	28	38	48	58	68	78	88	98
9	19	29	39	49	59	69	79	89	99

Figure 5.1: Illustration of local territories and dialect regions. Matrix (A) shows which territories are considered local when neighbors are 1 (blue) or 2 (blue and red) steps away for a corner (0), edge (6) and center territory (67). Matrix (B) shows how 4 dialects would be positioned in this matrix. Regional dialect 1=white, 2=red, 3=blue, 4=purple.

Dialects in males and females are treated slightly differently. For females, dialects can be set in one of two ways:

1. **Uniform Templates:** regional dialects are composed of a single song template that all females in a region share.

2. **Noisy Templates:** each female has a uniquely generated song template composed of her region's syllables.

Male dialects can be set in one of three ways:

1. **Same:** males begin with songs that are a copy of their resident females' song template (all males begin with 100% template matching accuracy).

2. **Similar:** males have uniquely generated songs that use only the syllables for the regional dialect of his resident female.

3. **None:** males have uniquely generated songs that use only the syllables for dialect 1. Note that Similar and None led to identical starting properties if there is only one regional dialect.

Finally, one of the previous limitations of the model was that learners could not judge which potential tutors had appropriate songs, while vertical tutors (fathers) were always likely to be males with appropriate songs, because they had bred. This made it difficult to disentangle whether vertical learning was superior to oblique learning in certain cases, or if social cues that reveal tutor fitness are critical for valuable learning in those cases. To address this, I implemented a new, optional feature that makes learners more likely to choose oblique tutors that bred in the previous time step.

All parameter sets were run for 4000 time steps and repeated 50 times based on our previous testing (see **Chapter 4**). For these experiments, simulations generally reached their equilibrium state by 400 time steps. To assess how well regional dialects were maintained or established, I visualized the average matching accuracy between the male song and female song template in each territory across the 50 iterations every 200 time steps as heat maps, though only noteworthy examples are shown. To quantitatively access dialect establishment and maintenance in the matrix for all experiments, I took the average match across the 50 iterations at the final time step for the following key locations:

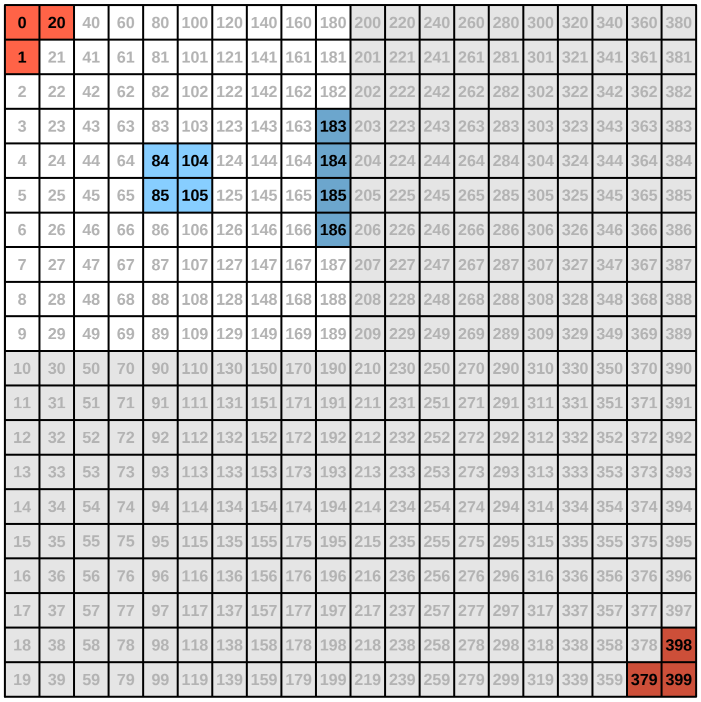
1. **Corner:** The outer corner of regional dialect 1 (i.e. light red squares in **Figure 5.2**).

2. **Far Corner:** The outer corner of the final regional dialect (i.e. dark red squares in **Figure 5.2**)

3. **Center:** The territory closest to the middle of regional dialect 1 (i.e. light blue squares in **Figure 5.2**).

4. **Edge:** The territory to the right of center at the edge of regional dialect 1 where it would contact the regional dialect to its right (i.e. dark blue squares in **Figure 5.2**).

Tables with all quantifications are located in **Appendix D**. Center and edge values were used to examine how well dialects were maintained in maintenance experiments, while corner and far corner were used to examine whether dialects established in establishment experiments.



0	20	40	60	80	100	120	140	160	180	200	220	240	260	280	300	320	340	360	380
1	21	41	61	81	101	121	141	161	181	201	221	241	261	281	301	321	341	361	381
2	22	42	62	82	102	122	142	162	182	202	222	242	262	282	302	322	342	362	382
3	23	43	63	83	103	123	143	163	183	203	223	243	263	283	303	323	343	363	383
4	24	44	64	84	104	124	144	164	184	204	224	244	264	284	304	324	344	364	384
5	25	45	65	85	105	125	145	165	185	205	225	245	265	285	305	325	345	365	385
6	26	46	66	86	106	126	146	166	186	206	226	246	266	286	306	326	346	366	386
7	27	47	67	87	107	127	147	167	187	207	227	247	267	287	307	327	347	367	387
8	28	48	68	88	108	128	148	168	188	208	228	248	268	288	308	328	348	368	388
9	29	49	69	89	109	129	149	169	189	209	229	249	269	289	309	329	349	369	389
10	30	50	70	90	110	130	150	170	190	210	230	250	270	290	310	330	350	370	390
11	31	51	71	91	111	131	151	171	191	211	231	251	271	291	311	331	351	371	391
12	32	52	72	92	112	132	152	172	192	212	232	252	272	292	312	332	352	372	392
13	33	53	73	93	113	133	153	173	193	213	233	253	273	293	313	333	353	373	393
14	34	54	74	94	114	134	154	174	194	214	234	254	274	294	314	334	354	374	394
15	35	55	75	95	115	135	155	175	195	215	235	255	275	295	315	335	355	375	395
16	36	56	76	96	116	136	156	176	196	216	236	256	276	296	316	336	356	376	396
17	37	57	77	97	117	137	157	177	197	217	237	257	277	297	317	337	357	377	397
18	38	58	78	98	118	138	158	178	198	218	238	258	278	298	318	338	358	378	398
19	39	59	79	99	119	139	159	179	199	219	239	259	279	299	319	339	359	379	399

Figure 5.2: Illustration of key locations for dialect quantification. White squares are territories that make up regional dialect 1, while grey squares belong to dialects 2-4. Light red shows territories used for corner measurements, while dark red shows those used for far corner measurements. Light blue territories were used for center measurements and dark blue territories were used for edge measurements.

5.3 Results

5.3.1 Maintaining existing dialects

I first wanted to know which parameter sets would allow existing dialects to persist in the matrix across many generations. I ran simulations where 1, 2, 4, 8, or 16 regional dialects were included in the matrix, and males knew a song that matched the dialect of the region in which they resided. Simulations with only a single "regional" dialect were included to measure the best average template-matching that could be achieved for different parameter combinations. Dialects could be maintained for some parameter sets even when there were 16 dialects (**Figure 5.3** and **Table D.1.1**), though the more dialects that existed in the matrix, the worse the average template matching became. For one-dialect simulations, the highest template-matching accuracy was achieved when vertical learning was enabled, tutor choice was global, and there was no nest-site fidelity (**Table D.1.1**). For simulations with multiple dialects, the best average template-matches occurred when males learned vertically, picked tutors locally, and exhibited high nest-site fidelity (**Figure 5.4B** and **Table D.1.1**). Interestingly, I saw a quilt-like pattern appear in simulations with these parameters, where the edges of a dialect region showed worse template matching than the center of the dialect region (**Figures 5.3** and **5.4B**). It seems likely that this pattern emerged because males on the edge of a dialect region could have learned from either a father or oblique tutors that resided in another dialect region. In the parameter sets with the next highest template-matching accuracy, males learned vertically and either chose tutors locally or exhibited high nest-site fidelity (**Figure 5.4D** and **Table D.1.1**). Interestingly, I did not see the quilt-pattern form in these simulations; there was little difference between the average template-matching at the edge of a dialect region and its center, which was similar to what occurred in the parameter sets where vertical learning was disabled and which showed the lowest template-matching accuracy (**Figure 5.4**). In general, simulations wherein females from each dialect region had a uniform template and simulations wherein males started the simulation with the same template as their resident females maintained higher average template-matching accuracy than those that did not (**Table D.1.1**). I also tested whether dialects could be maintained when there was no sexual selection pressure acting on song to actively perpetuate dialects. Dialects were either lost (e.g. **Figure 5.5B**) or very poorly maintained (e.g. **Figure 5.5D**) when females did not prefer specific song templates (**Table D.1.2**).

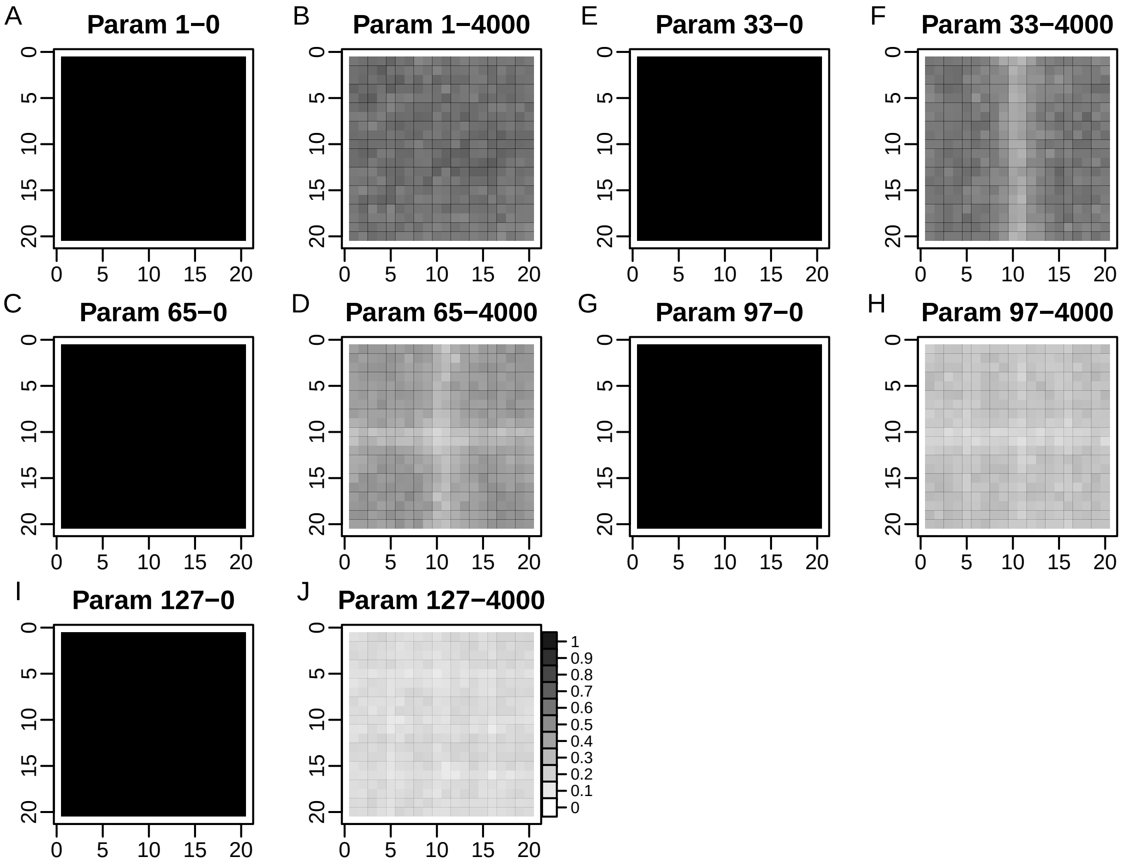


Figure 5.3: Example heat maps of template-matching accuracy in simulations where dialects were maintained for each number of regional dialects. Each square in a plot represents a territory. Colors are linearly scaled, where black is perfect template-matching accuracy. The x and y axes are numbered based on the rows and columns in the matrix respectively. Parameter set numbers match **Table D.1.1**. The second number in plot titles is the time step when the data was collected. Plots have 1 (A-B), 2 (C-D), 4 (E-F), 8 (G-H), or 16 (I-J) dialects. All other parameters are the same. Vertical learning was enabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

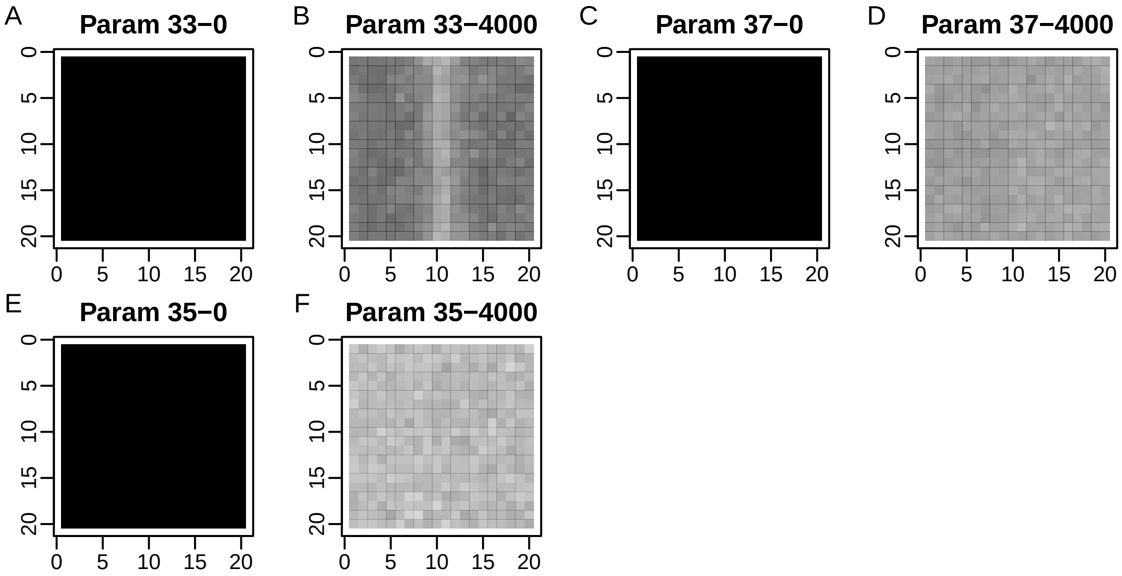


Figure 5.4: Example heat maps of patterns seen in template-matching accuracy. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.1.1**. All plots have two dialect regions. (A-B) have vertical learning, local tutor choice, and high nest-site fidelity. These plots show the quilt pattern and have the highest template-matching accuracy. (C-D) have vertical learning, global tutor choice, and high nest-site fidelity. These plots do not show the quilt pattern, but have moderate template-matching accuracy. (E-F) do not have vertical learning, but do have local tutor choice, and high nest-site fidelity. These plots do not show the quilt pattern and have the lowest template-matching accuracy.

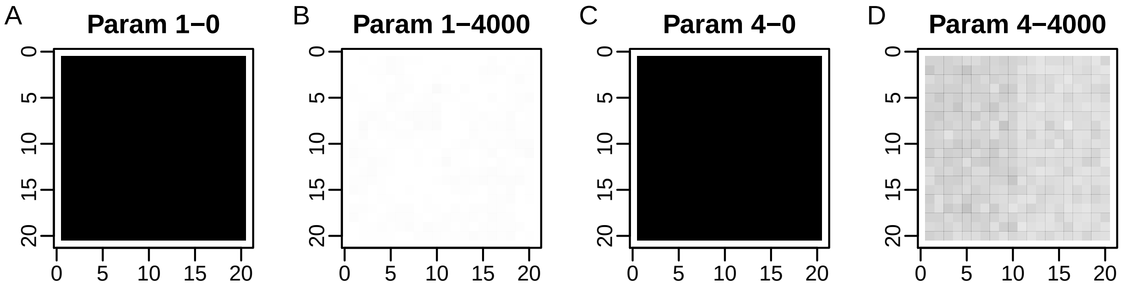


Figure 5.5: Example heat maps of regional dialect maintenance in the absence of sexual selection on song. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.1.2**. All plots have two dialect regions. (A-B) show the worst maintenance of regional dialects in this experiment. Vertical learning was enabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same). (C-D) show the best maintenance of regional dialects in this experiment. Vertical learning was disabled, tutor choice was global, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

One caveat to the above analysis was that when chicks learn from their fathers, they are learning from a male that was reproductively successful. In contrast, when males learned from oblique tutors, they could not distinguish between tutors that have been reproductively successful and those that have not. In real birds, it has been shown that fledglings attend to social interactions between males and other conspecific adults[32–34]. Thus, fledglings may be more likely to learn from socially successful males (those that attract females or win in male-male competitions). It could be that socially informed oblique tutor choice would allow the population to maintain dialects without vertical learning, and lead to higher accuracy template-matching with vertical learning. To test this possibility, I created parameter sets wherein that males that bred in the previous time step were far more likely to be chosen as oblique tutors than males that did not breed in the previous time step. This change allowed some parameter sets to evolve somewhat better template-matching accuracy while others evolved somewhat worse template-matching accuracy, but there was no consistent, overall pattern (**Table D.1.3**). Furthermore, socially informed oblique learning did not allow dialects to be maintained in the absence of vertical learning. An additional difference between oblique learning and vertical learning in these simulations was that chicks used the Add strategy when learning vertically, but they used the Conformity strategy when learning from oblique tutors. Therefore, I next tested whether using the Add strategy for oblique learning would allow dialects to be maintained in the absence of vertical learning. Dialects could be maintained if tutor choice was socially informed (**Figure 5.6** and **Table D.1.4**), though the template-matching accuracy was reduced compared to previous simulations (**Table D.1.1**). However, without social information, regional dialect 1 was lost in these simulations even when there was only one regional dialect (**Figure 5.6** and **Table D.1.4**). This was in stark contrast to previous simulations, wherein males learned only obliquely using the Conformity strategy with no social information but regional dialect 1 had been maintained (e.g. par 34 from **Table D.1.1**).

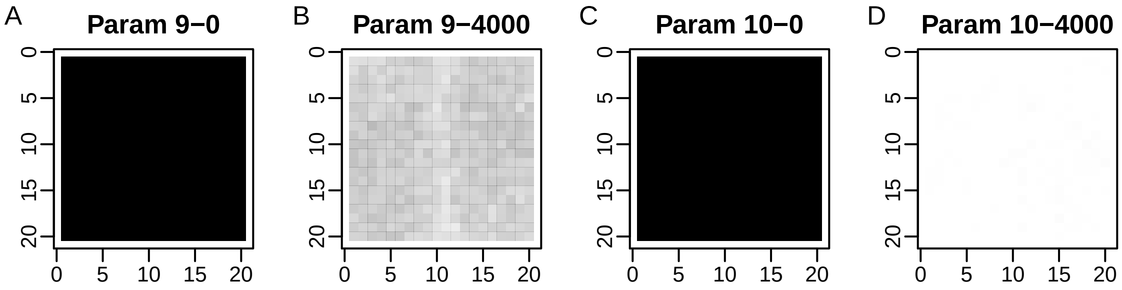


Figure 5.6: Example heat maps of regional dialect maintenance when learners used the Add strategy obliquely and picked tutors with or without social information. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.1.4**. All plots have two dialect regions, vertical learning disabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same). (A-B) show template-matching accuracy when tutor choice is socially informed, while (C-D) show template-matching accuracy when tutor choice is not socially informed.

In real birds, mixed dialect singers and birds that sing songs that are a hybrid between two dialects are rare[234–236]. In this model, learners evaluate oblique tutors on the level of syllables not songs; in other words, learners do not pick the most common song to learn, but rather the most common syllables, so mixed dialects songs could easily be formed. However, if a dialect is only made of one syllable, then there is no difference between discrimination of syllables and songs. I therefore tested whether the mechanics of dialect maintenance were altered when each dialect was composed of only one syllable. The most noticeable effects this had on our results was that it led to overall higher accuracy template matching (**Figure 5.7** and **Table D.1.5** versus **Figure 5.3C-F** and **Table D.1.1**), but there was also a small percentage of runs that did much worse than the majority (**Figure 5.8**). Otherwise, I found that dialect maintenance depended on the same factors: vertical learning and either high nest-site fidelity or local tutor choice (**Table D.1.5**).

It seemed possible that the shape of the matrix might affect dialect maintenance; if dialects had fewer territories next to one another, this should reduce contact and song learning between males that know different regional dialects. This could allow males in each region to match their region's dialect with higher accuracy. I compared parameter sets where the matrix dimensions were 20x20 (the default), 10x40, 5x80, and 4x100. Changing the matrix dimensions did not significantly affect my results (**Figure 5.9** and **Table D.1.6**). I also wondered if the reason that template-matching

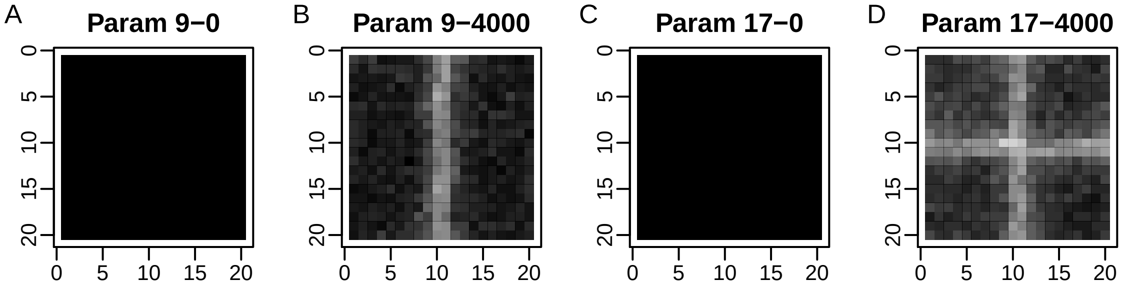


Figure 5.7: Example heat maps of dialect maintenance when dialects are one-syllable songs. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.1.5**. (A-B) have two dialect regions, while (C-D) have four dialect regions. Vertical learning was enabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

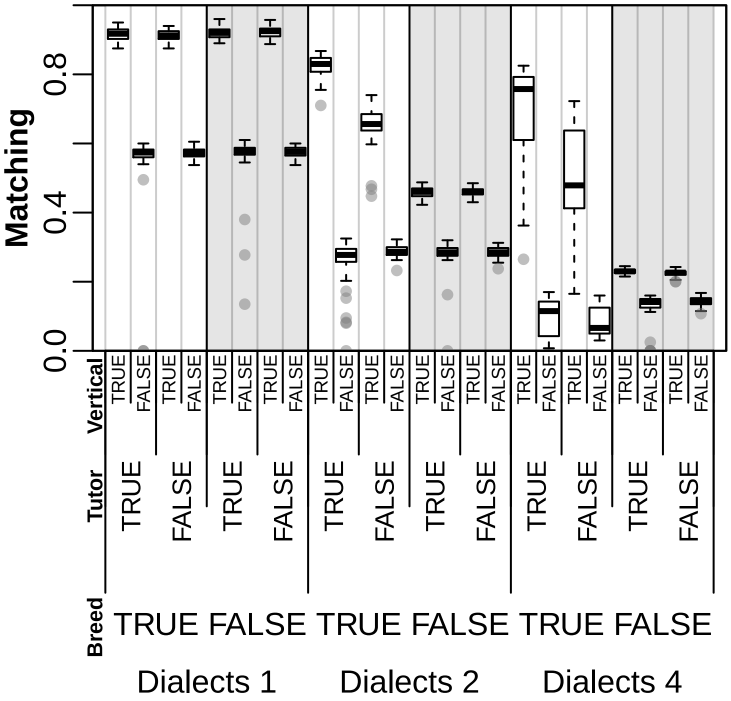


Figure 5.8: Boxplot of average template-matching accuracy for dialect maintenance across the entire matrix. Each box plot represents 50 individual iterations. Grey dots are outliers. Boxes are in the same order as parameters in **Table D.1.5**.

accuracy decreased as more regional dialects were added because there were more contact points between dialect regions, or whether it was simply because the population of birds that knew each dialect had decreased. To test the latter possibility, I ran two simulations where each regional dialect was known by 400 birds when there were 2 (20x40) and 4 (40x40) dialects. Creating larger dialect regions led to template-matching accuracy that was equivalent between the two larger matrices (**Figure 5.10** and **Table D.1.7**), and the values for the center square in these simulations were equivalent to those seen when there was only 1 dialect in a 20x20 matrix with otherwise equivalent parameters (**Table D.1.7**). Thus, the average template-matching accuracy only declined as the number of dialects in the matrix increased because there were fewer birds singing each dialect.

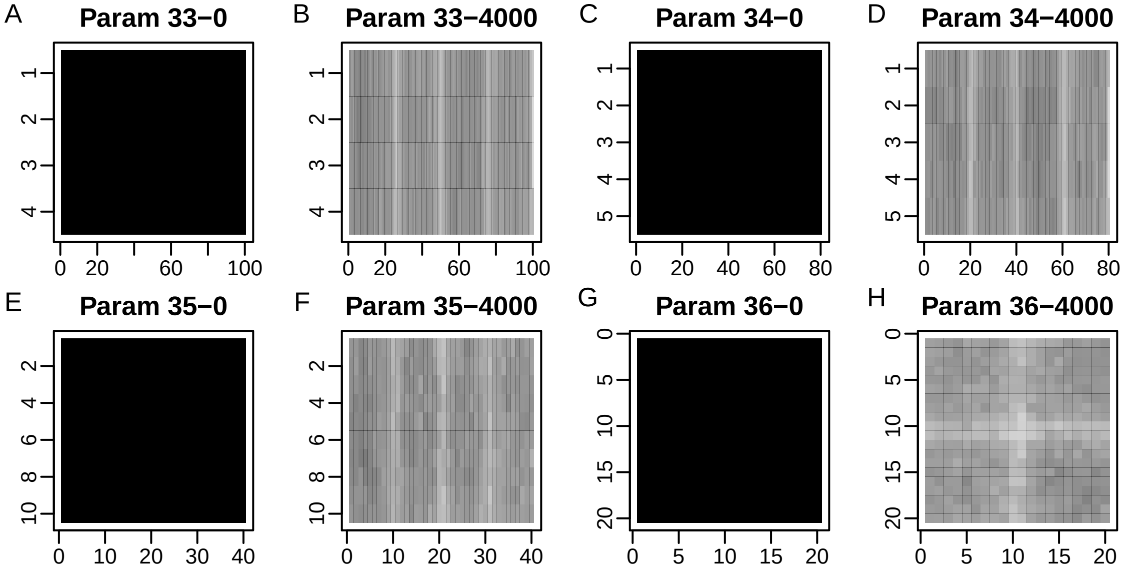


Figure 5.9: Example heat maps of regional dialect maintenance when the matrix dimensions are different. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.1.6** . All plots have four dialect regions. Vertical learning was enabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

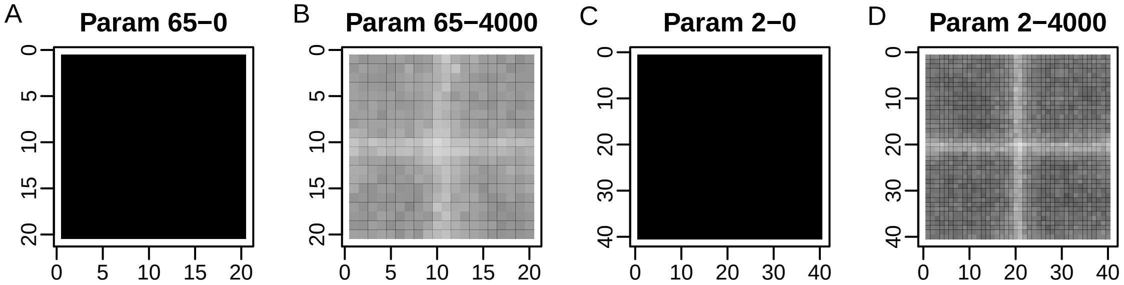


Figure 5.10: Example heat maps of regional dialect maintenance in the default matrix size compared to a larger matrix. Annotations are the same as in **Figure 5.3**. All plots have four dialect regions. (A-B) parameter set numbers match **Table D.1.1**, and (C-D) parameter set numbers match **Table D.1.7**. The larger matrix size allowed for higher accuracy template-matching. Vertical learning was enabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

5.3.2 Establishing new regional dialects

I was also interested in testing what parameters were necessary for new regional dialects to establish in males that did not previously have a dialect structure. To do this, I tested scenarios in which all males began the simulation knowing a variant of regional dialect 1 and females could know one of 2 or more regional dialects. I found that both vertical learning and high nest-site fidelity were required for establishing new regional dialects in males (**Figure 5.11** and **Table D.2.1**). Local tutor choice led to higher accuracy song template matching in these cases, but it was neither required nor sufficient alone or in combination with vertical learning to allow new male regional dialects to establish (**Table D.2.1**). This was unexpected, because either high nest-site fidelity or local tutor choice had been sufficient to maintain dialects (**Table D.1.1**). Similar to what I saw in dialect maintenance simulations, when females had a uniform song template instead of noisy templates, males were able to achieve somewhat higher template-matching accuracy, but this did not have any other significant effects on regional dialect establishment (**Table D.2.1**). When dialects did establish, territories on the boundaries between dialect regions had the worst matches, creating a quilt-like pattern between dialect regions previously seen in some maintenance parameter sets (**Figure 5.11B and F**).

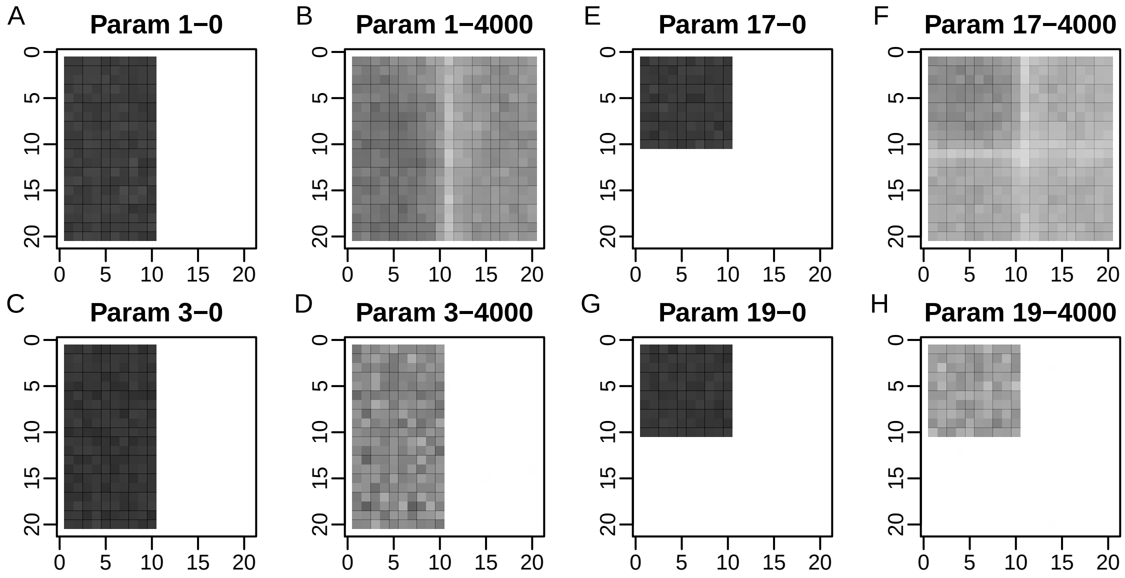


Figure 5.11: Example heat maps of template-matching accuracy in simulations that did and did not establish new regional dialects. Annotations are the same as in **Figure 5.3**. Plots (A-D) have 2 dialect regions, while plots (E-H) have four. The black area in (A, C, E, and G) is the region for dialect 1. Parameter set numbers match **Table D.2.1**. The parameter sets for top row of plots allowed vertical learning, while the bottom row did not. Tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

Enabling socially informed tutor choice did not allow new dialects to establish in the absence of vertical learning (**Table D.2.2**). However, I found that regional dialects could establish without vertical learning if learners used the Add strategy and oblique tutor choice was socially informed (**Figure 5.6** and **Table D.2.3**), though the matches to the female templates were much worse than in prior parameter sets (**Figure 5.11** and **Table D.2.1**). Furthermore, socially informed tutor choice with the Add strategy allowed for the establishment of dialects when there was no nest-site fidelity **Table D.2.1**. However, as seen in maintenance experiments, regional dialect 1 was lost in these simulations when tutor choice was not socially informed (**Figure 5.6** and **Table D.2.3**). Using one-syllable songs led to similar evolutionary dynamics (**Figure 5.13** and **Table D.2.4**). While it did not improve the template-matching accuracy for new dialects on average, the original dialect was maintained with higher template-matching accuracy (**Figure 5.13**), suggesting that choosing to learn the most common song instead of the most common syllables has a greater effect on maintenance than establishment. However, I also saw that for some parameter sets, the end values were split between two groups of higher and lower template-matching accuracy values (e.g. 20% of iterations for parameter sets 9 and 11 had >65% template-matching accuracy while the rest hovered around 45%, **Figure 5.14**). Together, these results suggest that sometimes picking the most common dialect leads to better establishment, though this alone is not a reliable way to establish a new dialect with high fidelity template-matching accuracy. Finally, changing the dimensions of the matrix did not seem to have a large effect on dialect establishment, as no consistent effect was seen with any specific set of dimensions (**Figure 5.15** and **Table D.2.5**).

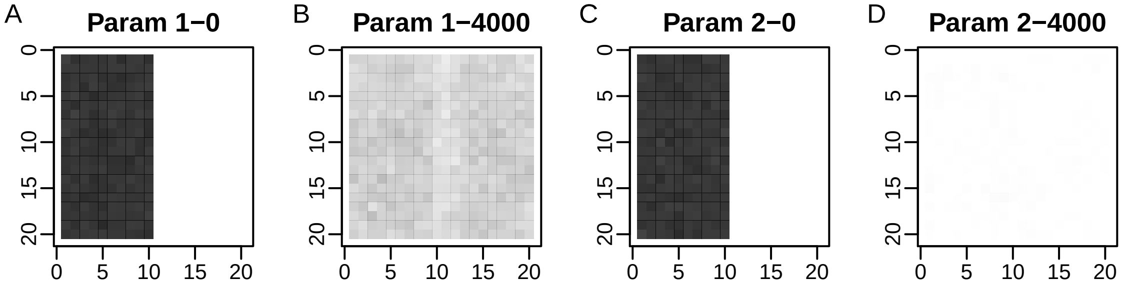


Figure 5.12: Example heat maps of regional dialect establishment or loss when learners used the Add strategy obliquely and picked tutors with or without social information respectively. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.2.3**. All plots have two dialect regions. Tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

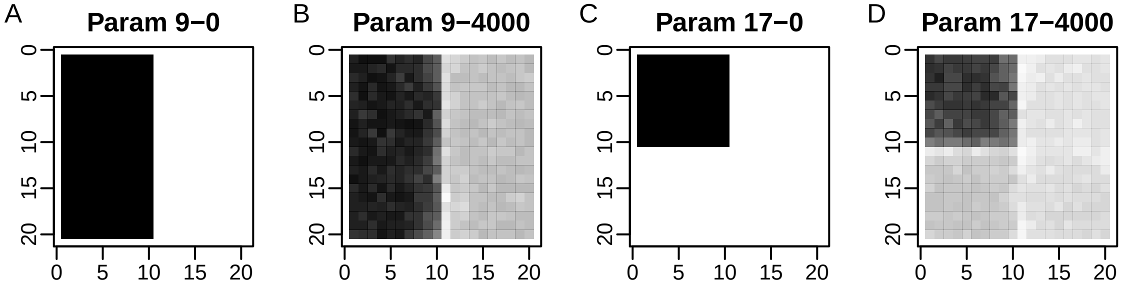


Figure 5.13: Example heat maps of dialect establishment when dialects are one-syllable songs. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.2.4**. (A-B) have two dialect regions, while (C-D) have four dialect regions. Vertical learning was enabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

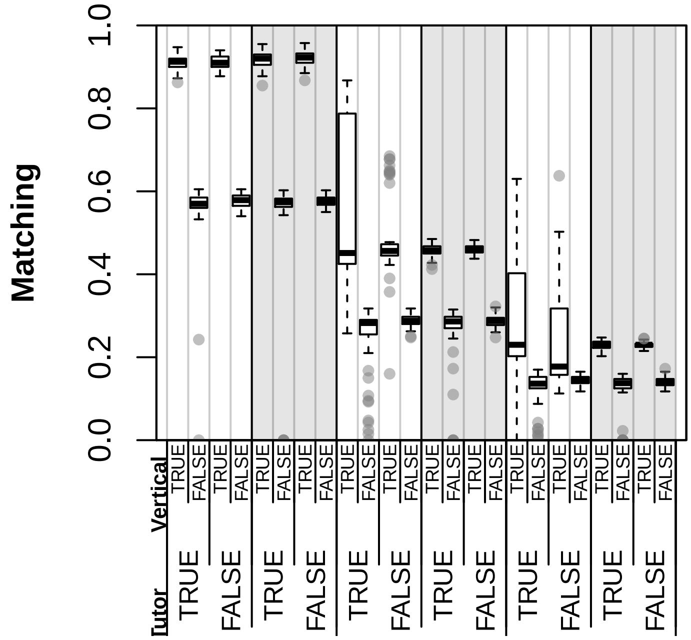


Figure 5.14: Boxplots of average template-matching accuracy for dialect establishment across the entire matrix. Annotations the same as in **Figure 5.7**. Boxes are in the same order as parameters in **Table D.2.4**.

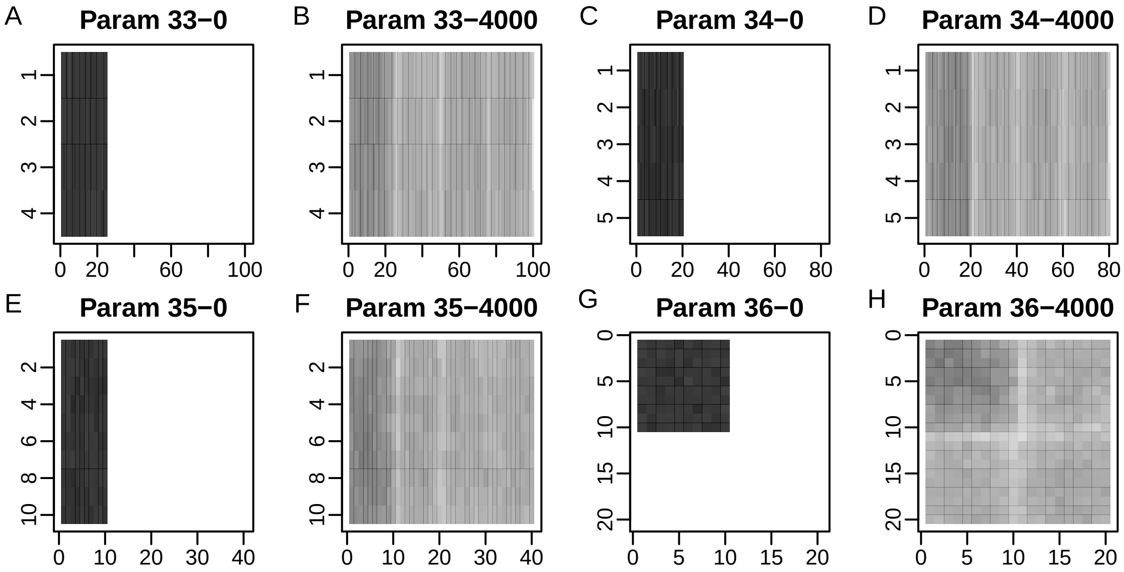


Figure 5.15: Example heat maps of regional dialect establishment when the matrix dimensions are different. Annotations are the same as in **Figure 5.3**. Parameter set numbers match **Table D.2.5**. All plots have four dialect regions. Vertical learning was enabled, tutor choice was local, there was high nest-site fidelity, females had uniform dialects, and male songs initially matched female song templates (Same).

5.4 Discussion

Previously, it was not known which behaviors would allow regional dialects to establish and be maintained. My model makes multiple predictions about what behaviors are critical for this process. First, it predicts that dialect establishment requires that learners preferentially pick reproductively successful tutors. If males first learned song from their father, high nest-site fidelity was required to establish dialects. In contrast, if males learned their song first from reproducetively successful oblique tutors, dialects were best established when males learned from neighboring tutors rather than any male in the population. Thus, dialect establishment required socially informed learning in combination with either spatially restricted dispersal behavior or tutor choice, depending on the whether males learned primarily vertically or obliquely. I could not test whether sexual selection on song was required to establish dialects, because female preference for specific song templates is the only way to intentionally establish dialects in this model; however, cessation of sexual selection on dialects post-establishment led to the loss of dialects. Second, once dialects had been established, socially informed learning was required to maintain them. Spatial interactions were less important for maintaining dialects, but the highest template-matching accuracy and the associated quilt-like pattern was only seen when vertical learning was paired with high nest-site fidelity and socially informed oblique learning was paired with local tutor choice.

Although this is the first research to my knowledge to predict that nest-site fidelity is critical to dialect formation and maintenance in populations where males rely on vertical learning, research in the yellow-naped amazon (*Amazona auropalliata*), a parrot, also suggested that limited dispersal (higher nest-site fidelity) was a key factor in maintaining dialects in contact calls[251]. One of my more interesting findings was that were two behavioral methods that allow dialects to form and be maintained: in one, males learned the entirety of a song vertically and then either spend their lives near their natal grounds or learn from local tutors post dispersal. In the other, males learned the entirety of a song from reproductively successful oblique tutors post-dispersal. Theoretically, the vertical method of dialect establishment and maintenance would allow genetically related males to share the same dialect and would be in line with the idea that dialects may act as a marker of kinship between subpopulations. This may be the case in species such as white crowned sparrows[252,253], zebra finches (*Taeniopygia guttata*)[254], and Belganese finches (*Lonchura domestica*)[254]. However, whether this this marker of male relatedness would be used by females to ensure mating within a subpopulation or to avoid mating with closely related individuals would have to be studied on a species-by-species case. The latter method of dialect establishment would better fit species in which it is known that males crystallize their final song after reaching their final breeding ground, such as corn buntings[242] and indigo buntings[255]. Furthermore, in rufous-collared sparrows, it has been shown that females have a strong preference to their natal dialect, while males respond equivalently to all dialects[256]. In this case, it may be more advantageous for males to use the oblique learning method and adopt the most common dialect around his breeding territory, because there may be more females that prefer the most common song, so males that sing it would have the best chance at finding a mate. If dialects are used in male-male interactions, then the either the oblique or vertical strategy for dialect establishment could be used in species with high nest-site fidelity, but the oblique strategy for establishment would be most valuable in species that disperse far from their natal territory. Finally, in northern cardinals (*Cardinalis cardinalis*), most males have low nest-site dispersal, while others disperse long distances form their natal territory[233]. It could be that both mechanisms of dialect maintenance are at play in this species.

While my model predicts that ongoing sexual selection on song is necessary for dialect establishment and maintenance, I was not able to disentangle which forms of sexual selection promote dialects. For simplicity, I frame selection pressure in this model in terms of female choice on song, but an equally valid interpretation is that males that better match their dialect are more likely to fend off intruding males and thus are most likely to breed successfully. It may be that the best way to disentangle these two possibilities is for field researchers to both assess female preferences and establish what role song competitions, such as song-matching, play in species with dialects to determine which of the possible selection pressures (or both) exists in those species. Female preferences have already been established in many cases as noted above, however less research as been done on male-male interactions in these species.

Overall, my results propose that there is not a single mechanism for dialect formation, and dialects need not coincide with genetic differences between individuals that sing different dialects, though they certainly could. While my results confirmed the reasonable hypothesis that sexual selection on song is necessary for dialect formation, they also highlight the importance of using social information to choose high quality tutors in dialect establishment, a facet that has not be considered in dialect research. In general, more field work will need to be done in a broader array of species to better understand which species use which strategies to form dialects and the role they play in female choice and male-male interactions.

Chapter 6

A higher resolution exploration of interactions between females preferences and male song-learning behavior

6.1 Introduction

In **Chapter 4**, I examined some of the possible interactions between female preferences and male song-learning strategies, with a focus on how these traits affect the evolution of the song learning window. However, there is a much wider array of proposed male learning strategies and female preferences which could affect song evolution than were examined there. For example, it has been suggested that song features could be propagated via cultural hitchhiking if females prefer to mate with the same or similar males that other females mated with[257]. Alternatively, larger repertoires might be the product of a female preference for rare syllables[258], while species with limited repertoires may be a result of a preference for common syllables[248,249]. Finally, it has been proposed that females prefer “sexy syllables” in canaries (*Serinus canaria*)[231,232]. For learning strategies, it has been proposed that males in some species may learn too many syllables as fledglings and then prune back their repertoires as adults[221,228]. Furthermore, while I previously explored how learning by conformity bias affects song, I did not examine how learning from multiple tutors based on either 1) simple chance or 2) stringently learning only the syllables that all other males sing affects song evolution. In this chapter, I begin to explore which combinations of male learning strategies and female preference lead to large or small syllable repertoires and which ones allow certain syllables to be propagated to future generations with high fidelity. Specifically, I want to test two possibilities. First, can high accuracy template-matching be maintained or can large syllable repertoires be formed in the absence of direct selection on song? Second, do different female preferences always lead to different evolutionary dynamics between syllable repertoire and the length of learning, or do some distinct females preferences have similar effects to one another on the evolution of song an song-learning?

6.2 Methods

6.2.1 Data and code availability

C# library, programs, and walk-through:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/Csharp-Library-and-Programs

R package, code manual, and walk-though:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/R-Package

Parameter generators and data visualizing code:

https://github.com/CreanzaLab/SongEvolutionModel/tree/master/PreferencesPaper

6.2.2 Model features

I used the **C#** version of the model for all experiments, though R was used to generate the parameter sets and plot the data. The model operates as described in **Chapters 4** and **5**, though I implemented three new learning strategies and five new female preferences.

The new learning strategies are:

1. **AllNone:** males listen to multiple tutors and will only learn syllables that all tutors sing.
2. **Percentage:** males listen to multiple tutors and are more likely to learn common syllables. Unlike Conformity, which uses conformity bias, this strategy uses a linear scale. So if 7 out of 8 tutors (87% of tutors) sing a syllable, a learner has a 87% chance to attempt to learn it.
3. **Prune:** chicks learn from their father and three oblique tutors in the first year of their life. In subsequent years, males listen to individual tutors and try to forget syllables that the tutor did not sing.

The new female preferences are:

1. **Common:** Females prefer syllables that are more common among the current population of singing males.
2. **Rare:** Females prefer syllables that are rarer among the current population of singing males.

3. **SexySylls:** Males that know specific syllables are more attractive, but they are not penalized for having extra syllables (unlike in the template-matching preference).

4. **Social:** Females prefer males that bred previously.

5. **Noise:** Females have no preference. All living males that sing at least one syllable are equally likely to breed.

Furthermore, I allowed female templates to be generated even when females do not explicitly prefer specific syllables. Thus, the template-matching box plots generated when females prefer larger syllable repertoires (Rep), Common or Rare syllables, males that had bred previously (Social) or nothing (Noise) reflect whether the song structure that was in the population at the beginning of the simulation was maintained. Finally, because SexySylls uses a different algorithm to calculate the match between male songs and female song templates, the template-matching box plots in this preference cannot be directly compared to those from other preferences. I only changed parameters relating to female preferences, male learning strategies, vertical learning and socially informed tutor choice. All other parameters were left as the model defaults (see the **C#** Walk-Through).

6.2.3 Analysis

All parameter sets were repeated 50 times and run for 2000 time steps. Data were collected every 20 time steps. For the AllNone strategy, songs were entirely lost 1-5% of the time. This ended the simulation prematurely, so for parameter sets where the simulations did not always finish, I coded a new program that would repeat the simulations until they reached completion 50 times. If an iteration of a parameter set reached a point where there were not enough males that still knew at least one syllable for breeding or tutor choice, that iteration was stopped and discarded.

6.3 Results

I was interested in examining how different female preferences interacted with different male song-learning strategies to affect the evolution of song and song-learning. Additionally, it seemed likely that preventing vertical learning or allowing for socially-informed tutor choice could affect interactions between female preferences and male song-learning strategy, so I tested how changing these parameters affected song and song-learning evolution as well. Because some female preferences led to similar results, their results were covered in the same section below.

6.3.1 Selection for non-song features

There were two preferences that led to a lack of direct selection on song. A 100% preference for Noise meant that all males were equally likely to be chosen (in other words, a lack of sexual selection), and a 100% preference for Social meant that males that had bred previously were more likely to breed again. Both of these selection forces led to similar outcomes. Syllable repertoires did not exceed 25 syllables for any male learning strategy (**Figure 6.1A** and **Figure 6.2A**). Vertical learning led to larger syllable repertoires, but there were minimal differences between simulations based on whether oblique tutor choice was socially informed or not. Males learned obliquely for one or two years depending on the learning strategy (**Figure 6.1B** and **Figure 6.2B**). Learning accuracy tended to be between 50% and 80%, which was lower than what I had seen previously when song was under direct sexual selection (**Figure 6.1C** and **Figure 6.2C**), compare to **Panel C** of all box plots below and in **Chapter 4**). Usually, the song structure that the simulation began with was lost; however, when the learning strategy was AllNone or Conformity and vertical learning was enabled, the song structure was maintained to a limited degree (>0%, **Figure 6.1D** and **Figure 6.2D**). Disabling vertical learning with the same learning strategies allowed for a modest amount of song structure to be maintained in the absence of vertical learning (10-20%, **Figure 6.1D** and **Figure 6.2D**).

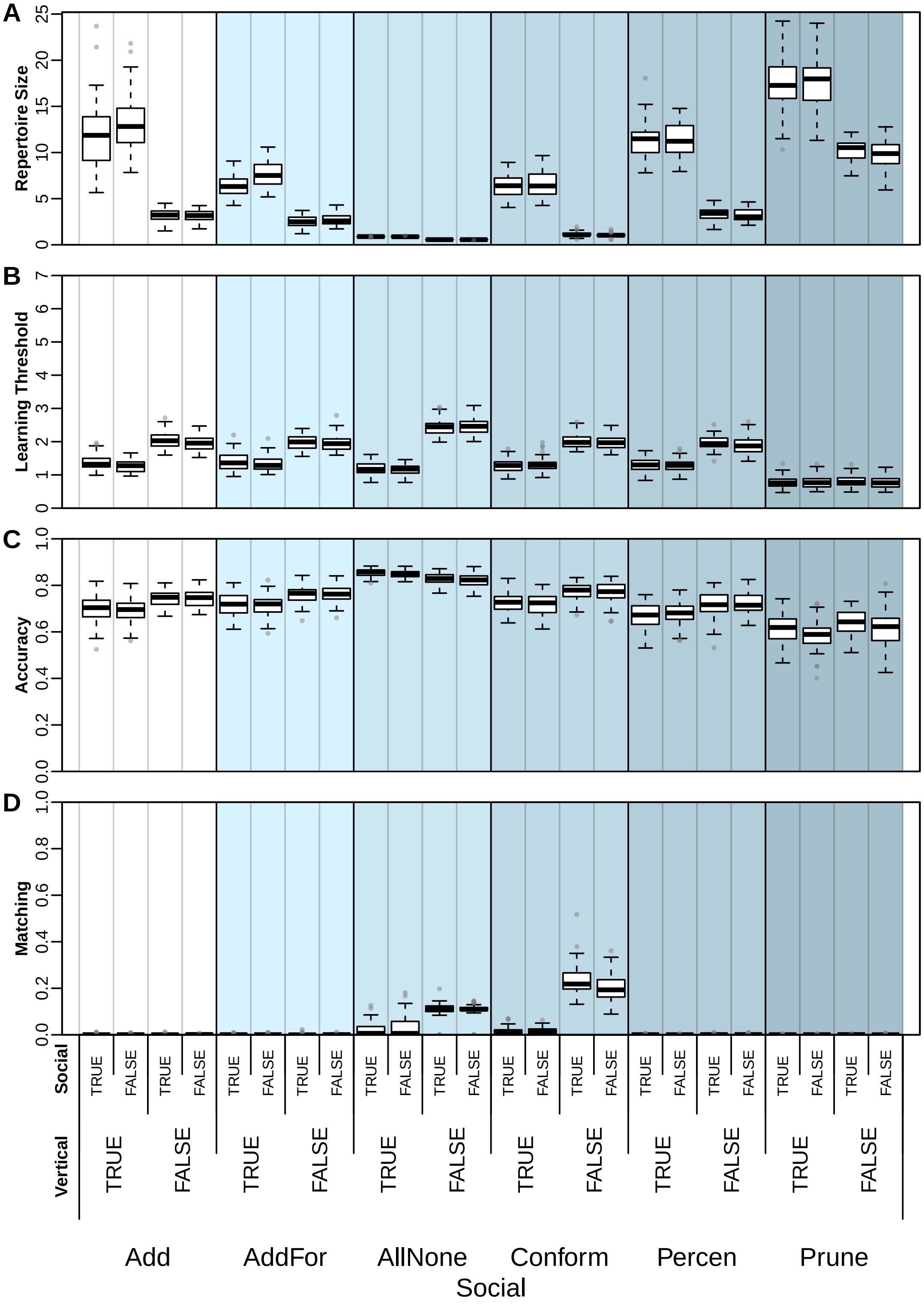


Figure 6.1: Song traits when females prefer males that bred previously. Each box plot represents 50 individual iterations. The same iterations were used in all four plots. Vertical is whether vertical learning was enabled. Social is whether social information was used to choose oblique tutors.

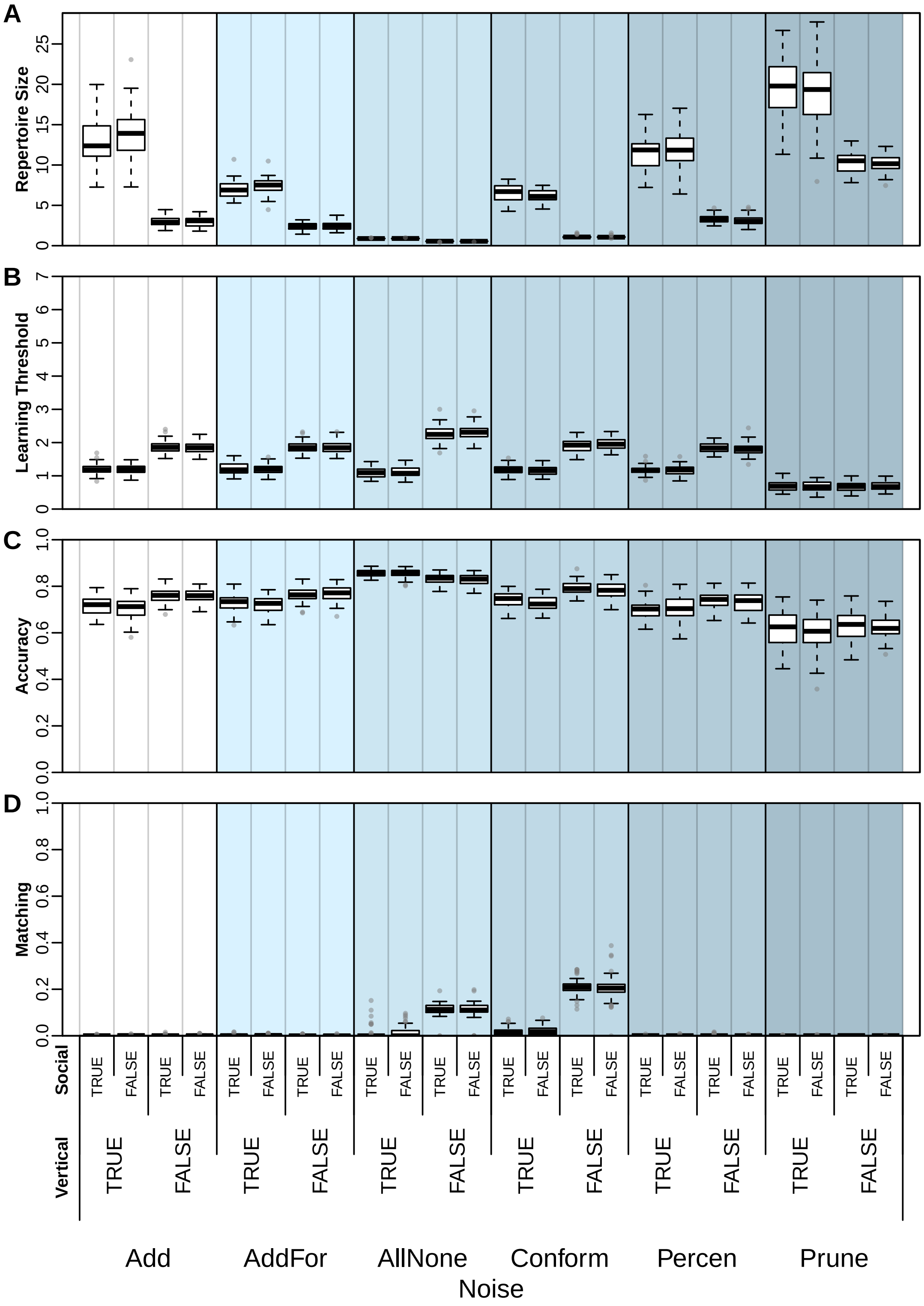


Figure 6.2: Song traits when females have no song preferences. Annotations the same as in **Figure 6.1**.

6.3.2 Selection for syllable frequency and number

I had expected that preference for Rare syllables and preferences for larger syllable repertoires might lead to similar evolutionary dynamics, as males in both contexts would benefit from having larger repertoires. This expectation was met (**Figure 6.3** and **Figure 6.5**); however, I had not expected that preferences for common syllables would also lead to similar evolutionary dynamics, because there was no reason to presume that larger repertoires would be more beneficial in this case (**Figure 6.4**). The Add learning strategy led to the largest syllable repertoires (>100 syllables) and Prune led to the second largest repertoires (~50 syllables) (**Figure 6.3A**, **Figure 6.5A**, and **Figure 6.4A**). Males learned for longer in these selection contexts than they had when there was no direct selection on song for all learning strategies except Prune (**Figure 6.3B**, **Figure 6.5B**, and **Figure 6.4B**). The only difference I noticed between these three selection contexts was that there was better maintenance of the initial song structure when males used the AllNone strategy in combination with vertical learning when females preferred common syllables instead of rare syllables or larger syllable repertoires (**Figure 6.3D**, **Figure 6.5D**, and **Figure 6.4D**).

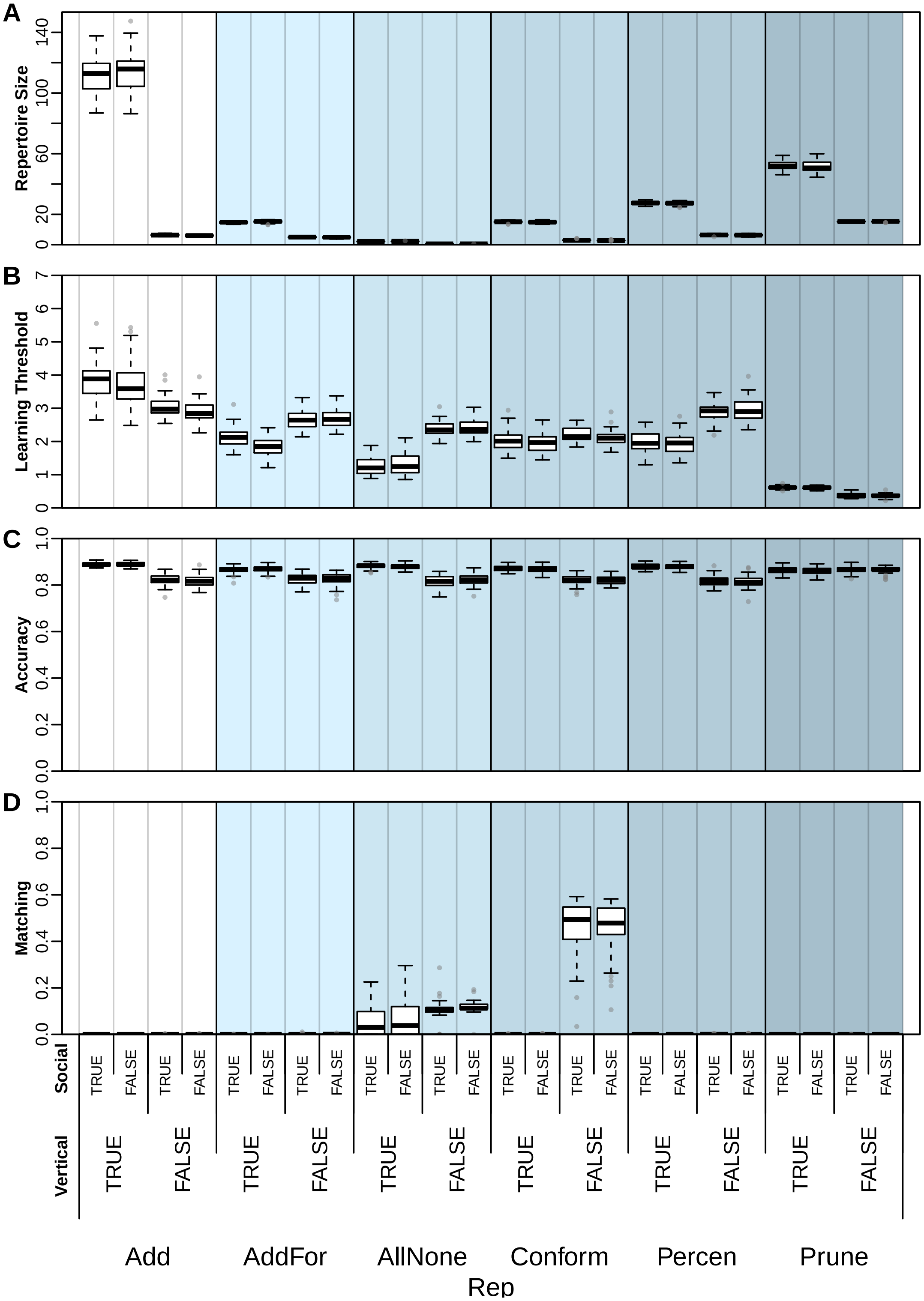


Figure 6.3: Song traits when females prefer larger syllable repertoires. Annotations the same as in **Figure 6.1**.

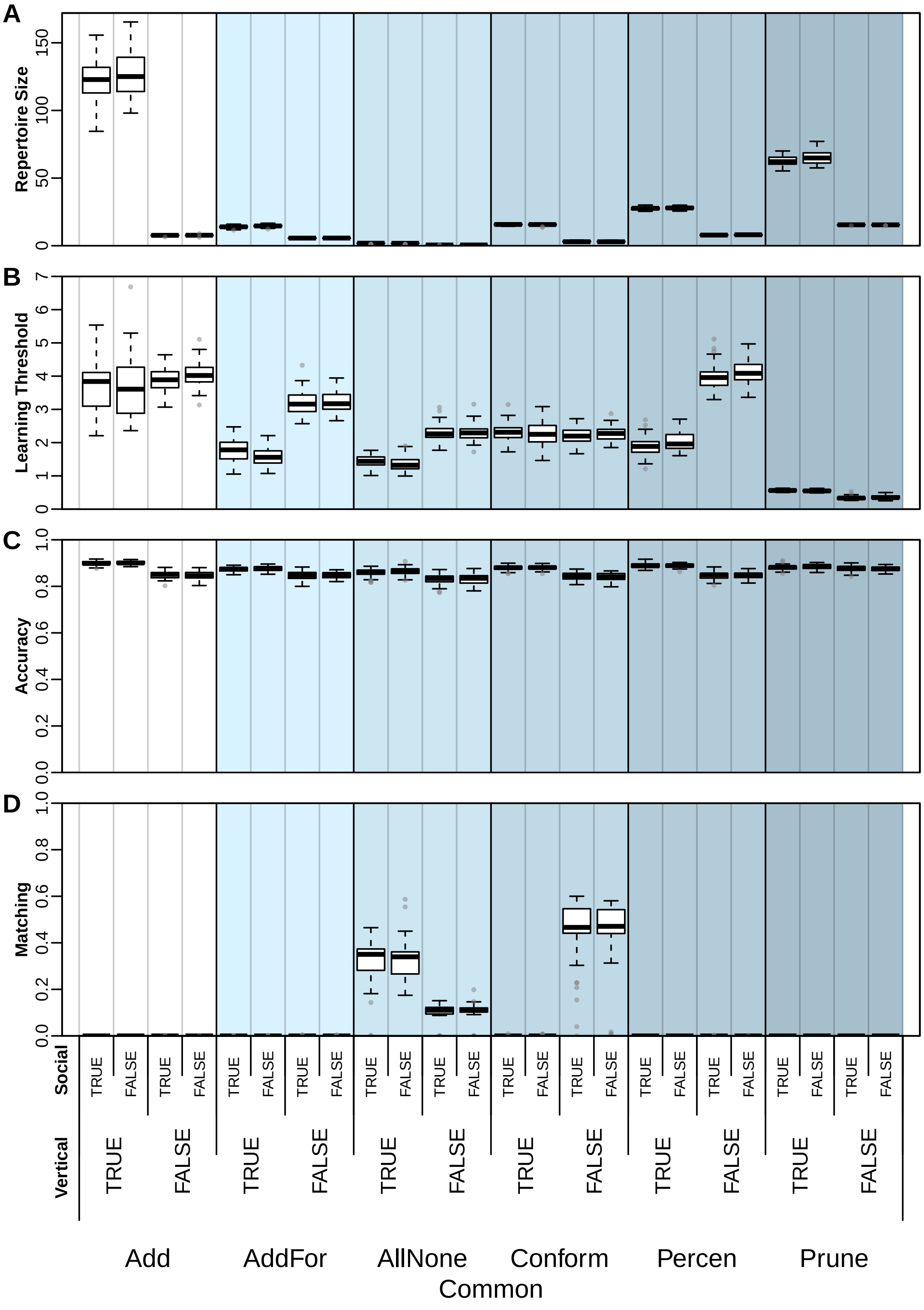


Figure 6.4: Song traits when females prefer common syllables. Annotations the same as in **Figure 6.1**.

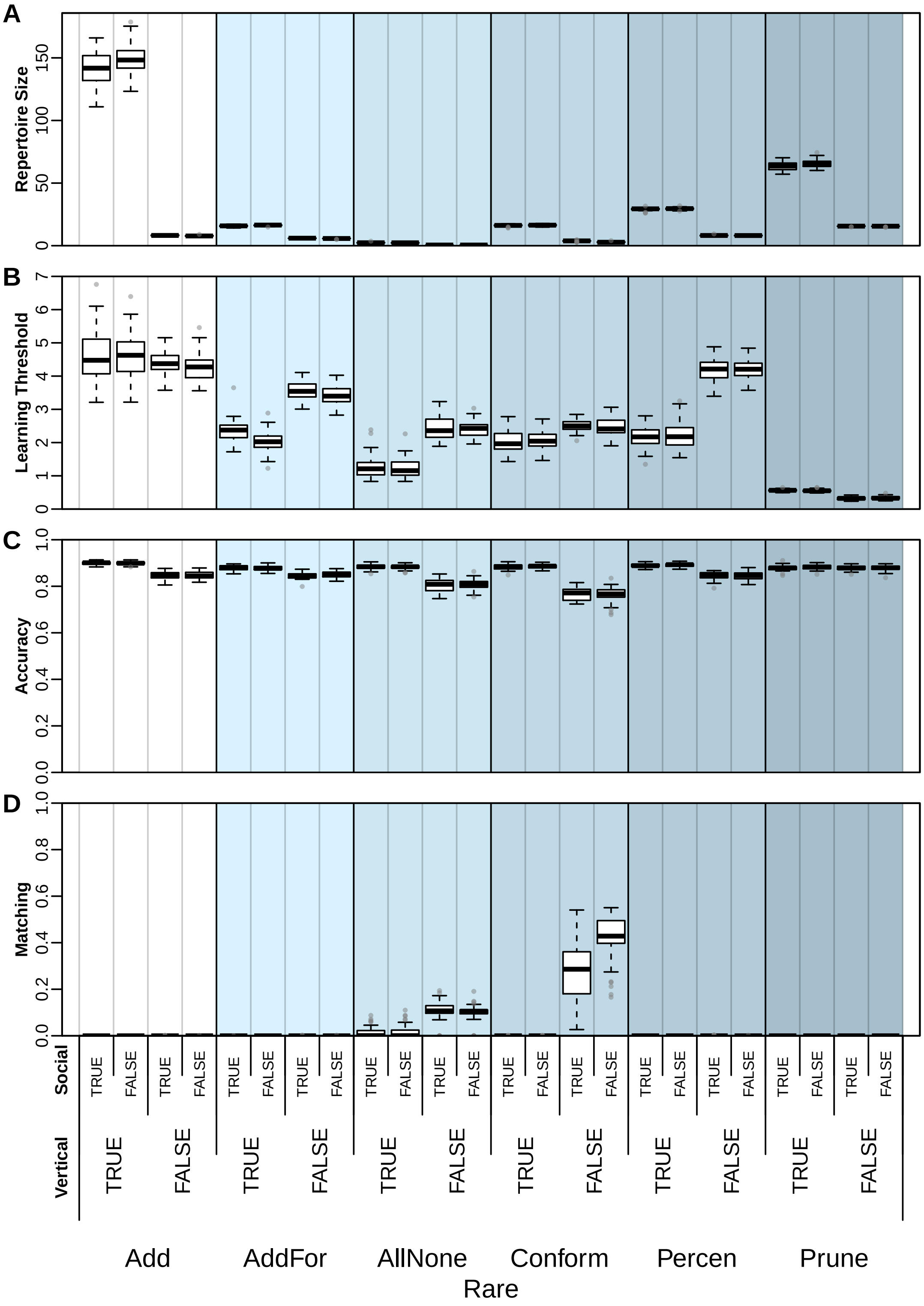


Figure 6.5: Song traits when females prefer rare syllables. Annotations the same as in **Figure 6.1**.

6.3.3 Selection for template-matching

When males needed to maintain the correct overall song structure (i.e. females prefer males that have all of the correct syllables and no extras), repertoire size evolved to 5 or fewer syllables for all learning strategies except Prune, where the syllable repertoire could grow as large as 20 syllables (**Figure 6.6A**). When vertical learning was enabled, males generally learned obliquely for a shorter period of time. There was an exception to this, where males learned for longer when the learning strategy was Prune and tutor choice was socially informed. (**Figure 6.6B**). Again, learning accuracy was within the range of values I had seen previously for all learning strategies except Prune, which led to lower learning accuracy (~50%, **Figure 6.6C**). The highest accuracy template matches were obtained when vertical learning was enabled and learning was socially informed for all learning strategies (**Figure 6.6D**). However, Conformity was only marginally affected by disabling vertical learning and uninformed oblique tutor choice, and the Prune strategy led to low accuracy template matching even in the best circumstances.

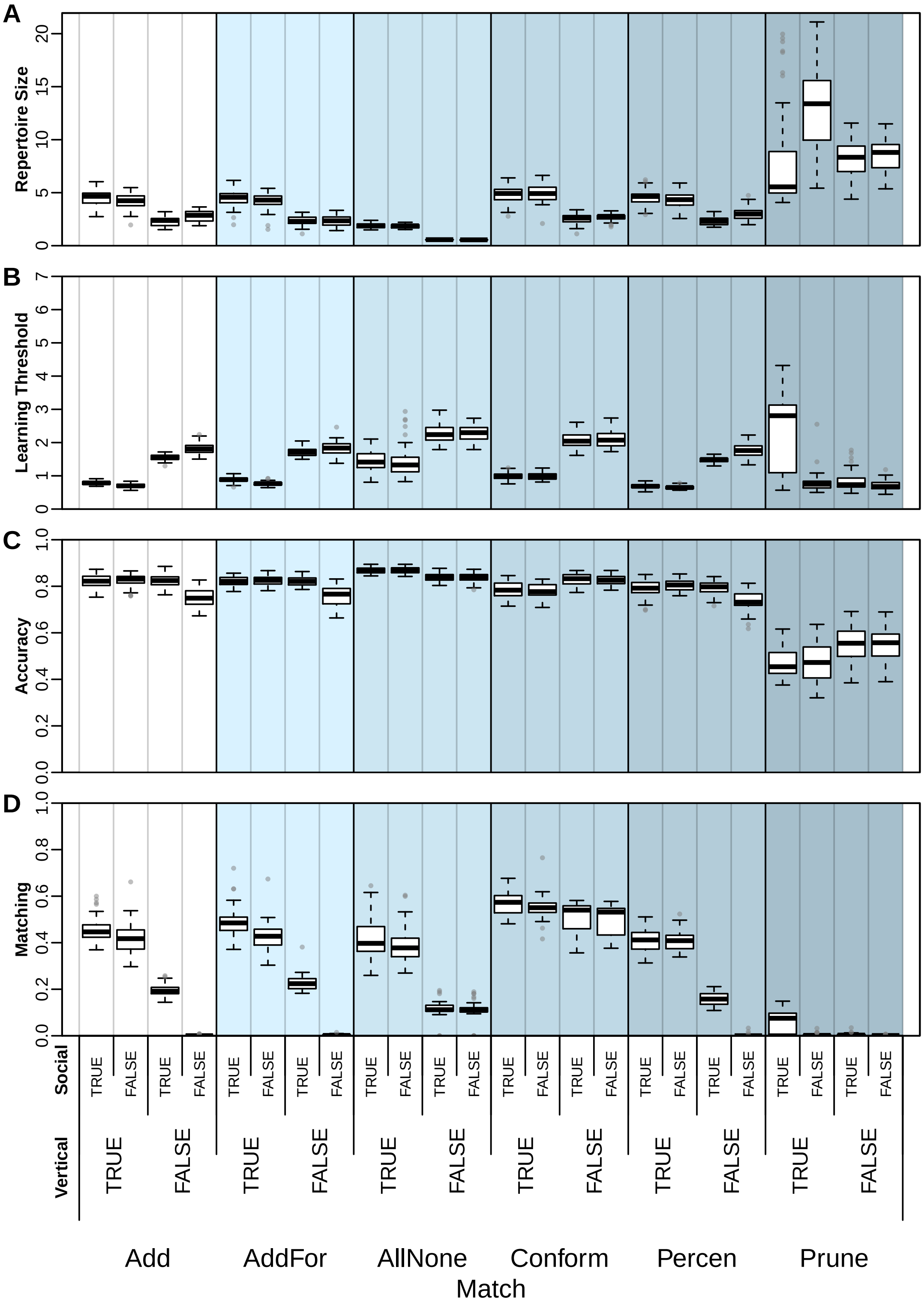


Figure 6.6: Song traits when females prefer songs that match their song templates. Annotations the same as in **Figure 6.1**.

6.3.4 Selection for specific syllables

When males that produced specific syllables were preferred, but they were not penalized for having extra syllables, syllable repertoires generally stayed small (<5 syllables); however, repertoire size reached an average of 50 syllables when the Add or Prune strategies were used in combination with vertical learning (**Figure 6.7A**). Learning thresholds were longer when vertical learning was disabled, as I had seen when females preferred template-matching (**Figure 6.7B**), and learning accuracy values were similar to those seen with other preferences (**Figure 6.7C**). When template-matching accuracy did not include a penalty for producing extra syllables, template-matching accuracy was higher overall, but showed a very similar pattern to the selection context where females preferred template-matching accuracy (**Figure 6.7D**). The only notable exception was that the Prune strategy led to higher accuracy template-matching in this context.

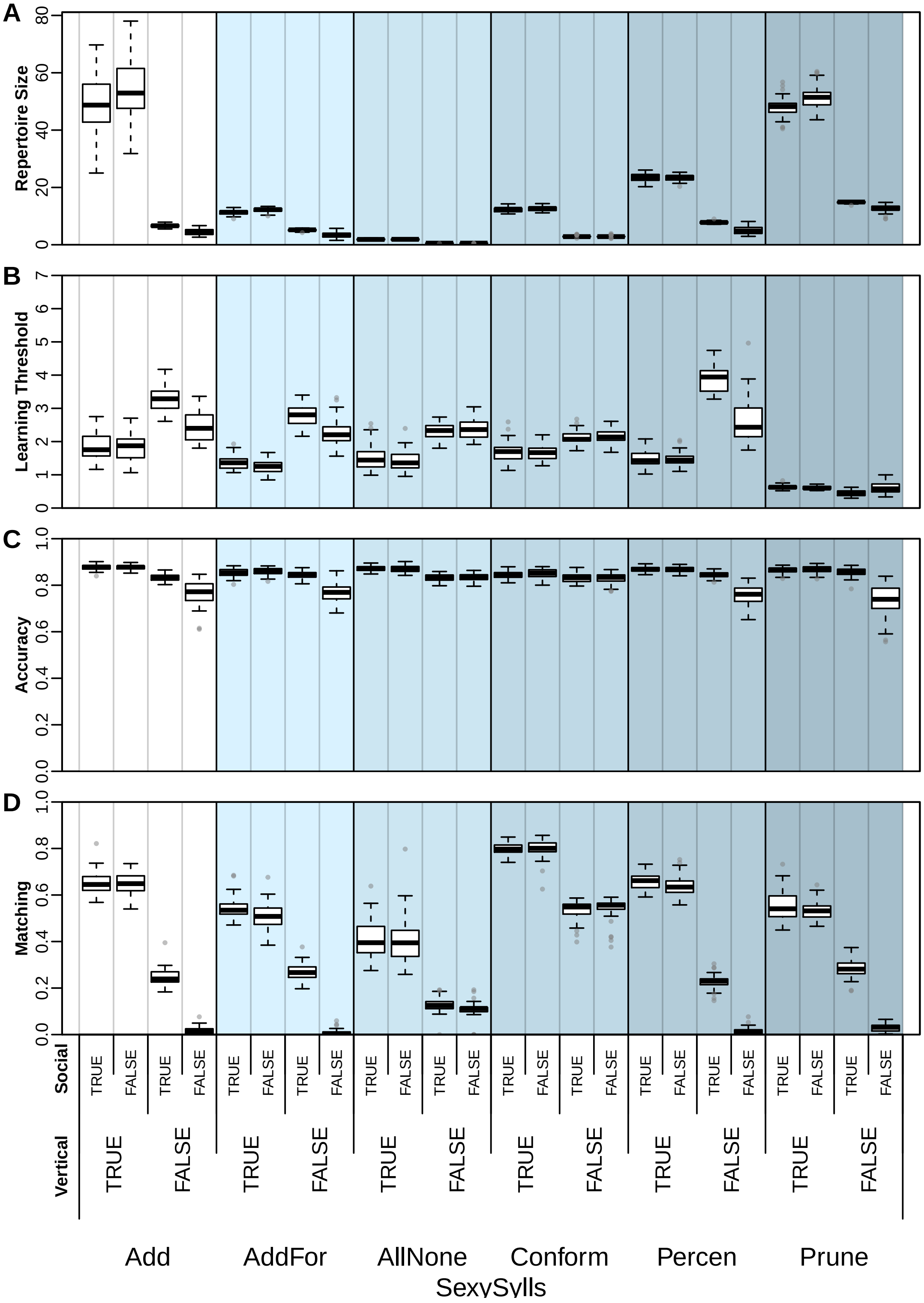


Figure 6.7: Song traits when females prefer songs that contain specific syllables. Annotations the same as in **Figure 6.1**.

6.3.5 Overall trends

There were several notable trends that occurred across the simulations regardless of the selection pressure acting in the simulation. The Prune learning strategy was supposed to allow birds to learn too many syllables early in life, so they could then reduce their repertoire to a subset of these syllables based on what other males were singing. However, birds with this learning strategy evolved learning windows shorter than one year, so they could avoid forgetting these extra syllables. Thus, the Prune strategy actually morphed into a modified Add strategy, where birds were guaranteed to learn from at least three males in their first year. For all other learning strategies, males typically engaged in oblique learning for one to four years, depending on the selection pressure and learning strategy. The AllNone and Conformity learning strategies allowed template-matching to be maintained to some degree in all selection contexts as long as vertical learning was disabled. However, in some selection contexts, vertical learning led to higher template-matching accuracy, while in others it led to a complete loss in template-matching. In general, socially informed oblique learning was not better than socially uninformed oblique learning, but there are several cases with matching strategies where this was important, as highlighted above.

6.4 Discussion

So far, these results have shown that several proposed female preferences can lead to remarkably similar outcomes on song evolution. Preference for larger repertoires, common syllables, or rare syllables led to syllable repertoires of similar sizes and limited maintenance of specific syllables in the population. This result was unexpected, and further investigation will be required to understand how these very different preferences can have such similar effects on song evolution. Furthermore, this result suggests that it will be difficult to tell which of these selective forces are responsible for the formation of large repertoires in real species without explicitly examining what preferences females in each species display and the way males respond to songs with varying levels of elaboration and common or rare syllables. Although it has been proposed that song features could be accidentally propagated through cultural hitchhiking if females copy other female's mate choices (Social), I found that this did not produce affects on song that were different from when females lacked any preferences (Noise). This implies that mate-choice copying does not play a large role in the evolution of either large syllable repertoires or highly stereotyped songs. In contrast, some seemingly similar selective forces led to different results. While female preferences for template-matching generally constrained the evolution of syllable repertoire size to the length of the female template or shorter, if males were expected to produce certain syllables but were not penalized for having extra syllables, then it was possible for males to evolve repertoires of 20 syllables or greater.

Male learning strategies also had pervasive effects across most if not all female preferences. The Add strategy led to the largest syllable repertoires in all cases except when females preferred template-matching. The Conformity strategy always led to the highest accuracy template-matching, even when there was no penalty for knowing extra syllables. The Add/Forget and Percentage learning strategies consistently led to repertoire sizes between 5 and 20 syllables, while the AllNone strategy consistently led to repertoires of less than 5 syllables. Thus, the interaction between male song learning strategy and the selective pressures on song are critical to shaping the overall syllable repertoire and the maintenance of specific song features across generations.

There are several key components that would need to be added to the model and several additional experiments to run to fully examine these interactions. Currently, female preferences remain static. Thus, it not clear what mix of female preferences best match each song learning strategy, nor is it clear how song would evolve in the face of mixed selection pressures. While this could theoretically be addressed with static preferences by running different simulations with mixes of the seven preferences in combination with the seven learning strategies, this could lead to the production of an overwhelming amount of data that is difficult to parse into meaningful conclusions. Instead, I propose that female preferences be refactored to act more like male song-learning traits. In other words, each individual female would have her own values for each preference that could potentially evolve by having females turn over as males do. In the model, there already exists a mechanism that picks fathers for new females when males die to allow for templates to evolve.

Female chicks could also inherit their song preference for their mother at the same time. Because the model is already designed to allow male traits to evolve, it would not be difficult to extend this flexibility to females.

Additionally, in the current model, song itself is not tied to any actual fitness trait. This means that the model ignores one of the assumptions of the birdsong field: that song acts as an honest signal of male fitness, and this is why females might prefer specific song features. The model is already designed to incorporate a survival punishment based on the length of the song-learning window, such that males with longer learning windows are more likely to die sooner. It would be trivial to add a survival reward based on traits like larger repertoire size or higher song-learning accuracy. This would allow the user to better examine how an association between song quality and fitness affects song evolution in the face of different sexual selection pressures. It would be especially interesting to see how this affects the evolution of song in the case of the Social selection preference, as males that live longer may be more likely to breed, and this may drive song evolution in the absence of explicit selection on song.

I would also like to perform "insult" experiments, which could be run in the model as currently designed. In these simulations, the population is first allowed to reach equilibrium while under the influence of one selection pressure. Post-equilibrium, the selection pressure is changed (the insult). This experiment tests how males respond to fast changes in female preferences and whether they can evolve to keep up with these changes or if the evolution of certain song characteristics makes it difficult for males to adapt to changing preferences. Several such experiments I performed early in model development suggested that males can evolve their song traits to extremes that prevent the population from responding to future changes in female preferences. Running these experiments would only require the creation of a new simulation wrapper and program in **C#** that could be derived from the existing invasion program.

Chapter 7

Conclusions and future directions

7.1 Synthesis and limitations

The aim of my thesis was to better understand song evolution based on data from real birds and data generated from a computational model informed by the current state of the literature. I first provided a novel solution to a longstanding controversy in the literature: Do female preferences for elaborate songs lead to greater reproductive success for males with large repertoires? If so, could these preferences have driven the evolution of elaborate song? My meta-analysis of 25 bird species (**Chapter 2** and[9]) suggested that strong sexual selection for individuals with larger repertoires exists in species that already have moderate to large syllable repertoires. This selection pressure was much weaker in species with small average syllable repertoires. I also examined whether there was a difference in the strength of the correlation between individual elaboration and reproductive success between song-plastic species (likely open-ended learners) and song-stable species (likely closed-ended learners), but I did not find significant evidence for this relationship. However, this work was done in a relatively small number of species, because the correlation between individual song elaboration and reproductive success has only been studied in a few species. Furthermore, it is the least well studied in species with large repertoires, where my results suggest that the correlation is likely to be the strongest. My dataset also had a limited number of song-stable species, so there may well be a difference between song-stable and song-plastic species that will be detectable when there are more data available.

Despite the limited amount of field data available, my simulation model corroborates the findings of my meta-analysis (**Chapters 4** and **6**). Simulations with female preference for larger repertoires developed large syllable repertoires, and large repertoires did not develop when there was no direct selection on song or when specific syllables were preferred (**Chapter 4**). However, I also found that preference for common or rare syllables could drive the evolution of larger repertoires

(**Chapter 6**). Additionally, there were many cases where learning after the first breeding season was associated with higher accuracy template-matching (especially with the Conformity learning strategy). Additionally, the Prune strategy lost oblique learning altogether to become a modified Add strategy, but allowed for the development of intermediate to large repertoires, suggesting that extensive learning in the first year of life is sufficient to develop large repertoires. This supports my prediction that extended learning can be valuable in multiple contexts, so longer learning does not always mean that repertoire size will increase with age. Male song-learning strategy in combination with length of learning may be a better predictor of the strength of the correlation between individual repertoire size and reproductive success.

In my co-author paper[70] (**Chapter 3**), we found that adult song plasticity was associated with larger repertoire sizes, and that it affected the rate of evolution for other song traits such as song duration and intersong interval. Furthermore, we found that larger repertoires with song stability and small repertoires with song plasticity were evolutionarily unstable states. For this research, we collected a larger number of species (67), but this is still a very small sample of songbird diversity. My model does not allow for the examination of how structural song features evolve, but I did see a trend regarding repertoire size: in cases where female preferences were able to drive males to evolve larger repertoires, longer learning thresholds also evolved to be longer in tandem (**Chapter 4**). When evaluating individual line traces over time, I saw that populations that reached larger syllable repertoires and then experienced a drop in the average learning threshold also experienced a simultaneous drop in syllable repertoire. Furthermore, for some learning strategies (e.g Add and Add/Forget), males would evolve away from oblique learning to keep their repertoires small in the face of female preference for higher song template-matching accuracy. This supports our empirical finding that intermediate to large syllable repertoires require typically longer learning to be maintained, but also underscores that the male learning strategy adds an additional layer of nuance as to how much additional time spent learning is valuable and how large syllable repertoires can become.

Overall, my completed research projects put together a theoretical framework of the interplay between sexual selection, learning behavior, and cultural evolution. If males with certain song traits (culture) are more likely to reproduce (a selective force), this interaction affects the length of time males are capable of learning. However, the framework is complicated by the fact that certain learning strategies tend to produce songs with certain features. Thus, the final expression of a species-typical song interacts with both 1) the sexual selection forces on song and 2) the male song learning strategy. In turn, the value of longer learning would depend on the interaction between those two features. Very little research has been done on male learning strategies, and my research motivates the importance of filling this gap. Additionally, my model predicts a key role for social information in tutor choice when males need to maintain specific song features, such as dialects or specific syllables (**Chapter 5**); males cannot learn from random, neighboring males and reliably maintain these features across generations. Instead, they must preferentially learn from males that will reproduce or have previously reproduced. The importance of socially informed tutor choice has been studied in humans[259–267], but it is a novel idea for the birdsong field that has received little attention in the past. That said, it is also possible that males could have preferences for specific songs among potential tutors, similar to how females have preferences for specific songs among potential mates. If male and female preferences align, then males would not necessarily need to rely on social information to learn song from appropriate tutors; they would be attracted to learn the same songs that females prefer for mating. However, it is known that male and female preferences do not always align[256]. Furthermore, in scenarios where there is a mix of females with different preferences, female preferences change over time, or male-male interactions are critical to mating success, males would most easily adapt by copying successful males.

7.2 Avenues for future research

My research indicates that the correlation between individual song elaboration and reproductive success should be studied in more species with large syllable repertoires. Initially, this may seem like a difficult proposition; recording the full repertoire from a male that produces hundreds of syllables requires hours of dedicated recording time for each individual[268,269]. However, several researchers have developed methods for estimating the overall repertoire size of an individual based on much shorter recordings. While not all methods to estimate repertoire size have been useful in species for which the true repertoire size is not known[270], simple techniques, such as rarefaction curves, could be highly valuable[271] or species-specific algorithms may be highly effective in some cases. For example, it was shown that the repertoire size of Marsh Warblers (*Acrocephalus palustris*) can be estimated based on 2 minutes of recording[268]. Indeed, since it is unlikely that females listen to individual males for hours at a time before making mate choice decisions, females may use other heuristics to estimate whether a male has a larger repertoire than his peers. Put another way, males with greater individual song elaboration may perform their songs differently than conspecific males with lesser individual song elaboration. It has been suggested that there is a relationship between the variety of songs produced during a period of singing and length of time spent singing versus being silent that can be examined between species, though this has not been rigorously refuted or supported[272,273]. I hypothesize that there could be within-species differences in the presentation of song depending on syllable repertoire size, such that males with larger repertoires could sing more often, have more unique syllables per song, or sing with greater song variety in the same amount of time compared to their peers. It was noted in the Great Reed Warbler (*Acrocephalus arundinaceus*) that males with larger repertoires continued to produce novel syllable types for a longer period of time[95]. From my own observations in quantifying the repertoires for House Finches (*Carpodacus mexicanus*) (**Table E.1.1**), I would hypothesize that males in this species with larger syllable repertoires have longer songs with more unique syllables per song and shorter intersong intervals than males with smaller syllable repertoires. Similarly, for Canada Warbler (*Wilsonia canadensis*) (**Table E.1.2**), I would hypothesize that males with larger syllable repertoires have shorter intersong intervals and potentially a greater number of longer songs with more syllables per song, though I did not have enough data points to be confident in the latter. Extensive recording of a small number of individuals for one species could be sufficient to develop species-specific algorithms for estimating repertoire size from relatively short recordings (ideally, no longer than 20 minutes). Such algorithms would make studies in large repertoire species more feasible and allow researchers to track a greater number of males each season.

Naturally, it would be valuable to know the length of the song-learning window or the song-stability state of males in more species, however, it is especially important for future research to examine the duration of song learning in females. The role of female song in bird ecology has been greatly underappreciated[7,8]. I was able to define the song-stability state for males of 69 species, but to my knowledge this has not been assessed in females for any species, including those that engage in duetting behavior. It may be that the learning program in males and females is very similar, but this is not necessarily the case. In males, testosterone surges associated with sexual maturity play a role in song production and crystallization[174,274]. Females do not experience heightened levels of testosterone, so perhaps the female song template is more flexible than that of the males. It has been shown that females can change their song preferences as adults[204], and that estradiol, the songbird equivalent of estrogen, affects song preferences in both males and females[275]. Furthermore, estradiol seems to play an important role in male neural plasticity in songbirds[276–278]. Thus, it is possible that female singing behavior is regulated more strongly by estradiol than testosterone, though, again, this has not been examined. Future research should first examine the song-learning program in duetting species, where the song behavior of males and females can more easily be studied and compared. Once between-sex similarities and differences have been established in these species, it will be important to confirm whether species in which females sing more rarely act similarly to duetting species.

Most studies examining the length of the song-learning window only compare first-year males to second-year males. Therefore, even though open-ended learning has been defined as the ability to modify song throughout life, it may not necessarily be the case that all species currently considered song-plastic or open-ended learners actually show lifelong changes to song. Indeed, studies in canaries showed that the majority of song plasticity is exhibited from year 1 to year 2 with much less seen in year 2 to year 3. My model also suggested that lifelong learning would be exceedingly rare, especially if longer learning imposes a survival cost (e.g. by being more metabolically costly or reducing time for activities like foraging). In fact, learning through the second breeding season was sufficient for populations using the Add strategy to evolve and maintain very large syllable repertoires (>300) when the maximum repertoire size possible for the population was 500 syllables. It has been suggested that species that display seasonal plasticity may do so to reduce the costs of extended learning windows. Could it also be the case that true life-long learning is rare in songbirds, because males reap the majority of the benefits of longer learning in their first few years of life? Future research should examine song changes in older males to verify if presumed open-ended learners truly modify their repertoires throughout life, or whether they simply have extended learning that diminishes or ceases late in life. Interestingly, it has been noted that species with delayed plumage and song maturation (which I labeled as song-stable) have longer lifespans[179,180]. It has not been examined whether there is a correlation between lifespan and the length of the song-learning window, but it may be that species that live longer spend more time learning.

For the majority of species, it is unknown what strategy males (or females) use to acquire song. In Darwin's finches[27], it has been shown that males learn vertically from their fathers with high fidelity, and it has been shown in some species with small to moderate repertoires that males pick their final song during their first breeding season either to conform to local neighbors or dialects[20,209–211]. Finally, some species have been shown to increase repertoire sharing with their neighbors during the breeding season or from one breeding season to the next[279–281], though this has also mostly been studied in species with small to moderate repertoires. Much less is known about the learning program for species with extremely large repertoires. My model suggests that a strategy where males attempt to learn any syllable that they hear from conspecifics allows for the evolution of very large syllable repertoires (>100) and mechanisms that prune the repertoire (i.e. the Add/Forget strategy) prevent repertoires from becoming extremely large. Laboratory or aviary studies in species with large repertoires would allow researchers to assess whether males tend to produce every syllable they learn, or whether males tend to prune their repertoires based on the songs of other males around them.

My data on dialects suggests that there are two mechanisms by which dialects form: vertical learning in combination with limited nest-site dispersal and oblique learning from reproductively successful tutors. It will be important to examine which of these two mechanisms of dialect formation and maintenance fit the life history of species with known dialects and whether the existence of dialects in some species cannot be attributed to either mechanism. Additionally, both of these mechanisms relied on learners picking tutors that had a history of reproductive success, but the social contexts in which birds pick oblique tutors has not been well studied[32–34]. My results regarding the interactions between female preferences and male learning strategies suggested that socially informed tutor choice was not important for the formation of large repertoires. Thus, it would be interesting to examine whether species with dialects or highly stereotyped songs, which also tend to have smaller syllable repertoires, rely more on social information to pick tutors than species with very large repertoires. It would also be interesting to examine whether birds that learn their song from reproductively successful tutors are more reproductively successful themselves. However, due to the difficultly in determining which potential tutors a male learned from, such experiments may be restricted to laboratory and aviary settings.

My song evolution model was designed with the intent that additional models could be easily integrated in the future, so that new questions could be explored. As I discussed in **Chapter 6**, I would like the model to allow female preferences to evolve and would like to add survival rewards that would link males trait positively to survival probability. There is one additional feature that could be added that would allow the model to examine the vast majority of hypotheses regarding birdsong evolution currently in the field. One of the model's current limitations is that the user cannot examine the distinction or interplay between selection based on female choice and selection based on male-male competition. To add this functionality, I would like there to be set of features added based on territoriality. Territories themselves could have a value for "quality" that could either affect the chance that chicks sired from that territory will survive to the next time step or affect the chance that a male will be included in the pool of males eligible to breed (or both). Males on lower-quality territories could challenge males on high-quality territories to try to take over better territories. The probability of success could be calculated using the existing function for picking which males breed, but additional features to calculate male-male competition-specific traits may be added (e.g. song-matching). This would allow the user to examine what happens if male competition selects for certain song traits while female preferences select for others.

There are three additional features that would add more nuance to the model, but it is not clear whether they would affect the overall outcomes of the simulations. 1) Currently, the learning penalty is prorated, such that males that can potentially learn for a long period of time are more heavily punished every time step (including time steps after they have ceased learning) than males that learn for a shorter period of time. This mechanism assumes that simply having the biological machinery for longer learning is more costly. However, it may be that the capacity to learn for a longer period itself is not more costly, but rather that learning always costs the same amount; leaning into adulthood simply extends the amount of time an individual spends accruing the costs of learning. In that case, all males that learned in a given time step should be penalized by the same value, and males would not be penalized in time steps after they stop learning. This feature would only require the addition of a new survival penalty equation and another parameter to toggle between the two styles of learning penalties. 2) Many songbird species are proposed to have a predisposition to song structures and features. It would be interesting to see whether incorporating a form of innate song structure (e.g. all chicks are born knowing a specific syllable) would affect song evolution in the model. It might also allow the model to examine evolution between completely innate song (as it is suggested to be in subocines) and song that requires learning from conspecifics (oscines). This feature would require one or more new parameters to define the initial song template for chicks and a small modification to the chick generating procedure. 3) Inheritance in the model as currently implemented is uni-parental. It is unknown whether male song-learning behavior is primarily inherited from fathers or if female preferences are primarily inherited from mothers, with the exception that female song templates are believed to be based on their father's template in many cases. Allowing for bi-parental inheritance would require that all sex-specific traits were generated for both sexes and the creation of a new inheritance mechanism that considered the trait values of both parents. This extension would require significant reworking of many model features related to chick generation and the data structures for male and female birds.

In total, I have provided possible solutions to controversies in the literature and provided evidence for theorized connections between sexual selection and the evolution of song-learning behavior. My empirical results were then confirmed by my modeling work. Furthermore, my model allowed me to explore questions that the field could not previously ask and proposed novel predictions about the evolution of birdsong. Finally, my model is flexible enough that it still has untapped potential to answer evolutionary questions as is and can be fairly easily modified to answer an even broader array of questions. My thesis work leaves the field poised to better understand the interactions between cultural evolution, sexual selection, and learned behaviors in the context of both bird evolution and wider evolutionary trends.

BIBLIOGRAPHY

[1] C.K. Catchpole and P.J.B. Slater. *Bird song: biological themes and variations*. Cambridge university press, 2003.

[2] S.T. Emlen. An experimental analysis of the parameters of bird song eliciting species recognition. *Behaviour*, 41(1-2):130–171, 1972.

[3] D.A. Nelson. The importance of invariant and distinctive features in species recognition of bird song. *The Condor*, 91(1):120–130, 1989.

[4] M.D. Beecher and E.A. Brenowitz. Functional aspects of song learning in songbirds. *Trends in ecology & evolution*, 20(3):143–149, 2005.

[5] J.R. Krebs. The significance of song repertoires: the beau geste hypothesis. *Animal Behaviour*, 25:475–478, 1977.

[6] R. Schmidt, V. Amrhein, H.P. Kunc, and M. Naguib. The day after: effects of vocal interactions on territory defence in nightingales. *Journal of Animal Ecology*, 76(1):168–173, 2007.

[7] N.E. Langmore. Functions of duet and solo songs of female birds. *Trends in Ecology & Evolution*, 13(4):136–140, 1998.

[8] K.J. Odom, M.L. Hall, K. Riebel, K.E. Omland, and N.E. Langmore. Female song is widespread and ancestral in songbirds. *Nature Communications*, 5:3379, 2014.

[9] C. Robinson and N. Creanza. Species-level repertoire size predicts a correlation between individual song elaboration and reproductive success. *Ecology and Evolution*, 9(14):8362–8377, 2019.

[10] D. Gil and M. Gahr. The honesty of bird song: multiple constraints for multiple traits. *Trends in Ecology & Evolution*, 17(3):133–141, 2002.

[11] K.A. Spencer, K.L. Buchanan, A.R. Goldsmith, and C.K. Catchpole. Song as an honest signal of developmental stress in the zebra finch (*Taeniopygia guttata*). *Hormones and behavior*, 44(2):132–139, 2003.

[12] S. Nowicki, W.A. Searcy, and S. Peters. Brain development, song learning and mate choice in birds: a review and experimental test of the “nutritional stress hypothesis”. *Journal of Comparative Physiology A*, 188(11-12):1003–1014, 2002.

[13] J.M. Reid, P. Arcese, A.L.E.V. Cassidy, A.B. Marr, J.N.M. Smith, and L.F. Keller. Hamilton and zuk meet heterozygosity? song repertoire size indicates inbreeding and immunity in song sparrows (*Melospiza melodia*). *Proceedings of the Royal Society B: Biological Sciences*, 272(1562):481–487, 2005.

[14] A.P. Møller. Parasite load reduces song output in a passerine bird. *Animal Behaviour*, 41(4):723–730, 1991.

[15] W.A. Searcy and M. Andersson. Sexual selection and the evolution of song. *Annual Review of Ecology and Systematics*, 17(1):507–533, 1986.

[16] F. Nottebohm and M.E. Nottebohm. Relationship between song repertoire and age in the canary, serinus canarius. *Zeitschrift für Tierpsychologie*, 46(3):298–305, 1978.

[17] D. Gil, J.L.S. Cobb, and P.J.B. Slater. Song characteristics are age dependent in the willow warbler, *Phylloscopus trochilus*. *Animal Behaviour*, 62(4):689–694, 2001.

[18] B.E. Byers and D.E. Kroodsma. Female mate choice and songbird song repertoires. *Animal Behaviour*, 77(1):13–22, 2009.

[19] M. Soma and L.Z. Garamszegi. Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. *Behavioral Ecology*, 22(2):363–371, 2011.

[20] W-C. Liu and D.E. Kroodsma. Song learning by chipping sparrows: when, where, and from whom. *The Condor*, 108(3):509–517, 2006.

[21] H. Hultsch. *Beziehungen zwischen Struktur, zeitlicher Variabiliffit und sozialem Einsatz des Gesangs der Nachtigall Luscinia megarhynchos*. PhD thesis, Free University of Berlin, 1980.

[22] A.F. Read and D.M. Weary. The evolution of bird song: comparative analyses. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 338(1284):165–187, 1992.

[23] S.A. Macdougall-Shackleton. Sexual selection and the evolution of song repertoires. In *Current ornithology*, pages 81–124. Springer, 1997.

[24] K.T. Snyder and N. Creanza. Polygyny is linked to accelerated birdsong evolution but not to larger song repertoires. *Nature communications*, 10(1):884, 2019.

[25] C. Darwin. *The descent of man and selection in relation to sex*, volume 1. D. Appleton, 1896.

[26] C.K. Catchpole. Bird song, sexual selection and female choice. *Trends in Ecology & Evolution*, 2(4):94–97, 1987.

[27] B.R. Grant and P.R. Grant. Cultural inheritance of song and its role in the evolution of darwin's finches. *Evolution*, 50(6):2471–2487, 1996.

[28] L.A. Eales. Song learning in zebra finches: some effects of song model availability on what is learnt and when. *Animal Behaviour*, 33(4):1293–1300, 1985.

[29] D.E. Kroodsma. Learning and the ontogeny of sound signals in birds. *Acoustic communication in birds*, 2:1–23, 1982.

[30] M.L. Kreutzer and L. Nagle. Song tutoring influences female song preferences in domesticated canaries. *Behaviour*, 134(1-2):89–104, 1997.

[31] K Riebel. Early exposure leads to repeatable preferences for male song in female zebra finches. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1461):2553–2558, 2000.

[32] Christopher N Templeton, Çağlar Akçay, S Elizabeth Campbell, and Michael D Beecher. Juvenile sparrows preferentially eavesdrop on adult song interactions. *Proceedings of the Royal Society B: Biological Sciences*, 277(1680):447–453, 2009.

[33] Michael D Beecher, John M Burt, Adrian L O'Loghlen, Christopher N Templeton, and S Elizabeth Campbell. Bird song learning in an eavesdropping context. *Animal Behaviour*, 73(6):929–935, 2007.

[34] Michael D Beecher. Birdsong learning as a social process. *Animal Behaviour*, 124:233–246, 2017.

[35] B. Ballentine, J. Hyman, and S. Nowicki. Vocal performance influences female response to male bird song: an experimental test. *Behavioral Ecology*, 15(1):163–168, 2004.

[36] S.M. Lyons, M. Beaulieu, and K.W. Sockman. Contrast influences female attraction to performance-based sexual signals in a songbird. *Biology letters*, 10(10):20140588, 2014.

[37] S. Nowicki and W.A. Searcy. Song function and the evolution of female preferences: why birds sing, why brains matter. *Annals of the New York Academy of Sciences*, 1016(1):704–723, 2004.

[38] E. Hofstad, Y. Espmark, A. Moksnes, T. Haugan, and M. Ingebrigtsen. The relationship between song performance and male quality in snow buntings (*Plectrophenax nivalis*). *Canadian Journal of Zoology*, 80(3):524–531, 2002.

[39] A.V. Badyaev, G.E. Hill, and B.V. Weckworth. Species divergence in sexually selected traits: increase in song elaboration is related to decrease in plumage ornamentation in finches. *Evolution*, 56(2):412–419, 2002.

[40] M. Soma and L. Z. Garamszegi. Evolution of patterned plumage as a sexual signal in estrildid finches. *Behavioral Ecology*, 29(3):676–685, 2018.

[41] G.E. Hill. Plumage coloration is a sexually selected indicator of male quality. *Nature*, 350(6316):337, 1991.

[42] D.F. Maynard. *The vocal behavior of long-tailed manakins (Chiroxiphia linearis): the role of vocalizations in mate attraction and male-male interactions*. PhD thesis, University of Windsor, 2012.

[43] P. Fiske, P.T. Rintamäki, and E. Karvonen. Mating success in lekking males: a meta-analysis. *Behavioral Ecology*, 9(4):328–338, 1998.

[44] J. Byers, E. Hebets, and J. Podos. Female mate choice based upon male motor performance. *Animal Behaviour*, 79(4):771–778, 2010.

[45] P. Marler. On innateness: are sparrow songs learned or innate. *The design of animal communication*, pages 293–318, 1999.

[46] D.E. Kroodsma and M. Konishi. A suboscine bird (eastern phoebe, sayornis phoebe) develops normal song without auditory feedback. *Animal Behaviour*, 42(3):477–487, 1991.

[47] P. Marler. Song learning: innate species differences in the learning process. In *The biology of learning*, pages 289–309. Springer, 1984.

[48] C.S. Whaling, M.M. Solis, A.J. Doupe, J.A. Soha, and P. Marler. Acoustic and neural bases for innate recognition of song. *Proceedings of the National Academy of Sciences*, 94(23):12694–12698, 1997.

[49] P. Marler and V. Sherman. Song structure without auditory feedback: emendations of the auditory template hypothesis. *Journal of Neuroscience*, 3(3):517–531, 1983.

[50] P. Marler. Three models of song learning: evidence from behavior. *Journal of neurobiology*, 33(5):501–516, 1997.

[51] M.S. Brainard and A.J. Doupe. What songbirds teach us about learning. *Nature*, 417(6886):351, 2002.

[52] E.A. Brenowitz and M.D. Beecher. Song learning in birds: diversity and plasticity, opportunities and challenges. *Trends in neurosciences*, 28(3):127–132, 2005.

[53] H. Williams. Birdsong and singing behavior. *Annals of the New York Academy of Sciences*, 1016(1):1–30, 2004.

[54] M. Cucco and G. Malacarne. Delayed maturation in passerine birds: an examination of plumage effects and some indications of a related effect in song. *Ethology Ecology & Evolution*, 12(3):291–308, 2000.

[55] P. Marler and S. Peters. Sparrows learn adult song and more from memory. *Science*, 213(4509):780–782, 1981.

[56] L.E. Vargas-Castro, N.V. Sánchez, and G. Barrantes. Song plasticity over time and vocal learning in clay-colored thrushes. *Animal cognition*, 18(5):1113–1123, 2015.

[57] M. Chaiken, J. Böhner, and P. Marler. Repertoire turnover and the timing of song acquisition in european starlings. *Behaviour*, 128(1-2):25–39, 1994.

[58] D.A. Nelson. External validity and experimental design: the sensitive phase for song learning. *Animal Behaviour*, 56(2):487–491, 1998.

[59] F. Nottebohm. The “critical period” for song learning. *Ibis*, 111(3):386–387, 1969.

[60] P. Marler and S. Peters. Sensitive periods for song acquisition from tape recordings and live tutors in the swamp sparrow, *Melospiza georgiana*. *Ethology*, 77(1):76–84, 1988.

[61] L.F. Baptista and L. Petrinovich. Social interaction, sensitive phases and the song template hypothesis in the white-crowned sparrow. *Animal Behaviour*, 32(1):172–181, 1984.

[62] D.E. Kroodsma and R. Pickert. Sensitive phases for song learning: Effects of social interaction and individual variation. *Animal Behaviour*, 32(2):389–394, 1984.

[63] T.J.S. Balsby and P. Hansen. Element repertoire: change and development with age in whitethroat sylvia communis song. *Journal of Ornithology*, 151(2):469–476, 2010.

[64] E.A. Brenowitz. Plasticity of the adult avian song control system. *Annals of the New York Academy of Sciences*, 1016(1):560–585, 2004.

[65] K. Aoki. A sexual-selection model for the evolution of imitative learning of song in polygynous birds. *The American Naturalist*, 134(4):599–612, 1989.

[66] M.N. Verzijden, R.F. Lachlan, and M.R. Servedio. Female mate-choice behavior and sympatric speciation. *Evolution*, 59(10):2097–2108, 2005.

[67] M.N. Verzijden, C. Ten Cate, M.R. Servedio, G.M. Kozak, J.W. Boughman, and E.I. Svensson. The impact of learning on sexual selection and speciation. *Trends in ecology & evolution*, 27(9):511–519, 2012.

[68] R.F. Lachlan and M.R. Servedio. Song learning accelerates allopatric speciation. *Evolution*, 58(9):2049–2063, 2004.

[69] J.T. Rowell and M.R. Servedio. Vocal communications and the maintenance of population specific songs in a contact zone. *PloS one*, 7(5):e35257, 2012.

[70] C. Robinson, K.T. Snyder, and N. Creanza. Correlated evolution between repertoire size and song plasticity predicts that sexual selection on song promotes open-ended learning. *eLife*, 8:e44454, 2019.

[71] C. Robinson and N. Creanza. Modeling the evolutionary interactions of sexual selection and learning strategies on the duration of song-learning in birds. *PLOS Computational Biology*, 2019.

[72] A. Roper and R. Zann. The onset of song learning and song tutor selection in fledgling zebra finches. *Ethology*, 112(5):458–470, 2006.

[73] R.F. Lachlan and P.J.B. Slater. Song learning by chaffinches: how accurate, and from where? *Animal Behaviour*, 65(5):957–969, 2003.

[74] O. Tchernichovski, P.P. Mitra, T. Lints, and F. Nottebohm. Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science*, 291(5513):2564–2569, 2001.

[75] K. Yasukawa, J.L. Blank, and C.B. Patterson. Song repertoires and sexual selection in the red-winged blackbird. *Behavioral Ecology and Sociobiology*, 7(3):233–238, 1980.

[76] C.K. Catchpole. Sexual selection and the evolution of complex songs among european warblers of the genus *Acrocephalus*. *Behaviour*, pages 149–166, 1980.

[77] S. Nowicki, S. Peters, and J. Podos. Song learning, early nutrition and sexual selection in songbirds. *American Zoologist*, 38(1):179–190, 1998.

[78] R.D. Howard. The influence of sexual selection and interspecific competition on mockingbird song (*Mimus polyglottos*). *Evolution*, pages 428–438, 1974.

[79] J.M. Reid, P. Arcese, A.L.E.V. Cassidy, S.M. Hiebert, J.N.M. Smith, P.K. Stoddard, A.B. Marr, and L.F. Keller. Song repertoire size predicts initial mating success in male song sparrows, *Melospiza melodia*. *Animal Behaviour*, 68(5):1055–1063, 2004.

[80] W.A. Searcy. Song repertoire and mate choice in birds. *American Zoologist*, 32(1):71–80, 1992.

[81] W.A. Searcy. Song repertoire size and female preferences in song sparrows. *Behavioral Ecology and Sociobiology*, 14(4):281–286, 1984.

[82] Jeremy A Pfaff, Liana Zanette, S.A. MacDougall-Shackleton, and E.A. MacDougall-Shackleton. Song repertoire size varies with hvc volume and is indicative of male quality in song sparrows (*Melospiza melodia*). *Proceedings of the Royal Society B: Biological Sciences*, 274(1621):2035–2040, 2007.

[83] D.E. Kroodsma. Reproductive development in a female songbird: differential stimulation by quality of male song. *Science*, 192(4239):574–575, 1976.

[84] K.L. Buchanan and C.K. Catchpole. Song as an indicator of male parental effort in the sedge warbler. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 267(1441):321–326, 2000.

[85] M. Eens, R. Pinxten, and R.F. Verheyen. Male song as a cue for mate choice in the european starling. *Behaviour*, pages 210–238, 1991.

[86] C. Catchpole, B. Leisler, and J. Dittami. Sexual differences in the responses of captive great reed warblers (*Acrocephalus arundinaceus*) to variation in song structure and repertoire size. *Ethology*, 73(1):69–77, 1986.

[87] C.K. Catchpole, J. Dittami, and B. Leisler. Differential responses to male song repertoires in female songbirds implanted with oestradiol. *Nature*, 312(5994):563, 1984.

[88] M.C. Baker, T.K. Bjerke, H. Lampe, and Y. Espmark. Sexual response of female great tits to variation in size of males' song repertoires. *The American Naturalist*, 128(4):491–498, 1986.

[89] H.M. Lampe and G-P. Saetre. Female pied flycatchers prefer males with larger song repertoires. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 262(1364):163–167, 1995.

[90] D.M. Logue and W. Forstmeier. Constrained performance in a communication network: implications for the function of song-type matching and for the evolution of multiple ornaments. *The American Naturalist*, 172(1):34–41, 2008.

[91] M. Treisman. Bird song dialects, repertoire size, and kin association. *Animal Behaviour*, 26:814–817, 1978.

[92] J.M. Williams and P.J.B. Slater. Modelling bird song dialects: the influence of repertoire size and numbers of neighbours. *Journal of Theoretical Biology*, 145(4):487–496, 1990.

[93] S. Nowicki, W.A. Searcy, and S. Peters. Quality of song learning affects female response to male bird song. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1503):1949–1954, 2002.

[94] M-C. Gontard-Danek and A.P. Møller. The strength of sexual selection: a meta-analysis of bird studies. *Behavioral ecology*, 10(5):476–486, 1999.

[95] C.K. Catchpole. Song repertoires and reproductive success in the great reed warbler *Acrocephalus arundinaceus*. *Behavioral Ecology and Sociobiology*, 19(6):439–445, 1986.

[96] S.M. Hiebert, P.K. Stoddard, and P. Arcese. Repertoire size, territory acquisition and reproductive success in the song sparrow. *Animal Behaviour*, 37:266–273, 1989.

[97] M. Andersson. Female choice selects for extreme tail length in a widowbird. *Nature*, 299(5886):818, 1982.

[98] S.R. Pryke, S. Andersson, and M.J. Lawes. Sexual selection of multiple handicaps in the red-collared widowbird: female choice of tail length but not carotenoid display. *Evolution*, 55(7):1452–1463, 2001.

[99] S. Andersson. Female preference for long tails in lekking jackson's widowbirds: experimental evidence. *Animal Behaviour*, 43(3):379–388, 1992.

[100] M. Griggio, F. Valera, A. Casas-Crivillé, H. Hoi, and A. Barbosa. White tail markings are an indicator of quality and affect mate preference in rock sparrows. *Behavioral Ecology and Sociobiology*, 65(4):655–664, 2011.

[101] D.F. Westneat. No evidence of current sexual selection on sexually dimorphic traits in a bird with high variance in mating success. *The American Naturalist*, 167(6):E171–E189, 2006.

[102] K.J. Thusius, K.A. Peterson, P.O. Dunn, and L.A. Whittingham. Male mask size is correlated with mating success in the common yellowthroat. *Animal Behaviour*, 62(3):435–446, 2001.

[103] J.R. Krebs and D.E. Kroodsma. Repertoires and geographical variation in bird song. In *Advances in the Study of Behavior*, volume 11, pages 143–177. Elsevier, 1980.

[104] J. Sundberg and A. Dixon. Old, colourful male yellowhammers, *Emberiza citrinella*, benefit from extra-pair copulations. *Animal Behaviour*, 52(1):113–122, 1996.

[105] J.L. Dickinson. Extrapair copulations in western bluebirds (*Sialia mexicana*): female receptivity favors older males. *Behavioral Ecology and Sociobiology*, 50(5):423–429, 2001.

[106] H. Kokko. Good genes, old age and life-history trade-offs. *Evolutionary Ecology*, 12(6):739–750, 1998.

[107] R. Brooks and D.J. Kemp. Can older males deliver the good genes? *Trends in ecology & evolution*, 16(6):308–313, 2001.

[108] K. Martin. Patterns and mechanisms for age-dependent reproduction and survival in birds. *American Zoologist*, 35(4):340–348, 1995.

[109] S. Kipper and S. Kiefer. Age-related changes in birds' singing styles: on fresh tunes and fading voices? In *Advances in the Study of Behavior*, volume 41, pages 77–118. Elsevier, 2010.

[110] J. Schroeder, S. Nakagawa, M. Rees, M-E. Mannarelli, and T. Burke. Reduced fitness in progeny from old parents in a natural population. *Proceedings of the National Academy of Sciences*, 112(13):4021–4025, 2015.

[111] Xeno-Canto Foundation, 2005. www.xeno-canto.org.

[112] S. Verhulst and J-Å. Nilsson. The timing of birds' breeding seasons: a review of experiments that manipulated timing of breeding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1490):399–410, 2007.

[113] B.C. Pinkowski. Breeding adaptations in the eastern bluebird. *The Condor*, 79(3):289–302, 1977.

[114] I. Newton and M. Marquiss. Seasonal trend in the breeding performance of sparrowhawks. *The Journal of Animal Ecology*, pages 809–829, 1984.

[115] A.L.A. Middleton. Influence of age and habitat on reproduction by the american goldfinch. *Ecology*, 60(2):418–432, 1979.

[116] M.T. Murphy. Ecological aspects of the reproductive biology of eastern kingbirds: geographic comparisons. *Ecology*, 64(4):914–928, 1983.

[117] M.T. Murphy. Temporal components of reproductive variability in eastern kingbirds (*Tyrannus tyrannus*). *Ecology*, 67(6):1483–1492, 1986.

[118] L.Z. Garamszegi and A.P. Møller. Extrapair paternity and the evolution of bird song. *Behavioral Ecology*, 15(3):508–519, 2004.

[119] M. Petrie and B. Kempenaers. Extra-pair paternity in birds: explaining variation between species and populations. *Trends in Ecology & Evolution*, 13(2):52–58, 1998.

[120] F.M. Wolf. *Meta-analysis: Quantitative methods for research synthesis*, volume 59. Sage, 1986.

[121] N. Nachar. The mann-whitney u: A test for assessing whether two independent samples come from the same distribution. *Tutorials in quantitative Methods for Psychology*, 4(1):13–20, 2008.

[122] H. Friedman. Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin*, 70(4):245, 1968.

[123] A.P. Field and R. Gillett. How to do a meta-analysis. *British Journal of Mathematical and Statistical Psychology*, 63(3):665–694, 2010.

[124] W. Viechtbauer. Conducting meta-analyses in r with the metafor package. *J Stat Softw*, 36(3):1–48, 2010.

[125] F. Song, K.S. Khan, J. Dinnes, and A.J. Sutton. Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy. *International journal of epidemiology*, 31(1):88–95, 2002.

[126] J.D. Hadfield. Mcmc methods for multi-response generalized linear mixed models: the mcmcglmm r package. *Journal of Statistical Software*, 33(2):1–22, 2010.

[127] S. Nakagawa and E.S.A. Santos. Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology*, 26(5):1253–1274, 2012.

[128] A.J. Wilson, D. Reale, M.N. Clements, M.M. Morrissey, E. Postma, C.A. Walling, L.E.B. Kruuk, and D.H. Nussey. An ecologists guide to the animal model. *Journal of Animal Ecology*, 79(1):13–26, 2010.

[129] D.A. Pastor and R.A. Lazowski. On the multilevel nature of meta-analysis: a tutorial, comparison of software programs, and discussion of analytic choices. *Multivariate behavioral research*, 53(1):74–89, 2018.

[130] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472, 1992.

[131] S.P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4):434–455, 1998.

[132] M. Plummer, N. Best, K. Cowles, and K. Vines. Coda: convergence diagnosis and output analysis for mcmc. *R news*, 6(1):7–11, 2006.

[133] J.K. Kruschke. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2):573, 2013.

[134] W. Jetz, G.H. Thomas, J.B. Joy, K. Hartmann, and A.O. Mooers. The global diversity of birds in space and time. *Nature*, 491(7424):444, 2012.

[135] L.J. Revell. phytools: an r package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223, 2012.

[136] H.F. Rivera-Gutierrez, R. Pinxten, and M. Eens. Difficulties when assessing birdsong learning programmes under field conditions: a re-evaluation of song repertoire flexibility in the great tit. *PloS one*, 6(1):e16003, 2011.

[137] B.E. Byers. Birdsong, migration and sexual selection: a skeptical view. *Animal Behaviour*, 82(5):e1–e3, 2011.

[138] M.D. Beecher, S.E. Campbell, and J.C. Nordby. Territory tenure in song sparrows is related to song sharing with neighbours, but not to repertoire size. *Animal behaviour*, 59(1):29–37, 2000.

[139] C.K. Catchpole. Variation in the song of the great reed warbler *Acrocephalus arundinaceus* in relation to mate attraction and territorial defence. *Animal Behaviour*, 31(4):1217–1225, 1983.

[140] M. Konishi. Birdsong: from behavior to neuron. *Annual review of neuroscience*, 8(1):125–170, 1985.

[141] A.L. O'Loghlen and S.I. Rothstein. Culturally correct song dialects are correlated with male age and female song preferences in wild populations of brown-headed cowbirds. *Behavioral Ecology and Sociobiology*, 36(4):251–259, 1995.

[142] M. Naguib. Effects of song overlapping and alternating on nocturnally singing nightingales. *Animal Behaviour*, 58(5):1061–1067, 1999.

[143] E.S.C. Scordato. Male competition drives song divergence along an ecological gradient in an avian ring species. *Evolution*, 72(11):2360–2377, 2018.

[144] K. Yasukawa. Female song sparrows prefer males that learn well. *BioScience*, 52(12):1064–1065, 2002.

[145] K.L. Buchanan and C.K. Catchpole. Female choice in the sedge warbler *Acrocephalus schoenobaenus* multiple cues from song and territory quality. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1381):521–526, 1997.

[146] S. Nowicki and W.A. Searcy. Song and mate choice in birds: how the development of behavior helps us understand function. *The Auk*, 122(1):1–14, 2005.

[147] D.C. Airey and T.J. DeVoogd. Greater song complexity is associated with augmented song system anatomy in zebra finches. *Neuroreport*, 11(10):2339–2344, 2000.

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

[149] M. Hoi-Leitner, H. Nechtelberger, and H. Hoi. Song rate as a signal for nest site quality in blackcaps (*Sylvia atricapilla*). *Behavioral Ecology and Sociobiology*, 37(6):399–405, 1995.

[150] N. Creanza, L. Fogarty, and M.W. Feldman. Cultural niche construction of repertoire size and learning strategies in songbirds. *Evolutionary ecology*, 30(2):285–305, 2016.

[151] A. Guolo and C. Varin. Random-effects meta-analysis: the number of studies matters. *Statistical methods in medical research*, 26(3):1500–1518, 2017.

[152] S.T. Emlen and L.W. Oring. Ecology, sexual selection, and the evolution of mating systems. *Science*, 197(4300):215–223, 1977.

[153] O. Vedder, J. Komdeur, M. van der Velde, E. Schut, and M.J.L. Magrath. Polygyny and extra-pair paternity enhance the opportunity for sexual selection in blue tits. *Behavioral ecology and sociobiology*, 65(4):741–752, 2011.

[154] C.R. Freeman-Gallant, N.T. Wheelwaright, K.E. Meiklejohn, S.L. States, and S.V. Sollecito. Little effect of extrapair paternity on the opportunity for sexual selection in savannah sparrows (*Passerculus sandwichenisis*). *Evolution*, 59(2):422–430, 2005.

[155] T.R. Birkhead and J.D. Biggins. Reproductive synchrony and extra-pair copulation in birds. *Ethology*, 74(4):320–334, 1987.

[156] S.M. Yezerinac and P.J. Weatherhead. Reproductive synchrony and extra-pair mating strategy in a socially monogamous bird, *Dendroica petechia*. *Animal Behaviour*, 54(6):1393–1403, 1997.

[157] D.E. Irwin. Song variation in an avian ring species. *Evolution*, 54(3):998–1010, 2000.

[158] D.J. Mountjoy and D.W. Leger. Vireo song repertoires and migratory distance: three sexual selection hypotheses fail to explain the correlation. *Behavioral Ecology*, 12(1):98–102, 2001.

[159] S.A. Collins, S.R. de Kort, J. Pérez-Tris, and J. Luis Tellería. Migration strategy and divergent sexual selection on bird song. *Proceedings of the Royal Society B: Biological Sciences*, 276(1656):585–590, 2008.

[160] K.W. Nordeen. Neural correlates of sensitive periods in avian song learning. *Annals of the New York Academy of Sciences*, 807(1):386–400, 1997.

[161] L.A. Eales. Song learning in female-raised zebra finches: another look at the sensitive phase. *Animal Behaviour*, 35(5):1356–1365, 1987.

[162] J. Böhner. Early acquisition of song in the zebra finch, *Taeniopygia guttata*. *Animal Behaviour*, 39(2):369–374, 1990.

[163] K. Immelmann. Song development in the zebra finch and other estrildid finches. *Bird vocalizations*, pages 61–77, 1969.

[164] F. Nottebohm. Birdsong as a model in which to study brain processes related to learning. *The Condor*, 86(3):227–236, 1984.

[165] F. Dowsett-Lemaire. The imitative range of the song of the marsh warbler *Acrocephalus palustris*, with special reference to imitations of african birds. *Ibis*, 121(4):453–468, 1979.

[166] J. Martens and P. Kessler. Territorial song and song neighbourhoods in the scarlet rosefinch *Carpodacus erythrinus*. *Journal of avian biology*, 31(3):399–411, 2000.

[167] W-C. Liu and F. Nottebohm. A learning program that ensures prompt and versatile vocal imitation. *Proceedings of the National Academy of Sciences*, 104(51):20398–20403, 2007.

[168] M. Adret-Hausberger, H-R. Güttinger, and F.W. Merkel. Individual life history and song repertoire changes in a colony of starlings (*Sturnus vulgaris*). *Ethology*, 84(4):265–280, 1990.

[169] Y.O. Espmark and H.M. Lampe. Variations in the song of the pied flycatcher within and between breeding seasons. *Bioacoustics*, 5(1-2):33–65, 1993.

[170] M. Hausberger, P.F. Jenkins, and J. Keene. Species-specificity and mimicry in bird song: are they paradoxes? a reevaluation of song mimicry in the european starling. *Behaviour*, pages 53–81, 1991.

[171] J.D. Mountjoy and R.E. Lemon. Extended song learning in wild european starlings. *Animal Behaviour*, 49(2):357–366, 1995.

[172] J.J. Price and D.H. Yuan. Song-type sharing and matching in a bird with very large song repertoires, the tropical mockingbird. *Behaviour*, pages 673–689, 2011.

[173] F. Nottebohm, M.E. Nottebohm, and L. Crane. Developmental and seasonal changes in canary song and their relation to changes in the anatomy of song-control nuclei. *Behavioral and neural biology*, 46(3):445–471, 1986.

[174] G.T. Smith, E.A. Brenowitz, M.D. Beecher, and J.C. Wingfield. Seasonal changes in testosterone, neural attributes of song control nuclei, and song structure in wild songbirds. *Journal of Neuroscience*, 17(15):6001–6010, 1997.

[175] A.D. Tramontin, N. Perfito, J.C. Wingfield, and E.A. Brenowitz. Seasonal growth of song control nuclei precedes seasonal reproductive development in wild adult song sparrows. *General and comparative endocrinology*, 122(1):1–9, 2001.

[176] W.A. Searcy and P. Marler. Interspecific differences in the response of female birds to song repertoires. *Zeitschrift für Tierpsychologie*, 66(2):128–142, 1984.

[177] L.Z. Garamszegi and M. Eens. Brain space for a learned task: strong intraspecific evidence for neural correlates of singing behavior in songbirds. *Brain Research Reviews*, 44(2-3):187–193, 2004.

[178] A.D. Tramontin and E.A. Brenowitz. Seasonal plasticity in the adult brain. *Trends in neurosciences*, 23(6):251–258, 2000.

[179] T.I. Draganoiu, A. Moreau, L. Ravaux, W. Bonckaert, and N. Mathevon. Song stability and neighbour recognition in a migratory songbird, the black redstart. *Behaviour*, 151(4):435–453, 2014.

[180] J.M. Trainer and R.J. Parsons. Delayed vocal maturation in polygynous yellow-rumped caciques. *The Wilson Journal of Ornithology*, 114(2):249–255, 2002.

[181] (Editor) P. Rodewald. *The Birds of North America.* Cornell Laboratory of Ornithology: Ithaca, NY, 2015. https://birdsna.org.

[182] J. Del Hoyo, A. Elliott, J. Sargatal, D.A. Christie, and E. de Juana. Handbook of the birds of the world alive. *Lynx Edicions, Barcelona*, 2014. http://www.hbw.com/.

[183] S.C. Griffith, I.P.F. Owens, and K.A. Thuman. Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molecular ecology*, 11(11):2195–2212, 2002.

[184] M. Pagel. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 255(1342):37–45, 1994.

[185] M. Pagel and A. Meade. Bayesian analysis of correlated evolution of discrete characters by reversible-jump markov chain monte carlo. *The American Naturalist*, 167(6):808–825, 2006.

[186] J.F. Clements, Cornell Laboratory of Ornithology, and American Birding Association. *The Clements checklist of the birds of the world.* Christopher Helm London, 2007.

[187] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[188] P.F. Jenkins. Cultural transmission of song patterns and dialect development in a free-living bird population. *Animal Behaviour*, 26:50–78, 1978.

[189] A. Eriksen, T. Slagsvold, and H.M. Lampe. Vocal plasticity — are pied flycatchers, *Ficedula Hypoleuca*, open-ended learners? *Ethology*, 117(3):188–198, 2011.

[190] P. Galeotti, N. Saino, E. Perani, R. Sacchi, and A.P. Møller. Age-related song variation in male barn swallows. *Italian Journal of Zoology*, 68(4):305–310, 2001.

[191] L.Z. Garamszegi, D. Heylen, A.P. Møller, M. Eens, and F. De Lope. Age-dependent health status and song characteristics in the barn swallow. *Behavioral Ecology*, 16(3):580–591, 2005.

[192] J.S. Nicholson, K.L. Buchanan, R.C. Marshall, and C.K. Catchpole. Song sharing and repertoire size in the sedge warbler, *Acrocephalus schoenobaenus*: changes within and between years. *Animal Behaviour*, 74(5):1585–1592, 2007.

[193] T. Garland Jr, A.W. Dickerman, C.M. Janis, and J.A. Jones. Phylogenetic analysis of covariance by computer simulation. *Systematic Biology*, 42(3):265–292, 1993.

[194] B.C. O'Meara, C. Ané, M.J. Sanderson, and P.C. Wainwright. Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5):922–933, 2006.

[195] R.B. Payne. Sexual selection, lek and arena behavior, and sexual size dimorphism in birds. *Ornithological monographs*, 33:iii–52, 1984.

[196] J.L. Wildenthal. Structure in primary song of the mockingbird (*Mimus polyglottos*). *The Auk*, pages 161–189, 1965.

[197] D.E. Kroodsma and R. Pickert. Environmentally dependent sensitive periods for avian vocal learning. *Nature*, 288(5790):477, 1980.

[198] R.G. Morrison and F. Nottebohm. Role of a telencephalic nucleus in the delayed song learning of socially isolated zebra finches. *Journal of neurobiology*, 24(8):1045–1064, 1993.

[199] T.K. Kelly, S. Ahmadiantehrani, A. Blattler, and S.E. London. Epigenetic regulation of transcriptional plasticity associated with developmental song learning. *Proceedings of the Royal Society B: Biological Sciences*, 285(1878):20180160, 2018.

[200] L. Wilbrecht, H. Williams, N. Gangadhar, and F. Nottebohm. High levels of new neuron addition persist when the sensitive period for song learning is experimentally prolonged. *Journal of Neuroscience*, 26(36):9135–9141, 2006.

[201] E.J. Nordeen and K.W. Nordeen. Neurogenesis and sensitive periods in avian song learning. *Trends in neurosciences*, 13(1):31–36, 1990.

[202] F. Nottebohm. The search for neural mechanisms that define the sensitive period for song learning in birds. *Netherlands Journal of Zoology*, 43(1-2):193–234, 1992.

[203] M. Goller and D. Shizuka. Evolutionary origins of vocal mimicry in songbirds. *Evolution Letters*, 2(4):417–426, 2018.

[204] L. Nagle and M.L. Kreutzer. Adult female domesticated canaries can modify their song preferences. *Canadian Journal of Zoology*, 75(8):1346–1350, 1997.

[205] P. E. Lowther and Cink C. L. *The Birds of North America*. Cornell Lab of Ornithology, Ithaca, NY, USA, 2006.

[206] Satoshi Kojima and Allison J Doupe. Song selectivity in the pallial-basal ganglia song circuit of zebra finches raised without tutor song exposure. *Journal of neurophysiology*, 98(4):2099–2109, 2007.

[207] A Qvarnström, Jon Haavie, SA Saether, D Eriksson, and T Pärt. Song similarity predicts hybridization in flycatchers. *Journal of evolutionary biology*, 19(4):1202–1209, 2006.

[208] Peter R Grant and B Rosemary Grant. Hybridization, sexual imprinting, and mate choice. *The American Naturalist*, 149(1):1–28, 1997.

[209] Peter Marler and Miwako Tamura. Song directs in three populations of white-crowned sparrows. *The Condor*, 64(5):368–377, 1962.

[210] Peter Marler and Miwako Tamura. Culturally transmitted patterns of vocal behavior in sparrows. *Science*, 146(3650):1483–1486, 1964.

[211] Robert F Lachlan, Oliver Ratmann, and Stephen Nowicki. Cultural conformity generates extremely stable traditions in bird song. *Nature communications*, 9(1):2417, 2018.

[212] Anastasia H Dalziell and Robert D Magrath. Fooling the experts: accurate vocal mimicry in the song of the superb lyrebird, menura novaehollandiae. *Animal Behaviour*, 83(6):1401–1410, 2012.

[213] Marcel Eens. Understanding the complex song of the european starling: an integrated ethological approach. *Advances in the Study of Behaviour*, 26:355–434, 1997.

[214] Kim C Derrickson and R Breitwisch. Northern mockingbird. *The Birds of North America*, 7:1–26, 1992.

[215] Beat Naef-Daenzer, Fritz Widmer, and Maria Nuber. Differential post-fledging survival of great and coal tits in relation to their condition and fledging date. *Journal of Animal ecology*, 70(5):730–738, 2001.

[216] Anne Loison, Bernt-Erik Sæther, Kurt Jerstad, and Ole Wiggo Røstad. Disentangling the sources of variation in the survival of the european dipper. *Journal of Applied Statistics*, 29(1-4):289–304, 2002.

[217] Lynne E. Stenzel, Gary W. Page, Jane C. Warriner, John S. Warriner, Douglas E. George, Carleton R. Eyster, Bernadette A. Ramer, and Kristina K. Neuman. Survival and natal dispersal of juvenile snowy plovers (charadrius alexandrinus) in central coastal california. *The Auk*, 124(3):1023–1036, 2007.

[218] GA Tyler and RE Green. Effects of weather on the survival and growth of corncrake crex crex chicks. *Ibis*, 146(1):69–76, 2004.

[219] Michael Begon, John L Harper, Colin R Townsend, et al. *Ecology. Individuals, populations and communities.* Blackwell scientific publications, 1986.

[220] Heather Williams. Models for song learning in the zebra finch: fathers or others? *Animal Behaviour*, 39(4):745–757, 1990.

[221] Douglas A Nelson. Song overproduction and selective attrition lead to song sharing in the field sparrow (spizella pusilla). *Behavioral Ecology and Sociobiology*, 30(6):415–424, 1992.

[222] DOUGLAS A BELL, PEPPER W TRAIL, and LUIS F BAPTISTA. Song learning and vocal tradition in nuttall's white-crowned sparrows. *Animal behaviour*, 55(4):939–956, 1998.

[223] RF Lachlan, VM Janik, and PJB Slater. The evolution of conformity-enforcing behaviour in cultural communication systems. *Animal Behaviour*, 68(3):561–570, 2004.

[224] Robert B. Payne and Laura L. Payne. *Social influences on vocal development*. Cambridge University Press, 1997.

[225] Ofer Tchernichovski, Fernando Nottebohm, Ching Elizabeth Ho, Bijan Pesaran, and Partha Pratim Mitra. A procedure for an automated measurement of song similarity. *Animal behaviour*, 59(6):1167–1176, 2000.

[226] B. Rosemary Grant and Peter R. Grant. Cultural inheritance of song and its role in the evolution of darwin's finches. *Evolution*, 50(6):2471–2487, 1996.

[227] Stephen J. Millington and Trevor D. Price. Song Inheritance and Mating Patterns in Darwin's Finches. *The Auk: Ornithological Advances*, 102(2):342–346, 04 1985.

[228] Douglas A Nelson and Peter Marler. Selection-based learning in bird song development. *Proceedings of the National Academy of Sciences*, 91(22):10498–10501, 1994.

[229] Peter K McGregor and John R Krebs. Song learning in adult great tits (parus major): effects of neighbours. *Behaviour*, pages 139–159, 1989.

[230] Fredrik Widemo. Alternative reproductive strategies in the ruff, philomachus pugnax: a mixed ess? *Animal Behaviour*, 56(2):329–336, 1998.

[231] Eric Vallet, Irina Beme, and Michel Kreutzer. Two-note syllables in canary songs elicit high levels of sexual display. *Animal behaviour*, 55(2):291–297, 1998.

[232] Stefan Leitner and Clive K Catchpole. Syllable repertoire and the size of the song control system in captive canaries (serinus canaria). *Journal of neurobiology*, 60(1):21–27, 2004.

[233] Robert E Lemon. How birds develop song dialects. *The Condor*, 77(4):385–406, 1975.

[234] PK McGregor. Song dialects in the corn bunting (emberiza calandra). *Zeitschrift für Tierpsychologie*, 54(3):285–297, 1980.

[235] Luis Felipe Baptista. *Song dialects and demes in sedentary populations of the white-crowned sparrow (Zonotrichia leucophrys nuttalli)*, volume 105. University of California Press, 1975.

[236] Klaus Conrads. Studien an fremddialekt-sängern und dialekt-mischsängern des ortolans (emberiza hortulana). *Journal für Ornithologie*, 117(4):438–450, 1976.

[237] Peter K McGregor and DBA Thompson. Constancy and change in local dialects of the corn bunting. *Ornis Scandinavica*, pages 153–159, 1988.

[238] DJ Goodfellow and PJ Slater. A model of bird song dialects. *Animal Behaviour*, 1986.

[239] DL Maney, EA MacDougall-Shackleton, SA MacDougall-Shackleton, GF Ball, and TP Hahn. Immediate early gene response to hearing song correlates with receptive behavior and depends on dialect in a female songbird. *Journal of Comparative Physiology A*, 189(9):667–674, 2003.

[240] Adrian L O'Loghlen and Stephen I Rothstein. Culturally correct song dialects are correlated with male age and female song preferences in wild populations of brown-headed cowbirds. *Behavioral Ecology and Sociobiology*, 36(4):251–259, 1995.

[241] Myron Charles Baker and Michael A Cunningham. The biology of bird-song dialects. *Behavioral and Brain Sciences*, 8(1):85–100, 1985.

[242] PETER K MCGREGOR, VERENA R WALFORD, and DAVID GC HARPER. Song inheritance and mating in a songbird with local dialects. *Bioacoustics*, 1(2-3):107–129, 1988.

[243] Robert M Zink and George F Barrowclough. Allozymes and song dialects: a reassessment. *Evolution*, 38(2):444–448, 1984.

[244] Angelika Poesel, Anthony C Fries, Lisa Miller, H Lisle Gibbs, Jill A Soha, and Douglas A Nelson. High levels of gene flow among song dialect populations of the puget sound white-crowned sparrow. *Ethology*, 123(9):581–592, 2017.

[245] Sara E Lipshutz, Isaac A Overcast, Michael J Hickerson, Robb T Brumfield, and Elizabeth P Derryberry. Behavioural response to song and genetic divergence in two subspecies of white-crowned sparrows (zonotrichia leucophrys). *Molecular ecology*, 26(11):3011–3027, 2017.

[246] Paul Handford and Fernando Nottebohm. Allozymic and morphological variation in population samples of rufous-collared sparrow, zonotrichia capensis, in relation to vocal dialects. *Evolution*, pages 802–817, 1976.

[247] Stephen C Lougheed, Paul Handford, and Allan J Baker. Mitochondrial dna hyperdiversity and vocal dialects in a subspecies transition of the rufous-collared sparrow. *The Condor*, 95(4):889–895, 1993.

[248] Robert B Payne, Laura L Payne, and Susan M Doehlert. Biological and cultural success of song memes in indigo buntings. *Ecology*, 69(1):104–117, 1988.

[249] AL O'Loghlen and MD Beecher. Mate, neighbour and stranger songs: a female song sparrow perspective. *Animal Behaviour*, 58(1):13–20, 1999.

[250] Trevor Price et al. *Speciation in birds*. Roberts and Co., 2008.

[251] Alejandro Salinas-Melgoza and Timothy F Wright. Evidence for vocal learning and limited dispersal as dual mechanisms for dialect maintenance in a parrot. *PloS one*, 7(11):e48667, 2012.

[252] Michael A Cunningham and Myron Charles Baker. Vocal learning in white-crowned sparrows: Sensitive phase and song dialects. *Behavioral Ecology and Sociobiology*, 13(4):259–269, 1983.

[253] Myron Charles Baker and L Richard Mewaldt. Song dialects as barriers to dispersal in white-crowned sparrows, zonotrichia leucophrys nuttalli. *Evolution*, pages 712–722, 1978.

[254] Nicky S Clayton and Ekki Pröve. Song discrimination in female zebra finches and bengalese finches. *Animal Behaviour*, 1989.

[255] Robert B Payne and David F Westneat. A genetic and behavioral analysis of mate choice and song neighborhoods in indigo buntings. *Evolution*, 42(5):935–947, 1988.

[256] Julie E Danner, Raymond M Danner, Frances Bonier, Paul R Martin, Thomas W Small, and Ignacio T Moore. Female, but not male, tropical sparrows respond more strongly to the local song dialect: implications for population divergence. *The American Naturalist*, 178(1):53–63, 2011.

[257] Grace Freed-Brown and David J White. Acoustic mate copying: female cowbirds attend to other females' vocalizations to modify their song preferences. *Proceedings of the Royal Society B: Biological Sciences*, 276(1671):3319–3325, 2009.

[258] Robert E Lemon, DM Weary, and KJ Norris. Male morphology and behavior correlate with reproductive success in the american redstart (setophaga ruticilla). *Behavioral Ecology and Sociobiology*, 29(6):399–403, 1992.

[259] Victoria Reyes-Garcia, Jose Luis Molina, James Broesch, Laura Calvet, Tomas Huanca, Judith Saus, Susan Tanner, William R Leonard, Thomas W McDade, TAPS Bolivian Study Team, et al. Do the aged and knowledgeable men enjoy more prestige? a test of predictions from the prestige-bias model of cultural transmission. *Evolution and Human Behavior*, 29(4):275–281, 2008.

[260] Nicole Creanza, Oren Kolodny, and Marcus W Feldman. Cultural evolutionary theory: How culture evolves and why it matters. *Proceedings of the National Academy of Sciences*, 114(30):7782–7789, 2017.

[261] Joseph Henrich. The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and human behavior*, 30(4):244–260, 2009.

[262] Joseph Henrich and James Broesch. On the nature of cultural transmission networks: evidence from fijian villages for adaptive learning biases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 366(1567):1139–1148, 2011.

[263] Maciej Chudek, Sarah Heller, Susan Birch, and Joseph Henrich. Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior*, 33(1):46–56, 2012.

[264] Joseph Henrich. Cultural transmission and the diffusion of innovations: Adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. *American Anthropologist*, 103(4):992–1013, 2001.

[265] Joseph Henrich and Francisco J Gil-White. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior*, 22(3):165–196, 2001.

[266] Alex Mesoudi. An experimental comparison of human social learning strategies: payoff-biased social learning is adaptive but underused. *Evolution and Human Behavior*, 32(5):334–342, 2011.

[267] Alex Mesoudi and Michael J O'Brien. The cultural transmission of great basin projectile-point technology ii: an agent-based computer simulation. *American Antiquity*, 73(4):627–644, 2008.

[268] Ben D Bell, Marta Borowiec, Jan Lontkowski, and Shirley Pledger. Short records of marsh warbler (acrocephalus palustris) song provide indices that correlate with nesting success. *Journal of Ornithology*, 145(1):8–15, 2004.

[269] Thorsten JS Balsby. Song activity and variability in relation to male quality and female choice in whitethroats sylvia communis. *Journal of avian biology*, 31(1):56–62, 2000.

[270] Carlos A Botero, Andrew E Mudge, Amanda M Koltz, Wesley M Hochachka, and Sandra L Vehrencamp. How reliable are the methods for estimating repertoire size? *Ethology*, 114(12):1227–1238, 2008.

[271] Kathleen R Peshek and Daniel T Blumstein. Can rarefaction be used to estimate song repertoire size in birds? *Current Zoology*, 57(3):300–306, 2011.

[272] Charles Hartshorne. The monotony-threshold in singing birds. *The Auk*, 73(2):176–192, 1956.

[273] Donald E Kroodsma. Continuity and versatility in bird song: support for the monotony–threshold hypothesis. *Nature*, 274(5672):681, 1978.

[274] G Troy Smith, Eliot A Brenowitz, and John C Wingfield. Roles of photoperiod and testosterone in seasonal plasticity of the avian song control system. *Journal of neurobiology*, 32(4):426–442, 1997.

[275] Donna Maney and Raphael Pinaud. Estradiol-dependent modulation of auditory processing and selectivity in songbirds. *Frontiers in neuroendocrinology*, 32(3):287–302, 2011.

[276] Liisa A Tremere, Jin Kwon Jeong, and Raphael Pinaud. Estradiol shapes auditory processing in the adult brain by regulating inhibitory transmission and plasticity-associated gene expression. *Journal of Neuroscience*, 29(18):5949–5963, 2009.

[277] Kiran K Soma, Anthony D Tramontin, Joy Featherstone, and Eliot A Brenowitz. Estrogen contributes to seasonal plasticity of the adult avian song control system. *Journal of neurobiology*, 58(3):413–422, 2004.

[278] Iñigo Azcoitia, Maria Angeles Arevalo, and Luis M Garcia-Segura. Neural-derived estradiol regulates brain plasticity. *Journal of chemical neuroanatomy*, 89:53–59, 2018.

[279] Wolfgang Forstmeier and Thorsten JS Balsby. Why mated dusky warblers sing so much: territory guarding and male quality announcement. *Behaviour*, 139(1):89–111, 2002.

[280] Martine Hausberger, M-A Richard-Yris, Laurence Henry, L Lepage, and I Schmidt. Song sharing reflects the social organization in a captive group of european starlings (sturnus vulgaris). *Journal of Comparative Psychology*, 109(3):222, 1995.

[281] Sarah Kiefer, Christina Sommer, Constance Scharff, and Silke Kipper. Singing the popular songs? nightingales share more song types with their breeding population in their second season than in their first. *Ethology*, 116(7):619–626, 2010.

[282] L.V. Hedges. How hard is hard science, how soft is soft science? the empirical cumulativeness of research. *American Psychologist*, 42(5):443, 1987.

[283] A.A. Veroniki, D. Jackson, W. Viechtbauer, R. Bender, J. Bowden, G. Knapp, O. Kuss, J.P.T. Higgins, D. Langan, and G. Salanti. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods*, 7(1):55–79, 2016.

[284] Cornell University Cornell Lab of Ornithology. *Macaulay Library: archive of animal sounds and video*. Cornell Laboratory of Ornithology: Ithaca, NY, 2009. https://www.macaulaylibrary.org/.

APPENDIX A

Bayesian multi-level phylogenetic meta-analysis

A.1 Meta-analysis field study references

[1] D. Hasselquist, Polygyny in great reed warblers: A long-term study of factors contributing to male fitness, Ecology, vol. 79, no. 7, pp. 23762390, 1998.

[2] C. K. Catchpole, Song repertoires and reproductive success in the great reed warbler *Acrocephalus arundinaceus*, Behav. Ecol. Sociobiol., vol. 19, pp. 439445, 1986.

[3] W. Forstmeier and B. Leisler, Repertoire size, sexual selection, and offspring viability in the great reed warbler: Changing patterns in space and time, Behav. Ecol., vol. 15, no. 4, pp. 555563, 2004.

[4] B. Leisler, J. Beier, G. Heine, and K.-H. Siebenrock, Age and other factors influencing mating status in German great reed warblers (*Acrocephalus arundinaceus*), Japanese J. Ornithol., vol. 44, no. 3, pp. 215216, 1995.

[5] D. Hasselquist, S. Bensch, and T. von Schantz, Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler, Nature, vol. 381, no. 6579, pp. 229232, 1996.

[6] S. Hamao and H. Eda-Fujiwara, Vocal mimicry by the black-browed reed warbler *Acrocephalus bistrigiceps*: Objective identification of mimetic sounds, Ibis (Lond. 1859)., vol. 146, no. 1, pp. 6168, 2004.

[7] B. D. Bell, M. Borowiec, J. Lontkowski, and S. Pledger, Short records of marsh warbler (*Acrocephalus palustris*) song provide indices that correlate with nesting success, J. Ornithol., vol. 145, no. 1, pp. 815, 2004.

[8] A. Darolov, J. Kritofk, H. Hoi, and M. Wink, Song complexity in male marsh warblers: Does it reflect male quality?, J. Ornithol., vol. 153, no. 2, pp. 431439, 2012.

[9] K. L. Buchanan and C. K. Catchpole, Female choice in the sedge warbler *Acrocephalus schoenobaenus*: multiple cues from song and territory quality, Proc. R. Soc. B Biol. Sci., vol. 264, no. 1381, pp. 521526, 1997.

[10] C. K. Catchpole, Sexual selection and the evolution of complex songs among European warblers of the genus *Acrocephalus*, Behaviour, 1980.

[11] K. L. Buchanan and C. K. Catchpole, Extra-pair paternity in the socially monogamous sedge warbler *Acrocephalus schoenobaenus* as revealed by multilocus DNA fingerprinting, Ibis (Lond. 1859)., vol. 142, no. 1, pp. 1220, 2000.

[12] R. C. Marshall, K. L. Buchanan, and C. K. Catchpole, Song and female choice for extrapair copulations in the sedge warbler, *Acrocephalus schoenobaenus*, Anim. Behav., vol. 73, no. 4, pp. 629635, 2007.

[13] K. Yasukawa, J. L. Blank, and C. B. Patterson, Song repertoires and sexual selection in the red-winged blackbird, Behav. Ecol. Sociobiol., vol. 7, no. 3, pp. 233238, 1980.

[14] U. Rehsteiner, H. Geisser, and H. Reyer, Singing and mating success in water pipits: one specific song element makes all the difference, Anim. Behav., vol. 55, no. 6, pp. 147181, 1998.

[15] R. N. Conner, M. E. Anderson, and J. G. Dickson, Relationships among territory size, habitat, song, and nesting success of Northern cardinals, Auk, vol. 103, no. January, pp. 2331, 1986.

[16] D. J. Mennill, A. V. Badyaev, L. M. Jonart, and G. E. Hill, Male house finches with elaborate songs have higher reproductive performance, Ethology, vol. 112, no. 2, pp. 174180, 2006.

[17] B. E. Byers, Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance, Behav. Ecol., vol. 18, no. 1, pp. 130136, 2007.

[18] K. M. Bouwman, R. E. van Dijk, J. J. Wijmenga, and J. Komdeur, Older male reed buntings are more successful at gaining extrapair fertilizations, Anim. Behav., vol. 73, no. 1, pp. 1527, 2007.

[19] G. Hegyi, E. Szllosi, S. Jenni-Eiermann, J. Trk, M. Eens, and L. Z. Garamszegi, Nutritional correlates and mate acquisition role of multiple sexual traits in male collared flycatchers, Naturwissenschaften, vol. 97, no. 6, pp. 567576, 2010.

[20] L. Z. Garamszegi and A. P. Mller, Extrapair paternity and the evolution of bird song, Behav. Ecol., vol. 15, no. 3, pp. 508519, 2004.

[21] H. M. Lampe and Y. O. Espmark, Mate choice in pied flycatchers *Ficedula hypoleuca*: can females use song to find high-quality males and territories?, Ibis (Lond. 1859)., vol. 145, no. 1, pp. E24E33, 2003.

[22] L. Z. Garamszegi et al., The design of complex sexual traits in male barn swallows: Associations between signal attributes, J. Evol. Biol., vol. 19, no. 6, pp. 20522066, 2006.

[23] C. Landgraf, K. Wilhelm, J. Wirth, M. Weiss, and S. Kipper, Affairs happen-To whom? A study on extrapair paternity in common nightingales, Curr. Zool., vol. 63, no. 4, pp. 421431, 2017.

[24] J. M. Reid et al., Song repertoire size predicts initial mating success in male song sparrows, *Melospiza melodia*, Anim. Behav., vol. 68, no. 5, pp. 10551063, 2004.

[25] W. A. Searcy, Song repertore size and female preferences in song sparrows, Behav. Ecol. Sociobiol., vol. 14, pp. 281286, 1984.

[26] S. M. Hiebert, P. K. Stoddard, and P. Arcese, Repertoire size, territory acquisition, and reproductive success in the song sparrow, Anim. Behav., vol. 37, no. PART 2, pp. 266273, 1989.

[27] J. M. Reid et al., Fitness correlates of song repertoire size in free-living song sparrows (*Melospiza melodia*), Am. Nat., vol. 165, no. 3, pp. 299310, 2005.

[28] C. E. Hill, . Akay, S. E. Campbell, and M. D. Beecher, Extrapair paternity, song, and genetic quality in song sparrows, Behav. Ecol., vol. 22, no. 1, pp. 7381, 2011.

[29] D. A. Potvin, P. W. Crawford, S. A. MacDougall-Shackleton, and E. A. MacDougall-Shackleton, Song repertoire size, not territory location, predicts reproductive success and territory tenure in a migratory songbird, Can. J. Zool., vol. 93, no. 8, pp. 627633, 2015.

[30] R. D. Howard, The influence of sexual selection and interspecific competition on mockingbird song (*Mimus polyglottus*), Evolution (N. Y)., vol. 28, no. 3, pp. 428438, 1974.

[31] A. Poesel, K. Foerster, and B. Kempenaers, The dawn song of the blue tit *Parus caeruleus* and its role in sexual selection, Ethology, vol. 107, no. 6, pp. 521531, 2001.

[32] J. Krebs, R. Ashcroft, and M. Webber, Song repertoires and territory defence in the great tit, Nature, vol. 271, no. 5645, pp. 539542, 1978.

[33] M. Lambrechts and A. A. Dhondt, Male quality, reproduction, and survival in the great tit (*Parus major*), Behav. Ecol. Sociobiol., vol. 19, no. 1, pp. 5763, 1986.

[34] C. McGregor, P., Krebs, J., Perrins, Song repertoires and lifetime reproductive success in the great tit (*Parus major*), Am. Soc. Nat., vol. 118, no. 2, pp. 149159, 1981.

[35] W. Forstmeier, B. Kempenaers, A. Meyer, and B. Leisler, A novel song parameter correlates with extra-pair paternity and reflects male longevity, Proc. R. Soc. B Biol. Sci., vol. 269, no. 1499, pp. 14791485, 2002.

[36] E. S. C. Scordato, Geographical and temporal variation in sexually selected traits: Environmental variation, multiple signals, and consequences for population divergence, The University of Chicago, 2012.

[37] D. Gil, P. J. B. Slater, and J. A. Graves, Extra-pair paternity and song characteristics in the willow warbler *Phylloscopus trochilus*, J. Avian Biol., vol. 38, no. 3, pp. 291297, 2007.

[38] D. Gil and P. J. B. Slater, Multiple song repertoire characteristics in the willow warbler (*Phylloscopus trochilus*): Correlations with female choice and offspring viability, Behav. Ecol. Sociobiol., vol. 47, no. 5, pp. 319326, 2000.

[39] T. Järvi, The evolution of song versatility in the willow warbler *Phylloscopus trochilus*: A case of evolution by intersexual selection explained by the 'female's choice of the best mate,' Ornis Scand., 1983.

[40] E. Hofstad, Y. Espmark, A. Moksnes, T. Haugan, and M. Ingebrigtsen, The relationship between song performance and male quality in snow buntings (*Plectrophenax nivalis*), Can. J. Zool., vol. 80, no. 3, pp. 524531, 2002.

[41] N. Dadwal and D. Bhatt, Relationship between song repertoire size with nesting success, territory size, and territorial conflict in pied bush chat (*Saxicola caprata*), Wilson J. Ornithol., vol. 129, no. 4, pp. 701712, 2017.

[42] R. E. Lemon, D. M. Weary, and K. J. Norris, Male morphology and behavior correlate with reproductive success in the American redstart (*Setophaga ruticilla*), Behav. Ecol. Sociobiol., pp. 399403, 1992.

[43] a. G. Horn, T. E. Dickinson, and J. B. Falls, Male quality and song repertoires in Western meadowlarks (*Sturnella-neglecta*), Can. J. Zool. Can. Zool., vol. 71, pp. 10591061, 1993.

[44] M. K. Aweida, Repertoires, territory size and mate attraction in Western meadowlarks, Condor, vol. 97, no. 4, pp. 10801083, 1995.

[45] D. J. Mountjoy and R. E. Lemon, Female choice for complex song in the European starling: a field experiment, Behav. Ecol. Sociobiol., vol. 38, no. 1, pp. 6571, 1996.

[46] M. Eens, R. Pinxten, and R. F. Verheyen, Male song as a cue for mate choice in the European starling, Behaviour, vol. 116, no. 34, pp. 210238, 1991.

[47] T. J. S. Balsby, Song activity and variability in relation to male quality and female choice in whitethroats *Sylvia communis*, J. Avian Biol., vol. 31, pp. 5662, 2000.

[48] A. Demko, Temporal and individual song variation in the canada warbler (*Cardellina canadensis*), no. March, pp. 1114, 2012.

[49] S. Nowicki, D. Hasselquist, S. Bensch, and S. Peters, Nestling growth and song repertoire size in great reed warblers: evidence for song learning as an indicator mechanism in mate choice., Proc. Biol. Sci., vol. 267, no. 1460, pp. 241924, 2000.

[50] F. Dowsett-Lemaire, The imitative range of the song of the marsh warbler *Acrocephalus palustris*, with special reference to imitations of african birds, Ibis (Lond. 1859)., vol. 121, no. 4, pp. 453468, 1979.

[51] J. S. Nicholson, K. L. Buchanan, R. C. Marshall, and C. K. Catchpole, Song sharing and repertoire size in the sedge warbler, *Acrocephalus schoenobaenus*: Changes within and between years, Anim. Behav., 2007.

[52] B. Y. P. Marler, P. Mundinger, M. S. U. E. Waser, and A. N. N. Lutjen, Song development in red-winged blackbirds (*Agelaius phoeniceus*), Anim. Behav., pp. 586606, 1972.

[53] R. E. Lemon, Geographic variation in the song of cardinals, Can. J. Zool., 1966.

[54] B. E. Byers, Geographic variation of song form within and among chestnut-sided warbler populations, Auk, vol. 113, no. 2, pp. 288299, 1996.

[55] M. Bessert-Nettelbeck, S. Kipper, C. Bartsch, and S. L. Voigt-Heucke, Similar, yet different: Male reed buntings (*Emberiza schoeniclus*) show high individual differences in song composition, rates of syllable sharing and use, J. Ornithol., 2014.

[56] L. Z. Garamszegi, J. Trk, G. Hegyi, E. Szllsi, B. Rosivall, and M. Eens, Age-dependent expression of song in the collared flycatcher, *Ficedula albicollis*, Ethology, vol. 113, no. 3, pp. 246256, 2007.

[57] A. Eriksen, T. Slagsvold, and H. M. Lampe, Vocal plasticity - are pied flycatchers, *Ficedula Hypoleuca*, open-ended learners?, Ethology, 2011.

[58] L. Z. Garamszegi, D. Heylen, A. P. Mller, M. Eens, and F. De Lope, Age-dependent health status and song characteristics in the barn swallow, Behav. Ecol., 2005.

[59] P. Galeotti, N. Saino, E. Perani, R. Sacchi, and A. R. Mller, Age-related song variation in male barn swallows, Ital. J. Zool., vol. 68, no. 4, pp. 305310, 2001.

[60] S. Kiefer et al., First-year common nightingales (*Luscinia megarhynchos*) have smaller song-type repertoire sizes than older males, Ethology, vol. 112, no. 12, pp. 12171224, 2006.

[61] J. C. Nordby, S. E. Campbell, and M. D. Beecher, Adult song sparrows do not alter their song repertoires, Ethology, 2002.

[62] A. Poesel, H. P. Kunc, K. Foerster, A. Johnsen, and B. Kempenaers, Early birds are sexy: male age, dawn song and extrapair paternity in blue tits, *Cyanistes* (formerly *Parus*) *caeruleus*, Anim. Behav., 2006.

[63] H. F. Rivera-Gutierrez, R. Pinxten, and M. Eens, Difficulties when assessing birdsong learning programmes under field conditions: A re-evaluation of song repertoire flexibility in the great tit, PLoS One, 2011.

[64] W. Forstmeier and T. J. S. Balsby, Why mated dusky warblers sing so much: Territory guarding and male quality announcement, Behaviour, 2002.

[65] D. Gil, J. L. S. Cobb, and P. J. B. Slater, Song characteristics are age dependent in the willow warbler, *Phylloscopus trochilus*, Anim. Behav., 2001.

[66] A. Horn, Structure of western meadowlark (*Sturnella neglecta*) song repertoires, Can. J. Zool., vol. 66, no. 2, pp. 284288, 1988.

[67] D. J. Mountjoy and R. E. Lemon, Extended song learning in wild European starlings, Anim. Behav., vol. 49, no. 2, pp. 357366, 1995.

[68] T. J. S. Balsby and P. Hansen, Element repertoire: Change and development with age in whitethroat *Sylvia communis* song, J. Ornithol., 2010.

[69] J. M. Moore, T. Szekely, J. Buki, and T. J. DeVoogd, Motor pathway convergence predicts syllable repertoire size in oscine birds, Proc. Natl. Acad. Sci., 2011.

[70] W. Forstmeier, D. Hasselquist, S. Bensch, and B. Leisler, Does song reflect age and viability? A comparison between two populations of the great reed warbler *Acrocephalus arundinaceus*, Behav. Ecol. Sociobiol., 2006.

[71] T. Szekely, C. K. Catchpole, A. Devoogd, Z. Marchl, and T. J. Devoogd, Evolutionary changes in a song control area of the brain (HVC) are associated with evolutionary changes in song repertoire among European warblers (*Sylviidae*)., Proceedings of the Royal Society B: Biological Sciences. 1996.

[72] S. Hamao, Syntactical complexity of songs in the black-browed reed warbler *Acrocephalus bistrigiceps*, Ornithol. Sci., vol. 7, no. 2, pp. 173177, 2008.

[73] L. Z. Garamszegi, Bird song and parasites, Behav. Ecol. Sociobiol., vol. 59, no. 2, pp. 167180, 2005.

[74] R. E. Irwin, Directional sexual selection cannot explain variation in song repertoire size in the new world blackbirds (*Icterinae*), Ethology, 1990.

[75] M. Soma and L. Z. Garamszegi, Rethinking birdsong evolution: Meta-analysis of the relationship between song complexity and reproductive success, Behav. Ecol., vol. 22, no. 2, pp. 363371, 2011.

[76] M. E. Anderson and R. N. Conner, Northern cardinal song in three forest habitats in Eastern Texas., Wilson Bull., 1985.

[77] T. T. Tracy and M. C. Baker, Geographic variation of syllables of house finch songs, Auk, vol. 116, no. 3, pp. 666676, 1999.

[78] L. Z. Garamszegi et al., Estimating the complexity of bird song by using capture-recapture approaches from community ecology, Behavioral Ecology and Sociobiology. 2005.

[79] L. Z. Garamszegi, T. Boulinier, A. P. Mller, J. Trk, G. Michl, and J. D. Nichols, The estimation of size and change in composition of avian song repertoires, Anim. Behav., 2002.

[80] Y. O. Espmark and H. M. Lampe, Variations in the song of the pied flycatcher within and between breeding seasons, Bioacoustics, 1993.

[81] H. M. Lampe and Y. O. Espmark, Singing activity and song pattern of the redwing *Turdus iliacus* during the breeding season, Ornis Scand., 1987.

[82] P. Galeotti, N. Saino, R. Sacchi, and A. P. Moller, Song correlates with social context, testosterone and body condition in male barn swallows, Anim. Behav., 1997.

[83] D. J. Borror, Song variation in Maine song-sparrows, Wilson Bull., vol. 77, no. 1, pp. 537, 1965.

[84] J. L. Wildenthal, Structure in primary song of the mockingbird (*Mimus polyglottos*), Auk, 1965.

[85] W. Forstmeier, Individual reproductive strategies in the dusky warbler (*Phylloscopus fuscatus*): Female and male perspectives, Thesis, p. 112, 2001.

[86] T. Jrvi, T. Radester, and S. Jakobsson, The song of the willow warbler *Phylloscopus trochilus* with special reference to singing behaviour in agonistic situations, Ornis Scand., 1980.

[87] M. Eens, Understanding the complex song of the European starling: an integrated ethological approach, Adv. Study Behav., 1997.

[88] T. J. S. Balsby and T. Dabelsteen, The meaning of song repertoire size and song length to male whitethroats *Sylvia communis*, Behav. Processes, 2001.

[89] S. A. MacDougall-Shackleton, Sexual Selection and the Evolution of Song Repertoires, in Current Ornithology, V. Nolan, E. D. Ketterson, and C. F. Thompson, Eds. Boston; MA: Springer US, 1997, pp. 81124.

[90] T. J. Devoogd, J. R. Krebs, S. D. Healy, and A. Purvis, Relations between song repertoire size and the volume of brain nuclei related to song: comparative evolutionary analyses amongst oscine birds., Proc. Biol. Sci., 1993.

[91] C. W. Dobson and R. E. Lemon, Re-examination of monotony threshold hypothesis in bird song, Nature, 1975.

[92] C. K. Catchpole and P. K. McGregor, Sexual selection, song complexity and plumage dimorphism in European buntings of the genus *Emberiza*, Anim. Behav., 1985.

[93] E. del Hoyo; J., Elliott, A.; Sargatal, J.; Christie, D.A.; de Juana, Ed., Handbook of the Birds of the World Alive. Barcelona: Lynx Edicions, 2017.

A.2 Syllable repertoire as a discrete variable

Species average syllable repertoire size should be cautiously treated as a continuous variable, because each researcher could define syllables slightly differently. In addition, there is a precedent for conducting this type of meta-analysis by separating birds into discrete groups based on song traits. In their analysis, Soma and Garamszegi[19] separated species into either 5 bins based on species average song repertoire size, or 3 bins based on species average unique syllables per song, to see if these measures of average species aong elaboration could predict which species show a correlation between individual song elaboration and reproductive success. Binning by species average song repertoire size did lead to significant differences, but this was largely discounted, because species average song repertoire size did not account for a significant proportion of the variability in their dataset. Species average syllables per song did not lead to significant results or account for a significant amount of variability. Therefore, we retested this hypothesis by dividing species in the species average syllable repertoire size dataset into 2 bins, “smaller” and “larger.” There was no clear value at which to divide species into larger and smaller syllable repertoire sizes, so we tested all species average syllable repertoire sizes as the threshold between the bins. We found the posterior mean for the larger species average repertoire bin was significantly separated from 0 for the majority of the repertoire size thresholds, predicting weak correlations when the threshold value was small (e.g. threshold$\geq$18.5, z=0.333) and strong correlations when the threshold was large (threshold$\geq$216, z=0.611) (**Figures A.2.1** and **2.6** and **Table A.2.1**). The smaller species average syllable repertoire bin was never significantly separated from 0 (**Figures A.2.1** and **2.6** and **Table A.2.1**). BEST analysis confirmed it was highly likely that there was a real difference between the larger and smaller bins for all thresholds between 10.75 and 216 syllables and that this difference was substantial (differences in z ranging from 0.238 to 0.376 depending on which significantly different threshold was examined) **Table A.2.2**. Three species average syllable repertoire size thresholds were of particular interest: 1) at 18.5 syllables the meta-analytic mean for the smaller bin was the most negative, 2) 38 syllables led to the model with the most significant evidence of a strong correlation in the larger repertoire bin, and 3) 216 syllables led to the greatest difference between the meta-analytic means of the smaller and larger species average syllable repertoire bins. We used only these three thresholds for the remaining analyses in this section. These results were not significantly affected by using the song stability dataset (**Table A.2.3**), jackknife removal of each individual species in turn (**Tables A.2.4** to **A.2.6**), or using territory-controlled measurements (**Tables A.2.7** and **A.2.8**).

Table A.2.1: Relatively larger syllable repertoires are predictive of a correlation between individual song elaboration and reproductive success for most tested thresholds. There was significant evidence for this correlation in the larger repertoire group when the threshold was 18.15 to 216. There was not significant evidence for this correlation in the smaller repertoire group for any threshold. Performed in the syllable repertoire dataset. Asterisks (*) denote significant groups.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<7.46	Smaller	2	4	0.161	[−0.378;0.695]	0.538
≥7.46	Larger	23	82	0.249	[−0.025;0.535]	0.067
<7.9	Smaller	3	6	0.054	[−0.382;0.526]	0.812
≥7.9	Larger	22	80	0.263	[0.007;0.551]	0.05
<10.75	Smaller	4	11	0.03	[−0.369;0.436]	0.88
≥10.75	Larger	21	75	0.284	[0.011;0.569]	0.042*
<14	Smaller	5	13	0.116	[−0.263;0.489]	0.515
≥14	Larger	20	73	0.271	[03.00;0.543]	0.047*
<17.43	Smaller	6	14	0.042	[−0.297;0.388]	0.804
≥17.43	Larger	19	72	0.294	[0.045;0.554]	0.029*
<18.15	Smaller	7	16	0.016	[−0.313;0.337]	0.914
≥18.15	Larger	18	70	0.311	[0.051;0.567]	0.021*
<18.5	Smaller	8	17	−0.034	[−0.338;0.267]	0.82
≥18.5	Larger	17	69	0.333	[0.093;0.574]	0.01*
<20.1	Smaller	9	19	−0.027	[−0.317;0.264]	0.857
≥20.1	Larger	16	67	0.344	[0.115;0.597]	0.009*
<22.5	Smaller	10	21	−0.01	[−0.268;0.258]	0.946
≥22.5	Larger	15	65	0.358	[0.134;0.594]	0.005*
<25.3	Smaller	11	34	0.053	[−0.215;0.309]	0.652
≥25.3	Larger	14	52	0.388	[0.121;0.629]	0.005*
<28.4	Smaller	12	35	0.056	[−0.199;0.316]	0.627
≥28.4	Larger	13	51	0.393	[0.152;0.651]	0.003*
<31.4	Smaller	13	36	0.085	[−0.164;0.329]	0.454
≥31.4	Larger	12	50	0.393	[0.143;0.632]	0.005*
<35.1	Smaller	14	38	0.071	[−0.146;0.3]	0.484
≥35.1	Larger	11	48	0.453	[0.225;0.688]	0.002*
<38	Smaller	15	42	0.098	[−0.121;0.324]	0.336
≥38	Larger	10	44	0.482	[0.244;0.732]	0.001*
<41.3	Smaller	16	56	0.124	[−0.125;0.369]	0.276
≥41.3	Larger	9	30	0.491	[0.204;0.775]	0.002*
<41.95	Smaller	17	58	0.133	[−0.118;0.373]	0.244
≥41.95	Larger	8	28	0.501	[0.208;0.802]	0.002*
<55	Smaller	18	64	0.165	[−0.097;0.434]	0.178
≥55	Larger	7	22	0.486	[0.155;0.824]	0.006*
<86	Smaller	20	71	0.175	[−0.072;0.44]	0.147
≥86	Larger	5	15	0.553	[0.197;0.928]	0.005*
<216	Smaller	21	78	0.185	[−0.068;0.459]	0.141
≥216	Larger	4	8	0.611	[0.194;1.051]	0.006*
<241	Smaller	22	79	0.206	[−0.072;0.478]	0.116
≥241	Larger	3	7	0.536	[0.068;1.049]	0.029*
<367.5	Smaller	23	84	0.22	[−0.049;0.497]	0.093
≥367.5	Larger	2	2	0.722	[0.014;1.44]	0.046*

Table A.2.2: Larger and smaller syllable repertoire size groups have significantly different correlations between individual song elaboration and reproductive success for most thresholds. BEST analysis predicts that the two groups had significantly different means for the thresholds 10.75 to 86. Performed in the syllable repertoire dataset. Asterisks (*) denote models with significantly different groups.

Threshold	BEST Mean	95% CredInt	%<0
7.46	0.125	[−0.337;0.587]	22.7
7.90	0.257	[−0.093;0.607]	6.6
10.75	0.26	[0.086;0.434]	0.3*
14	0.272	[0.117;0.427]	0.1*
17.43	0.307	[0.15;0.465]	0.1*
18.15	0.315	[0.169;0.461]	0.1*
18.50	0.358	[0.198;0.519]	0.1*
20.10	0.36	[0.201;0.518]	0.1*
22.50	0.373	[0.225;0.522]	0.1*
25.30	0.314	[0.165;0.463]	0.1*
28.40	0.337	[0.19;0.485]	0.1*
31.40	0.329	[0.179;0.479]	0.1*
35.10	0.376	[0.232;0.521]	0.1*
38	0.365	[0.216;0.515]	0.1*
41.30	0.324	[0.142;0.506]	0.1*
41.95	0.331	[0.138;0.525]	0.1*
55	0.238	[0.041;0.435]	1.3*
86	0.294	[0.144;0.444]	0.1*
216	0.246	[−0.028;0.52]	3.6*
241	0.189	[−0.085;0.462]	7.9
367.50	9.115	[−44.729;62.958]	32.6

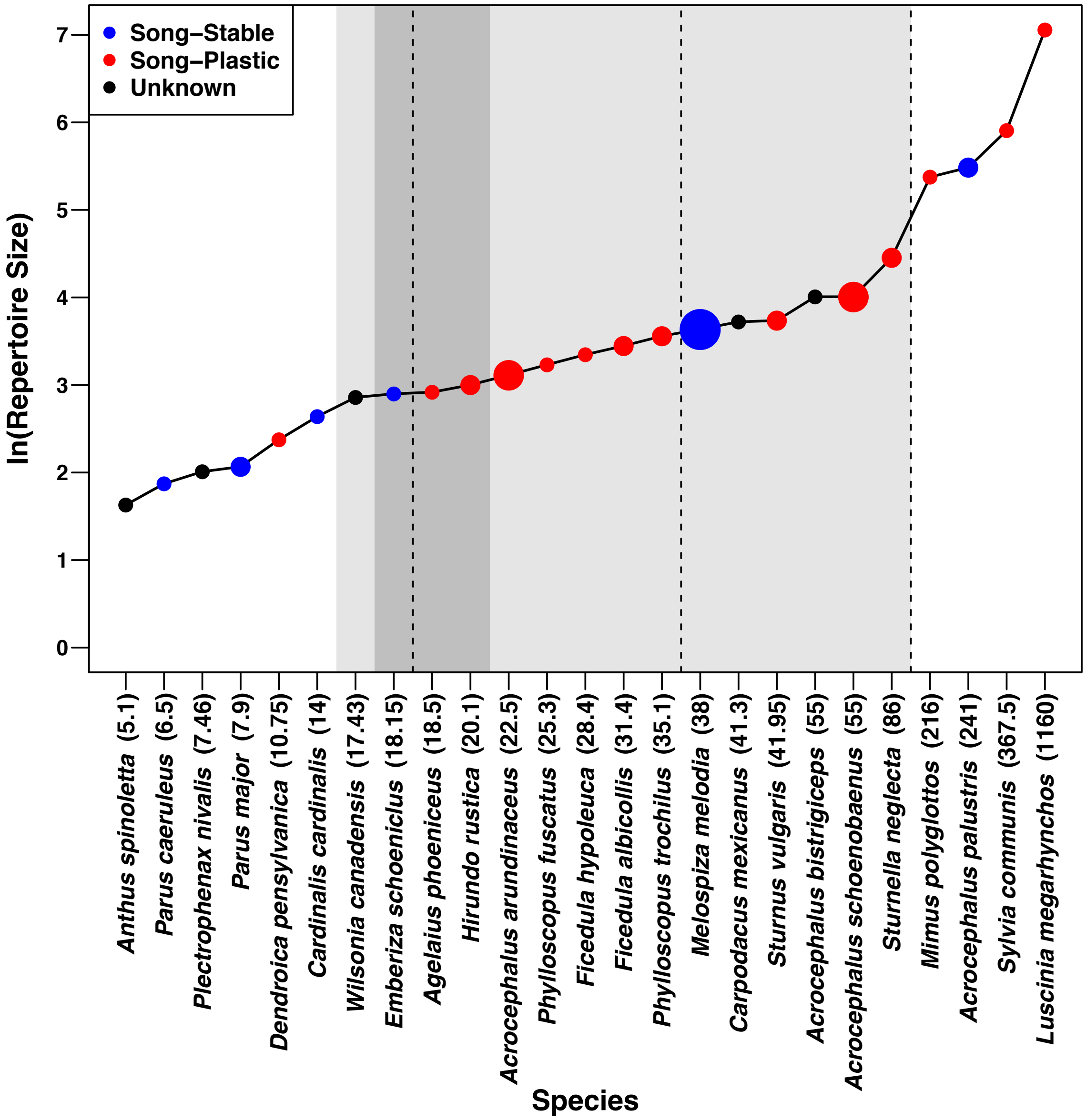


Figure A.2.1: Distribution of species average repertoire sizes. Species plotted in order of increasing average syllable repertoire size. Species average syllable repertoire size (without the log transformation) is included next to species name in parentheses. Red circles denote song-plastic species, and blue circles denote song-stable species. Black circles denote species for which no song stability information was available. Circle size increases as more studies were included for a single species. We tested each species average syllable repertoire size between 7.46 and 367.5 as a threshold between smaller and larger species average syllable repertoires in the Bayesian meta-analyses. The darker grey region denotes species average syllable repertoire thresholds for which the meta-analytic mean for the smaller group was less than 0.05, while the lighter grey region denotes species average syllable repertoire thresholds for which p_{MCMC}<0.025 for the larger species average syllable repertoire group. Dashed lines show the thresholds used for all subsequent species average syllable repertoire size analyses ($\geq$18.5, 38, and 216).

Table A.2.3: Relatively larger syllable repertoires are predictive of a correlation between individual song elaboration and reproductive success for all tested thresholds. There was not significant evidence for this correlation in the smaller repertoire group for any tested threshold. Performed in the song stability dataset. Asterisks (*) denote significant groups.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<18.5	Smaller	5	11	-0.016	$[-0.375;0.343]$	0.937
≥ 18.5	Larger	15	66	0.336	[0.099;0.579]	0.009*
<38	Smaller	12	36	0.111	$[-0.114;0.352]$	0.31
≥ 38	Larger	8	41	0.486	[0.228;0.738]	0.002*
<216	Smaller	16	69	0.216	$[-0.041;0.47]$	0.091
≥ 216	Larger	4	8	0.637	[0.205;1.105]	0.006*

Table A.2.4: Removal of individual species does not significantly affect the syllable repertoire results. Performed in the syllable repertoire dataset. Threshold ≥ 18.5. Asterisks (*) denote significant groups.

Removed	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Acrocephalus arundinaceus	Smaller	8	17	−0.037	[−0.364;0.27]	0.813
Acrocephalus arundinaceus	Larger	16	56	0.353	[0.093;0.608]	0.013*
Acrocephalus bistrigiceps	Smaller	8	17	−0.038	[−0.345;0.257]	0.784
Acrocephalus bistrigiceps	Larger	16	68	0.33	[0.101;0.547]	0.007*
Acrocephalus palustris	Smaller	8	17	−0.045	[−0.359;0.241]	0.762
Acrocephalus palustris	Larger	16	64	0.32	[0.092;0.555]	0.01*
Acrocephalus schoenobaenus	Smaller	8	17	−0.055	[−0.363;0.219]	0.697
Acrocephalus schoenobaenus	Larger	16	63	0.335	[0.122;0.543]	0.005*
Agelaius phoeniceus	Smaller	8	17	−0.037	[−0.32;0.25]	0.798
Agelaius phoeniceus	Larger	16	67	0.341	[0.107;0.563]	0.008*
Anthus spinoletta	Smaller	7	15	−0.048	[−0.354;0.277]	0.757
Anthus spinoletta	Larger	17	69	0.333	[0.097;0.576]	0.012*
Cardinalis cardinalis	Smaller	7	16	0.001	[−0.28;0.289]	0.988
Cardinalis cardinalis	Larger	17	69	0.331	[0.105;0.561]	0.01*
Carpodacus mexicanus	Smaller	8	17	−0.034	[−0.327;0.258]	0.81
Carpodacus mexicanus	Larger	16	67	0.334	[0.088;0.568]	0.012*
Dendroica pensylvanica	Smaller	7	15	−0.098	[−0.407;0.208]	0.505
Dendroica pensylvanica	Larger	17	69	0.336	[0.087;0.589]	0.015*
Emberiza schoeniclus	Smaller	7	16	0.008	[−0.263;0.29]	0.931
Emberiza schoeniclus	Larger	17	69	0.329	[0.122;0.533]	0.008*
Ficedula albicollis	Smaller	8	17	−0.011	[−0.287;0.259]	0.93
Ficedula albicollis	Larger	16	67	0.37	[0.138;0.591]	0.005*
Ficedula hypoleuca	Smaller	8	17	−0.049	[−0.368;0.246]	0.761
Ficedula hypoleuca	Larger	16	68	0.315	[0.071;0.555]	0.023*
Hirundo rustica	Smaller	8	17	−0.032	[−0.327;0.259]	0.827
Hirundo rustica	Larger	16	67	0.35	[0.124;0.591]	0.008*
Luscinia megarhynchos	Smaller	8	17	−0.037	[−0.358;0.252]	0.807
Luscinia megarhynchos	Larger	16	68	0.32	[0.084;0.567]	0.017*
Melospiza melodia	Smaller	8	17	−0.026	[−0.333;0.258]	0.863
Melospiza melodia	Larger	16	55	0.317	[0.069;0.56]	0.017*
Mimus polyglottos	Smaller	8	17	−0.035	[−0.328;0.247]	0.807
Mimus polyglottos	Larger	16	68	0.318	[0.091;0.548]	0.012*
Parus caeruleus	Smaller	7	15	−0.058	[−0.36;0.234]	0.678
Parus caeruleus	Larger	17	69	0.334	[0.105;0.575]	0.012*
Parus major	Smaller	7	12	−0.037	[−0.361;0.276]	0.8
Parus major	Larger	17	69	0.339	[0.098;0.58]	0.011*
Phylloscopus fuscatus	Smaller	8	17	−0.033	[−0.336;0.255]	0.821
Phylloscopus fuscatus	Larger	16	68	0.336	[0.115;0.577]	0.011*
Phylloscopus trochilus	Smaller	8	17	−0.025	[−0.352;0.267]	0.879
Phylloscopus trochilus	Larger	16	65	0.328	[0.085;0.576]	0.015*
Plectrophenax nivalis	Smaller	7	15	−0.016	[−0.321;0.286]	0.919
Plectrophenax nivalis	Larger	17	69	0.332	[0.097;0.565]	0.011*
Sturnella neglecta	Smaller	8	17	−0.039	[−0.343;0.246]	0.793
Sturnella neglecta	Larger	16	62	0.326	[0.096;0.576]	0.013*
Sturnus vulgaris	Smaller	8	17	−0.034	[−0.315;0.246]	0.815
Sturnus vulgaris	Larger	16	63	0.316	[0.092;0.544]	0.011*
Sylvia communis	Smaller	8	17	−0.036	[−0.328;0.263]	0.805
Sylvia communis	Larger	16	68	0.326	[0.096;0.557]	0.011*
Wilsonia canadensis	Smaller	7	15	−0.024	[−0.308;0.273]	0.864
Wilsonia canadensis	Larger	17	69	0.338	[0.117;0.574]	0.01*

Table A.2.5: Removal of individual species does not significantly affect the syllable repertoire results. Performed in the syllable repertoire dataset. Threshold ≥ 38. Asterisks (*) denote significant groups.

Removed	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Acrocephalus arundinaceus	Smaller	14	29	0.046	[−0.187;0.272]	0.651
Acrocephalus arundinaceus	Larger	10	44	0.468	[0.242;0.718]	0.002*
Acrocephalus bistrigiceps	Smaller	15	42	0.097	[−0.108;0.317]	0.327
Acrocephalus bistrigiceps	Larger	9	43	0.49	[0.252;0.723]	0.001*
Acrocephalus palustris	Smaller	15	42	0.093	[−0.118;0.312]	0.352
Acrocephalus palustris	Larger	9	39	0.487	[0.239;0.743]	0.001*
Acrocephalus schoenobaenus	Smaller	15	42	0.102	[−0.089;0.291]	0.256
Acrocephalus schoenobaenus	Larger	9	38	0.531	[0.307;0.759]	0*
Agelaius phoeniceus	Smaller	14	40	0.078	[−0.136;0.289]	0.424
Agelaius phoeniceus	Larger	10	44	0.454	[0.224;0.682]	0.001*
Anthus spinoletta	Smaller	14	40	0.094	[−0.138;0.319]	0.366
Anthus spinoletta	Larger	10	44	0.474	[0.233;0.722]	0.002*
Cardinalis cardinalis	Smaller	14	41	0.101	[−0.108;0.31]	0.295
Cardinalis cardinalis	Larger	10	44	0.465	[0.239;0.699]	0.001*
Carpodacus mexicanus	Smaller	15	42	0.088	[−0.129;0.311]	0.37
Carpodacus mexicanus	Larger	9	42	0.478	[0.24;0.735]	0.001*
Dendroica pensylvanica	Smaller	14	40	0.06	[−0.182;0.296]	0.564
Dendroica pensylvanica	Larger	10	44	0.466	[0.218;0.731]	0.003*
Emberiza schoeniclus	Smaller	14	41	0.102	[−0.098;0.282]	0.254
Emberiza schoeniclus	Larger	10	44	0.459	[0.256;0.669]	0.001*
Ficedula albicollis	Smaller	14	40	0.137	[−0.09;0.373]	0.199
Ficedula albicollis	Larger	10	44	0.491	[0.258;0.759]	0.001*
Ficedula hypoleuca	Smaller	14	41	0.072	[−0.15;0.295]	0.476
Ficedula hypoleuca	Larger	10	44	0.475	[0.246;0.726]	0.001*
Hirundo rustica	Smaller	14	40	0.097	[−0.136;0.32]	0.35
Hirundo rustica	Larger	10	44	0.483	[0.238;0.735]	0.001*
Luscinia megarhynchos	Smaller	15	42	0.094	[−0.129;0.324]	0.361
Luscinia megarhynchos	Larger	9	43	0.47	[0.224;0.72]	0.001*
Melospiza melodia	Smaller	15	42	0.087	[−0.142;0.304]	0.399
Melospiza melodia	Larger	9	30	0.473	[0.195;0.728]	0.003*
Mimus polyglottos	Smaller	15	42	0.094	[−0.12;0.312]	0.338
Mimus polyglottos	Larger	9	43	0.459	[0.219;0.693]	0.002*
Parus caeruleus	Smaller	14	40	0.086	[−0.135;0.313]	0.397
Parus caeruleus	Larger	10	44	0.477	[0.22;0.717]	0.002*
Parus major	Smaller	14	37	0.106	[−0.123;0.349]	0.333
Parus major	Larger	10	44	0.486	[0.232;0.744]	0.002*
Phylloscopus fuscatus	Smaller	14	41	0.1	[−0.116;0.328]	0.331
Phylloscopus fuscatus	Larger	10	44	0.483	[0.243;0.729]	0.001*
Phylloscopus trochilus	Smaller	14	38	0.077	[−0.149;0.293]	0.462
Phylloscopus trochilus	Larger	10	44	0.47	[0.234;0.712]	0.002*
Plectrophenax nivalis	Smaller	14	40	0.097	[−0.119;0.319]	0.329
Plectrophenax nivalis	Larger	10	44	0.466	[0.225;0.695]	0.002*
Sturnella neglecta	Smaller	15	42	0.083	[−0.15;0.298]	0.414
Sturnella neglecta	Larger	9	37	0.473	[0.213;0.726]	0.002*
Sturnus vulgaris	Smaller	15	42	0.096	[−0.113;0.316]	0.342
Sturnus vulgaris	Larger	9	38	0.466	[0.222;0.72]	0.002*
Sylvia communis	Smaller	15	42	0.097	[−0.119;0.321]	0.351
Sylvia communis	Larger	9	43	0.475	[0.236;0.721]	0.002*
Wilsonia canadensis	Smaller	14	40	0.093	[−0.124;0.309]	0.348
Wilsonia canadensis	Larger	10	44	0.47	[0.232;0.707]	0.001*

Table A.2.6: Removal of individual species does not significantly affect the syllable repertoire results. Performed in the syllable repertoire dataset. Threshold $\geq$ 216. Asterisks (*) denote significant groups.

Removed	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Acrocephalus arundinaceus	Smaller	20	65	0.181	[−0.091;0.462]	0.162
Acrocephalus arundinaceus	Larger	4	8	0.627	[0.16;1.067]	0.007*
Acrocephalus bistrigiceps	Smaller	20	77	0.192	[−0.054;0.436]	0.103
Acrocephalus bistrigiceps	Larger	4	8	0.62	[0.187;1.034]	0.004*
Acrocephalus palustris	Smaller	21	78	0.18	[−0.075;0.415]	0.122
Acrocephalus palustris	Larger	3	3	0.782	[0.247;1.338]	0.006*
Acrocephalus schoenobaenus	Smaller	20	72	0.177	[−0.058;0.409]	0.114
Acrocephalus schoenobaenus	Larger	4	8	0.627	[0.221;1.062]	0.004*
Agelaius phoeniceus	Smaller	20	76	0.181	[−0.058;0.42]	0.124
Agelaius phoeniceus	Larger	4	8	0.59	[0.161;1.017]	0.008*
Anthus spinoletta	Smaller	20	76	0.19	[−0.072;0.45]	0.134
Anthus spinoletta	Larger	4	8	0.615	[0.163;1.037]	0.007*
Cardinalis cardinalis	Smaller	20	77	0.195	[−0.046;0.433]	0.099
Cardinalis cardinalis	Larger	4	8	0.606	[0.19;1.012]	0.004*
Carpodacus mexicanus	Smaller	20	76	0.172	[−0.085;0.428]	0.168
Carpodacus mexicanus	Larger	4	8	0.611	[0.174;1.053]	0.009*
Dendroica pensylvanica	Smaller	20	76	0.162	[−0.101;0.424]	0.191
Dendroica pensylvanica	Larger	4	8	0.608	[0.171;1.052]	0.01*
Emberiza schoeniclus	Smaller	20	77	0.201	[−0.022;0.412]	0.07
Emberiza schoeniclus	Larger	4	8	0.594	[0.206;1.002]	0.004*
Ficedula albicollis	Smaller	20	76	0.227	[−0.033;0.494]	0.073
Ficedula albicollis	Larger	4	8	0.61	[0.194;1.037]	0.006*
Ficedula hypoleuca	Smaller	20	77	0.157	[−0.113;0.412]	0.216
Ficedula hypoleuca	Larger	4	8	0.598	[0.164;1.028]	0.008*
Hirundo rustica	Smaller	20	76	0.192	[−0.085;0.451]	0.134
Hirundo rustica	Larger	4	8	0.617	[0.194;1.059]	0.006*
Luscinia megarhynchos	Smaller	21	78	0.183	[−0.085;0.436]	0.144
Luscinia megarhynchos	Larger	3	7	0.579	[0.123;1.032]	0.015*
Melospiza melodia	Smaller	20	64	0.159	[−0.076;0.416]	0.169
Melospiza melodia	Larger	4	8	0.592	[0.178;1.031]	0.008*
Mimus polyglottos	Smaller	21	78	0.189	[−0.061;0.443]	0.12
Mimus polyglottos	Larger	3	7	0.537	[0.068;0.999]	0.021*
Parus caeruleus	Smaller	20	76	0.177	[−0.076;0.448]	0.155
Parus caeruleus	Larger	4	8	0.605	[0.181;1.051]	0.008*
Parus major	Smaller	20	73	0.196	[−0.067;0.469]	0.129
Parus major	Larger	4	8	0.614	[0.193;1.065]	0.006*
Phylloscopus fuscatus	Smaller	20	77	0.188	[−0.06;0.448]	0.129
Phylloscopus fuscatus	Larger	4	8	0.613	[0.194;1.055]	0.005*
Phylloscopus trochilus	Smaller	20	74	0.172	[−0.092;0.446]	0.167
Phylloscopus trochilus	Larger	4	8	0.597	[0.174;1.024]	0.009*
Plectrophenax nivalis	Smaller	20	76	0.192	[−0.063;0.437]	0.111
Plectrophenax nivalis	Larger	4	8	0.609	[0.186;1.04]	0.008*
Sturnella neglecta	Smaller	20	71	0.164	[−0.096;0.404]	0.168
Sturnella neglecta	Larger	4	8	0.608	[0.189;1.067]	0.007*
Sturnus vulgaris	Smaller	20	72	0.154	[−0.092;0.396]	0.189
Sturnus vulgaris	Larger	4	8	0.61	[0.185;1.008]	0.005*
Sylvia communis	Smaller	21	78	0.185	[−0.069;0.444]	0.136
Sylvia communis	Larger	3	7	0.604	[0.161;1.066]	0.01*
Wilsonia canadensis	Smaller	20	76	0.191	[−0.06;0.443]	0.115
Wilsonia canadensis	Larger	4	8	0.614	[0.174;1.045]	0.006*

Table A.2.7: Relatively larger syllable repertoires are predictive of a correlation between individual song elaboration and reproductive success for all tested thresholds. There was not significant evidence for this correlation in the smaller repertoire group for any threshold. Performed in the syllable repertoire dataset with territory-controlled measurements. Asterisks (*) denote significant groups.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<18.5	Smaller	8	17	−0.039	[−0.316;0.25]	0.778
≥18.5	Larger	17	69	0.307	[0.088;0.542]	0.013*
<38	Smaller	15	42	0.084	[−0.136;0.283]	0.374
≥38	Larger	10	44	0.448	[0.212;0.667]	0.002*
<216	Smaller	21	78	0.176	[−0.076;0.424]	0.134
≥216	Larger	4	8	0.476	[0.086;0.902]	0.023*

Table A.2.8: Larger and smaller syllable repertoire size groups have significantly different correlations between individual song elaboration and reproductive success for most thresholds. Performed in syllable repertoire dataset with territory-controlled measurements. Asterisks (*) denote models with significantly different groups.

Threshold	BEST Mean	95% CredInt	%<0
18.50	0.345	[0.187;0.504]	0.1*
38	0.361	[0.22;0.502]	0.1*
216	0.183	[−0.054;0.42]	6.2

Measurements of average syllable repertoire size are inherently noisy, because different investigators define syllables differently, and syllables are markedly different between species. Therefore, we repeated the threshold analysis using the highest and lowest values that we encountered for each species in the literature. While switching the repertoire size values did change the overall order of species, it did not have a major effect on the position of individual species, with most birds changing their ranking order by 2 positions or fewer (**Figures A.2.1** and **A.2.2**). Of note, the position of *Emberiza schoeniclus* was noticeably affected by the repertoire size estimate used (Original position=8, Min position=5, Max position=13), as was the position of *Phylloscopus trochilus* (Original position=15, Min position=15, Max position=20). Using the minimum or maximum values did not significantly affect the results (**Tables A.2.9** to **A.2.12**), suggesting that these results do not depend on the precise repertoire size measurements used or require species to be in the exact order they were arranged; As long as species with smaller repertoires are generally early in the ordering from smallest to largest average species syllable repertoires, and species with larger repertoires are late in that order, the same outcome is achieved.

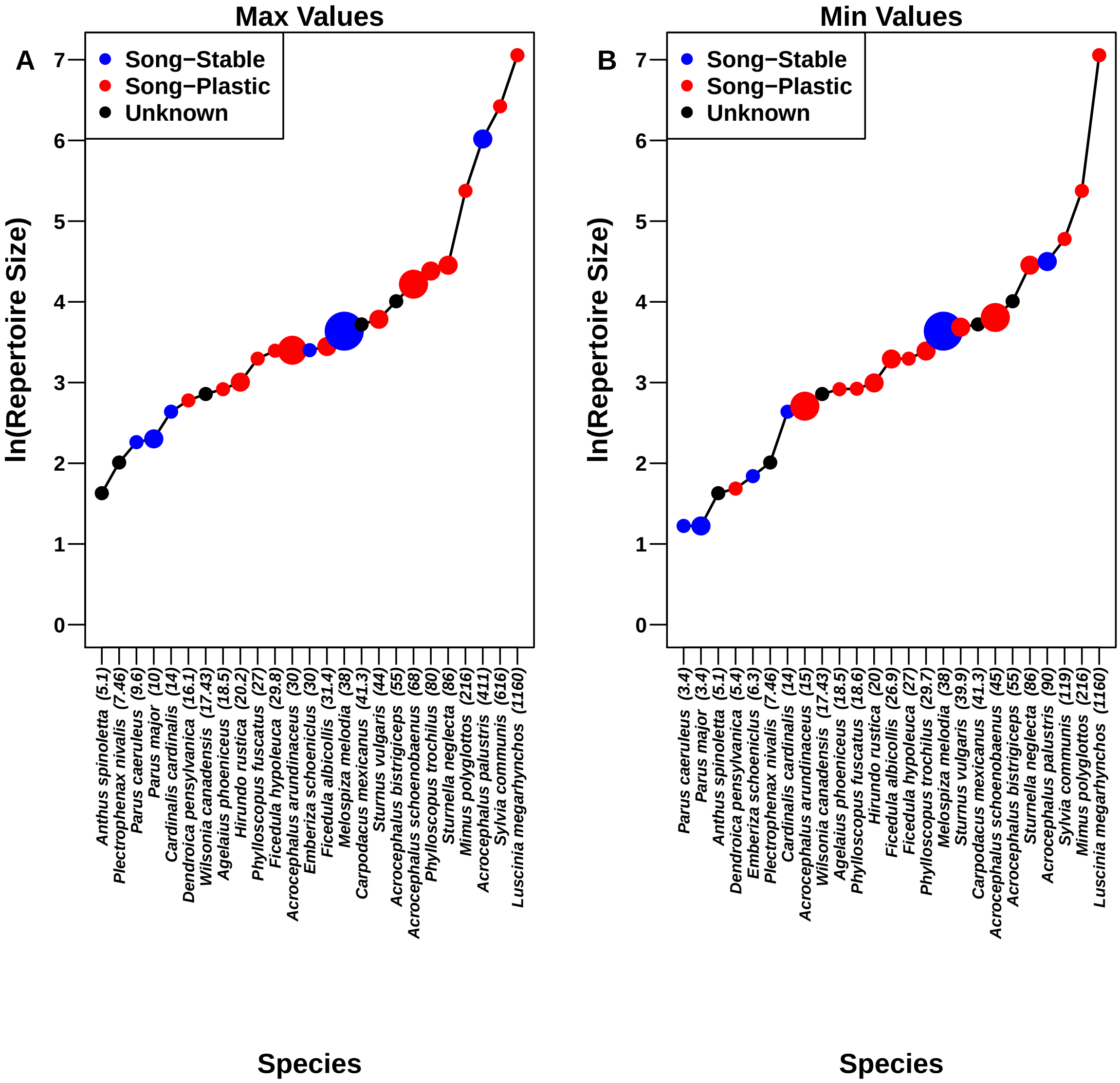


Figure A.2.2: Species plotted in order of increasing syllable repertoire size. Repertoire size (without the log transformation) is included next to species name in parentheses. Red circles denote song-plastic species, and blue circles denote song-stable species. Black circles denote species for which no song stability information was available. Circle size increases as more studies were included for a single species. (A) shows the order when the maximum estimate values in the literature were used to rank species, while (B) shows the order when the minimum values in the literature were used to rank species.

Table A.2.9: Using the maximum values in the literature for species average syllable repertoire size did not significantly affect the results. Performed in the syllable repertoire dataset. Asterisks (*) denote significant groups.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<7.46	Smaller	1	2	0.067	[−0.599;0.784]	0.848
≥7.46	Larger	24	84	0.25	[−0.024;0.534]	0.064
<9.6	Smaller	2	4	−0.044	[−0.574;0.466]	0.853
≥9.6	Larger	23	82	0.261	[−0.015;0.528]	0.052
<10	Smaller	3	6	0.054	[−0.382;0.526]	0.812
≥10	Larger	22	80	0.263	[0.007;0.551]	0.05
<14	Smaller	4	11	0.03	[−0.369;0.436]	0.88
≥14	Larger	21	75	0.284	[0.011;0.569]	0.042*
<16.1	Smaller	5	12	−0.051	[−0.409;0.329]	0.779
≥16.1	Larger	20	74	0.311	[0.039;0.584]	0.028*
<17.43	Smaller	6	14	0.042	[−0.297;0.388]	0.804
≥17.43	Larger	19	72	0.294	[0.045;0.554]	0.029*
<18.5	Smaller	7	16	0.016	[−0.313;0.337]	0.914
≥18.5	Larger	18	70	0.311	[0.051;0.567]	0.021*
<20.2	Smaller	8	18	0.018	[−0.296;0.318]	0.894
≥20.2	Larger	17	68	0.322	[0.062;0.574]	0.017*
<27	Smaller	9	20	0.03	[−0.253;0.311]	0.816
≥27	Larger	16	66	0.333	[0.084;0.58]	0.013*
<29.8	Smaller	10	21	0.034	[−0.241;0.316]	0.794
≥29.8	Larger	15	65	0.337	[0.1;0.589]	0.011*
<30	Smaller	11	22	0.068	[−0.201;0.337]	0.598
≥30	Larger	14	64	0.337	[0.1;0.586]	0.012*
<31.4	Smaller	13	36	0.085	[−0.164;0.329]	0.454
≥31.4	Larger	12	50	0.393	[0.143;0.632]	0.005*
<38	Smaller	14	38	0.071	[−0.146;0.3]	0.484
≥38	Larger	11	48	0.453	[0.225;0.688]	0.002*
<41.3	Smaller	15	52	0.102	[−0.144;0.357]	0.379
≥41.3	Larger	10	34	0.457	[0.184;0.728]	0.003*
<44	Smaller	16	54	0.112	[−0.134;0.364]	0.335
≥44	Larger	9	32	0.463	[0.181;0.741]	0.003*
<55	Smaller	17	60	0.148	[−0.121;0.415]	0.232
≥55	Larger	8	26	0.446	[0.125;0.755]	0.007*
<68	Smaller	18	61	0.15	[−0.103;0.428]	0.224
≥68	Larger	7	25	0.454	[0.14;0.778]	0.007*
<80	Smaller	19	67	0.157	[−0.117;0.411]	0.202
≥80	Larger	6	19	0.487	[0.154;0.825]	0.006*
<86	Smaller	20	71	0.175	[−0.072;0.44]	0.147
≥86	Larger	5	15	0.553	[0.197;0.928]	0.005*
<216	Smaller	21	78	0.185	[−0.068;0.459]	0.141
≥216	Larger	4	8	0.611	[0.194;1.051]	0.006*
<411	Smaller	22	79	0.206	[−0.072;0.478]	0.116
≥411	Larger	3	7	0.536	[0.068;1.049]	0.029*
<616	Smaller	23	84	0.22	[−0.049;0.497]	0.093
≥616	Larger	2	2	0.722	[0.014;1.44]	0.046*

Table A.2.10: Using the minimum values in the literature for species average syllable repertoire size did not significantly affect the results. Performed in the syllable repertoire dataset. Asterisks (*) denote significant groups.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<5.1	Smaller	2	7	0.117	[−0.454;0.695]	0.65
≥5.1	Larger	23	79	0.256	[−0.026;0.553]	0.068
<5.4	Smaller	3	9	0.098	[−0.355;0.56]	0.65
≥5.4	Larger	22	77	0.266	[−0.01;0.566]	0.056
<6.3	Smaller	4	11	0.185	[−0.222;0.612]	0.356
≥6.3	Larger	21	75	0.254	[−0.02;0.542]	0.063
<7.46	Smaller	5	12	0.082	[−0.278;0.475]	0.646
≥7.46	Larger	20	74	0.276	[0.017;0.545]	0.039*
<14	Smaller	6	14	0.039	[−0.306;0.373]	0.818
≥14	Larger	19	72	0.294	[0.057;0.557]	0.023*
<15	Smaller	7	15	−0.015	[−0.322;0.303]	0.915
≥15	Larger	18	71	0.316	[0.077;0.552]	0.013*
<17.43	Smaller	8	28	0.059	[−0.238;0.368]	0.674
≥17.43	Larger	17	58	0.329	[0.064;0.584]	0.017*
<18.5	Smaller	9	30	0.041	[−0.25;0.341]	0.747
≥18.5	Larger	16	56	0.354	[0.094;0.611]	0.013*
<18.6	Smaller	10	32	0.043	[−0.243;0.324]	0.722
≥18.6	Larger	15	54	0.371	[0.119;0.641]	0.011*
<20	Smaller	11	33	0.047	[−0.24;0.316]	0.694
≥20	Larger	14	53	0.376	[0.123;0.635]	0.008*
<26.9	Smaller	12	35	0.056	[−0.199;0.316]	0.627
≥26.9	Larger	13	51	0.393	[0.152;0.651]	0.003*
<27	Smaller	13	37	0.049	[−0.186;0.273]	0.629
≥27	Larger	12	49	0.447	[0.213;0.677]	0.002*
<29.7	Smaller	14	38	0.071	[−0.146;0.3]	0.484
≥29.7	Larger	11	48	0.453	[0.225;0.688]	0.002*
<38	Smaller	15	42	0.098	[−0.121;0.324]	0.336
≥38	Larger	10	44	0.482	[0.244;0.732]	0.001*
<39.9	Smaller	16	56	0.124	[−0.125;0.369]	0.276
≥39.9	Larger	9	30	0.491	[0.204;0.775]	0.002*
<41.3	Smaller	17	62	0.159	[−0.113;0.422]	0.196
≥41.3	Larger	8	24	0.479	[0.167;0.82]	0.005*
<45	Smaller	18	64	0.165	[−0.097;0.434]	0.178
≥45	Larger	7	22	0.486	[0.155;0.824]	0.006*
<55	Smaller	19	70	0.176	[−0.079;0.429]	0.14
≥55	Larger	6	16	0.514	[0.176;0.871]	0.006*
<86	Smaller	20	71	0.175	[−0.072;0.44]	0.147
≥86	Larger	5	15	0.553	[0.197;0.928]	0.005*
<90	Smaller	21	78	0.185	[−0.068;0.459]	0.141
≥90	Larger	4	8	0.611	[0.194;1.051]	0.006*
<119	Smaller	22	83	0.2	[−0.072;0.444]	0.114
≥119	Larger	3	3	0.801	[0.234;1.334]	0.006*
<216	Smaller	23	84	0.206	[−0.049;0.476]	0.108
≥216	Larger	2	2	0.874	[0.209;1.539]	0.011*

Table A.2.11: Groups of species with relatively larger and smaller syllable repertoire sizes were significantly different. Performed in the syllable repertoire dataset, using the maximum reported species average syllable repertoire size for each species. Asterisks (*) denote significant groups.

Threshold	BEST Mean	95% CredInt	%<0
9.60	0.326	[−0.245;0.897]	6.8
10	0.257	[−0.093;0.607]	6.6
14	0.26	[0.086;0.434]	0.3*
16.10	0.301	[0.118;0.484]	0.1*
17.43	0.307	[0.15;0.465]	0.1*
18.50	0.315	[0.169;0.461]	0.1*
20.20	0.319	[0.173;0.464]	0.1*
27	0.333	[0.192;0.475]	0.1*
29.80	0.356	[0.218;0.495]	0.1*
30	0.339	[0.198;0.481]	0.1*
31.40	0.329	[0.179;0.479]	0.1*
38	0.376	[0.232;0.521]	0.1*
41.30	0.325	[0.155;0.495]	0.1*
44	0.317	[0.14;0.495]	0.1*
55	0.231	[0.045;0.417]	0.7*
68	0.237	[0.048;0.426]	0.9*
80	0.273	[0.108;0.438]	0.1*
86	0.294	[0.144;0.444]	0.1*
216	0.246	[−0.028;0.52]	3.6*
411	0.189	[−0.085;0.462]	7.9
616	9.115	[−44.729;62.958]	32.6

Table A.2.12: Groups of species with relatively larger and smaller syllable repertoire sizes were significantly different. Performed in the syllable repertoire dataset, using the minimum reported species average syllable repertoire size for each species. Asterisks (*) denote significant groups.

Threshold	BEST Mean	95% CredInt	%<0
5.40	0.196	[0.021;0.371]	1.4*
6.30	0.221	[0.072;0.37]	0.4*
7.46	0.279	[0.087;0.471]	0.3*
14	0.325	[0.149;0.501]	0.1*
15	0.362	[0.183;0.542]	0.1*
17.43	0.26	[0.107;0.413]	0.1*
18.50	0.285	[0.136;0.433]	0.1*
18.60	0.292	[0.142;0.441]	0.1*
20	0.313	[0.161;0.464]	0.1*
26.90	0.337	[0.19;0.485]	0.1*
27	0.384	[0.24;0.528]	0.1*
29.70	0.376	[0.232;0.521]	0.1*
38	0.365	[0.216;0.515]	0.1*
39.90	0.324	[0.142;0.506]	0.1*
41.30	0.225	[0.045;0.404]	0.7*
45	0.238	[0.041;0.435]	1.3*
55	0.273	[0.122;0.425]	0.1*
86	0.294	[0.144;0.444]	0.1*
90	0.246	[−0.028;0.52]	3.6*
119	0.357	[−1.669;2.382]	12.2
216	−12.852	[−109.569;83.864]	38.5

Considering the result that the exact ordering of species could change and similar results would be obtained, we wanted to assess the probability of getting significant results when the data was split into two groups arbitrarily instead of based on published species average syllable repertoire size. We generated 500 permutations of the data, randomly assigning each species to one of two groups and performing a meta-analysis with each permutation. We found that 4.4% of the randomized groups (8.8% of the two group models) showed significant (p_{MCMC}<0.025) evidence of a real correlation between song elaboration and reproductive success, which was also significantly different from the other group in that same model ($\text{BEST}_{\%<0}$<2.5%). In contrast, 14 out of 21 thresholded models from the real data were significant (thresholds 18.15 to 216 syllables); in the real data, the middle range of thresholds was significant while the largest and smallest thresholds were not. (**Figure A.2.3A**). Interestingly, the arbitrary significant groups were enriched with species with larger repertoires and had fewer species with smaller repertoires relative to the arbitrary nonsignificant groups (**Figure A.2.3B**). If the p_{MCMC} threshold for significant results was reduced to 0.013 — allowing only a 2.5% chance of type I error based on the arbitrary grouping results — the main analysis results are still significant for 13 out of 21 species average syllable repertoire size thresholds (18.5 to 216 syllables).

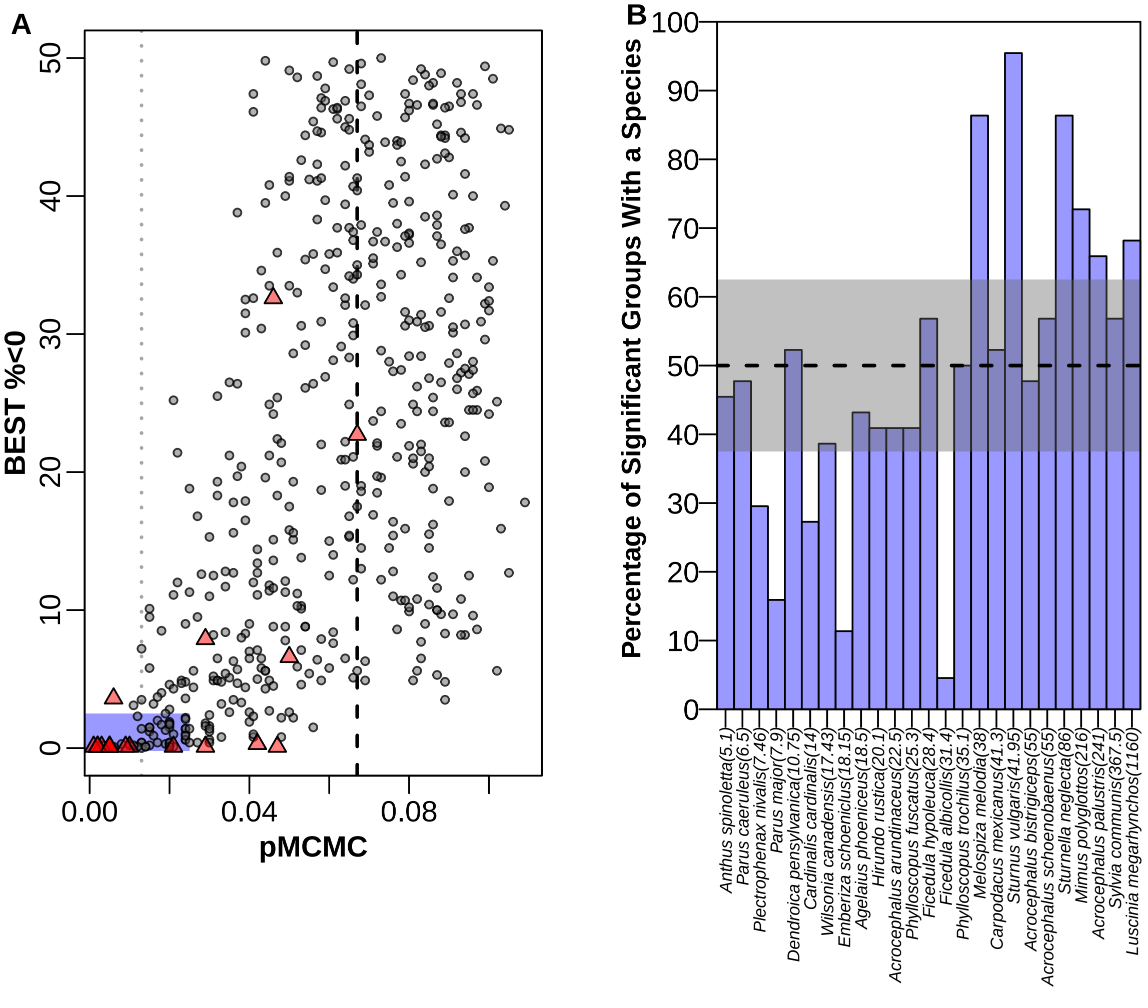


Figure A.2.3: Distribution of model statistics when the two constituent groups are formed randomly using the species average syllable repertoire dataset. (A) Black dashed line denotes the meta-analytic mean for the full population in the syllable repertoire dataset. Each grey dot represents the group from a randomized model with the lowest p_{MCMC}. Each red triangle represents the p_{MCMC} for the larger species average syllable repertoire group for one of the thresholds tested in the dichotomized species average syllable repertoire size analyses. The light blue rectangle shows the region where a model must fall to have a group with significant evidence for a correlation between individual song elaboration and reproductive success ($p_{MCMC}<0.025$) that was also significantly different from the other group in that same model ($\text{BEST}_{\%<0}<2.5\%$). In total, 4.4% of all randomized groups (8.8% of the 2-group models) were significant based on this criterion. In contrast, 14 out of 21 models from the real data were significant (thresholds 18.15 to 216 syllables). The grey dashed line indicates the p_{MCMC} threshold at which 2.5% percent of the randomized data would be considered significant ($p_{MCMC}<=0.013$). At this significance threshold, 13 out of 21 models from the real data were still significant (thresholds 18.5 to 216 syllables). (B) Histogram of species in the group with a significant p_{MCMC} from the randomized models present in the blue region of (A) (44 models). Species are ordered from smallest to largest average syllable repertoire size, with the literature reported value in parentheses after the species name. The black dashed line shows the chance each species had to be randomly assigned into group 1 or group 2. The grey shaded rectangle shows where the data would be expected to fall due to chance 95% of the time. Significant arbitrary groups contained several species with repertoire sizes less than 25 syllables less often than would be expected by random chance (*Parus caeruleus*, *Plectrophenax nivalis*, *Cardinalis cardinalis*, and *Emberiza schoeniclus*). These groups were also enriched above chance levels in most species with repertoires 86 syllables (*Sturnella neglecta*, *Mimus polyglottos*, *Acrocephalus palustris*, *Luscinia megarhynchos*).

Although song stability did not provide predictive value alone, it was possible that it would provide additional predictive information in combination with species average syllable repertoire size. We tested for the existence of this interaction by including an interaction term between species average syllable repertoire size and song stability in the Bayesian meta-analysis; however, breaking the species into four groups led to very small sample sizes that make the results of the model difficult to interpret (**Table A.2.13**). The results tentatively suggest that there may be some additional predictive power of these two variables in combination; species with intermediate to large average syllable repertoires combined with plastic songs were most likely to show a correlation between individual song elaboration and reproductive success, while species with smaller average syllable repertoires and stable songs were least likely to show this correlation. A larger number of species will need to be studied to make any conclusive statements.

Table A.2.13: Combined meta-analysis of species average syllable repertoire size and song stability. Performed in the song stability dataset. Asterisks (*) denote significant groups.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<18.5	SmallerStable	4	9	-0.137	$[-0.575;0.283]$	0.52
≥ 18.5	LargerStable	2	19	0.435	[0.011;0.827]	0.036*
<18.5	SmallerPlastic	1	2	0.385	$[-0.283;1.042]$	0.24
≥ 18.5	LargerPlastic	13	47	0.308	[0.039;0.583]	0.028*
<38	SmallerStable	4	9	-0.113	$[-0.5;0.223]$	0.523
≥ 38	LargerStable	2	19	0.436	[0.106;0.784]	0.015*
<38	SmallerPlastic	8	27	0.187	$[-0.065;0.448]$	0.137
≥ 38	LargerPlastic	6	22	0.535	[0.243;0.841]	0.002*
<216	SmallerStable	5	23	0.077	$[-0.325;0.468]$	0.657
≥ 216	LargerStable	1	5	0.452	$[-0.208;1.109]$	0.158
<216	SmallerPlastic	11	46	0.265	$[-0.014;0.557]$	0.062
≥ 216	LargerPlastic	3	3	0.843	[0.249;1.442]	0.008*

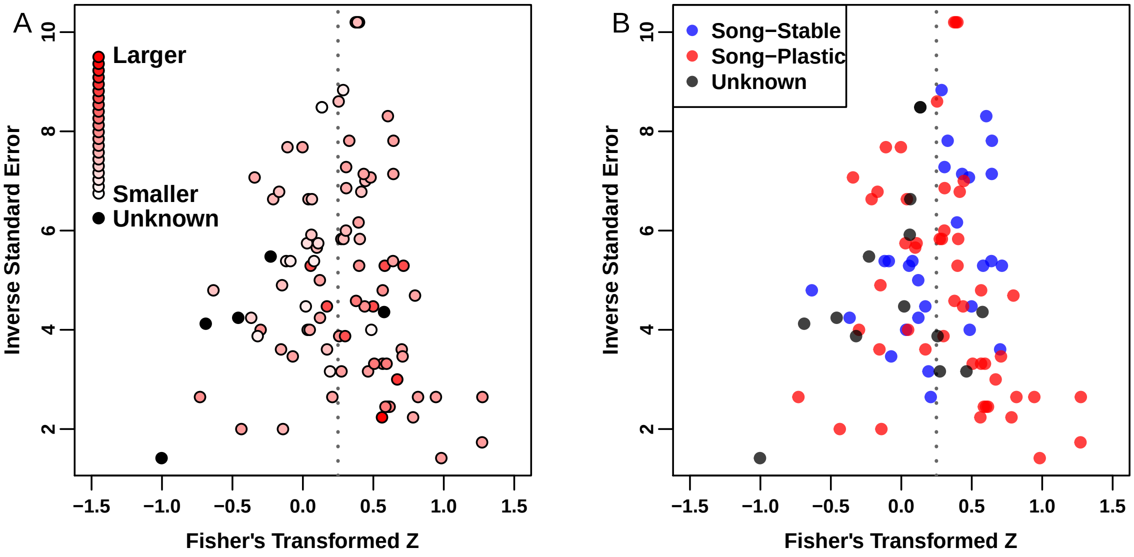


Figure A.3.1: Using territory-controlled measurements did not significantly affect funnel-plot asymetry. Funnel plots show the 91 measurements of the correlation between individual song elaboration and reproductive success from the full dataset used in the Bayesian meta-analysis presented in the main text, here performed with territory-controlled measurements. The grey dotted line represents the mean Fisher's transformed Z. (A) Circle color becomes more red as the average syllable repertoire of the species increases. Black circles show measurements from species for which the syllable repertoire size is unknown. (B) Blue circles indicate measures from song-stable species, while red circles indicate measurements from song-plastic species. Black circles denote species for which no song stability information was available. Regression testing on the full dataset (z=0.8393, *p*=0.4013), syllable repertoire dataset (z=1.4602, *p*=0.1442), or song stability dataset (z=1.3952, *p*=0.1630) revealed no significant funnel plot asymmetry. Ranked correlation testing on the full dataset (τ=0.0168, *p*=0.8156), syllable repertoire dataset (τ=0.0472, *p*=0.5233), or song stability dataset (τ=0.0345, *p*=0.6598) also revealed no significant funnel plot asymmetry.

Table A.3.1: Population meta-analysis performed in all three datasets with territory-controlled measurements.

Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Full Dataset	27	91	0.188	[−0.154;0.53]	0.197
Repertoire Dataset	25	86	0.214	[−0.041;0.458]	0.082
Song Stability dataset	20	77	0.24	[−0.007;0.476]	0.05

Table A.3.2: Using multiple predictive trees in place of a consensus tree did not significantly affect any results. For each model below, 100 trees were randomly selected from the sample of 1000 used to generate the consensus tree. Each tree was run twice, so 200 MCMC chains were generated for each model. Reported below are values calculated using the quantile method. We obtained similar results using the hdr method, however, this methods threw warnings, because the hrd credibility intervals could not be calculated for the residual variance. p_{MCMC} not reported, because `mulTree.summary` does not provide this value.

Model Group	Post Mean	95% CredInt
Population	0.199	[−0.004;0.402]
Stable	0.147	[−0.206;0.465]
Plastic	0.317	[0.062;0.599]
Intercept	−0.355	[−0.751;0.018]
Slope	0.174	[0.079;0.273]
Smaller <18.5	−0.022	[−0.279;0.229]
Larger ≥ 18.5	0.332	[0.131;0.542]
Smaller <38	0.076	[−0.126;0.273]
Larger ≥ 38	0.463	[0.249;0.687]
Smaller <216	0.174	[−0.04;0.385]
Larger ≥ 216	0.615	[0.225;1.032]

Table A.3.3: Continuous species average syllable repertoire size model meta-analysis with each species removed in turn. Removal of any one species did not significantly affect the data. Asterisks (*) denote significant slopes.

Removed	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Acrocephalus arundinaceus	Intercept	24	73	−0.382	[−0.817;0.064]	0.079
Acrocephalus arundinaceus	Slope	24	73	0.179	[0.075;0.283]	0.001*
Acrocephalus bistrigiceps	Intercept	24	85	−0.358	[−0.777;0.053]	0.079
Acrocephalus bistrigiceps	Slope	24	85	0.176	[0.074;0.277]	0.001*
Acrocephalus palustris	Intercept	24	81	−0.424	[−0.86;0.006]	0.052
Acrocephalus palustris	Slope	24	81	0.198	[0.085;0.314]	0.002*
Acrocephalus schoenobaenus	Intercept	24	80	−0.365	[−0.784;0.028]	0.072
Acrocephalus schoenobaenus	Slope	24	80	0.177	[0.076;0.281]	0.002*
Agelaius phoeniceus	Intercept	24	84	−0.34	[−0.756;0.057]	0.093
Agelaius phoeniceus	Slope	24	84	0.168	[0.065;0.272]	0.002*
Anthus spinoletta	Intercept	24	84	−0.391	[−0.843;0.045]	0.074
Anthus spinoletta	Slope	24	84	0.183	[0.075;0.292]	0.001*
Cardinalis cardinalis	Intercept	24	85	−0.324	[−0.709;0.08]	0.094
Cardinalis cardinalis	Slope	24	85	0.169	[0.073;0.266]	0.001*
Carpodacus mexicanus	Intercept	24	84	−0.364	[−0.777;0.039]	0.072
Carpodacus mexicanus	Slope	24	84	0.176	[0.076;0.277]	0.002*
Dendroica pensylvanica	Intercept	24	84	−0.448	[−0.863;−0.017]	0.032
Dendroica pensylvanica	Slope	24	84	0.196	[0.1;0.298]	0*
Emberiza schoeniclus	Intercept	24	85	−0.31	[−0.698;0.06]	0.096
Emberiza schoeniclus	Slope	24	85	0.165	[0.07;0.261]	0.001*
Ficedula albicollis	Intercept	24	84	−0.306	[−0.699;0.072]	0.113
Ficedula albicollis	Slope	24	84	0.169	[0.075;0.262]	0.002*
Ficedula hypoleuca	Intercept	24	85	−0.38	[−0.814;0.036]	0.071
Ficedula hypoleuca	Slope	24	85	0.177	[0.077;0.28]	0.001*
Hirundo rustica	Intercept	24	84	−0.351	[−0.764;0.072]	0.089
Hirundo rustica	Slope	24	84	0.174	[0.074;0.279]	0.001*
Luscinia megarhynchos	Intercept	24	85	−0.362	[−0.805;0.053]	0.084
Luscinia megarhynchos	Slope	24	85	0.176	[0.073;0.286]	0.002*
Melospiza melodia	Intercept	24	72	−0.35	[−0.785;0.033]	0.083
Melospiza melodia	Slope	24	72	0.168	[0.065;0.271]	0.002*
Mimus polyglottos	Intercept	24	85	−0.322	[−0.714;0.1]	0.106
Mimus polyglottos	Slope	24	85	0.163	[0.064;0.264]	0.002*
Parus caeruleus	Intercept	24	84	−0.395	[−0.825;0.013]	0.061
Parus caeruleus	Slope	24	84	0.184	[0.082;0.287]	0.001*
Parus major	Intercept	24	81	−0.353	[−0.799;0.075]	0.106
Parus major	Slope	24	81	0.175	[0.07;0.283]	0.002*
Phylloscopus fuscatus	Intercept	24	85	−0.356	[−0.775;0.055]	0.081
Phylloscopus fuscatus	Slope	24	85	0.175	[0.075;0.275]	0.001*
Phylloscopus trochilus	Intercept	24	82	−0.344	[−0.761;0.059]	0.091
Phylloscopus trochilus	Slope	24	82	0.169	[0.07;0.27]	0.002*
Plectrophenax nivalis	Intercept	24	84	−0.33	[−0.749;0.072]	0.105
Plectrophenax nivalis	Slope	24	84	0.168	[0.065;0.268]	0.002*
Sturnella neglecta	Intercept	24	79	−0.371	[−0.802;0.037]	0.076
Sturnella neglecta	Slope	24	79	0.179	[0.073;0.287]	0.002*
Sturnus vulgaris	Intercept	24	80	−0.361	[−0.756;0.034]	0.064
Sturnus vulgaris	Slope	24	80	0.172	[0.078;0.271]	0.001*
Sylvia communis	Intercept	24	85	−0.357	[−0.79;0.049]	0.082
Sylvia communis	Slope	24	85	0.175	[0.075;0.28]	0.001*
Wilsonia canadensis	Intercept	24	84	−0.337	[−0.749;0.051]	0.093
Wilsonia canadensis	Slope	24	84	0.172	[0.074;0.272]	0.002*

Table A.3.4: Continuous species average syllable repertoire size model meta-analyses, wherein 3 to 9 species with the largest average syllable repertoires were removed. Asterisks (*) denote significant slopes.

# Removed	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
3	Intercept	21	78	−0.4	[−0.912;0.149]	0.13
3	Slope	21	78	0.189	[0.027;0.341]	0.024*
4	Intercept	20	71	−0.423	[−1.068;0.194]	0.17
4	Slope	20	71	0.196	[−0.002;0.387]	0.054
5	Intercept	19	65	−0.452	[−1.14;0.238]	0.172
5	Slope	19	65	0.204	[−0.008;0.428]	0.065
6	Intercept	18	64	−0.466	[−1.176;0.296]	0.188
6	Slope	18	64	0.21	[−0.021;0.446]	0.079
7	Intercept	17	58	−0.363	[−1.08;0.386]	0.301
7	Slope	17	58	0.167	[−0.079;0.408]	0.166
8	Intercept	16	56	−0.328	[−1.119;0.482]	0.377
8	Slope	16	56	0.152	[−0.119;0.425]	0.242
9	Intercept	15	42	−0.141	[−1.114;0.727]	0.742
9	Slope	15	42	0.077	[−0.25;0.394]	0.599

Table A.3.5: Continuous species average syllable repertoire size model meta-analyses, wherein 3 to 9 species with the smallest average syllable repertoires were removed. Asterisks (*) denote significant slopes.

# Removed	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
3	Intercept	22	80	−0.412	[−0.901;0.075]	0.088
3	Slope	22	80	0.187	[0.071;0.314]	0.003*
4	Intercept	21	75	−0.446	[−0.985;0.131]	0.108
4	Slope	21	75	0.194	[0.056;0.326]	0.006*
5	Intercept	20	73	−0.592	[−1.155;0]	0.041
5	Slope	20	73	0.229	[0.09;0.365]	0.002*
6	Intercept	19	72	−0.493	[−1.079;0.076]	0.082
6	Slope	19	72	0.207	[0.071;0.342]	0.005*
7	Intercept	18	70	−0.416	[−1.022;0.174]	0.158
7	Slope	18	70	0.192	[0.055;0.332]	0.007*
8	Intercept	17	69	−0.323	[−0.914;0.269]	0.268
8	Slope	17	69	0.171	[0.028;0.311]	0.017*
9	Intercept	16	67	−0.261	[−0.921;0.359]	0.41
9	Slope	16	67	0.155	[0.004;0.307]	0.044*

Table A.3.6: Continuous species average syllable repertoire size model meta-analysis when the maximum or minimum values for syllable repertoire size reported in the literature were used. Asterisks (*) denote significant slopes.

Min/Max	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Max	Intercept	25	86	−0.356	[−0.78;0.067]	0.09
Max	Slope	25	86	0.165	[0.072;0.264]	0.001*
Min	Intercept	25	86	−0.282	[−0.659;0.078]	0.128
Min	Slope	25	86	0.166	[0.068;0.259]	0.002*

Table A.3.7: Continuous species average syllable repertoire size model meta-analysis in the repertoire dataset with territory-controlled measurements. Asterisks (*) denote significant slopes.

Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Intercept	25	86	−0.356	[−0.753;0.056]	0.081
Slope	25	86	0.175	[0.075;0.272]	0.001*

Species (Rep) [citation]	**nBirds**	**Fisher's Z**	**[95% CI]**
Anthus spinoletta (5.1) [14]	75	0.135	[−0.096 ; 0.366]
Anthus spinoletta (5.1) [14]	75	0.135	[−0.096 ; 0.366]
Parus caeruleus (6.5) [31]	19	0.034	[−0.456 ; 0.524]
Parus caeruleus (6.5) [31]	19	0.486	[−0.004 ; 0.975]
Plectrophenax nivalis (7.46) [40]	18	−0.321	[−0.827 ; 0.185]
Plectrophenax nivalis (7.46) [40]	23	0.02	[−0.418 ; 0.459]
Parus major (7.9) [33]	32	−0.119	[−0.482 ; 0.245]
Parus major (7.9) [33]	13	0.193	[−0.427 ; 0.813]
Parus major (7.9) [33]	32	−0.089	[−0.453 ; 0.275]
Parus major (7.9) [33]	32	0.079	[−0.285 ; 0.443]
Parus major (7.9) [34]	81	0.286	[0.064 ; 0.508]
Dendroica pensylvanica (10.75) [17]	36	0.109	[−0.232 ; 0.45]
Dendroica pensylvanica (10.75) [17]	36	0.03	[−0.312 ; 0.371]
Cardinalis cardinalis (14) [15]	21	−0.367	[−0.829 ; 0.095]
Wilsonia canadensis (17.43) [48]	38	0.06	[−0.271 ; 0.391]
Wilsonia canadensis (17.43) [48]	47	0.064	[−0.232 ; 0.359]
Emberiza schoeniclus (18.15) [18]	26	−0.635	[−1.044 ; −0.227]
Agelaius phoeniceus (18.5) [13]	49	−0.169	[−0.458 ; 0.12]
Agelaius phoeniceus (18.5) [13]	49	0.415	[0.126 ; 0.704]
Hirundo rustica (20.1) [22]	16	0.171	[−0.373 ; 0.715]
Hirundo rustica (20.1) [20]	27	−0.148	[−0.548 ; 0.252]
Acrocephalus arundinaceus (22.5) [1]	107	0.375	[0.183 ; 0.567]
Acrocephalus arundinaceus (22.5) [1]	107	0.398	[0.206 ; 0.59]
Acrocephalus arundinaceus (22.5) [1]	107	0.387	[0.194 ; 0.579]
Acrocephalus arundinaceus (22.5) [2]	37	0.273	[−0.063 ; 0.609]
Acrocephalus arundinaceus (22.5) [2]	37	0.288	[−0.048 ; 0.624]
Acrocephalus arundinaceus (22.5) [2]	37	0.404	[0.068 ; 0.74]
Acrocephalus arundinaceus (22.5) [3]	35	0.099	[−0.248 ; 0.445]
Acrocephalus arundinaceus (22.5) [3]	62	−0.11	[−0.365 ; 0.146]
Acrocephalus arundinaceus (22.5) [3]	52	0.443	[0.163 ; 0.723]
Acrocephalus arundinaceus (22.5) [3]	47	0.04	[−0.256 ; 0.335]
Acrocephalus arundinaceus (22.5) [3]	39	0.305	[−0.021 ; 0.632]
Acrocephalus arundinaceus (22.5) [3]	47	−0.211	[−0.506 ; 0.085]
Acrocephalus arundinaceus (22.5) [5]	77	0.254	[0.026 ; 0.482]
Phylloscopus fuscatus (25.3) [35]	7	−0.141	[−1.121 ; 0.839]
Ficedula hypoleuca (28.4) [21]	50	0.307	[0.022 ; 0.593]
Ficedula albicollis (31.4) [19]	53	−0.343	[−0.62 ; −0.066]
Ficedula albicollis (31.4) [20]	16	−0.155	[−0.699 ; 0.388]
Phylloscopus trochilus (35.1) [38]	62	−0.003	[−0.258 ; 0.252]
Phylloscopus trochilus (35.1) [38]	25	0.796	[0.378 ; 1.214]
Phylloscopus trochilus (35.1) [38]	19	0.048	[−0.442 ; 0.538]
Phylloscopus trochilus (35.1) [39]	8	0.782	[−0.095 ; 1.659]
Melospiza melodia (38) [24]	72	0.603	[0.367 ; 0.839]
Melospiza melodia (38) [24]	41	0.395	[0.077 ; 0.713]
Melospiza melodia (38) [25]	21	0.121	[−0.341 ; 0.583]
Melospiza melodia (38) [25]	15	−0.072	[−0.638 ; 0.494]
Melospiza melodia (38) [25]	10	0.209	[−0.532 ; 0.95]
Melospiza melodia (38) [26]	16	0.702	[0.158 ; 1.245]
Melospiza melodia (38) [22]	54	0.642	[0.368 ; 0.916]
Melospiza melodia (38) [22]	53	0.48	[0.203 ; 0.757]
Melospiza melodia (38) [22]	54	0.432	[0.157 ; 0.706]
Melospiza melodia (38) [22]	32	0.639	[0.275 ; 1.003]
Melospiza melodia (38) [28,20]	28	0.12	[−0.272 ; 0.512]
Melospiza melodia (38) [29]	64	0.643	[0.392 ; 0.894]
Melospiza melodia (38) [29]	56	0.307	[0.037 ; 0.576]
Melospiza melodia (38) [29]	64	0.329	[0.078 ; 0.58]
Carpodacus mexicanus (41.3) [16]	13	0.462	[−0.157 ; 1.082]
Carpodacus mexicanus (41.3) [16]	13	0.274	[−0.346 ; 0.894]
Sturnus vulgaris (41.95) [45]	15	0.708	[0.142 ; 1.273]
Sturnus vulgaris (41.95) [45]	7	−0.437	[−1.417 ; 0.543]
Sturnus vulgaris (41.95) [46]	6	1.272	[0.14 ; 2.403]
Sturnus vulgaris (41.95) [46]	5	0.982	[−0.404 ; 2.368]
Sturnus vulgaris (41.95) [46]	10	0.944	[0.203 ; 1.685]
Sturnus vulgaris (41.95) [46]	10	0.817	[0.076 ; 1.558]
Acrocephalus bistrigiceps (55) [6]	18	0.257	[−0.249 ; 0.764]
Acrocephalus schoenobaenus (55) [9]	26	0.566	[0.158 ; 0.975]
Acrocephalus schoenobaenus (55) [9]	31	0.399	[0.029 ; 0.769]
Acrocephalus schoenobaenus (55) [9]	23	0.438	[0 ; 0.877]
Acrocephalus schoenobaenus (55) [10]	10	1.274	[0.533 ; 2.015]
Acrocephalus schoenobaenus (55) [11]	19	−0.3	[−0.79 ; 0.19]
Acrocephalus schoenobaenus (55) [12]	10	−0.73	[−1.471 ; 0.011]
Sturnella neglecta (86) [43]	9	0.601	[−0.199 ; 1.401]
Sturnella neglecta (86) [43]	14	0.566	[−0.025 ; 1.157]
Sturnella neglecta (86) [43]	9	0.615	[−0.185 ; 1.416]
Sturnella neglecta (86) [43]	14	0.594	[0.003 ; 1.185]
Sturnella neglecta (86) [43]	9	0.586	[−0.214 ; 1.386]
Sturnella neglecta (86) [43]	14	0.506	[−0.085 ; 1.097]
Sturnella neglecta (86) [44]	24	0.377	[−0.05 ; 0.805]
Mimus polyglottos (216) [30]	18	0.3	[−0.207 ; 0.806]
Acrocephalus palustris (241) [7]	23	0.17	[−0.268 ; 0.609]
Acrocephalus palustris (241) [7]	23	0.498	[0.06 ; 0.937]
Acrocephalus palustris (241) [8]	31	0.581	[0.21 ; 0.951]
Acrocephalus palustris (241) [8]	31	0.714	[0.344 ; 1.084]
Acrocephalus palustris (241) [8]	31	0.055	[−0.315 ; 0.425]
Sylvia communis (367.5) [47]	12	0.67	[0.016 ; 1.323]
Luscinia megarhynchos (1160) [23]	8	0.561	[−0.315 ; 1.438]

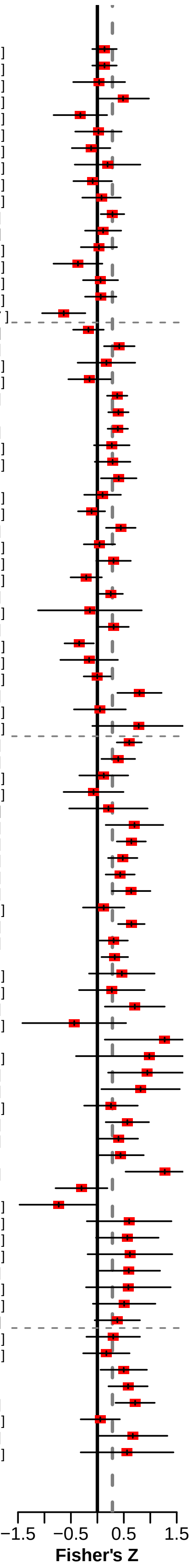


Figure A.3.2: Forest plot of the syllable repertoire dataset with territory-controlled measurements. Columns shows the study reference, species studied, the number of birds used to generate a measurement, the Fisher's Z form of the estimate, and its 95% confidence intervals. Ticks in the boxes mark the Fisher's Z and black horizontal lines show the confidence interval. Grey, dashed vertical line shows the population mean. The grey, dashed horizontal lines show the thresholds used for subsequent analysis. When the same study is listed in more than one row on the plot, multiple different metrics of reproductive success were obtained from that study. Results of the Bayesian meta-analysis and BEST analysis for discrete species average syllable repertoire size are in **Tables A.2.1** and **A.2.2** respectively, and meta-analysis results for continuous species average syllable repertoire size are in **Table A.3.7**.

Species (Rep) [citation]	**nBirds**	**Fisher's Z**	**[95% CI]**
Acrocephalus palustris (241) [7]	23	0.17	[−0.268 ; 0.609]
Acrocephalus palustris (241) [7]	23	0.498	[0.06 ; 0.937]
Acrocephalus palustris (241) [8]	31	0.581	[0.21 ; 0.951]
Acrocephalus palustris (241) [8]	31	0.714	[0.344 ; 1.084]
Acrocephalus palustris (241) [8]	31	0.055	[−0.315 ; 0.425]
Cardinalis cardinalis (14) [15]	21	−0.367	[−0.829 ; 0.095]
Emberiza schoeniclus (18.15) [18]	26	−0.635	[−1.044 ; −0.227]
Melospiza melodia (38) [24]	72	0.603	[0.367 ; 0.839]
Melospiza melodia (38) [24]	41	0.395	[0.077 ; 0.713]
Melospiza melodia (38) [25]	21	0.121	[−0.341 ; 0.583]
Melospiza melodia (38) [25]	15	−0.072	[−0.638 ; 0.494]
Melospiza melodia (38) [25]	10	0.209	[−0.532 ; 0.95]
Melospiza melodia (38) [26]	16	0.702	[0.158 ; 1.245]
Melospiza melodia (38) [22]	54	0.642	[0.368 ; 0.916]
Melospiza melodia (38) [22]	53	0.48	[0.203 ; 0.757]
Melospiza melodia (38) [22]	54	0.432	[0.157 ; 0.706]
Melospiza melodia (38) [22]	32	0.639	[0.275 ; 1.003]
Melospiza melodia (38) [28,20]	28	0.12	[−0.272 ; 0.512]
Melospiza melodia (38) [29]	64	0.643	[0.392 ; 0.894]
Melospiza melodia (38) [29]	56	0.307	[0.037 ; 0.576]
Melospiza melodia (38) [29]	64	0.329	[0.078 ; 0.58]
Parus caeruleus (6.5) [31]	19	0.034	[−0.456 ; 0.524]
Parus caeruleus (6.5) [31]	19	0.486	[−0.004 ; 0.975]
Parus major (7.9) [33]	32	−0.119	[−0.482 ; 0.245]
Parus major (7.9) [33]	13	0.193	[−0.427 ; 0.813]
Parus major (7.9) [33]	32	−0.089	[−0.453 ; 0.275]
Parus major (7.9) [33]	32	0.079	[−0.285 ; 0.443]
Parus major (7.9) [34]	81	0.286	[0.064 ; 0.508]
Acrocephalus arundinaceus (22.5) [1]	107	0.375	[0.183 ; 0.567]
Acrocephalus arundinaceus (22.5) [1]	107	0.398	[0.206 ; 0.59]
Acrocephalus arundinaceus (22.5) [1]	107	0.387	[0.194 ; 0.579]
Acrocephalus arundinaceus (22.5) [2]	37	0.273	[−0.063 ; 0.609]
Acrocephalus arundinaceus (22.5) [2]	37	0.288	[−0.048 ; 0.624]
Acrocephalus arundinaceus (22.5) [2]	37	0.404	[0.068 ; 0.74]
Acrocephalus arundinaceus (22.5) [3]	35	0.099	[−0.248 ; 0.445]
Acrocephalus arundinaceus (22.5) [3]	62	−0.11	[−0.365 ; 0.146]
Acrocephalus arundinaceus (22.5) [3]	52	0.443	[0.163 ; 0.723]
Acrocephalus arundinaceus (22.5) [3]	47	0.04	[−0.256 ; 0.335]
Acrocephalus arundinaceus (22.5) [3]	39	0.305	[−0.021 ; 0.632]
Acrocephalus arundinaceus (22.5) [3]	47	−0.211	[−0.506 ; 0.085]
Acrocephalus arundinaceus (22.5) [5]	77	0.254	[0.026 ; 0.482]
Acrocephalus schoenobaenus (55) [9]	26	0.566	[0.158 ; 0.975]
Acrocephalus schoenobaenus (55) [9]	31	0.399	[0.029 ; 0.769]
Acrocephalus schoenobaenus (55) [9]	23	0.438	[0 ; 0.877]
Acrocephalus schoenobaenus (55) [10]	10	1.274	[0.533 ; 2.015]
Acrocephalus schoenobaenus (55) [11]	19	−0.3	[−0.79 ; 0.19]
Acrocephalus schoenobaenus (55) [12]	10	−0.73	[−1.471 ; 0.011]
Agelaius phoeniceus (18.5) [13]	49	−0.169	[−0.458 ; 0.12]
Agelaius phoeniceus (18.5) [13]	49	0.415	[0.126 ; 0.704]
Dendroica pensylvanica (10.75) [17]	36	0.109	[−0.232 ; 0.45]
Dendroica pensylvanica (10.75) [17]	36	0.03	[−0.312 ; 0.371]
Ficedula albicollis (31.4) [19]	53	−0.343	[−0.62 ; −0.066]
Ficedula albicollis (31.4) [20]	16	−0.155	[−0.699 ; 0.388]
Ficedula hypoleuca (28.4) [21]	50	0.307	[0.022 ; 0.593]
Hirundo rustica (20.1) [22]	16	0.171	[−0.373 ; 0.715]
Hirundo rustica (20.1) [20]	27	−0.148	[−0.548 ; 0.252]
Luscinia megarhynchos (1160) [23]	8	0.561	[−0.315 ; 1.438]
Mimus polyglottos (216) [30]	18	0.3	[−0.207 ; 0.806]
Phylloscopus fuscatus (25.3) [35]	7	−0.141	[−1.121 ; 0.839]
Phylloscopus trochilus (35.1) [38]	62	−0.003	[−0.258 ; 0.252]
Phylloscopus trochilus (35.1) [38]	25	0.796	[0.378 ; 1.214]
Phylloscopus trochilus (35.1) [38]	19	0.048	[−0.442 ; 0.538]
Phylloscopus trochilus (35.1) [39]	8	0.782	[−0.095 ; 1.659]
Sturnella neglecta (86) [43]	9	0.601	[−0.199 ; 1.401]
Sturnella neglecta (86) [43]	14	0.566	[−0.025 ; 1.157]
Sturnella neglecta (86) [43]	9	0.615	[−0.185 ; 1.416]
Sturnella neglecta (86) [43]	14	0.594	[0.003 ; 1.185]
Sturnella neglecta (86) [43]	9	0.586	[−0.214 ; 1.386]
Sturnella neglecta (86) [43]	14	0.506	[−0.085 ; 1.097]
Sturnella neglecta (86) [44]	24	0.377	[−0.05 ; 0.805]
Sturnus vulgaris (41.95) [45]	15	0.708	[0.142 ; 1.273]
Sturnus vulgaris (41.95) [45]	7	−0.437	[−1.417 ; 0.543]
Sturnus vulgaris (41.95) [46]	6	1.272	[0.14 ; 2.403]
Sturnus vulgaris (41.95) [46]	5	0.982	[−0.404 ; 2.368]
Sturnus vulgaris (41.95) [46]	10	0.944	[0.203 ; 1.685]
Sturnus vulgaris (41.95) [46]	10	0.817	[0.076 ; 1.558]
Sylvia communis (367.5) [47]	12	0.67	[0.016 ; 1.323]

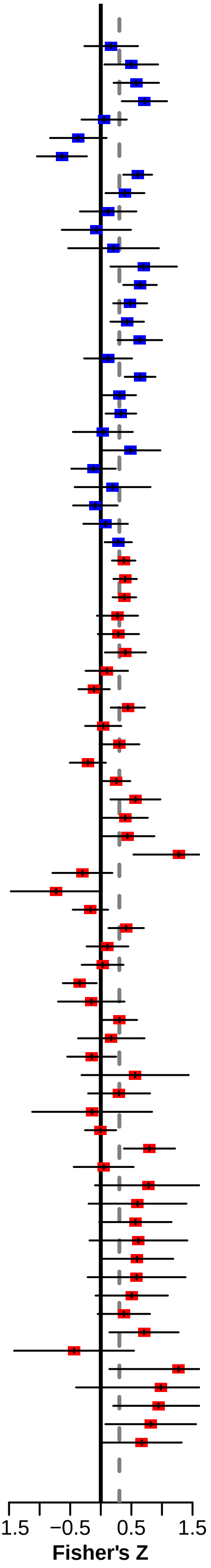


Figure A.3.3: Forest plot of the song stability dataset with territory-controlled measurements. Columns shows the study reference, species studied, the number of birds used to generate a measurement, the Fisher's Z form of the estimate, and its 95% confidence intervals. Blue boxes mark measurements in the song-stable group, while red boxes mark measurements in the song-plastic group. Ticks in the boxes mark the Fisher's Z and black horizontal lines show the confidence interval. The grey, dashed vertical line shows the population mean. When the same study is listed in more than one row on the plot, multiple different metrics of reproductive success were obtained from that study. Results of the Bayesian meta-analysis and BEST analysis are in **Tables A.3.11** and **A.3.12** respectively.

Table A.3.8: Increasing repertoire size with age was not a better predictor of the correlation between reproductive success and individual repertoire size than song stability. Performed in the song stability dataset. Three species that were labeled as song plastic (*Dendroica pensylvanica*, *Mimus polyglottos*, and *Phylloscopus fuscatus*) were put in the Not Increasing group based on information available in the literature.

Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Increasing	11	45	0.278	[−0.008;0.571]	0.056
Not Increasing	9	32	0.238	[−0.092;0.564]	0.14

Table A.3.9: Increasing repertoire with age BEST results in the song stability dataset. There was a reasonable probability that there was no difference between species with repertoires that increased with age and those whose repertoires do not. Three species that were labeled as song plastic (*Dendroica pensylvanica*, *Mimus polyglottos*, and *Phylloscopus fuscatus*) were put in the Not Increasing group based on information available in the literature.

BEST Mean	95% CredInt	%<0
0.096	[−0.087;0.278]	15.6

Table A.3.10: Switching the song stability category of most species did not significantly the song stability results. Performed in the song stability dataset. Asterisks (*) denote significant groups.

Switched	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Acrocephalus arundinaceus	Plastic	13	36	0.331	[0.036;0.633]	0.028*
Acrocephalus arundinaceus	Stable	7	41	0.168	[−0.167;0.495]	0.286
Acrocephalus palustris	Plastic	15	54	0.31	[0.034;0.594]	0.028*
Acrocephalus palustris	Stable	5	23	0.149	[−0.226;0.511]	0.39
Acrocephalus schoenobaenus	Plastic	13	43	0.305	[0.011;0.583]	0.035*
Acrocephalus schoenobaenus	Stable	7	34	0.194	[−0.151;0.514]	0.233
Agelaius phoeniceus	Plastic	13	47	0.332	[0.056;0.62]	0.021*
Agelaius phoeniceus	Stable	7	30	0.13	[−0.223;0.471]	0.423
Cardinalis cardinalis	Plastic	15	50	0.31	[0.034;0.594]	0.028*
Cardinalis cardinalis	Stable	5	27	0.149	[−0.226;0.511]	0.39
Dendroica pensylvanica	Plastic	13	47	0.305	[0.026;0.598]	0.034*
Dendroica pensylvanica	Stable	7	30	0.182	[−0.174;0.509]	0.276
Emberiza schoeniclus	Plastic	15	50	0.31	[0.034;0.594]	0.028*
Emberiza schoeniclus	Stable	5	27	0.149	[−0.226;0.511]	0.39
Ficedula albicollis	Plastic	13	47	0.35	[0.079;0.632]	0.016*
Ficedula albicollis	Stable	7	30	0.092	[−0.248;0.428]	0.557
Ficedula hypoleuca	Plastic	13	48	0.303	[0.027;0.588]	0.034*
Ficedula hypoleuca	Stable	7	29	0.187	[−0.162;0.525]	0.255
Hirundo rustica	Plastic	13	47	0.327	[0.058;0.62]	0.022*
Hirundo rustica	Stable	7	30	0.137	[−0.207;0.485]	0.408
Luscinia megarhynchos	Plastic	13	48	0.294	[0.023;0.577]	0.036*
Luscinia megarhynchos	Stable	7	29	0.194	[−0.162;0.539]	0.252
Melospiza melodia	Plastic	15	63	0.31	[0.034;0.594]	0.028*
Melospiza melodia	Stable	5	14	0.149	[−0.226;0.511]	0.39
Mimus polyglottos	Plastic	13	48	0.282	[0.01;0.57]	0.044*
Mimus polyglottos	Stable	7	29	0.22	[−0.127;0.57]	0.194
Parus caeruleus	Plastic	15	51	0.31	[0.034;0.594]	0.028*
Parus caeruleus	Stable	5	26	0.149	[−0.226;0.511]	0.39
Parus major	Plastic	15	54	0.31	[0.034;0.594]	0.028*
Parus major	Stable	5	23	0.149	[−0.226;0.511]	0.39
Phylloscopus fuscatus	Plastic	13	48	0.313	[0.036;0.594]	0.027*
Phylloscopus fuscatus	Stable	7	29	0.148	[−0.216;0.505]	0.383
Phylloscopus trochilus	Plastic	13	45	0.313	[0.033;0.606]	0.03*
Phylloscopus trochilus	Stable	7	32	0.174	[−0.173;0.518]	0.293
Sturnella neglecta	Plastic	13	42	0.298	[0.016;0.595]	0.038*
Sturnella neglecta	Stable	7	35	0.2	[−0.146;0.547]	0.235
Sturnus vulgaris	Plastic	13	43	0.287	[−0.005;0.568]	0.043*
Sturnus vulgaris	Stable	7	34	0.211	[−0.148;0.556]	0.216
Sylvia communis	Plastic	13	48	0.296	[0.021;0.579]	0.036*
Sylvia communis	Stable	7	29	0.189	[−0.172;0.526]	0.269

Table A.3.11: The strength of the correlation between individual song elaboration and reproductive success for song stability was separated from zero with weak significance. Performed in the song stability dataset meta-analysis with territory-controlled measurements. Asterisks (*) denote significant groups.

Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
Stable	14	49	0.144	[−0.201;0.487]	0.378
Plastic	6	28	0.279	[0.014;0.544]	0.036*

Table A.3.12: Song-stable and song-plastic species did not form two significantly different groups. Performed in the song stability dataset meta-analysis with territory-controlled measurements.

BEST Mean	95% CredInt	%<0
0.053	[−0.125;0.23]	28.6

Some of the metrics of reproductive success in this dataset are indicative of mate choice, such as pairing date, and others are more indicative of reproductive output, such as number of offspring. In a previous meta-analysis, Byers and Kroodsma[18] specifically addressed mate choice and individual song elaboration. Thus, we wanted to test how these results would be affected if only proxies of mate choice were analyzed, omitting proxies of reproductive output (number of offspring). This reduced the syllable repertoire dataset to 53 measurements, but all 25 species were still represented. In this model, the posterior mean was not significantly separated from zero for the correlation between individual song elaboration and success in becoming a social or genetic mate, though it was trending in this direction with the largest threshold of 216 (**Table A.3.13**). Previous analysis of the metrics of mating success had suggested that the metric for genetic mate choice (extra-pair paternity) actually showed a negative relationship between individual male repertoire size and reproductive success **Table 2.2**, which hints at the possibility that females choose social mates and extra-pair mates based on different criteria. We ran an additional model that excluded both measures related to number of offspring and extra-pair paternity, so that it only included proxies of social mate choice. This reduced the dataset to 42 measurements in 19 species. For this model, there was a trend towards a posterior mean separated from zero between individual song elaboration and success in becoming a social mate when the threshold was set at 38 or 216 (**Table A.3.14**). Interestingly, for both the dataset with social and genetic mate metrics and the dataset with social mate metrics only, the smaller repertoire size group had a smaller meta-analytic mean for all three thresholds, and BEST analysis still predicted that it was highly likely that there was a difference between the real means of the larger and smaller species average syllable repertoire size groups only when the threshold was set at 38 (**Tables A.3.15** and **A.3.16**). While we did not observe a significant correlation between song elaboration and mate choice, the trends in the

BEST results leave open the possibility that, given a larger dataset, there would be a significant association between song elaboration and social mate choice.

Table A.3.13: Relatively larger syllable repertoire size was not predictive of mating success. Performed in the dataset without offspring measurements.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<18.5	Smaller	6	7	0.071	[−0.524;0.635]	0.753
≥18.5	Larger	17	46	0.272	[−0.254;0.831]	0.195
<38	Smaller	13	21	0.078	[−0.49;0.657]	0.654
≥38	Larger	10	32	0.407	[−0.21;0.972]	0.104
<216	Smaller	19	47	0.18	[−0.401;0.803]	0.366
≥216	Larger	4	6	0.54	[−0.125;1.243]	0.087

Table A.3.14: Relatively larger species averge repertoire size was not predictive of social mating success. Performed in the dataset without offspring or extra-pair paternity measurements.

Threshold	Group	#Species	#Measure	Post Mean	95% CredInt	p_{MCMC}
<18.5	Smaller	4	4	0.14	[−0.531;0.824]	0.536
≥18.5	Larger	15	38	0.355	[−0.195;0.971]	0.097
<38	Smaller	10	15	0.151	[−0.557;0.834]	0.35
≥38	Larger	9	27	0.522	[−0.171;1.221]	0.072
<216	Smaller	16	38	0.285	[−0.378;0.949]	0.148
≥216	Larger	3	4	0.605	[−0.146;1.389]	0.076

Table A.3.15: Larger and smaller species average syllable repertoire size groups have significantly different correlations between individual elaboration and mating success only at the threshold of $\geq$ 38 syllables. BEST analysis performed in the no offspring dataset. Asterisks (*) denotes the model with significantly different groups.

Threshold	BEST Mean	95% CredInt	%<0
18.50	0.293	[−0.083;0.669]	4.3*
38	0.322	[0.108;0.537]	0.2*
216	0.227	[−0.196;0.651]	11.9

Table A.3.16: Larger and smaller syllable repertoire size groups have significantly different correlations between individual elaboration and social mating success only at the threshold of $\geq$ 38 syllables. BEST analysis performed in the dataset without offspring or extra-pair paternity measurements. Asterisk (*) denotes the model with significantly different groups.

Threshold	BEST Mean	95% CredInt	%<0
18.50	0.287	[−0.307;0.881]	9.8
38	0.294	[0.071;0.518]	0.5*
216	0.174	[−0.567;0.915]	20.9

APPENDIX B

Random effects meta-analysis

B.1 Random effects meta-analysis methods

Dataset formation for the traditional meta-analysis

I gathered data as described in the main text for the Bayesian meta-analysis (**Chapter 2**), but this traditional style of meta-analysis could not accommodate multiple measurement types from the same study. Therefore, we sorted the measurements into three categories depending on the way reproductive success was measured as detailed below. In studies where r values were obtained for multiple variables belonging to one category, only one of these measures was included in the primary dataset, so that each study could contribute only one r value per category. In these cases, measurements were chosen for each category in descending order of the number of available measurements for each variable, setting the variable priority as follows:

Latency to Pairing Date: 1) latency to pairing date, 2) latency to laying date, and 3) latency to hatching date (21 studies, 14 bird species).

Number of Offspring: 1) number of fledglings, 2) clutch size, and 3) recruits (16 studies, 9 bird species).

Number of Females: 1) number of females and 2) measurements of extra-pair paternity (17 studies, 13 bird species).

Traditional meta-analysis

With the dataset containing species for which we had information on 1) individual song elaboration and reproductive success, 2) song stability stability, and 3) species average syllable repertoire size estimates, we performed a meta-analyses using the **R** packages `meta` and `metafor`, which uses Hedges and colleagues' method[123,282]. Additionally, we used the information in Birds of

North America (birdsna.org) and Handbook of Birds of the World (hbw.com) to classify the migratory status of each bird species as sedentary, migratory, or mixed. Species that contained both sedentary and migratory individuals or populations (mixed) were labeled as migratory species in the final analysis to dichotomize the data. The literature on the role of song elaboration in reproductive success argues against the possibility that there is a common effect size across bird species; therefore, a fixed effect model would be inappropriate for this dataset. Instead, and in accordance with Soma and Garamszegi[19], we used a random-effects model, which would not assume that there is a common effect size across species. We calculated heterogeneity using the restricted maximum likelihood method (REML), as this has been suggested to be an appropriate method for meta-analysis of both continuous and dichotomous data[127,283]. Six meta-analyses were performed, two for each of the three dataset categories. In each of these analyses, the species were divided into one of two sets of subpopulations: 1) species with larger or smaller average syllable repertoire sizes with a threshold of 38 syllables or 2) song-stable and song-plastic species, as a proxy for the length of the song-learning window. Because the datasets were tested twice to examine these two subpopulation sets, we used a Bonferroni-corrected significance threshold (α=0.025).

Threshold testing

To determine the effect of the threshold value used to dichotomize species based on their average syllable repertoire size, we repeated the meta-analysis within each dataset category using each species' repertoire size as the threshold value in turn. The smallest repertoire threshold used was the value at which at least two studies were present in the smaller repertoire subpopulation. The largest threshold was the value at which there were at least two studies in the larger repertoire subpopulation. For each threshold value, we calculated whether a smaller versus larger syllable repertoire predicted an association between individual song elaboration and reproductive success. We used the same Bonferroni-corrected significance threshold for this procedure as in the meta-analyses (α=0.025).

Controlling for phylogenetic relationships

These results could be confounded by phylogenetic relationships, since closely related species can have related behaviors by chance, and thus may not represent independent samples. To control for phylogenetic effects, we performed a phylogenetically-controlled ANOVA (`phylANOVA`) using **R** package `phytools`[135]. With a list of each species in a given dataset category, we extracted a set of 1000 trees via the phylogeny subsets tool on birdtree.org[134]. We then created a consensus tree in **R** using the mean edge length method via the `consensus.edges` function. Use of a species-level phylogeny required each species to be present only once in a dataset category. Therefore, we averaged the literature r values (the correlation between a given reproductive success category and individual song elaboration) for each species that was represented by multiple studies in a dataset category. Literature r values were plotted as a continuous trait. The `phylANOVA` function was used to test for differences between song stability, species average syllable repertoire size, and migratory status while controlling for phylogeny. A Bonferroni-corrected significance threshold was used (α=0.017 for 3 ANOVAs).

Interacting variables

There could be be an interaction between species average syllable repertoire size and song stability. To examine this possibility, we divided the data in the latency to pairing date and number of offspring categories by song stability and compared the average literature r values between species with larger and smaller syllable repertoire sizes in song-stable versus song-plastic species via a Welch's two sample t-test. We repeated this process with the subpopulation variables reversed. Four t-tests were performed, so we used a Bonferroni-corrected significance threshold (α=0.0125).

Jackknife analysis

I performed a jackknife analysis to determine whether the potential miscategorization of a species' song stability could influence these results. Jackknife testing was performed first by changing the song stability grouping of one study at a time and repeating the meta-analysis to examine whether this regrouping changed the significance of the results. However, there were several cases where a single bird species was represented by multiple studies. Logically, if one of these species' song stability was miscategorized, this would change the grouping of all studies about that species. Therefore, we ran a second jackknife, wherein regrouping was performed for each bird species in turn. We used the same Bonferroni-corrected significance threshold for this procedure as in the main meta-analyses (α=0.025).

B.2 Random effects meta-analysis results

Population-Level meta-analysis

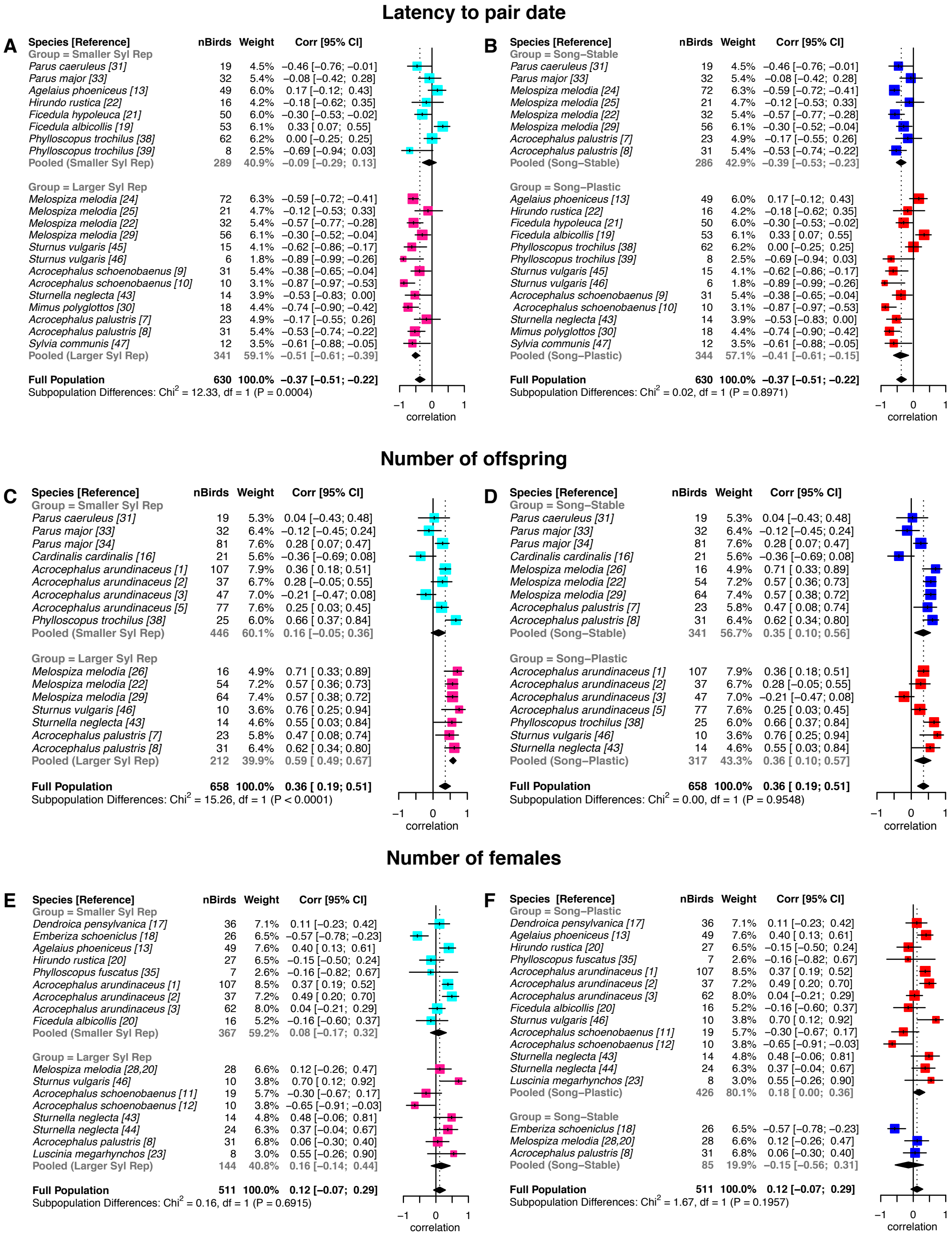
To examine whether individual song elaboration was correlated with only some indicators of reproductive success, we split the dataset into three categories: latency to pairing date (14 species, 21 studies), number of offspring (9 species, 16 studies), and number of females (13 species, 17 studies). We tested for a significant association between individual song elaboration and reproductive success in each reproductive success category. The mean effect size for **pairing date** was significant (r=0.37, I^2=74.7%, τ^2=0.1082, z=4.47, $p<0.0001$), as was the effect size for the **number of offspring** (r=0.35, I^2=74.6%, τ^2=0.1024, z=3.99, p=0.0001). In contrast, **number of females** did not have a significant effect size (r=0.12, I^2=68.3%, τ^2=0.0968, z=1.22, p=0.22).

However, the results of small, random-effects meta-analyses can be easily swayed by one or a few studies that are weighted more heavily in the analysis due to large sample sizes[151]. To examine whether any study in this analysis had this capability, we compared the range of weights in each category (W_{range}) to the weight value all studies would have if they were weighted equally (W_{equal}). W_{range} did not substantially ediviate from W_{equal} for any category (**pair date:** W_{equal}=4.76, W_{range}=[1.8; 6.3]; **offspring:** W_{equal}=6.25, W_{range}=[3.6; 7.9]; **females:** W_{equal}=5.88, W_{range}=[2.6; 8.5]). Therefore, we concluded that the data would not be skewed by any individual study due to large sample size. Furthermore, some of the studies reported the correlation between song elaboration and reproductive success before and after controlling for territory quality in the individual birds studied. We repeated these meta-analyses using a secondary dataset that exchanged non-territory-controlled measures with territory-controlled measures (only possible for three measures), which only marginally affected the results (**pair date:** r=0.34; **offspring:** r=0.36; **females:** r=0.10).

These results showed that while the effect size for each bird species is assumed to be different, the average effect size across the bird species in this study is small to moderate and negative for latency to pairing date, small to moderate and positive for number of offspring, and there is no correlation between repertoire size and the number of females a male acquires. Importantly, for latency to pairing date a negative correlation indicates higher reproductive success (individuals with greater song elaboration acquire mates faster), whereas for number of offspring a positive correlation indicates higher reproductive success (individuals with greater song elaboration produce more offspring). These results are consistent with previous studies: there is a weak correlation between individual song elaboration and reproductive success across species. However, this overall weak correlation may mask a phenomenon in which individual song elaboration is correlated with reproductive success in some species, but not in others. Thus, we sought to discover whether aspects of a species song behavior could be used divide the full set of species into subsets with different patterns of sexual selection pressure on individual song elaboration. We propose two variables as candidates for creating these subsets: species average syllable repertoire size and the duration of song-learning (closed-ended vs. open-ended learning), as proxied by song stability (stable songs vs. plastic songs).

Species average syllable repertoire size meta-analysis

Greater individual song elaboration may signal higher male intelligence, health, or fitness, decreased parasite load, increased access to resources, or other factors. Female preference for larger repertoires could in turn create a selection pressure that increases the average syllable repertoire size of the species over time. Therefore, we first split the full population into two subpopulations based on each species' average syllable repertoire size, using less than 38 syllables as the threshold for smaller syllable repertoires. Other possible thresholds were tested as explained below (**Figure B.2.3**). We reran the meta-analysis on all three dataset categories, despite the fact that on a full population level there was not a significant effect size for the correlation between individual song elaboration and the number of females attracted. There was a significant difference in the effect sizes between species with larger versus smaller average syllable repertoires for pairing date (**Figure B.2.1A**). While the average effect size for species with larger average syllable repertoires was significant, the average effect size for species with smaller average syllable repertoires was not (**Figure B.2.1A**). There was also a significant difference in the effect sizes between species with larger versus smaller average syllable repertoires for number of offspring (**Figure B.2.1C**). Once again, the effect size was only significant for species with larger average syllable repertoires and not for species with smaller average syllable repertoires (**Figure B.2.1C**). There was no significant difference between the effect sizes of species with larger versus smaller average syllable repertoires for number of females (**Figure B.2.1E**). Using the territory-controlled dataset did not significantly affect these results (**Figure B.2.2A, C, and E**).



Latency to pair date

A

Species [Reference]	nBirds	Weight	Corr [95% CI]
Group = Smaller Syl Rep			
Parus caeruleus [31]	19	4.5%	−0.46 [−0.76; −0.01]
Parus major [33]	32	5.4%	−0.08 [−0.42; 0.28]
Agelaius phoeniceus [13]	49	6.0%	0.17 [−0.12; 0.43]
Hirundo rustica [22]	16	4.2%	−0.18 [−0.62; 0.35]
Ficedula hypoleuca [21]	50	6.0%	−0.30 [−0.53; −0.02]
Ficedula albicollis [19]	53	6.1%	0.33 [0.07; 0.55]
Phylloscopus trochilus [38]	62	6.2%	0.00 [−0.25; 0.25]
Phylloscopus trochilus [39]	8	2.5%	−0.69 [−0.94; 0.03]
Pooled (Smaller Syl Rep)	**289**	**40.9%**	**−0.09 [−0.29; 0.13]**
Group = Larger Syl Rep			
Melospiza melodia [24]	72	6.3%	−0.59 [−0.72; −0.41]
Melospiza melodia [25]	21	4.7%	−0.12 [−0.53; 0.33]
Melospiza melodia [22]	32	5.4%	−0.57 [−0.77; −0.28]
Melospiza melodia [29]	56	6.1%	−0.30 [−0.52; −0.04]
Sturnus vulgaris [45]	15	4.1%	−0.62 [−0.86; −0.17]
Sturnus vulgaris [46]	6	1.8%	−0.89 [−0.99; −0.26]
Acrocephalus schoenobaenus [9]	31	5.4%	−0.38 [−0.65; −0.04]
Acrocephalus schoenobaenus [10]	10	3.1%	−0.87 [−0.97; −0.53]
Sturnella neglecta [43]	14	3.9%	−0.53 [−0.83; 0.00]
Mimus polyglottos [30]	18	4.4%	−0.74 [−0.90; −0.42]
Acrocephalus palustris [7]	23	4.9%	−0.17 [−0.55; 0.26]
Acrocephalus palustris [8]	31	5.4%	−0.53 [−0.74; −0.22]
Sylvia communis [47]	12	3.5%	−0.61 [−0.88; −0.05]
Pooled (Larger Syl Rep)	**341**	**59.1%**	**−0.51 [−0.61; −0.39]**
Full Population	**630**	**100.0%**	**−0.37 [−0.51; −0.22]**

Subpopulation Differences: $\text{Chi}^2 = 12.33$, df = 1 (P = 0.0004)

B

Species [Reference]	nBirds	Weight	Corr [95% CI]
Group = Song–Stable			
Parus caeruleus [31]	19	4.5%	−0.46 [−0.76; −0.01]
Parus major [33]	32	5.4%	−0.08 [−0.42; 0.28]
Melospiza melodia [24]	72	6.3%	−0.59 [−0.72; −0.41]
Melospiza melodia [25]	21	4.7%	−0.12 [−0.53; 0.33]
Melospiza melodia [22]	32	5.4%	−0.57 [−0.77; −0.28]
Melospiza melodia [29]	56	6.1%	−0.30 [−0.52; −0.04]
Acrocephalus palustris [7]	23	4.9%	−0.17 [−0.55; 0.26]
Acrocephalus palustris [8]	31	5.4%	−0.53 [−0.74; −0.22]
Pooled (Song–Stable)	**286**	**42.9%**	**−0.39 [−0.53; −0.23]**
Group = Song–Plastic			
Agelaius phoeniceus [13]	49	6.0%	0.17 [−0.12; 0.43]
Hirundo rustica [22]	16	4.2%	−0.18 [−0.62; 0.35]
Ficedula hypoleuca [21]	50	6.0%	−0.30 [−0.53; −0.02]
Ficedula albicollis [19]	53	6.1%	0.33 [0.07; 0.55]
Phylloscopus trochilus [38]	62	6.2%	0.00 [−0.25; 0.25]
Phylloscopus trochilus [39]	8	2.5%	−0.69 [−0.94; 0.03]
Sturnus vulgaris [45]	15	4.1%	−0.62 [−0.86; −0.17]
Sturnus vulgaris [46]	6	1.8%	−0.89 [−0.99; −0.26]
Acrocephalus schoenobaenus [9]	31	5.4%	−0.38 [−0.65; −0.04]
Acrocephalus schoenobaenus [10]	10	3.1%	−0.87 [−0.97; −0.53]
Sturnella neglecta [43]	14	3.9%	−0.53 [−0.83; 0.00]
Mimus polyglottos [30]	18	4.4%	−0.74 [−0.90; −0.42]
Sylvia communis [47]	12	3.5%	−0.61 [−0.88; −0.05]
Pooled (Song–Plastic)	**344**	**57.1%**	**−0.41 [−0.61; −0.15]**
Full Population	**630**	**100.0%**	**−0.37 [−0.51; −0.22]**

Subpopulation Differences: $\text{Chi}^2 = 0.02$, df = 1 (P = 0.8971)

Number of offspring

C

Species [Reference]	nBirds	Weight	Corr [95% CI]
Group = Smaller Syl Rep			
Parus caeruleus [31]	19	5.3%	0.04 [−0.43; 0.48]
Parus major [33]	32	6.4%	−0.12 [−0.45; 0.24]
Parus major [34]	81	7.6%	0.28 [0.07; 0.47]
Cardinalis cardinalis [16]	21	5.6%	−0.36 [−0.69; 0.08]
Acrocephalus arundinaceus [1]	107	7.9%	0.36 [0.18; 0.51]
Acrocephalus arundinaceus [2]	37	6.7%	0.28 [−0.05; 0.55]
Acrocephalus arundinaceus [3]	47	7.0%	−0.21 [−0.47; 0.08]
Acrocephalus arundinaceus [5]	77	7.6%	0.25 [0.03; 0.45]
Phylloscopus trochilus [38]	25	6.0%	0.66 [0.37; 0.84]
Pooled (Smaller Syl Rep)	**446**	**60.1%**	**0.16 [−0.05; 0.36]**
Group = Larger Syl Rep			
Melospiza melodia [26]	16	4.9%	0.71 [0.33; 0.89]
Melospiza melodia [22]	54	7.2%	0.57 [0.36; 0.73]
Melospiza melodia [29]	64	7.4%	0.57 [0.38; 0.72]
Sturnus vulgaris [46]	10	3.6%	0.76 [0.25; 0.94]
Sturnella neglecta [43]	14	4.6%	0.55 [0.03; 0.84]
Acrocephalus palustris [7]	23	5.8%	0.47 [0.08; 0.74]
Acrocephalus palustris [8]	31	6.4%	0.62 [0.34; 0.80]
Pooled (Larger Syl Rep)	**212**	**39.9%**	**0.59 [0.49; 0.67]**
Full Population	**658**	**100.0%**	**0.36 [0.19; 0.51]**

Subpopulation Differences: $\text{Chi}^2 = 15.26$, df = 1 (P < 0.0001)

D

Species [Reference]	nBirds	Weight	Corr [95% CI]
Group = Song–Stable			
Parus caeruleus [31]	19	5.3%	0.04 [−0.43; 0.48]
Parus major [33]	32	6.4%	−0.12 [−0.45; 0.24]
Parus major [34]	81	7.6%	0.28 [0.07; 0.47]
Cardinalis cardinalis [16]	21	5.6%	−0.36 [−0.69; 0.08]
Melospiza melodia [26]	16	4.9%	0.71 [0.33; 0.89]
Melospiza melodia [22]	54	7.2%	0.57 [0.36; 0.73]
Melospiza melodia [29]	64	7.4%	0.57 [0.38; 0.72]
Acrocephalus palustris [7]	23	5.8%	0.47 [0.08; 0.74]
Acrocephalus palustris [8]	31	6.4%	0.62 [0.34; 0.80]
Pooled (Song–Stable)	**341**	**56.7%**	**0.35 [0.10; 0.56]**
Group = Song–Plastic			
Acrocephalus arundinaceus [1]	107	7.9%	0.36 [0.18; 0.51]
Acrocephalus arundinaceus [2]	37	6.7%	0.28 [−0.05; 0.55]
Acrocephalus arundinaceus [3]	47	7.0%	−0.21 [−0.47; 0.08]
Acrocephalus arundinaceus [5]	77	7.6%	0.25 [0.03; 0.45]
Phylloscopus trochilus [38]	25	6.0%	0.66 [0.37; 0.84]
Sturnus vulgaris [46]	10	3.6%	0.76 [0.25; 0.94]
Sturnella neglecta [43]	14	4.6%	0.55 [0.03; 0.84]
Pooled (Song–Plastic)	**317**	**43.3%**	**0.36 [0.10; 0.57]**
Full Population	**658**	**100.0%**	**0.36 [0.19; 0.51]**

Subpopulation Differences: $\text{Chi}^2 = 0.00$, df = 1 (P = 0.9548)

Number of females

E

Species [Reference]	nBirds	Weight	Corr [95% CI]
Group = Smaller Syl Rep			
Dendroica pensylvanica [17]	36	7.1%	0.11 [−0.23; 0.42]
Emberiza schoeniclus [18]	26	6.5%	−0.57 [−0.78; −0.23]
Agelaius phoeniceus [13]	49	7.6%	0.40 [0.13; 0.61]
Hirundo rustica [20]	27	6.5%	−0.15 [−0.50; 0.24]
Phylloscopus fuscatus [35]	7	2.6%	−0.16 [−0.82; 0.67]
Acrocephalus arundinaceus [1]	107	8.5%	0.37 [0.19; 0.52]
Acrocephalus arundinaceus [2]	37	7.2%	0.49 [0.20; 0.70]
Acrocephalus arundinaceus [3]	62	8.0%	0.04 [−0.21; 0.29]
Ficedula albicollis [20]	16	5.2%	−0.16 [−0.60; 0.37]
Pooled (Smaller Syl Rep)	**367**	**59.2%**	**0.08 [−0.17; 0.32]**
Group = Larger Syl Rep			
Melospiza melodia [28,20]	28	6.6%	0.12 [−0.26; 0.47]
Sturnus vulgaris [46]	10	3.8%	0.70 [0.12; 0.92]
Acrocephalus schoenobaenus [11]	19	5.7%	−0.30 [−0.67; 0.17]
Acrocephalus schoenobaenus [12]	10	3.8%	−0.65 [−0.91; −0.03]
Sturnella neglecta [43]	14	4.8%	0.48 [−0.06; 0.81]
Sturnella neglecta [44]	24	6.3%	0.37 [−0.04; 0.67]
Acrocephalus palustris [8]	31	6.8%	0.06 [−0.30; 0.40]
Luscinia megarhynchos [23]	8	3.0%	0.55 [−0.26; 0.90]
Pooled (Larger Syl Rep)	**144**	**40.8%**	**0.16 [−0.14; 0.44]**
Full Population	**511**	**100.0%**	**0.12 [−0.07; 0.29]**

Subpopulation Differences: $\text{Chi}^2 = 0.16$, df = 1 (P = 0.6915)

F

Species [Reference]	nBirds	Weight	Corr [95% CI]
Group = Song–Plastic			
Dendroica pensylvanica [17]	36	7.1%	0.11 [−0.23; 0.42]
Agelaius phoeniceus [13]	49	7.6%	0.40 [0.13; 0.61]
Hirundo rustica [20]	27	6.5%	−0.15 [−0.50; 0.24]
Phylloscopus fuscatus [35]	7	2.6%	−0.16 [−0.82; 0.67]
Acrocephalus arundinaceus [1]	107	8.5%	0.37 [0.19; 0.52]
Acrocephalus arundinaceus [2]	37	7.2%	0.49 [0.20; 0.70]
Acrocephalus arundinaceus [3]	62	8.0%	0.04 [−0.21; 0.29]
Ficedula albicollis [20]	16	5.2%	−0.16 [−0.60; 0.37]
Sturnus vulgaris [46]	10	3.8%	0.70 [0.12; 0.92]
Acrocephalus schoenobaenus [11]	19	5.7%	−0.30 [−0.67; 0.17]
Acrocephalus schoenobaenus [12]	10	3.8%	−0.65 [−0.91; −0.03]
Sturnella neglecta [43]	14	4.8%	0.48 [−0.06; 0.81]
Sturnella neglecta [44]	24	6.3%	0.37 [−0.04; 0.67]
Luscinia megarhynchos [23]	8	3.0%	0.55 [−0.26; 0.90]
Pooled (Song–Plastic)	**426**	**80.1%**	**0.18 [0.00; 0.36]**
Group = Song–Stable			
Emberiza schoeniclus [18]	26	6.5%	−0.57 [−0.78; −0.23]
Melospiza melodia [28,20]	28	6.6%	0.12 [−0.26; 0.47]
Acrocephalus palustris [8]	31	6.8%	0.06 [−0.30; 0.40]
Pooled (Song–Stable)	**85**	**19.9%**	**−0.15 [−0.56; 0.31]**
Full Population	**511**	**100.0%**	**0.12 [−0.07; 0.29]**

Subpopulation Differences: $\text{Chi}^2 = 1.67$, df = 1 (P = 0.1957)

Figure B.2.1: Large species average syllable repertoires are predictive of moderate to strong correlations between individual song elaboration and reproductive success. Each panels shows the forest plot for the correlation between individual song elaboration and reproductive success with full model data present at the bottom of each analysis. The full population was split in half based on species syllable repertoire size or song stability. When pairing date was used as the metric for reproductive success (A-B), more negative correlations indicate higher reproductive success. When number of offspring or number of females was used as the metric for reproductive success (C-F), more positive correlations indicate higher reproductive success. Because all datasets were tested twice, a Bonferroni-corrected significance threshold was used in determining significance (α=0.025).

(A) There was a significant difference in effect size between species with smaller average syllable repertoires and larger average syllable repertoires (**Smaller:** N=8, r=0.09, I^2=64.9%, τ^2=0.0592, **Larger:** N=13, r=0.51, I^2=41.9%, τ^2=0.0262, χ^2=23.33, p=0.0004). The effect size was not significant for species with **smaller average syllable repertoires** (z=0.77, p=0.44). The effect size was significant for species with **larger average syllable repertoires** (z=7.23, $p<0.0001$).

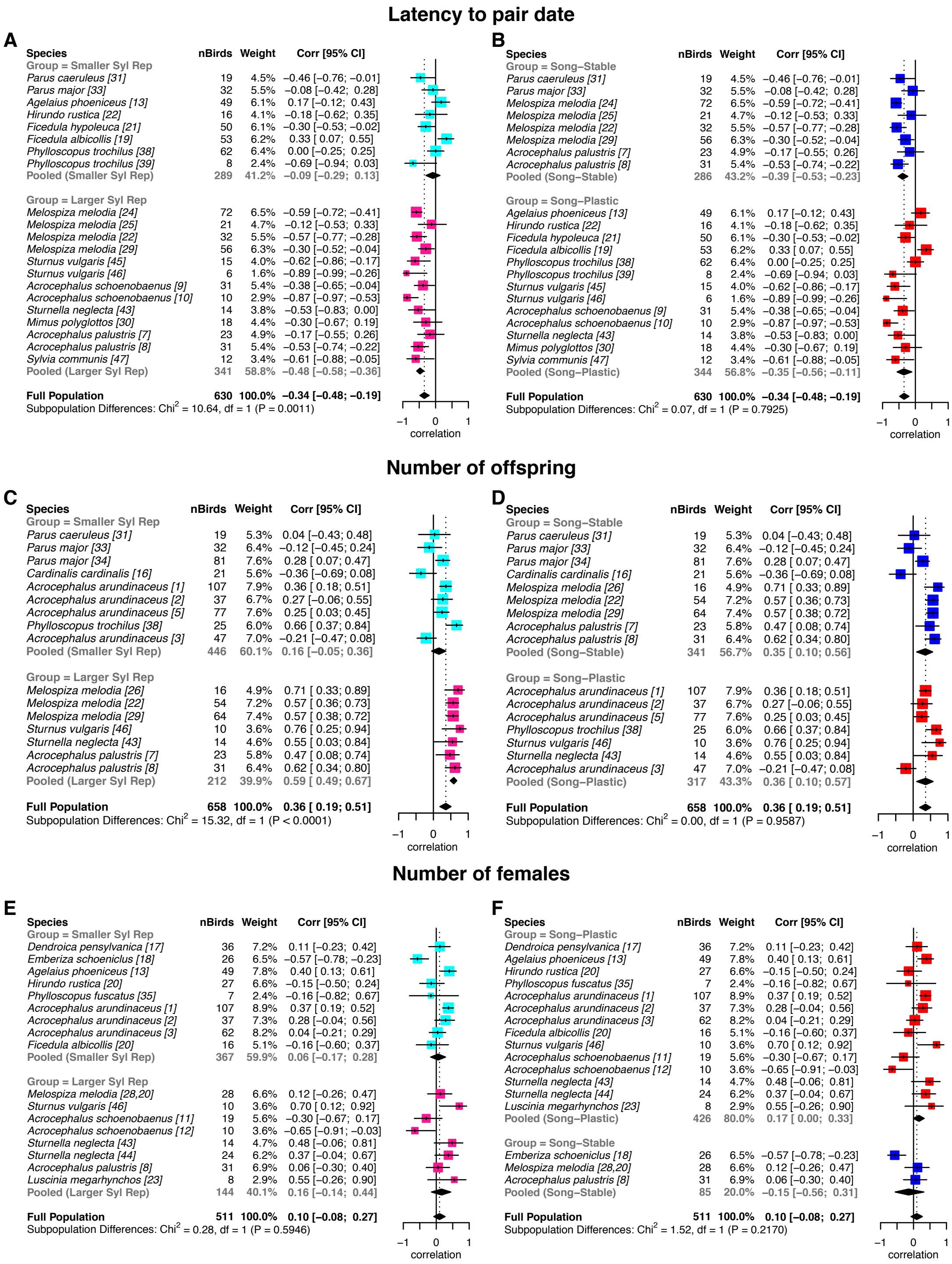
(B) There was no significant difference in effect size between song-stable and song-plastic species (**Stable:** N=8, r=0.41, I^2=78.8%, τ^2=0.1935, **Plastic:** N=13, r=0.39, I^2=48.7%, τ^2=0.0303, χ^2=0.02, p=0.89).

(C) There was a significant difference in effect size between species with smaller average syllable repertoires and species with larger average syllable repertoires (**Smaller:** N=8, r=0.16, I^2=72.5%, τ^2=0.0777, **Larger:** N=8, r=0.59, $I^2<0.1$%, $\tau^2<0.0001$, χ^2=15.26, p=0.0001. The effect size was not significant for species with **smaller average syllable repertoires** (z=1.49, p=0.14). The effect size was significant for species with **larger average syllable repertoires** (z=9.29, $p<0.0001$).

(D) There was no significant difference in effect size between song-stable and song-plastic species (**Stable:** N=9, r=0.36, I^2=73.2%, τ^2=0.0968, **Plastic:** N=7, r=0.35, I^2=77.1%, τ^2=0.1276, $\chi^2<0.01$, p=0.95).

(E) There was no significant difference in effect size between species with smaller average syllable repertoires and larger average syllable repertoires (**Smaller:** N=7, r=0.08, I^2=75.9%, τ^2=0.1064, **Larger:** N=10, r=0.12, I^2=59%, τ^2=0.1167, χ^2=0.16, p=0.69).

(F) There was no significant difference in effect size between song-stable and song-plastic species (**Stable:** N=4, r=0.12, I^2=60.6%, τ^2=0.0685, **Plastic:** N=13, r=0.18, I^2=77.1%, τ^2=0.1282, χ^2=1.67, p=0.20).



Latency to pair date

A

Species	nBirds	Weight	Corr [95% CI]
Group = Smaller Syl Rep			
Parus caeruleus [31]	19	4.5%	−0.46 [−0.76; −0.01]
Parus major [33]	32	5.5%	−0.08 [−0.42; 0.28]
Agelaius phoeniceus [13]	49	6.1%	0.17 [−0.12; 0.43]
Hirundo rustica [22]	16	4.1%	−0.18 [−0.62; 0.35]
Ficedula hypoleuca [21]	50	6.1%	−0.30 [−0.53; −0.02]
Ficedula albicollis [19]	53	6.2%	0.33 [0.07; 0.55]
Phylloscopus trochilus [38]	62	6.4%	0.00 [−0.25; 0.25]
Phylloscopus trochilus [39]	8	2.4%	−0.69 [−0.94; 0.03]
Pooled (Smaller Syl Rep)	**289**	**41.2%**	**−0.09 [−0.29; 0.13]**
Group = Larger Syl Rep			
Melospiza melodia [24]	72	6.5%	−0.59 [−0.72; −0.41]
Melospiza melodia [25]	21	4.7%	−0.12 [−0.53; 0.33]
Melospiza melodia [22]	32	5.5%	−0.57 [−0.77; −0.28]
Melospiza melodia [29]	56	6.3%	−0.30 [−0.52; −0.04]
Sturnus vulgaris [45]	15	4.0%	−0.62 [−0.86; −0.17]
Sturnus vulgaris [46]	6	1.6%	−0.89 [−0.99; −0.26]
Acrocephalus schoenobaenus [9]	31	5.4%	−0.38 [−0.65; −0.04]
Acrocephalus schoenobaenus [10]	10	2.9%	−0.87 [−0.97; −0.53]
Sturnella neglecta [43]	14	3.8%	−0.53 [−0.83; 0.00]
Mimus polyglottos [30]	18	4.4%	−0.30 [−0.67; 0.19]
Acrocephalus palustris [7]	23	4.9%	−0.17 [−0.55; 0.26]
Acrocephalus palustris [8]	31	5.4%	−0.53 [−0.74; −0.22]
Sylvia communis [47]	12	3.4%	−0.61 [−0.88; −0.05]
Pooled (Larger Syl Rep)	**341**	**58.8%**	**−0.48 [−0.58; −0.36]**
Full Population	**630**	**100.0%**	**−0.34 [−0.48; −0.19]**

Subpopulation Differences: $\text{Chi}^2 = 10.64$, df = 1 (P = 0.0011)

B

Species	nBirds	Weight	Corr [95% CI]
Group = Song–Stable			
Parus caeruleus [31]	19	4.5%	−0.46 [−0.76; −0.01]
Parus major [33]	32	5.5%	−0.08 [−0.42; 0.28]
Melospiza melodia [24]	72	6.5%	−0.59 [−0.72; −0.41]
Melospiza melodia [25]	21	4.7%	−0.12 [−0.53; 0.33]
Melospiza melodia [22]	32	5.5%	−0.57 [−0.77; −0.28]
Melospiza melodia [29]	56	6.3%	−0.30 [−0.52; −0.04]
Acrocephalus palustris [7]	23	4.9%	−0.17 [−0.55; 0.26]
Acrocephalus palustris [8]	31	5.4%	−0.53 [−0.74; −0.22]
Pooled (Song–Stable)	**286**	**43.2%**	**−0.39 [−0.53; −0.23]**
Group = Song–Plastic			
Agelaius phoeniceus [13]	49	6.1%	0.17 [−0.12; 0.43]
Hirundo rustica [22]	16	4.1%	−0.18 [−0.62; 0.35]
Ficedula hypoleuca [21]	50	6.1%	−0.30 [−0.53; −0.02]
Ficedula albicollis [19]	53	6.2%	0.33 [0.07; 0.55]
Phylloscopus trochilus [38]	62	6.4%	0.00 [−0.25; 0.25]
Phylloscopus trochilus [39]	8	2.4%	−0.69 [−0.94; 0.03]
Sturnus vulgaris [45]	15	4.0%	−0.62 [−0.86; −0.17]
Sturnus vulgaris [46]	6	1.6%	−0.89 [−0.99; −0.26]
Acrocephalus schoenobaenus [9]	31	5.4%	−0.38 [−0.65; −0.04]
Acrocephalus schoenobaenus [10]	10	2.9%	−0.87 [−0.97; −0.53]
Sturnella neglecta [43]	14	3.8%	−0.53 [−0.83; 0.00]
Mimus polyglottos [30]	18	4.4%	−0.30 [−0.67; 0.19]
Sylvia communis [47]	12	3.4%	−0.61 [−0.88; −0.05]
Pooled (Song–Plastic)	**344**	**56.8%**	**−0.35 [−0.56; −0.11]**
Full Population	**630**	**100.0%**	**−0.34 [−0.48; −0.19]**

Subpopulation Differences: $\text{Chi}^2 = 0.07$, df = 1 (P = 0.7925)

Number of offspring

C

Species	nBirds	Weight	Corr [95% CI]
Group = Smaller Syl Rep			
Parus caeruleus [31]	19	5.3%	0.04 [−0.43; 0.48]
Parus major [33]	32	6.4%	−0.12 [−0.45; 0.24]
Parus major [34]	81	7.6%	0.28 [0.07; 0.47]
Cardinalis cardinalis [16]	21	5.6%	−0.36 [−0.69; 0.08]
Acrocephalus arundinaceus [1]	107	7.9%	0.36 [0.18; 0.51]
Acrocephalus arundinaceus [2]	37	6.7%	0.27 [−0.06; 0.55]
Acrocephalus arundinaceus [5]	77	7.6%	0.25 [0.03; 0.45]
Phylloscopus trochilus [38]	25	6.0%	0.66 [0.37; 0.84]
Acrocephalus arundinaceus [3]	47	7.0%	−0.21 [−0.47; 0.08]
Pooled (Smaller Syl Rep)	**446**	**60.1%**	**0.16 [−0.05; 0.36]**
Group = Larger Syl Rep			
Melospiza melodia [26]	16	4.9%	0.71 [0.33; 0.89]
Melospiza melodia [22]	54	7.2%	0.57 [0.36; 0.73]
Melospiza melodia [29]	64	7.4%	0.57 [0.38; 0.72]
Sturnus vulgaris [46]	10	3.6%	0.76 [0.25; 0.94]
Sturnella neglecta [43]	14	4.6%	0.55 [0.03; 0.84]
Acrocephalus palustris [7]	23	5.8%	0.47 [0.08; 0.74]
Acrocephalus palustris [8]	31	6.4%	0.62 [0.34; 0.80]
Pooled (Larger Syl Rep)	**212**	**39.9%**	**0.59 [0.49; 0.67]**
Full Population	**658**	**100.0%**	**0.36 [0.19; 0.51]**

Subpopulation Differences: $\text{Chi}^2 = 15.32$, df = 1 ($P < 0.0001$)

D

Species	nBirds	Weight	Corr [95% CI]
Group = Song–Stable			
Parus caeruleus [31]	19	5.3%	0.04 [−0.43; 0.48]
Parus major [33]	32	6.4%	−0.12 [−0.45; 0.24]
Parus major [34]	81	7.6%	0.28 [0.07; 0.47]
Cardinalis cardinalis [16]	21	5.6%	−0.36 [−0.69; 0.08]
Melospiza melodia [26]	16	4.9%	0.71 [0.33; 0.89]
Melospiza melodia [22]	54	7.2%	0.57 [0.36; 0.73]
Melospiza melodia [29]	64	7.4%	0.57 [0.38; 0.72]
Acrocephalus palustris [7]	23	5.8%	0.47 [0.08; 0.74]
Acrocephalus palustris [8]	31	6.4%	0.62 [0.34; 0.80]
Pooled (Song–Stable)	**341**	**56.7%**	**0.35 [0.10; 0.56]**
Group = Song–Plastic			
Acrocephalus arundinaceus [1]	107	7.9%	0.36 [0.18; 0.51]
Acrocephalus arundinaceus [2]	37	6.7%	0.27 [−0.06; 0.55]
Acrocephalus arundinaceus [5]	77	7.6%	0.25 [0.03; 0.45]
Phylloscopus trochilus [38]	25	6.0%	0.66 [0.37; 0.84]
Sturnus vulgaris [46]	10	3.6%	0.76 [0.25; 0.94]
Sturnella neglecta [43]	14	4.6%	0.55 [0.03; 0.84]
Acrocephalus arundinaceus [3]	47	7.0%	−0.21 [−0.47; 0.08]
Pooled (Song–Plastic)	**317**	**43.3%**	**0.36 [0.10; 0.57]**
Full Population	**658**	**100.0%**	**0.36 [0.19; 0.51]**

Subpopulation Differences: $\text{Chi}^2 = 0.00$, df = 1 (P = 0.9587)

Number of females

E

Species	nBirds	Weight	Corr [95% CI]
Group = Smaller Syl Rep			
Dendroica pensylvanica [17]	36	7.2%	0.11 [−0.23; 0.42]
Emberiza schoeniclus [18]	26	6.5%	−0.57 [−0.78; −0.23]
Agelaius phoeniceus [13]	49	7.8%	0.40 [0.13; 0.61]
Hirundo rustica [20]	27	6.6%	−0.15 [−0.50; 0.24]
Phylloscopus fuscatus [35]	7	2.4%	−0.16 [−0.82; 0.67]
Acrocephalus arundinaceus [1]	107	8.9%	0.37 [0.19; 0.52]
Acrocephalus arundinaceus [2]	37	7.3%	0.28 [−0.04; 0.56]
Acrocephalus arundinaceus [3]	62	8.2%	0.04 [−0.21; 0.29]
Ficedula albicollis [20]	16	5.1%	−0.16 [−0.60; 0.37]
Pooled (Smaller Syl Rep)	**367**	**59.9%**	**0.06 [−0.17; 0.28]**
Group = Larger Syl Rep			
Melospiza melodia [28,20]	28	6.6%	0.12 [−0.26; 0.47]
Sturnus vulgaris [46]	10	3.6%	0.70 [0.12; 0.92]
Acrocephalus schoenobaenus [11]	19	5.6%	−0.30 [−0.67; 0.17]
Acrocephalus schoenobaenus [12]	10	3.6%	−0.65 [−0.91; −0.03]
Sturnella neglecta [43]	14	4.7%	0.48 [−0.06; 0.81]
Sturnella neglecta [44]	24	6.2%	0.37 [−0.04; 0.67]
Acrocephalus palustris [8]	31	6.9%	0.06 [−0.30; 0.40]
Luscinia megarhynchos [23]	8	2.9%	0.55 [−0.26; 0.90]
Pooled (Larger Syl Rep)	**144**	**40.1%**	**0.16 [−0.14; 0.44]**
Full Population	**511**	**100.0%**	**0.10 [−0.08; 0.27]**

Subpopulation Differences: $\text{Chi}^2 = 0.28$, df = 1 (P = 0.5946)

F

Species	nBirds	Weight	Corr [95% CI]
Group = Song–Plastic			
Dendroica pensylvanica [17]	36	7.2%	0.11 [−0.23; 0.42]
Agelaius phoeniceus [13]	49	7.8%	0.40 [0.13; 0.61]
Hirundo rustica [20]	27	6.6%	−0.15 [−0.50; 0.24]
Phylloscopus fuscatus [35]	7	2.4%	−0.16 [−0.82; 0.67]
Acrocephalus arundinaceus [1]	107	8.9%	0.37 [0.19; 0.52]
Acrocephalus arundinaceus [2]	37	7.3%	0.28 [−0.04; 0.56]
Acrocephalus arundinaceus [3]	62	8.2%	0.04 [−0.21; 0.29]
Ficedula albicollis [20]	16	5.1%	−0.16 [−0.60; 0.37]
Sturnus vulgaris [46]	10	3.6%	0.70 [0.12; 0.92]
Acrocephalus schoenobaenus [11]	19	5.6%	−0.30 [−0.67; 0.17]
Acrocephalus schoenobaenus [12]	10	3.6%	−0.65 [−0.91; −0.03]
Sturnella neglecta [43]	14	4.7%	0.48 [−0.06; 0.81]
Sturnella neglecta [44]	24	6.2%	0.37 [−0.04; 0.67]
Luscinia megarhynchos [23]	8	2.9%	0.55 [−0.26; 0.90]
Pooled (Song–Plastic)	**426**	**80.0%**	**0.17 [0.00; 0.33]**
Group = Song–Stable			
Emberiza schoeniclus [18]	26	6.5%	−0.57 [−0.78; −0.23]
Melospiza melodia [28,20]	28	6.6%	0.12 [−0.26; 0.47]
Acrocephalus palustris [8]	31	6.9%	0.06 [−0.30; 0.40]
Pooled (Song–Stable)	**85**	**20.0%**	**−0.15 [−0.56; 0.31]**
Full Population	**511**	**100.0%**	**0.10 [−0.08; 0.27]**

Subpopulation Differences: $\text{Chi}^2 = 1.52$, df = 1 (P = 0.2170)

Figure B.2.2: Using territory-quality-controlled measurements did not significantly affect any results. These meta-analyses were conducted identically to those explained in **Figure B.2.1**, however three measurements were exchanged with territory-controlled counterparts. For **latency to pairing date**, the full population had a significant effect size (N=21, r=-0.34, I^2=72.5%, τ^2=0.092, z=-4.32, $p<0.0001$). For **number of offspring**, the full population had a significant effect size (N=16, r=0.36, I^2=74.6%, τ^2=0.10, z=3.38, $p<0.0001$). For **number of females**, the full population did not have a significant effect size (N=17, r=0.10, I^2=65.4%, τ^2=0.082, z=1.11, p=0.27).

(A) There was a significant difference in effect size between species with smaller average syllable repertoires and larger average syllable repertoires (**Smaller:** N=8, r=-0.09, I^2=64.9%, τ^2=0.059, **Larger:** N=13, r=-0.48, $I^2<35.9\%$, $\tau^2<0.0.17$, χ^2=10.64, p=0.0011). The effect size was not significant for species with **smaller average syllable repertoires** (r=-0.09, z=-0.77, p=0.44). The effect size was significant for species with **larger average syllable repertoires** (r=-0.48, z=-7.23, $p<0.0001$)

(B) There was no significant difference in effect size between song-stable and song-plastic species (**Stable:** N=8, r=-0.39, I^2=48.7%, τ^2=0.030, **Plastic:** N=13, r=-0.35, I^2=74.9%, τ^2=0.16, χ^2=0.07, p=0.79).

(C) There was a significant difference in effect size between species with smaller average syllable repertoires and larger average syllable repertoires (**Smaller:** N=8, r=0.16, I^2=72.5%, τ^2=0.078, **Larger:** N=8, r=0.59, $I^2<0.01\%$, $\tau^2<0.01$, χ^2=15.32, $p<0.0001$). The effect size was trending significant for species with **smaller average syllable repertoires** (r=0.16, z=2.03, p=0.0443). The effect size was significant for species with **larger average syllable repertoires** (r=0.59, z=9.29, $p<0.0001$).

(D) There was no significant difference in effect size between song-stable and song-plastic species (**Stable:** N=9, r=0.35, I^2=77.1%, τ^2=0.13, **Plastic:** N=7, r=-0.35, I^2=73.3%, τ^2=0.10, $\chi^2<0.01$, p=0.96).

(E) There was no significant difference in effect size between species with smaller average syllable repertoires and larger average syllable repertoires (**Smaller:** N=9, r=0.06, I^2=72.5%, τ^2=0.085, **Larger:** N=10, r=0.16, I^2=59%, τ^2=0.12, χ^2=0.10, p=0.75).

(F) There was no significant difference in effect size between song-stable and song-plastic species (**Stable:** N=13, r=0.17, I^2=77.1%, τ^2=0.14, **Plastic:** N=4, r=-0.15, I^2=56.4%, τ^2=0.0505, χ^2=1.38, p=0.24).

These results reveal that species average syllable repertoire size could potentially be used as a metric to predict the importance of individual song elaboration in reproductive success. The data predict that on average, species with larger syllable repertoire sizes will show significant correlations between individual song elaboration and reproductive success (pair $\text{date}_{95\%\text{ConInt}}$=[$-0.61$;$-0.39$], $\text{offspring}_{95\%\text{ConInt}}$=[0.49;0.67]), while species with smaller average syllable repertoires will not (pair $\text{date}_{95\%\text{ConInt}}$=[$-0.29$;0.13], $\text{offspring}_{95\%\text{ConInt}}$=[$-0.05$;0.36]) (**Figure B.2.1A, C, and E**). This is not to say that bird species with smaller repertoires would never show a correlation between individual song elaboration and reproductive success, but rather the data suggest that smaller average syllable repertoires are not predictive of such a correlation, while larger syllable average repertoires are.

It cannot be ignored, however, that the selection of a species repertoire threshold at 38 syllables might not be meaningful for the species in this study or species in general. Therefore, we reran the meta-analysis for all three categories with a range of thresholds, and compiled the p-values. If we had chosen a realistic threshold value, we would expect to see that repertoire threshold similar to 38 syllables would also lead to significant group differences, while those that were further away from 38 syllables would not be significant, creating a U-shaped curve of p-values with a trough near 38 syllables.

Indeed, this analysis revealed a U-shaped p-value pattern for latency to pairing date and number of offspring (**Figure B.2.3A and B**). On the other hand, for number of females there had been no significant difference between species with larger and smaller average syllable repertoires when using a threshold of 38 syllables ($ln(38) \approx 3.6$), and we did not expect any other threshold value would lead to significant results. These expectations were confirmed; when the number of females dataset category was used to test alternative threshold values, no threshold led to significant results (**Figure B.2.3C**). The overall plot did not form a U-shaped distribution around 38 syllables, but rather the points were scattered randomly. To determine whether the threshold values for the latency to pairing date and number of offspring dataset categories were driven by the most positive or most negative correlations in a given dataset category, we repeated the analysis with either the two most positive or two most negative values omitted. These truncated category datasets created a similar U-shaped pattern to the full category datasets, which were centered around 38 syllables (**Figure B.2.3D-G**). Therefore, it is unlikely that the threshold was shifted one way or another by unusually strong or weak correlations. These data suggest that 38 syllables is near the implied, real threshold value where a bird species has a large enough syllable repertoire that one can predict that there will be a correlation between individual song elaboration and reproductive success.

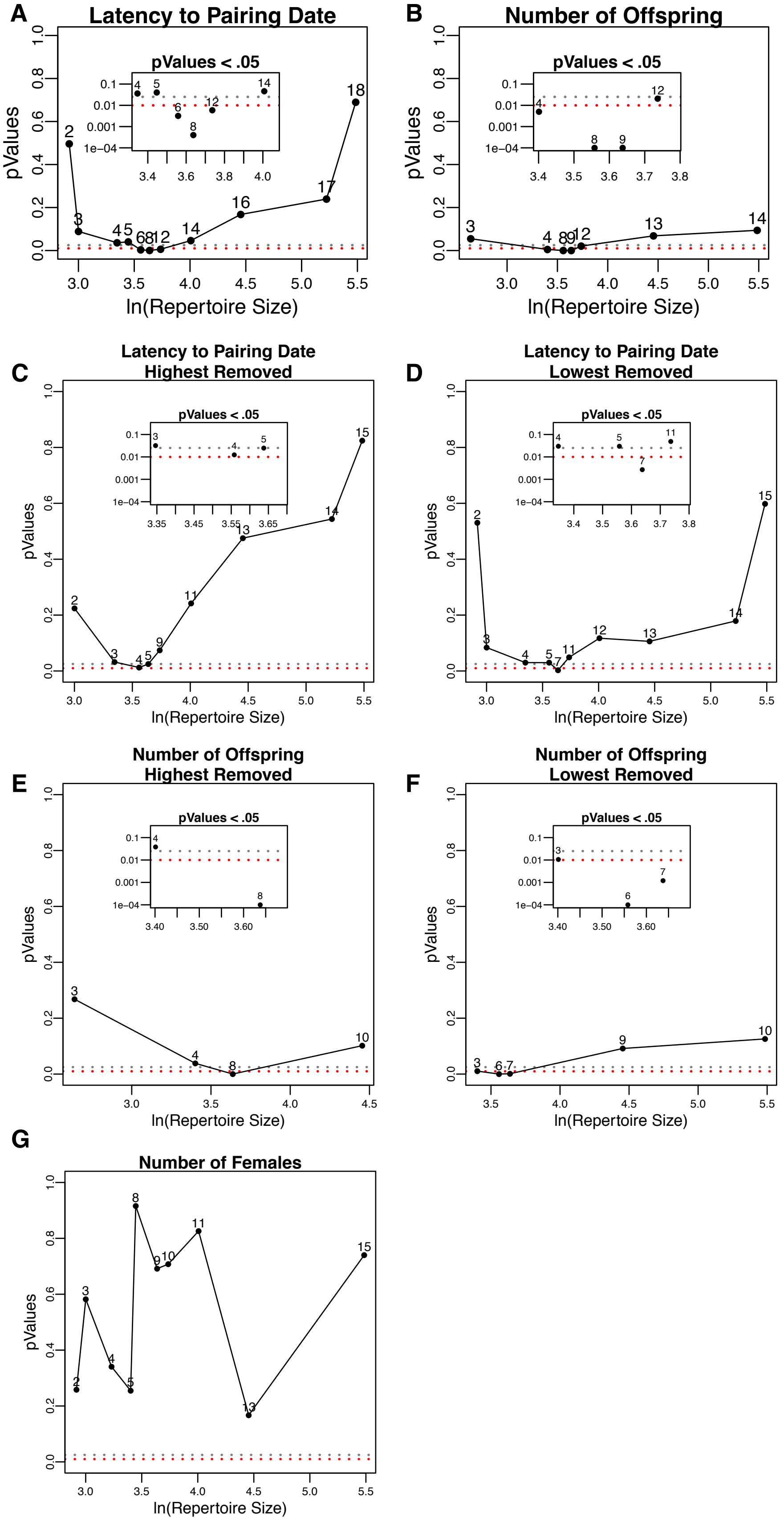

A
Latency to Pairing Date
pValues
ln(Repertoire Size)
pValues < .05
B
Number of Offspring
pValues
ln(Repertoire Size)
pValues < .05
C
Latency to Pairing Date
Highest Removed
pValues
ln(Repertoire Size)
pValues < .05
D
Latency to Pairing Date
Lowest Removed
pValues
ln(Repertoire Size)
pValues < .05
E
Number of Offspring
Highest Removed
pValues
ln(Repertoire Size)
pValues < .05
F
Number of Offspring
Lowest Removed
pValues
ln(Repertoire Size)
pValues < .05
G
Number of Females
pValues
ln(Repertoire Size)
0.0
0.2
0.4
0.6
0.8
1.0
3.0
3.5
4.0
4.5
5.0
5.5
0.1
0.01
0.001
1e−04

Figure B.2.3: Plots of *p*-values when different syllable repertoire sizes are used as the threshold value. Insets show *p*-values$<$0.05 plotted on a log_{10} scale. In both the main plot and the inset, the grey dotted line marks a significance threshold of 0.025, and the red dotted line marks a significance threshold of 0.01. Numbers above points are the number of studies/measurements present in the smaller average syllable repertoire group at that threshold value. Data were plotted from the latency to pairing date (A, C, D), number of offspring (B, E, F), or number of females (G) categories. (A, B, G) show the full dataset. Other panels show the *p*-values when two most positive (C, E) or two most negative (D, F) correlations are removed from the datasets. The most significant repertoire size thresholds are around 38 syllables for pairing date and number of offspring, even when the most positive or most negative correlations are removed (A-F). Number of females instead shows a random scattering of non-significant *p*-values (G).

Song stability meta-analysis

I also tested whether length of the song-learning window could predict the strength of the correlation between between individual song elaboration and reproductive success by using song stability as a proxy. For open-ended learners, repertoire size can be an indicator of age, and open-ended learning is correlated with larger species average repertoire sizes. Therefore, it is plausible that, in song-plastic species, individual song elaboration is more important for female choice than in song-stable species. Therefore, we hypothesized that elaboration would correlate with reproductive success more in species which show song plasticity in adulthood than in song-stable species. To address this, we split the dataset into two groups based on song stability and reran the meta-analysis for all three categories of reproductive success. There were no significant differences between song-plastic and song-stable species for any of the variables of reproductive success (**Figure B.2.1B, D, and F**). These unexpected findings show that, in the species sampled here, song stability was not a meaningful metric in predicting which bird would show strong correlations between individual song elaboration and reproductive success. Using the territory-controlled dataset did not significantly affect these results (**Figure B.2.2B, D, and F**).

Prediction intervals

This study was completed using a random-effects model; therefore, the population and subpopulation effect sizes, which are averages of the individual effect sizes in each species, are less meaningful than the prediction intervals *around* these averages, wherein the individual effect sizes of each constituent species would be expected to fall. For each category, the prediction interval for the full set of studies is very wide, encompassing predictions of positive, zero, and negative correlations. As expected, the prediction intervals for the song-plastic and song-stable subpopulations are similar to one another and similar to the full population for all categories (**Figure B.2.4A-C**). For number of females, the prediction intervals for the species with smaller and larger average syllable repertoire subpopulations were similar to one another and to the full population (**Figure B.2.44C**). In contrast, while the prediction intervals for the species with smaller average syllable repertoires subpopulation are similar to the full population prediction intervals for both latency to pairing date and number of offspring, the prediction intervals for the species with larger average syllable repertoires subpopulation are tighter and indicate a stronger correlation between individual song elaboration and reproductive success: a negative correlation for latency to pairing date and a positive correlation for number of offspring (**Figure B.2.4A and B**). Together, these data predict that in any given species with a larger average syllable repertoire size, there will be a small to large correlation between individual song elaboration and reproductive success when measured by latency to pairing date or number of offspring produced. On the other hand, within the subpopulation of birds with smaller syllable repertoire sizes, individual song elaboration does not reliably correlate with these measures of reproductive success. However, it is important to note that prediction intervals can be calculated to be erroneously small when datasets sets are small[151]; the real prediction intervals for species with larger repertoires are likely wider than those presented here. It will be interesting to see whether the relationships suggested by this analysis will be supported or refuted as additional data becomes available.

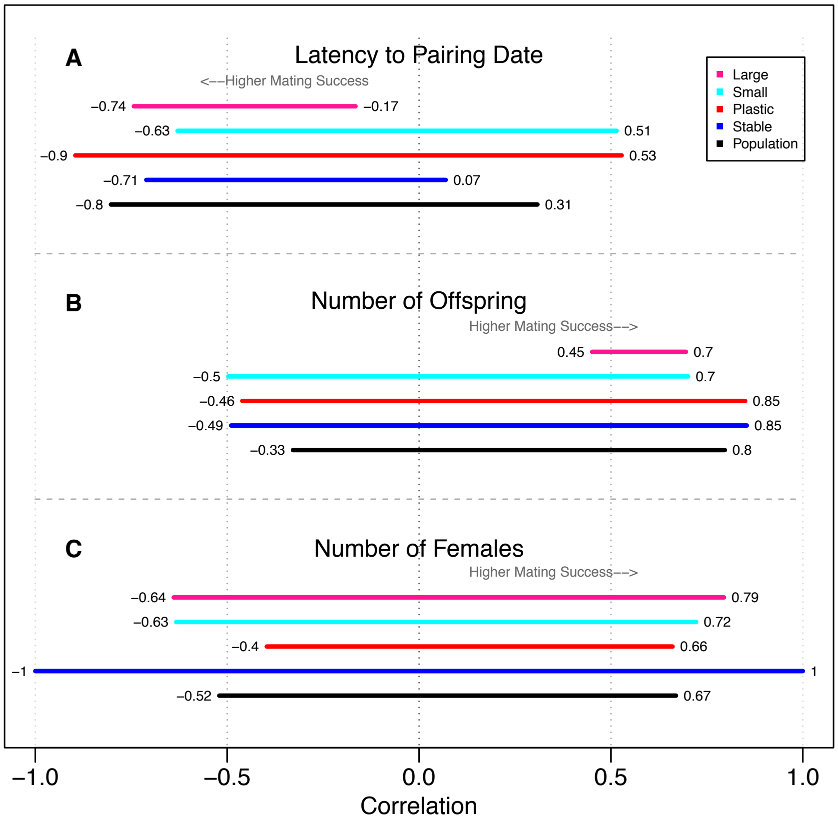


Figure B.2.4: Only the subpopulation with species that exhibit larger average syllable repertoire sizes show prediction intervals that are shorter than the full population prediction intervals. (A-C) Lines show the prediction intervals for the full population, the song stability groups, and species average syllable repertoire groups for each of the reproductive success categories. (A) For latency to pairing date, all subpopulations display similar, wide prediction intervals to the full population except the for the species with larger average syllable repertoires group. This subpopulation comprises only the negative half of the correlation range (earlier pairing, higher reproductive success). The expected correlations between individual song elaboration and reproductive success for all members of this group ranges from small to larger (PredInt=[−0.17;−0.74]). (B) The same pattern exists for number of offspring, where the prediction intervals for the species with larger average syllable repertoires subpopulation is shorter than the full population. This subpopulation comprises only the positive half of the population range (more offspring, higher reproductive success), with the expected correlations between individual song elaboration and reproductive success for all members of this group ranging from moderate to large (PredInt=[0.45;0.72]). (C) For number of females, there were no meaningful differences between the full population and any subpopulation.

Jackknife analysis

It is possible that a longitudinal study of repertoire size changes in a limited number of birds is not sufficient to accurately determine the whether a species generally modifies its adult song or not, and thus we may have miscategorized a species' song stability state. To account for this, we performed two regrouping jackknife analyses. In one, we switched the song stability grouping for each individual study in turn and reran the meta-analysis. In the other, we switched the grouping for all studies examining a given species (represented by up to 3 studies in one category). The study jackknife for latency to pairing date and number of offspring did bot yield any significant results (**Table B.2.1**). For number of females, reclassification of two studies (Bucannon, and Catchpole 2000) [11] and (Marshal, Bucannon, and Catchpole 2007) [12] on *Acrocephalus schoenobaenus* to song-stable led to significant differences between groups pre-Bonferroni correction, but not post-Bonferroni correction (*p*=0.0432 and *p*=0.0419) (**Table B.2.1**). In the species jackknife, regrouping of any individual species did not consistently change these results. However, regrouping of *Acrocephalus arundinaceus* to song-stable in the number of offspring dataset category led to significant differences between groups (*p*=0.0101), and regrouping of *A. schoenobaenus* in the number of females category led to significant difefrences between groups (*p*=0.006) **Table B.2.2**. It should be noted that there is unambiguous information available for the two mentioned *Acrocephalus* species: for *A. arundinaceus*, 56 individual males were studied for up to 5 years, and repertoires were found to change over time [1]. For *A. schoenobaenus*, 8 individuals followed within a single season increased repertoire sharing with their neighbors over time, and 6 individuals followed for up to 4 years showed changes in syllable repertoire size across years with a trend towards increasing repertoire sizes [40]. Furthermore, re-categorizing any single species did not change the results in more than one indicator of reproductive success. This makes us confident that these species were correctly grouped as song-plastic species. Together, these results suggest that it is unlikely that incorrect categorization of any one species is responsible for the lack of significance in the original analysis of the song stability subpopulations. These data support the result that song stability by itself is likely not a metric that can be used to predict whether individual song elaboration is correlated with reproductive success in a given bird species.

Table B.2.1: Possible miscategorization of song stability for one study would not change the outcome of the meta-analyses. The Bonferroni-corrected significance threshold (α=0.025) from the main meta-analyses was used to determine significance. The song stability state (song-plastic or song-stable) associated with each study was individually switched, and the meta-analysis was repeated to obtain a new *p*-value. Re-categorization of no individual study led to significant results. Numbered references correspond to the list of supplemental references in **Appendix A.1**.

Latency to Pairing Date		Number of Offspring		Number of Females	
Ref	*p*-value	Ref	*p*-value	Ref	*p*-value
[7]	0.8501	[1]	0.9131	[1]	0.5131
[8]	0.7322	[2]	0.8169	[2]	0.6326
[9]	0.8598	[3]	0.4206	[3]	0.1404
[10]	0.5815	[5]	0.7702	[8]	0.2665
[13]	0.371	[7]	0.8405	[11]	0.0432
[19]	0.29	[8]	0.6052	[12]	0.0419
[21]	0.7554	[16]	0.4333	[13]	0.5192
[22]	0.7124	[26]	0.551	[17]	0.1937
[24]	0.5371	[22]	0.6536	[18]	0.8605
[25]	0.8175	[29]	0.6462	[20]	0.0815
[22]	0.6704	[31]	0.7352	[20]	0.0621
[29]	0.9372	[33]	0.5448	[28;20]	0.1902
[30]	0.6921	[34]	0.905	[23]	0.4287
[31]	0.8463	[38]	0.5588	[35]	0.1249
[33]	0.6814	[43]	0.8691	[43]	0.498
[38]	0.461	[46]	0.6462	[44]	0.4319
[39]	0.9317			[46]	0.7164
[43]	0.9857				
[45]	0.9056				
[46]	0.7962				
[47]	0.9533				

Table B.2.2: Possible miscategorization of song stability for one species would not change the outcome of the meta-analyses. The Bonferroni-corrected significance threshold (α=0.025) from the main meta-analyses was used to determine significance. The song stability state of each species was individually switched, and the meta-analysis was rerun to obtain a new *p*-values. Miscategorization of *Acrocephalus arudinaceus* (3 studies) would have led to significant results for number of offspring (*p*=0.0101). Miscategorization of *A. schoenobaenus* (2 studies) would have led to significant results for number of females (textitp=0.006). Asterisks (*) denote significant differences between groups.

Latency to Pairing Date		Number of Offspring		Number of Females	
Species	*p*-value	Species	*p*-value	Species	*p*-value
Parus major	0.6814	*Parus major*	0.4444	*Emberiza schoeniclus*	0.8605
Agelaius phoeniceus	0.371	*Cardinalis cardinalis*	0.4333	*Agelaius phoeniceus*	0.5192
Hirundo rustica	0.7124	*Acrocephalus arundinaceus*	0.0101*	*Hirundo rustica*	0.0621
Ficedula hypoleuca	0.7554	*Phylloscopus trochilus*	0.5588	*Phylloscopus fuscatus*	0.1249
Ficedula albicollis	0.29	*Melospiza melodia*	0.1454	*Acrocephalus arundinaceus*	0.8945
Phylloscopus trochilus	0.6088	*Sturnus vulgaris*	0.6462	*Ficedula albicollis*	0.0815
Melospiza melodia	0.5903	*Sturnella neglecta*	0.8691	*Melospiza melodia*	0.1902
Sturnus vulgaris	0.6022	*Acrocephalus palustris*	0.5115	*Sturnus vulgaris*	0.7164
Acrocephalus schoenobaenus	0.6363			*Acrocephalus schoenobaenus*	0.006*
Sturnella neglecta	0.9857			*Sturnella neglecta*	0.7491
Mimus polyglottos	0.6921			*Acrocephalus palustris*	0.5773
Acrocephalus palustris	0.8502			*Luscinia megarhynchos*	0.1957
Sylvia communis	0.8971				

Possible interactions

In this dataset, both song-stable and song-plastic species could have either smaller or larger average syllable repertoire sizes. Therefore, we wanted to examine whether there was an interaction between these two variables. We divided the data in the latency to pairing date and number of offspring dataset catagories by one of these subpopulations (e.g. song stability) and tested for differences in the literature r values between the other subpopulation variable (species average syllable repertoire size). Four Welch's two sample t-tests were performed on these data, so we used a Bonferroni-corrected significance threshold (α=0.0125). The only significant interaction was revealed in latency to pairing date, where song-plastic species with larger species average syllable repertoires were significantly different from song-plastic species with smaller species average syllable repertoires (**Table B.2.3**). There was also an interaction in number of offspring that was trending significant, where song-stable species with larger species average syllable repertoires were significantly different from song-stable species with smaller species average syllable repertoires (**Table B.2.3**). These data leave open the possibility that there is some interaction between song stability and species average syllable repertoire size that yields greater predictive power, but larger sample sizes would be needed to make any definite conclusions.

Table B.2.3: Song stability may add some predictive value in combination with species average syllable repertoire size. Data from the latency to pairing date or number of offspring dataset categories were stratified by either song stability or species syllable repertoire size. A Welch's two sample t-test was used to examine whether there were a differences between groups for the other subpopulation variable. Because four t-tests were performed, a Bonferroni-corrected significance threshold was used (α=0.0125). The degrees of freedom for latency to pairing date were 11, 6, 11, and 6 in descending order. The degrees of freedom for number of offspring were 5, 7, 5, and 7 in descending order. Asterisks (*) denote significant differences between groups.

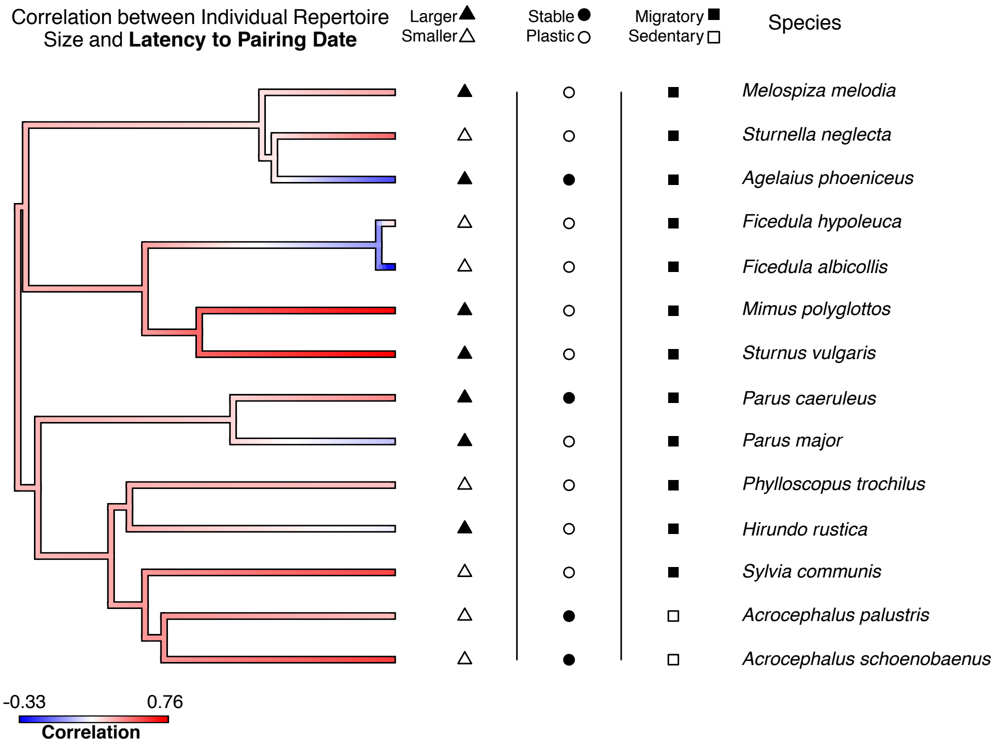
Latency to Pairing Date						Number of Offspring					
Class	Group	N	Mean	t	*p*-value	Class	Group	N	Mean	t	*p*-value
Plastic	Large	7	−0.663	−3.37	0.012*	Plastic	Large	2	0.656	2.2	0.0874
	Small	6	−0.11				Small	5	0.268		
Stable	Large	6	−0.3812	−0.583	0.63	Stable	Large	5	0.5886	4.51	0.015
	Small	2	−0.271				Small	4	−0.041		
Large	Plastic	7	−0.663	2.58	0.027	Large	Plastic	2	0.6555	−0.06	0.64
	Stable	6	−0.381				Stable	5	0.5886		
Small	Plastic	6	−0.271	−0.664	0.56	Small	Plastic	5	0.268	−1.59	0.156
	Stable	2	−0.11				Stable	4	−0.0412		

Controlling for phylogenetic relationships

A phylogenetic ANOVA was used to examine whether these results remained significant after accounting for phylogeny. Soma and Garamszegi[19] had previously shown that controlling for phylogeny led to a poorer fit model when the data was restricted in this manner. However, their dataset was not broken into three separate reproductive success categories. Interestingly, some research suggests that migratory birds and subspecies have larger average syllable repertoires than their sedentary counterparts. Therefore, we performed three phylogenetic ANOVAs on these datasets: one for species syllable repertoire size, a second for song stability, and a third for migratory status. For latency to pairing date, species average syllable repertoire size produced significantly different subpopulations after correcting for phylogeny, while song stability and migratory status did not (**Figure B.2.5A**). On the other hand, for number of offspring, migratory status did lead to significantly different subpopulations, species syllable repertoire was trending significant post-Bonferroni-correction, and song stability was not significant (**Figure B.2.5B**).

Conclusions from these `phylANOVAs` need to be interpreted with caution, because the sample sizes are small, but the data suggest that species with larger average syllable repertoires have larger correlations between individual song elaboration and latency to pairing date than species with smaller average syllable repertoires even after controlling for phylogenetic relatedness. The correlations between individual song elaboration and number of offspring were significantly different between species with larger and smaller average syllable repertoires after correcting for phylogenetic relatedness, but these differences did not remain significant after Bonferroni correction. Studies in more species will be required to definitively answer this question.

A



Phylogenically–Controlled ANOVA Results:
Bonferroni–Corrected Alpha = 0.017

Repertoire Size F–Value = 13.244 p–value = 0.0068 Means: Smaller: 0.123 Larger: 0.572
Learning Style F–Value = 0.032 p–value = 0.8818 Means: Stable: 0.358 Plastic: 0.322
Migratory Style F–Value = 0.123 p–value = 0.8042 Means: Sedentary: 0.271 Migratory: 0.36

B

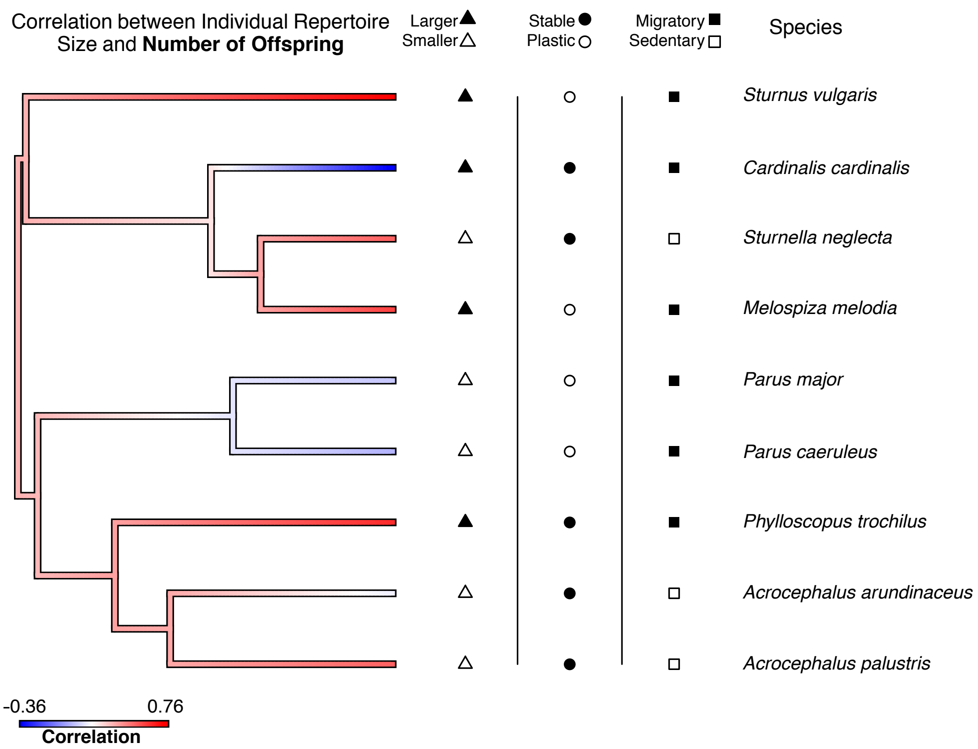


Figure B.2.5: Species average syllable repertoire size subpopulation differences persist after controlling for phylogeny in a traditional meta-analysis. (AB) Trees show the phylogeny between bird species and the continuous character map of the correlation between individual song elaboration and either the latency to pairing date (A) or number of offspring (B). Triangles represent syllable repertoire size, circles represent song stability, and squares represent migratory status. Only for latency to pairing date (A), the correlations were multiplied by -1, so that redder colors indicate greater reproductive success on both trees. A Bonferroni-corrected significance threshold (α=0.017) was used for the following phylogenetic ANOVAs. The degrees of freedom for each test are the number of species minus one.

(A) For latency to pairing date, there was a significant difference for the correlations between individual song elaboration and reproductive success for species with **smaller and larger average syllable repertoires** after phylogenetic control ($\text{Mean}_{\text{smaller}}$=0.123, $\text{Mean}_{\text{larger}}$=0.577, *p*=0.0068). There was still no significant difference between **song-stable and song-plastic** species after phylogenetic control ($\text{Mean}_{\text{plastic}}$=0.322, $\text{Mean}_{\text{stable}}$=0.358, *p*=0.88). There was no significant difference between **sedentary and migratory** species after phylogenetic control (Mean_{sed}=0.271, Mean_{mig}=0.36, F=0.123, *p*=0.80).

(B) For number of offspring, there was no significant difference after Bonferroni-correction for the correlations between individual song elaboration and reproductive success for species with **smaller and larger average syllable repertoires** after phylogenetic control ($\text{Mean}_{\text{smaller}}$=0.118, $\text{Mean}_{\text{larger}}$=0.619, *p*=0.039). There was still no significant difference between **song-stable and song-plastic** species after phylogentic control ($\text{Mean}_{\text{plastic}}$=0.184, $\text{Mean}_{\text{stable}}$=0.536, *p*=0.18). There was a significant difference between **sedentary and migratory** species after phylogenetic control (Mean_{mig}=0.551, Mean_{sed}=0.082, *p*=0.007).

APPENDIX C

Correlated evolution

C.1 Robustness testing

Reclassifying *Melospiza lincolnii* as labeled song-stable did not significantly affect our results (**Tables C.1.1** and **C.1.2**).

Table C.1.1: `PhylANOVA` results for all song traits when *Melospiza lincolnii* is labeled song-stable. Song-Stable and Song-Plastic columns show means. Song traits are sorted from most to least significant. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote traits with significantly different groups.

Song Trait	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Syllable Rep	1.9098	3.9792	41.1603	0.0071	<0.001*
Song Rep	1.094	4.0131	39.4823	0.0083	<0.001*
Syll\Song	1.2556	2.2962	9.2658	0.01	0.094
Duration	0.7736	1.2927	2.0783	0.0125	0.42
Continuity	−1.3453	−1.0286	2.1537	0.0167	0.474
Interval	1.6075	1.218	1.3879	0.025	0.567
Song Rate	1.8969	2.0971	0.6079	0.05	0.713

Table C.1.2: `Brownie` results for song traits when *Melospiza lincolnii* is labeled Song-Stable. Song traits are sorted from most to least significant. Asterisks (*) denote traits where the two-rate model fit the data significantly better than the one-rate model.

Song Trait	One Rate	Two Rates	p-Value
Syll\Song	−110.6482	−100.7673	<0.001*
Interval	−45.2842	−40.5004	0.002*
Duration	−71.2042	−66.3122	0.002*
Song Rep	−113.5829	−113.3919	0.537
Syllable Rep	−120.2983	−120.1567	0.595

Using the minimum or maximum values did not change the `PhylANOVA` results for any song trait when species were split into song-stable and song-plastic groups **Table C.1.3**, though when the early song-stable, delayed-song stable, and song-plastic dataset was used, using the maximum values for song repertoire led to non-significant results **Tables C.1.4** and **C.1.5**. When species were split into those with adult song plasticity and adult song stability, most `Brownie` analyses did not change **Tables C.1.6** and **C.1.7** However, it was originally suggested that syllables per song was evolving significantly faster in song-plastic lineages when we used the median value, but using the minimum values led to non-significant results **Table C.1.6**. When species were split into three states (early song stability, delayed song stability, and song plasticity), we found that the three-rate model was not better than the one-rate model when the minimum values for song repertoire were used, while the three-rate models for syllable repertoire became significant when either the maximum or the minimum values were used **Table C.1.7**.

Table C.1.3: `PhylANOVA` results for song traits when either the maximum or minimum reported values are used and birds are divided into song-stable and song-plastic. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote traits with significantly different groups.

Song Trait	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Syllable Rep Min	1.8106	3.7433	39.2342	0.0071	<0.001*
Syllable Rep Max	1.9176	4.0608	41.99	0.0071	<0.001*
Song Rep Min	0.8721	3.3024	37.8962	0.0083	<0.001*
Song Rep Max	1.181	4.211	34.4445	0.0083	<0.001*
Syll\Song Max	1.2971	2.47	9.714	0.01	0.087
Syll\Song Min	1.1881	1.8088	5.6204	0.01	0.217
Duration Max	0.7895	1.3853	2.4515	0.0125	0.389
Interval Max	1.6908	1.2852	1.7593	0.025	0.512
Duration Min	0.7503	1.1234	1.3346	0.0125	0.54
Interval Min	1.4121	0.9398	0.8895	0.025	0.661

Table C.1.4: `PhylANOVA` results for song traits when either the maximum or minimum reported values are used and birds are divided into early song-stable, delayed song-stable, and song-plastic. Song traits are sorted from most to least significant. Early, Delayed, and Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote traits with significantly different groups.

Song Trait	Early	Delayed	Plastic	F-Value	Corrected α	p-Value
Syllable Rep Min	1.5451	1.9802	3.7433	16.317	0.0071	0.003*
Syllable Rep Max	1.697	2.0233	4.0608	17.209	0.0071	0.002*
Song Rep Min	0.5887	1.1625	3.3024	14.2085	0.0083	0.006*
Song Rep Max	0.7427	1.5669	4.211	13.059	0.0083	0.011
Syll\Song Min	1.1987	1.1694	1.8088	2.4789	0.01	0.398
Syll\Song Max	1.3186	1.3015	2.47	3.7714	0.01	0.243

Table C.1.5: Post-hoc pairwise `phylANOVA` tests for significant song traits when either the maximum or minimum reported values are used and birds are divided into early song-stable, delayed song-stable, and song-plastic. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote traits with significantly different groups.

Song Trait	State 1	State 2	T-Value	p-Value
Syllable Rep Min	Plastic	Delayed	4.6745	0.014*
Syllable Rep Min	Early	Plastic	4.6194	0.003*
Syllable Rep Min	Early	Delayed	0.8659	0.581
Syllable Rep Max	Plastic	Delayed	4.9639	0.01*
Syllable Rep Max	Early	Plastic	4.5646	0.003*
Syllable Rep Max	Early	Delayed	0.5967	0.697
Song Rep Min	Plastic	Delayed	4.3491	0.022*
Song Rep Min	Early	Plastic	4.4142	0.012*
Song Rep Min	Early	Delayed	0.8988	0.614

Table C.1.6: `Brownie` results for song traits when either the maximum or minimum reported values are used and birds are divided into song-stable and song-plastic. Song traits are sorted from most to least significant. Asterisks (*) denote traits where the two-rate model fit the data significantly better than the one-rate model.

Song Trait	One Rate	Two Rates	p-Value
Interval Max	−43.8289	−38.1264	<0.001*
Duration Max	−74.1132	−68.1884	<0.001*
Syll\Song Max	−119.0363	−105.4808	<0.001*
Duration Min	−65.7687	−62.6047	0.012*
Interval Min	−58.6807	−56.4811	0.036*
Syll\Song Min	−88.5573	−87.264	0.108
Song Rep Max	−120.3993	−119.6825	0.231
Syllable Rep Max	−125.0196	−124.3595	0.251
Song Rep Min	−100.8545	−100.5988	0.475
Syllable Rep Min	−118.6497	−118.4524	0.53

Table C.1.7: `Brownie` results for song traits when either the maximum or minimum reported values are used and birds are divided into early song-stable, delayed song-stable, and song-plastic. Song traits are sorted from most to least significant. Asterisks (*) denote traits where the three-rate model fit the data significantly better than the one-rate model.

Song Trait	One Rate	Three Rates	p-Value
Syll\Song Max	−105.7229	−91.4766	<0.001*
Song Rep Max	−107.1249	−103.2302	0.005*
Syll\Song Min	−75.3007	−72.5126	0.018*
Syllable Rep Min	−106.2125	−104.0387	0.037*
Syllable Rep Max	−111.8435	−109.7931	0.043*
Song Rep Min	−88.7243	−86.9529	0.06

Exclusion of individual families did not significantly alter any of the `phylANOVA` results (**Tables C.1.8** to **C.1.13**). In general, exclusion of individual families did not affect the Brownie results (**Tables C.1.14** to **C.1.19**) or the `BayesTraits` results (see **Figure 6 - figure supplement 8-9 and Figure 7 - figure supplement 1** from Robinson, Snyder, and Creanza 2019[70]). However, there were several notable exceptions. In several cases, removing a single family altered the significance of our findings. Removal of the Muscicapidae (3 species) from the Brownie analysis of inter-song interval created a two-rate model that did not fit the data significantly better than the one-rate model (**Table C.1.15**). For song duration, removal of Fringillidae (3 species) led a two-rate model that was not significantly better fit than the one-rate model (**Table C.1.16**).

Table C.1.8: `PhylANOVA` results for syllable repertoire when each bird family is omitted. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote significantly different groups.

Removed Family	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Acrocephalidae	1.8266	4.0232	47.6941	0.0071	<0.001*
Icteridae	1.8195	3.9652	40.9489	0.0071	<0.001*
Mimidae	1.8807	3.5707	32.426	0.0071	<0.001*
Muscicapidae	1.888	3.8686	36.7766	0.0071	<0.001*
Parulidae	1.8817	4.099	46.2305	0.0071	<0.001*
Passerellidae	1.9468	3.9792	32.4733	0.0071	<0.001*
Fringillidae	1.8177	3.9474	37.5394	0.0071	0.003*

Table C.1.9: `PhylANOVA` results for interval when each bird family is omitted. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote significantly different groups.

Removed Family	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Muscicapidae	1.6075	1.0449	2.6556	0.025	0.326
Acrocephalidae	1.8406	1.3251	3.0278	0.025	0.404
Icteridae	1.6075	1.0659	2.5898	0.025	0.479
Parulidae	1.5815	1.1387	1.4568	0.025	0.595
Fringillidae	1.5821	1.218	1.1582	0.025	0.628
Passerellidae	1.4435	1.218	0.3135	0.025	0.725
Mimidae	1.6075	1.6071	0	0.025	1

Table C.1.10: `PhylANOVA` results for duration when each bird family is omitted. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote significantly different groups.

Removed Family	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Mimidae	0.7736	1.5596	4.8068	0.0125	0.155
Muscicapidae	0.7736	1.3495	2.183	0.0125	0.354
Passerellidae	0.8056	1.2927	1.1871	0.0125	0.403
Acrocephalidae	0.62	1.1605	2.3871	0.0125	0.409
Parulidae	0.7993	1.4276	2.6748	0.0125	0.421
Icteridae	0.7736	1.3928	2.7298	0.0125	0.435
Fringillidae	0.7747	0.9735	0.4281	0.0125	0.784

Table C.1.11: `PhylANOVA` results for syllables per song when each bird family is omitted. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote significantly different groups.

Removed Family	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Muscicapidae	1.2556	2.342	8.7962	0.01	0.086
Acrocephalidae	1.1915	2.3503	11.5839	0.01	0.091
Parulidae	1.2432	2.4329	10.9929	0.01	0.095
Passerellidae	1.3391	2.2962	5.8932	0.01	0.1
Fringillidae	1.2335	2.3554	9.4058	0.01	0.119
Icteridae	1.2556	2.3636	9.7574	0.01	0.13
Mimidae	1.2556	1.792	4.3591	0.01	0.222

Table C.1.12: `PhylANOVA` results for song rate when each bird family is omitted. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote significantly different groups.

Removed Family	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Muscicapidae	1.8969	2.1768	1.0364	0.05	0.564
Mimidae	1.8969	1.7406	0.7394	0.05	0.608
Passerellidae	1.8818	2.0971	0.4666	0.05	0.669
Icteridae	1.8969	2.1642	0.9848	0.05	0.67
Acrocephalidae	1.9649	2.1804	0.7139	0.05	0.688
Fringillidae	1.9034	2.0971	0.5401	0.05	0.748
Parulidae	1.8829	2.0942	0.5365	0.05	0.76

Table C.1.13: `PhylANOVA` results for song repertoire when each bird family is omitted. Song traits are sorted from most to least significant. Song-Stable and Song-Plastic columns show means. Corrected α indicates the threshold for significance with the Holm-Bonferroni correction. Asterisks (*) denote traits different groups.

Removed Family	Song-Stable	Song-Plastic	F-Value	Corrected α	p-Value
Acrocephalidae	1.033	3.722	31.0779	0.0083	<0.001*
Fringillidae	1.0383	3.7006	29.3831	0.0083	<0.001*
Mimidae	1.1055	3.3384	22.2449	0.0083	<0.001*
Muscicapidae	1.1055	3.6384	26.0734	0.0083	<0.001*
Passerellidae	1.218	4.0131	26.4254	0.0083	<0.001*
Icteridae	0.9628	4.0616	43.9031	0.0083	0.002*
Parulidae	1.0972	4.0946	37.4843	0.0083	0.002*

Table C.1.14: `Brownie` results for syllable repertoire when each bird family is omitted. Song traits are sorted from most to least significant. Asterisks (*) denote cases where the two-rate model fit the data significantly better than the one-rate model.

Removed Family	One Rate	Two Rates	p-Value
Passerellidae	−99.3059	−97.9546	0.1
Acrocephalidae	−109.715	−109.2164	0.318
Mimidae	−111.7997	−111.489	0.43
Fringillidae	−110.9574	−110.7158	0.487
Icteridae	−113.4483	−113.2436	0.522
Parulidae	−115.6046	−115.401	0.523
Muscicapidae	−111.2268	−111.0404	0.542

Table C.1.15: `Brownie` results for interval when each bird family is omitted. Song traits are Song traits are sorted from most to least significant. Asterisks (*) denote cases where the two-rate model fit the data significantly better than the one-rate model.

Removed Family	One Rate	Two Rates	p-Value
Acrocephalidae	−38.372	−31.0345	<0.001*
Icteridae	−43.5138	−38.6556	0.002*
Fringillidae	−44.2357	−39.9282	0.003*
Parulidae	−43.2215	−39.0435	0.004*
Mimidae	−37.7323	−34.1732	0.008*
Passerellidae	−40.7556	−37.7238	0.014*
Muscicapidae	−35.1535	−33.3576	0.058

Table C.1.16: `Brownie` results for duration when each bird family is omitted. Song traits are Song traits are sorted from most to least significant. Asterisks (*) denote cases where the two-rate model fit the data significantly better than the one-rate model.

Removed Family	One Rate	Two Rates	p-Value
Acrocephalidae	−63.9978	−57.4808	<0.001*
Muscicapidae	−68.4218	−62.5466	<0.001*
Icteridae	−69.2203	−64.2909	0.002*
Parulidae	−67.9632	−63.404	0.003*
Mimidae	−64.2702	−60.1031	0.004*
Passerellidae	−63.0661	−60.5806	0.026*
Fringillidae	−53.9545	−52.6407	0.105

Table C.1.17: `Brownie` results for syllables per song when each bird family is omitted. Song traits are Song traits are sorted from most to least significant. Asterisks (*) denote cases where the two-rate model fit the data significantly better than the one-rate model.

Removed Family	One Rate	Two Rates	p-Value
Acrocephalidae	−102.1934	−89.5897	<0.001*
Fringillidae	−103.9832	−94.7314	<0.001*
Icteridae	−107.6333	−97.659	<0.001*
Muscicapidae	−106.7467	−95.6359	<0.001*
Parulidae	−105.5031	−96.1153	<0.001*
Passerellidae	−96.8595	−88.2423	<0.001*
Mimidae	−71.5741	−70.2893	0.109

Table C.1.18: `Brownie` results for song rate when each bird family is omitted. Song traits are Song traits are sorted from most to least significant. Asterisks (*) denote cases where the two-rate model fit the data significantly better than the one-rate model.

Removed Family	One Rate	Two Rates	p-Value
Acrocephalidae	−36.7513	−28.3824	<0.001*
Icteridae	−41.9163	−36.9836	0.002*
Fringillidae	−42.4949	−38.1179	0.003*
Parulidae	−41.5382	−37.2602	0.003*
Muscicapidae	−35.7804	−32.5759	0.011*
Passerellidae	−39.2489	−36.1562	0.013*
Mimidae	−33.6912	−30.9239	0.019*

Table C.1.19: `Brownie` results for song repertoire when each bird family is omitted. Song traits are Song traits are sorted from most to least significant. Asterisks (*) denote cases where the two-rate model fit the data significantly better than the one-rate model.

Removed Family	One Rate	Two Rates	p-Value
Icteridae	−104.254	−103.5109	0.223
Mimidae	−106.4131	−106.0144	0.372
Fringillidae	−100.0817	−99.7318	0.403
Acrocephalidae	−103.331	−103.0443	0.449
Parulidae	−107.5562	−107.3195	0.491
Muscicapidae	−106.4831	−106.2634	0.507
Passerellidae	−97.4923	−97.3567	0.602

At low threshold values of syllables per song, song-plasticity with low syllables per song is an unstable state, with high rates of transition towards either higher syllables per song or towards song stability (77% of runs significant in this range); however, when mimid species are removed, gaining more syllables per song in the song-plastic state becomes far more likely. At moderate values of syllables per song, there are elevated rates of transition both towards more syllables per song and towards fewer syllables per song in song-plastic species (100% of runs significant in this range); there is a comparable trend when mimid species are removed. At high threshold values, more syllables per song with song-stability is an unstable state (63% of runs significant in this range)(see **Figure 6 - figure supplement 2** from Robinson, Snyder, and Creanza 2019[70]); however, with mimids removed, more syllables per song with song plasticity becomes the most unstable state (see **Figure 6 - figure supplement 9** from Robinson, Snyder, and Creanza 2019[70]).

Due to the strong effect that the inclusion of *Mimidae* had on the Brownie analysis of syllables per song, we performed a second jackknife analysis, in which each of the four mimid species was removed in turn. Exclusion of *Toxostoma rufum* or *Dumetella carolinensis* had little effect on the results (**Table C.1.20**), leading to significant support for the two-rate model. However, exclusion of *Mimus polyglottos* (p=0.074) or of *Mimus gilvus* (p=0.509) led to a two-rate model that did not fit the data significantly better than the one-rate model (**Table C.1.20**). Therefore, we concluded that the two *Mimus* species drove the estimated evolutionary rate of syllables per song in song-plastic species to be much greater than in song-stable species, and that faster evolution for syllables per song may not necessarily be a universal trend for song-plastic species.

Members of *Mimidae* are renowned for their vocal mimicry, frequently exhibiting improvisation and invention of syllables beyond simple imitation, and thus they lack the generally stereotyped song structure shown in other oscine families. Furthermore, mimids often have periods of continuous singing with minimal repetition of elements and irregular syllable spacing. Thus, quantification of song duration or number of syllables per song for mimids could be highly susceptible to listener perception[196]. Therefore, though we acknowledge that mimids are an important case study in extended learning durations, our results for the evolutionary rate of syllables per song might be more meaningful across all bird species when mimids are excluded, in which case we find the rate of evolution of syllables per song is be independent of adult song plasticity.

Table C.1.20: `Brownie` results for syllables per song when each mimid sepcies is omitted. Song traits are Song traits are sorted from most to least significant. Asterisks (*) denote cases where the two-rate model fit the data significantly better than the one-rate model.

Removed Mimid	One Rate	Two Rates	p-Value
Toxostoma rufum	−108.0309	−98.3481	<0.001*
Dumetella carolinensis	−108.1329	−98.3281	<0.001*
Mimus polyglottos	−89.0035	−87.4045	0.074
Mimus gilvus	−82.2923	−82.0747	0.509

APPENDIX D

Quantifying dialect establishment and maintenance

D.1 Maintenance

Table D.1.1: Dialects are maintained when there is vertical learning and either high nest site fidelity or local tutor choice. Par # is the parameter number (for figure references). MDial is whether males began the simulation with a song identical to their resident female's song (Same) or with a uniquely generated song made from syllables belonging to their region's dialect (Similar). Breed is whether chicks where sired by local fathers (high nest site fidelity) or fathers from anywhere in the matrix (no nest site fidelity). Tutor is whether tutors were chosen from a local (TRUE) or global (FALSE) pool of males. Vertical is whether chicks engaged in vertical learning from their fathers. Uniform is whether all females had the same region dialect template (TRUE) or variants of that template ("Noisy Templates", FALSE). Dialects is the number of female regional dialects. Center is a measurement of the template-matching accuracy in the middle of regional dialect 1. Edge is a measurement of the template-matching accuracy on the boundary between regional dialects 1 and the dialect to its right.

Par #	MDial	Breed	Tutor	Vertical	Uniform	Dialects	Center	Edge
1	Same	TRUE	TRUE	TRUE	TRUE	1	0.583	0.570
2	Same	TRUE	TRUE	TRUE	FALSE	1	0.549	0.497
3	Same	TRUE	TRUE	FALSE	TRUE	1	0.488	0.428
4	Same	TRUE	TRUE	FALSE	FALSE	1	0.467	0.430
5	Same	TRUE	FALSE	TRUE	TRUE	1	0.748	0.775
6	Same	TRUE	FALSE	TRUE	FALSE	1	0.682	0.676
7	Same	TRUE	FALSE	FALSE	TRUE	1	0.560	0.604
8	Same	TRUE	FALSE	FALSE	FALSE	1	0.485	0.513
9	Same	FALSE	TRUE	TRUE	TRUE	1	0.691	0.722
10	Same	FALSE	TRUE	TRUE	FALSE	1	0.622	0.608
11	Same	FALSE	TRUE	FALSE	TRUE	1	0.563	0.595
12	Same	FALSE	TRUE	FALSE	FALSE	1	0.461	0.461
13	Same	FALSE	FALSE	TRUE	TRUE	1	0.844	0.822
14	Same	FALSE	FALSE	TRUE	FALSE	1	0.720	0.755
15	Same	FALSE	FALSE	FALSE	TRUE	1	0.564	0.573
16	Same	FALSE	FALSE	FALSE	FALSE	1	0.467	0.487
17	Similar	TRUE	TRUE	TRUE	TRUE	1	0.537	0.560
18	Similar	TRUE	TRUE	TRUE	FALSE	1	0.498	0.498
19	Similar	TRUE	TRUE	FALSE	TRUE	1	0.414	0.476
20	Similar	TRUE	TRUE	FALSE	FALSE	1	0.518	0.502
21	Similar	TRUE	FALSE	TRUE	TRUE	1	0.709	0.695
22	Similar	TRUE	FALSE	TRUE	FALSE	1	0.676	0.680

23	Similar	TRUE	FALSE	FALSE	TRUE	1	0.487	0.511
24	Similar	TRUE	FALSE	FALSE	FALSE	1	0.495	0.479
25	Similar	FALSE	TRUE	TRUE	TRUE	1	0.692	0.710
26	Similar	FALSE	TRUE	TRUE	FALSE	1	0.647	0.629
27	Similar	FALSE	TRUE	FALSE	TRUE	1	0.542	0.528
28	Similar	FALSE	TRUE	FALSE	FALSE	1	0.549	0.451
29	Similar	FALSE	FALSE	TRUE	TRUE	1	0.792	0.779
30	Similar	FALSE	FALSE	TRUE	FALSE	1	0.729	0.740
31	Similar	FALSE	FALSE	FALSE	TRUE	1	0.461	0.438
32	Similar	FALSE	FALSE	FALSE	FALSE	1	0.460	0.474
33	Same	TRUE	TRUE	TRUE	TRUE	2	0.534	0.338
34	Same	TRUE	TRUE	TRUE	FALSE	2	0.498	0.309
35	Same	TRUE	TRUE	FALSE	TRUE	2	0.232	0.292
36	Same	TRUE	TRUE	FALSE	FALSE	2	0.199	0.204
37	Same	TRUE	FALSE	TRUE	TRUE	2	0.396	0.359
38	Same	TRUE	FALSE	TRUE	FALSE	2	0.322	0.311
39	Same	TRUE	FALSE	FALSE	TRUE	2	0.287	0.257
40	Same	TRUE	FALSE	FALSE	FALSE	2	0.239	0.260
41	Same	FALSE	TRUE	TRUE	TRUE	2	0.359	0.381
42	Same	FALSE	TRUE	TRUE	FALSE	2	0.278	0.297
43	Same	FALSE	TRUE	FALSE	TRUE	2	0.232	0.248
44	Same	FALSE	TRUE	FALSE	FALSE	2	0.223	0.235
45	Same	FALSE	FALSE	TRUE	TRUE	2	0.337	0.333
46	Same	FALSE	FALSE	TRUE	FALSE	2	0.338	0.334
47	Same	FALSE	FALSE	FALSE	TRUE	2	0.295	0.298
48	Same	FALSE	FALSE	FALSE	FALSE	2	0.234	0.205
49	Similar	TRUE	TRUE	TRUE	TRUE	2	0.560	0.373
50	Similar	TRUE	TRUE	TRUE	FALSE	2	0.458	0.294
51	Similar	TRUE	TRUE	FALSE	TRUE	2	0.235	0.218
52	Similar	TRUE	TRUE	FALSE	FALSE	2	0.240	0.228
53	Similar	TRUE	FALSE	TRUE	TRUE	2	0.378	0.349
54	Similar	TRUE	FALSE	TRUE	FALSE	2	0.298	0.296
55	Similar	TRUE	FALSE	FALSE	TRUE	2	0.228	0.226
56	Similar	TRUE	FALSE	FALSE	FALSE	2	0.192	0.183
57	Similar	FALSE	TRUE	TRUE	TRUE	2	0.328	0.315
58	Similar	FALSE	TRUE	TRUE	FALSE	2	0.291	0.310
59	Similar	FALSE	TRUE	FALSE	TRUE	2	0.227	0.219
60	Similar	FALSE	TRUE	FALSE	FALSE	2	0.218	0.227
61	Similar	FALSE	FALSE	TRUE	TRUE	2	0.343	0.323
62	Similar	FALSE	FALSE	TRUE	FALSE	2	0.303	0.290
63	Similar	FALSE	FALSE	FALSE	TRUE	2	0.252	0.247
64	Similar	FALSE	FALSE	FALSE	FALSE	2	0.233	0.261
65	Same	TRUE	TRUE	TRUE	TRUE	4	0.398	0.282
66	Same	TRUE	TRUE	TRUE	FALSE	4	0.360	0.253
67	Same	TRUE	TRUE	FALSE	TRUE	4	0.142	0.131

68	Same	TRUE	TRUE	FALSE	FALSE	4	0.076	0.089
69	Same	TRUE	FALSE	TRUE	TRUE	4	0.216	0.208
70	Same	TRUE	FALSE	TRUE	FALSE	4	0.190	0.166
71	Same	TRUE	FALSE	FALSE	TRUE	4	0.049	0.041
72	Same	TRUE	FALSE	FALSE	FALSE	4	0.049	0.057
73	Same	FALSE	TRUE	TRUE	TRUE	4	0.214	0.227
74	Same	FALSE	TRUE	TRUE	FALSE	4	0.138	0.139
75	Same	FALSE	TRUE	FALSE	TRUE	4	0.136	0.129
76	Same	FALSE	TRUE	FALSE	FALSE	4	0.105	0.110
77	Same	FALSE	FALSE	TRUE	TRUE	4	0.240	0.234
78	Same	FALSE	FALSE	TRUE	FALSE	4	0.150	0.148
79	Same	FALSE	FALSE	FALSE	TRUE	4	0.034	0.029
80	Same	FALSE	FALSE	FALSE	FALSE	4	0.041	0.050
81	Similar	TRUE	TRUE	TRUE	TRUE	4	0.448	0.286
82	Similar	TRUE	TRUE	TRUE	FALSE	4	0.314	0.207
83	Similar	TRUE	TRUE	FALSE	TRUE	4	0.119	0.101
84	Similar	TRUE	TRUE	FALSE	FALSE	4	0.079	0.074
85	Similar	TRUE	FALSE	TRUE	TRUE	4	0.221	0.216
86	Similar	TRUE	FALSE	TRUE	FALSE	4	0.152	0.145
87	Similar	TRUE	FALSE	FALSE	TRUE	4	0.030	0.028
88	Similar	TRUE	FALSE	FALSE	FALSE	4	0.028	0.030
89	Similar	FALSE	TRUE	TRUE	TRUE	4	0.202	0.204
90	Similar	FALSE	TRUE	TRUE	FALSE	4	0.111	0.113
91	Similar	FALSE	TRUE	FALSE	TRUE	4	0.122	0.129
92	Similar	FALSE	TRUE	FALSE	FALSE	4	0.122	0.123
93	Similar	FALSE	FALSE	TRUE	TRUE	4	0.191	0.217
94	Similar	FALSE	FALSE	TRUE	FALSE	4	0.167	0.165
95	Similar	FALSE	FALSE	FALSE	TRUE	4	0.036	0.036
96	Similar	FALSE	FALSE	FALSE	FALSE	4	0.047	0.041
97	Same	TRUE	TRUE	TRUE	TRUE	8	0.245	0.208
98	Same	TRUE	TRUE	TRUE	FALSE	8	0.231	0.188
99	Same	TRUE	TRUE	FALSE	TRUE	8	0.057	0.060
100	Same	TRUE	TRUE	FALSE	FALSE	8	0.051	0.053
101	Same	TRUE	FALSE	TRUE	TRUE	8	0.149	0.116
102	Same	TRUE	FALSE	TRUE	FALSE	8	0.140	0.104
103	Same	TRUE	FALSE	FALSE	TRUE	8	0.016	0.015
104	Same	TRUE	FALSE	FALSE	FALSE	8	0.016	0.012
105	Same	FALSE	TRUE	TRUE	TRUE	8	0.152	0.150
106	Same	FALSE	TRUE	TRUE	FALSE	8	0.094	0.101
107	Same	FALSE	TRUE	FALSE	TRUE	8	0.038	0.036
108	Same	FALSE	TRUE	FALSE	FALSE	8	0.032	0.030
109	Same	FALSE	FALSE	TRUE	TRUE	8	0.092	0.086
110	Same	FALSE	FALSE	TRUE	FALSE	8	0.104	0.115
111	Same	FALSE	FALSE	FALSE	TRUE	8	0.013	0.014
112	Same	FALSE	FALSE	FALSE	FALSE	8	0.012	0.013

113	Similar	TRUE	TRUE	TRUE	TRUE	8	0.286	0.236
114	Similar	TRUE	TRUE	TRUE	FALSE	8	0.219	0.173
115	Similar	TRUE	TRUE	FALSE	TRUE	8	0.056	0.058
116	Similar	TRUE	TRUE	FALSE	FALSE	8	0.040	0.041
117	Similar	TRUE	FALSE	TRUE	TRUE	8	0.152	0.141
118	Similar	TRUE	FALSE	TRUE	FALSE	8	0.120	0.091
119	Similar	TRUE	FALSE	FALSE	TRUE	8	0.021	0.021
120	Similar	TRUE	FALSE	FALSE	FALSE	8	0.030	0.026
121	Similar	FALSE	TRUE	TRUE	TRUE	8	0.096	0.099
122	Similar	FALSE	TRUE	TRUE	FALSE	8	0.033	0.046
123	Similar	FALSE	TRUE	FALSE	TRUE	8	0.068	0.064
124	Similar	FALSE	TRUE	FALSE	FALSE	8	0.040	0.060
125	Similar	FALSE	FALSE	TRUE	TRUE	8	0.109	0.119
126	Similar	FALSE	FALSE	TRUE	FALSE	8	0.084	0.084
127	Similar	FALSE	FALSE	FALSE	TRUE	8	0.013	0.019
128	Similar	FALSE	FALSE	FALSE	FALSE	8	0.026	0.021
129	Same	TRUE	TRUE	TRUE	TRUE	16	0.152	0.146
130	Same	TRUE	TRUE	TRUE	FALSE	16	0.123	0.113
131	Same	TRUE	TRUE	FALSE	TRUE	16	0.011	0.011
132	Same	TRUE	TRUE	FALSE	FALSE	16	0.024	0.021
133	Same	TRUE	FALSE	TRUE	TRUE	16	0.059	0.054
134	Same	TRUE	FALSE	TRUE	FALSE	16	0.056	0.054
135	Same	TRUE	FALSE	FALSE	TRUE	16	0.008	0.008
136	Same	TRUE	FALSE	FALSE	FALSE	16	0.000	0.001
137	Same	FALSE	TRUE	TRUE	TRUE	16	0.060	0.060
138	Same	FALSE	TRUE	TRUE	FALSE	16	0.051	0.055
139	Same	FALSE	TRUE	FALSE	TRUE	16	0.023	0.022
140	Same	FALSE	TRUE	FALSE	FALSE	16	0.019	0.023
141	Same	FALSE	FALSE	TRUE	TRUE	16	0.096	0.100
142	Same	FALSE	FALSE	TRUE	FALSE	16	0.041	0.043
143	Same	FALSE	FALSE	FALSE	TRUE	16	0.004	0.006
144	Same	FALSE	FALSE	FALSE	FALSE	16	0.014	0.014
145	Similar	TRUE	TRUE	TRUE	TRUE	16	0.154	0.131
146	Similar	TRUE	TRUE	TRUE	FALSE	16	0.117	0.116
147	Similar	TRUE	TRUE	FALSE	TRUE	16	0.024	0.029
148	Similar	TRUE	TRUE	FALSE	FALSE	16	0.023	0.036
149	Similar	TRUE	FALSE	TRUE	TRUE	16	0.069	0.057
150	Similar	TRUE	FALSE	TRUE	FALSE	16	0.051	0.038
151	Similar	TRUE	FALSE	FALSE	TRUE	16	0.009	0.010
152	Similar	TRUE	FALSE	FALSE	FALSE	16	0.011	0.012
153	Similar	FALSE	TRUE	TRUE	TRUE	16	0.094	0.095
154	Similar	FALSE	TRUE	TRUE	FALSE	16	0.053	0.055
155	Similar	FALSE	TRUE	FALSE	TRUE	16	0.033	0.031
156	Similar	FALSE	TRUE	FALSE	FALSE	16	0.014	0.011
157	Similar	FALSE	FALSE	TRUE	TRUE	16	0.062	0.066

158	Similar	FALSE	FALSE	TRUE	FALSE	16	0.036	0.030
159	Similar	FALSE	FALSE	FALSE	TRUE	16	0.010	0.009
160	Similar	FALSE	FALSE	FALSE	FALSE	16	0.014	0.014

Table D.1.2: Dialects are not well maintained in the absence of sexual selection. Annotations are the same as in **Table D.1.1**. Females had 0% preference for songs that matched their template, females had uniform dialects, and male songs initially matched female song templates (Same).

Par #	Breed	Tutor	Vertical	Dial	Center	Edge
1	TRUE	TRUE	TRUE	2	0.003	0.003
2	TRUE	TRUE	FALSE	2	0.114	0.119
3	TRUE	FALSE	TRUE	2	0.170	0.200
4	TRUE	FALSE	FALSE	2	0.178	0.170
5	FALSE	TRUE	TRUE	2	0.116	0.101
6	FALSE	TRUE	FALSE	2	0.125	0.117
7	FALSE	FALSE	TRUE	2	0.121	0.123
8	FALSE	FALSE	FALSE	2	0.112	0.118
9	TRUE	TRUE	TRUE	4	0.009	0.003
10	TRUE	TRUE	FALSE	4	0.059	0.061
11	TRUE	FALSE	TRUE	4	0.051	0.051
12	TRUE	FALSE	FALSE	4	0.056	0.062
13	FALSE	TRUE	TRUE	4	0.054	0.051
14	FALSE	TRUE	FALSE	4	0.063	0.074
15	FALSE	FALSE	TRUE	4	0.059	0.049
16	FALSE	FALSE	FALSE	4	0.051	0.050

Table D.1.3: Learning from reproductively successful tutors did not have a consistent effect on the maintenance of regional dialects. Annotations are the same as in **Table D.1.1**. Females had a uniform template, male songs initially matched female song templates (Same), and tutor choice was socially informed.

Par #	Breed	Tutor	Vertical	Dialects	Center	Edge
1	TRUE	TRUE	TRUE	1	0.591	0.561
2	TRUE	TRUE	FALSE	1	0.537	0.606
3	TRUE	FALSE	TRUE	1	0.836	0.803
4	TRUE	FALSE	FALSE	1	0.568	0.569
5	FALSE	TRUE	TRUE	1	0.749	0.746
6	FALSE	TRUE	FALSE	1	0.548	0.549
7	FALSE	FALSE	TRUE	1	0.868	0.847
8	FALSE	FALSE	FALSE	1	0.615	0.620
9	TRUE	TRUE	TRUE	2	0.619	0.342
10	TRUE	TRUE	FALSE	2	0.248	0.304
11	TRUE	FALSE	TRUE	2	0.358	0.349

12	TRUE	FALSE	FALSE	2	0.249	0.233
13	FALSE	TRUE	TRUE	2	0.398	0.399
14	FALSE	TRUE	FALSE	2	0.319	0.306
15	FALSE	FALSE	TRUE	2	0.344	0.342
16	FALSE	FALSE	FALSE	2	0.273	0.228
17	TRUE	TRUE	TRUE	4	0.449	0.295
18	TRUE	TRUE	FALSE	4	0.128	0.134
19	TRUE	FALSE	TRUE	4	0.205	0.205
20	TRUE	FALSE	FALSE	4	0.069	0.075
21	FALSE	TRUE	TRUE	4	0.156	0.161
22	FALSE	TRUE	FALSE	4	0.188	0.162
23	FALSE	FALSE	TRUE	4	0.245	0.230
24	FALSE	FALSE	FALSE	4	0.081	0.092
25	TRUE	TRUE	TRUE	8	0.281	0.274
26	TRUE	TRUE	FALSE	8	0.036	0.031
27	TRUE	FALSE	TRUE	8	0.168	0.126
28	TRUE	FALSE	FALSE	8	0.025	0.031
29	FALSE	TRUE	TRUE	8	0.133	0.137
30	FALSE	TRUE	FALSE	8	0.066	0.068
31	FALSE	FALSE	TRUE	8	0.088	0.086
32	FALSE	FALSE	FALSE	8	0.023	0.036
33	TRUE	TRUE	TRUE	16	0.146	0.125
34	TRUE	TRUE	FALSE	16	0.033	0.031
35	TRUE	FALSE	TRUE	16	0.050	0.040
36	TRUE	FALSE	FALSE	16	0.010	0.008
37	FALSE	TRUE	TRUE	16	0.073	0.067
38	FALSE	TRUE	FALSE	16	0.018	0.015
39	FALSE	FALSE	TRUE	16	0.057	0.061
40	FALSE	FALSE	FALSE	16	0.017	0.015

Table D.1.4: Socially uninformed tutor choice using the Add strategy led to the loss of all dialects. Annotations are the same as in **Table D.1.1**. Social is whether learners were more likely to choose tutors that had bred in the last time step. Females had a uniform template, male songs initially matched female song templates (Same), and vertical learning was disabled.

par #	Breed	Tutor	Social	Dialects	Center	Edge
1	TRUE	TRUE	TRUE	1	0.244	0.240
2	TRUE	TRUE	FALSE	1	0.000	0.000
3	TRUE	FALSE	TRUE	1	0.267	0.225
4	TRUE	FALSE	FALSE	1	0.002	0.003
5	FALSE	TRUE	TRUE	1	0.279	0.270
6	FALSE	TRUE	FALSE	1	0.006	0.003
7	FALSE	FALSE	TRUE	1	0.298	0.332
8	FALSE	FALSE	FALSE	1	0.002	0.001

9	TRUE	TRUE	TRUE	2	0.193	0.154
10	TRUE	TRUE	FALSE	2	0.000	0.003
11	TRUE	FALSE	TRUE	2	0.167	0.146
12	TRUE	FALSE	FALSE	2	0.000	0.002
13	FALSE	TRUE	TRUE	2	0.249	0.138
14	FALSE	TRUE	FALSE	2	0.001	0.003
15	FALSE	FALSE	TRUE	2	0.182	0.171
16	FALSE	FALSE	FALSE	2	0.001	0.000
17	TRUE	TRUE	TRUE	4	0.148	0.132
18	TRUE	TRUE	FALSE	4	0.002	0.003
19	TRUE	FALSE	TRUE	4	0.101	0.093
20	TRUE	FALSE	FALSE	4	0.001	0.002
21	FALSE	TRUE	TRUE	4	0.218	0.120
22	FALSE	TRUE	FALSE	4	0.000	0.008
23	FALSE	FALSE	TRUE	4	0.120	0.100
24	FALSE	FALSE	FALSE	4	0.000	0.001
25	TRUE	TRUE	TRUE	8	0.103	0.084
26	TRUE	TRUE	FALSE	8	0.000	0.001
27	TRUE	FALSE	TRUE	8	0.007	0.006
28	TRUE	FALSE	FALSE	8	0.004	0.003
29	FALSE	TRUE	TRUE	8	0.138	0.100
30	FALSE	TRUE	FALSE	8	0.002	0.003
31	FALSE	FALSE	TRUE	8	0.066	0.068
32	FALSE	FALSE	FALSE	8	0.002	0.001
33	TRUE	TRUE	TRUE	16	0.066	0.050
34	TRUE	TRUE	FALSE	16	0.003	0.003
35	TRUE	FALSE	TRUE	16	0.001	0.001
36	TRUE	FALSE	FALSE	16	0.001	0.001
37	FALSE	TRUE	TRUE	16	0.096	0.075
38	FALSE	TRUE	FALSE	16	0.000	0.002
39	FALSE	FALSE	TRUE	16	0.017	0.014
40	FALSE	FALSE	FALSE	16	0.002	0.002

Table D.1.5: Using one-syllable songs led to higher template-matching accuracy, but did not change any evolutionary dynamics. Annotations are the same as in **Table D.1.1**. Females had a uniform template, male songs initially matched female song templates (Same), and socially informed tutor choice was disabled.

	Breed	Tutor	Vertical	Dialects	Center	Edge
1	TRUE	TRUE	TRUE	1	0.900	0.915
2	TRUE	TRUE	FALSE	1	0.560	0.535
3	TRUE	FALSE	TRUE	1	0.920	0.920
4	TRUE	FALSE	FALSE	1	0.600	0.625
5	FALSE	TRUE	TRUE	1	0.915	0.915
6	FALSE	TRUE	FALSE	1	0.555	0.550
7	FALSE	FALSE	TRUE	1	0.935	0.925
8	FALSE	FALSE	FALSE	1	0.510	0.630
9	TRUE	TRUE	TRUE	2	0.890	0.530
10	TRUE	TRUE	FALSE	2	0.395	0.360
11	TRUE	FALSE	TRUE	2	0.665	0.470
12	TRUE	FALSE	FALSE	2	0.230	0.205
13	FALSE	TRUE	TRUE	2	0.335	0.345
14	FALSE	TRUE	FALSE	2	0.260	0.335
15	FALSE	FALSE	TRUE	2	0.355	0.365
16	FALSE	FALSE	FALSE	2	0.305	0.310
17	TRUE	TRUE	TRUE	4	0.785	0.495
18	TRUE	TRUE	FALSE	4	0.150	0.110
19	TRUE	FALSE	TRUE	4	0.595	0.375
20	TRUE	FALSE	FALSE	4	0.075	0.070
21	FALSE	TRUE	TRUE	4	0.215	0.215
22	FALSE	TRUE	FALSE	4	0.110	0.125
23	FALSE	FALSE	TRUE	4	0.275	0.315
24	FALSE	FALSE	FALSE	4	0.155	0.130

Table D.1.6: Changing the matrix shape had little effect on dialect maintenance. Annotations are the same as in **Table D.1.1**. Dim is the dimensions of the matrix. Females had a uniform template and male songs initially matched female song templates (Same).

Par #	**Breed**	**Tutor**	**Vertical**	**Dim**	**Dialects**	**Corner**	**Far Corner**
1	TRUE	TRUE	TRUE	4x100	2	0.490	0.283
2	TRUE	TRUE	TRUE	5x80	2	0.495	0.312
3	TRUE	TRUE	TRUE	10x40	2	0.558	0.318
4	TRUE	TRUE	TRUE	20x20	2	0.526	0.378
5	TRUE	TRUE	FALSE	4x100	2	0.270	0.192
6	TRUE	TRUE	FALSE	5x80	2	0.289	0.233
7	TRUE	TRUE	FALSE	10x40	2	0.286	0.275
8	TRUE	TRUE	FALSE	20x20	2	0.305	0.264

9	TRUE	FALSE	TRUE	4x100	2	0.362	0.324
10	TRUE	FALSE	TRUE	5x80	2	0.375	0.348
11	TRUE	FALSE	TRUE	10x40	2	0.360	0.329
12	TRUE	FALSE	TRUE	20x20	2	0.372	0.355
13	TRUE	FALSE	FALSE	4x100	2	0.256	0.249
14	TRUE	FALSE	FALSE	5x80	2	0.262	0.238
15	TRUE	FALSE	FALSE	10x40	2	0.219	0.275
16	TRUE	FALSE	FALSE	20x20	2	0.296	0.271
17	FALSE	TRUE	TRUE	4x100	2	0.348	0.350
18	FALSE	TRUE	TRUE	5x80	2	0.316	0.315
19	FALSE	TRUE	TRUE	10x40	2	0.366	0.355
20	FALSE	TRUE	TRUE	20x20	2	0.391	0.356
21	FALSE	TRUE	FALSE	4x100	2	0.328	0.236
22	FALSE	TRUE	FALSE	5x80	2	0.284	0.259
23	FALSE	TRUE	FALSE	10x40	2	0.275	0.241
24	FALSE	TRUE	FALSE	20x20	2	0.293	0.261
25	FALSE	FALSE	TRUE	4x100	2	0.364	0.375
26	FALSE	FALSE	TRUE	5x80	2	0.355	0.371
27	FALSE	FALSE	TRUE	10x40	2	0.341	0.373
28	FALSE	FALSE	TRUE	20x20	2	0.384	0.370
29	FALSE	FALSE	FALSE	4x100	2	0.239	0.246
30	FALSE	FALSE	FALSE	5x80	2	0.241	0.231
31	FALSE	FALSE	FALSE	10x40	2	0.260	0.278
32	FALSE	FALSE	FALSE	20x20	2	0.287	0.293
33	TRUE	TRUE	TRUE	4x100	4	0.444	0.303
34	TRUE	TRUE	TRUE	5x80	4	0.436	0.281
35	TRUE	TRUE	TRUE	10x40	4	0.437	0.303
36	TRUE	TRUE	TRUE	20x20	4	0.376	0.311
37	TRUE	TRUE	FALSE	4x100	4	0.129	0.111
38	TRUE	TRUE	FALSE	5x80	4	0.148	0.134
39	TRUE	TRUE	FALSE	10x40	4	0.119	0.140
40	TRUE	TRUE	FALSE	20x20	4	0.100	0.101
41	TRUE	FALSE	TRUE	4x100	4	0.205	0.206
42	TRUE	FALSE	TRUE	5x80	4	0.211	0.212
43	TRUE	FALSE	TRUE	10x40	4	0.211	0.201
44	TRUE	FALSE	TRUE	20x20	4	0.220	0.215
45	TRUE	FALSE	FALSE	4x100	4	0.042	0.037
46	TRUE	FALSE	FALSE	5x80	4	0.059	0.065
47	TRUE	FALSE	FALSE	10x40	4	0.075	0.078
48	TRUE	FALSE	FALSE	20x20	4	0.025	0.030
49	FALSE	TRUE	TRUE	4x100	4	0.281	0.262
50	FALSE	TRUE	TRUE	5x80	4	0.248	0.256
51	FALSE	TRUE	TRUE	10x40	4	0.219	0.213
52	FALSE	TRUE	TRUE	20x20	4	0.190	0.182
53	FALSE	TRUE	FALSE	4x100	4	0.162	0.149

54	FALSE	TRUE	FALSE	5x80	4	0.157	0.135
55	FALSE	TRUE	FALSE	10x40	4	0.121	0.140
56	FALSE	TRUE	FALSE	20x20	4	0.118	0.101
57	FALSE	FALSE	TRUE	4x100	4	0.192	0.193
58	FALSE	FALSE	TRUE	5x80	4	0.212	0.201
59	FALSE	FALSE	TRUE	10x40	4	0.218	0.213
60	FALSE	FALSE	TRUE	20x20	4	0.209	0.189
61	FALSE	FALSE	FALSE	4x100	4	0.063	0.059
62	FALSE	FALSE	FALSE	5x80	4	0.040	0.037
63	FALSE	FALSE	FALSE	10x40	4	0.041	0.044
64	FALSE	FALSE	FALSE	20x20	4	0.041	0.035

Table D.1.7: Increasing the matrix size led to higher template-matching accuracy. Annotations are the same as in **Table D.1.1**. The first three rows of Par # refer to the parameter sets in **Table D.1.1**. Dim is the dimensions of the matrix. Females had a uniform template, male songs initially matched female song templates (Same), vertical learning was enabled, tutor choice was local, and there was high nest site fidelity.

Par #	Dialects	Dim	Center	Edge
1	1	20x20	0.583	0.570
49	2	20x20	0.534	0.338
97	4	20x20	0.398	0.282
11	2	20x40	0.533	0.308
2	4	40x40	0.601	0.308

Table D.2.1: Dialects establish when there is high nest site fidelity and vertical learning. Annotations are the same as in **Table D.1.1**. Corner is a measurement of the template-matching accuracy at the outside corner of regional dialect 1. Far Corner is a measurement of the template-matching accuracy at the outside corner of the final regional dialect in the matrix. Asterisks (*) denote cases where the final regional dialect achieved at least 5% template-matching accuracy.

Par #	Breed	Tutor	Vertical	Uniform	Dialects	Corner	Far Corner
1	TRUE	TRUE	TRUE	TRUE	2	0.498	0.437*
2	TRUE	TRUE	TRUE	FALSE	2	0.411	0.334*
3	TRUE	TRUE	FALSE	TRUE	2	0.498	0
4	TRUE	TRUE	FALSE	FALSE	2	0.432	0
5	TRUE	FALSE	TRUE	TRUE	2	0.555	0.225*
6	TRUE	FALSE	TRUE	FALSE	2	0.467	0.165*
7	TRUE	FALSE	FALSE	TRUE	2	0.425	0.001
8	TRUE	FALSE	FALSE	FALSE	2	0.533	0
9	FALSE	TRUE	TRUE	TRUE	2	0.595	0.007
10	FALSE	TRUE	TRUE	FALSE	2	0.582	0
11	FALSE	TRUE	FALSE	TRUE	2	0.482	0
12	FALSE	TRUE	FALSE	FALSE	2	0.423	0
13	FALSE	FALSE	TRUE	TRUE	2	0.695	0
14	FALSE	FALSE	TRUE	FALSE	2	0.604	0
15	FALSE	FALSE	FALSE	TRUE	2	0.478	0
16	FALSE	FALSE	FALSE	FALSE	2	0.497	0.001
17	TRUE	TRUE	TRUE	TRUE	4	0.442	0.326*
18	TRUE	TRUE	TRUE	FALSE	4	0.346	0.269*
19	TRUE	TRUE	FALSE	TRUE	4	0.366	0
20	TRUE	TRUE	FALSE	FALSE	4	0.341	0
21	TRUE	FALSE	TRUE	TRUE	4	0.283	0.198*
22	TRUE	FALSE	TRUE	FALSE	4	0.234	0.129*
23	TRUE	FALSE	FALSE	TRUE	4	0.473	0
24	TRUE	FALSE	FALSE	FALSE	4	0.437	0
25	FALSE	TRUE	TRUE	TRUE	4	0.616	0
26	FALSE	TRUE	TRUE	FALSE	4	0.563	0.003
27	FALSE	TRUE	FALSE	TRUE	4	0.368	0
28	FALSE	TRUE	FALSE	FALSE	4	0.430	0
29	FALSE	FALSE	TRUE	TRUE	4	0.648	0
30	FALSE	FALSE	TRUE	FALSE	4	0.555	0.001
31	FALSE	FALSE	FALSE	TRUE	4	0.421	0
32	FALSE	FALSE	FALSE	FALSE	4	0.471	0
33	TRUE	TRUE	TRUE	TRUE	8	0.258	0.226*
34	TRUE	TRUE	TRUE	FALSE	8	0.193	0.195*
35	TRUE	TRUE	FALSE	TRUE	8	0.293	0
36	TRUE	TRUE	FALSE	FALSE	8	0.230	0

37	TRUE	FALSE	TRUE	TRUE	8	0.154	0.112*
38	TRUE	FALSE	TRUE	FALSE	8	0.111	0.103*
39	TRUE	FALSE	FALSE	TRUE	8	0.335	0
40	TRUE	FALSE	FALSE	FALSE	8	0.339	0
41	FALSE	TRUE	TRUE	TRUE	8	0.518	0.015
42	FALSE	TRUE	TRUE	FALSE	8	0.347	0.022
43	FALSE	TRUE	FALSE	TRUE	8	0.292	0.002
44	FALSE	TRUE	FALSE	FALSE	8	0.273	0
45	FALSE	FALSE	TRUE	TRUE	8	0.649	0.011
46	FALSE	FALSE	TRUE	FALSE	8	0.479	0.007
47	FALSE	FALSE	FALSE	TRUE	8	0.402	0
48	FALSE	FALSE	FALSE	FALSE	8	0.413	0
49	TRUE	TRUE	TRUE	TRUE	16	0.150	0.133*
50	TRUE	TRUE	TRUE	FALSE	16	0.127	0.097*
51	TRUE	TRUE	FALSE	TRUE	16	0.210	0
52	TRUE	TRUE	FALSE	FALSE	16	0.197	0
53	TRUE	FALSE	TRUE	TRUE	16	0.113	0.073*
54	TRUE	FALSE	TRUE	FALSE	16	0.110	0.045
55	TRUE	FALSE	FALSE	TRUE	16	0.344	0
56	TRUE	FALSE	FALSE	FALSE	16	0.339	0
57	FALSE	TRUE	TRUE	TRUE	16	0.341	0.027
58	FALSE	TRUE	TRUE	FALSE	16	0.230	0.032
59	FALSE	TRUE	FALSE	TRUE	16	0.192	0
60	FALSE	TRUE	FALSE	FALSE	16	0.150	0.001
61	FALSE	FALSE	TRUE	TRUE	16	0.454	0.008
62	FALSE	FALSE	TRUE	FALSE	16	0.333	0.012
63	FALSE	FALSE	FALSE	TRUE	16	0.318	0
64	FALSE	FALSE	FALSE	FALSE	16	0.237	0

Table D.2.2: Learning from reproductively successful tutors using the Conformity strategy was not sufficient to establish regional dialects. Annotations are the same as in **Tables D.1.1** and **D.2.1**. Females had a uniform template and vertical learning was disabled.

Par #	Breed	Tutor	Dialects	Corner	Far Corner
1	TRUE	TRUE	2	0.524	0
2	TRUE	FALSE	2	0.478	0
3	FALSE	TRUE	2	0.597	0
4	FALSE	FALSE	2	0.535	0
5	TRUE	TRUE	4	0.353	0.004
6	TRUE	FALSE	4	0.478	0.001
7	FALSE	TRUE	4	0.424	0
8	FALSE	FALSE	4	0.508	0
9	TRUE	TRUE	8	0.321	0.006
10	TRUE	FALSE	8	0.353	0

11	FALSE	TRUE	8	0.299	0.011
12	FALSE	FALSE	8	0.414	0
13	TRUE	TRUE	16	0.265	0
14	TRUE	FALSE	16	0.359	0
15	FALSE	TRUE	16	0.172	0.003
16	FALSE	FALSE	16	0.367	0

Table D.2.3: Learning from reproductively successful tutors using the Add strategy was sufficient to establish regional dialects. Annotations are the same as in **Tables D.1.1** and **D.2.1**. Social is whether learners were more likely to choose tutors that had bred in the last time step. Females had a uniform template and vertical learning was disabled.

Par #	**Breed**	**Tutor**	**Social**	**Dialects**	**Corner**	**Far Corner**
1	TRUE	TRUE	TRUE	2	0.161	0.165*
2	TRUE	TRUE	FALSE	2	0.000	0.001
3	TRUE	FALSE	TRUE	2	0.135	0.174*
4	TRUE	FALSE	FALSE	2	0.003	0
5	FALSE	TRUE	TRUE	2	0.201	0.173*
6	FALSE	TRUE	FALSE	2	0.000	0.001
7	FALSE	FALSE	TRUE	2	0.188	0.177*
8	FALSE	FALSE	FALSE	2	0.007	0.002
9	TRUE	TRUE	TRUE	4	0.100	0.128*
10	TRUE	TRUE	FALSE	4	0.002	0
11	TRUE	FALSE	TRUE	4	0.077	0.068*
12	TRUE	FALSE	FALSE	4	0.004	0
13	FALSE	TRUE	TRUE	4	0.197	0.162*
14	FALSE	TRUE	FALSE	4	0.004	0.001
15	FALSE	FALSE	TRUE	4	0.138	0.094*
16	FALSE	FALSE	FALSE	4	0.006	0
17	TRUE	TRUE	TRUE	8	0.112	0.09*
18	TRUE	TRUE	FALSE	8	0.002	0
19	TRUE	FALSE	TRUE	8	0.005	0.006
20	TRUE	FALSE	FALSE	8	0.002	0.004
21	FALSE	TRUE	TRUE	8	0.141	0.116*
22	FALSE	TRUE	FALSE	8	0.001	0.003
23	FALSE	FALSE	TRUE	8	0.067	0.059*
24	FALSE	FALSE	FALSE	8	0.001	0.006
25	TRUE	TRUE	TRUE	16	0.042	0.049
26	TRUE	TRUE	FALSE	16	0.000	0.001
27	TRUE	FALSE	TRUE	16	0.001	0.004
28	TRUE	FALSE	FALSE	16	0.001	0.003
29	FALSE	TRUE	TRUE	16	0.098	0.082*
30	FALSE	TRUE	FALSE	16	0.003	0
31	FALSE	FALSE	TRUE	16	0.015	0.02

32	FALSE	FALSE	FALSE	16	0.004	0.002

Table D.2.4: Using one-syllable did not affect average dialect establishment. Annotations are the same as in **Table D.1.1**. Females had a uniform template and socially informed tutor choice was disabled.

Par #	Breed	Tutor	Vertical	Dialects	Corner	Far Corner
1	TRUE	TRUE	TRUE	2	0.907	0.26*
2	TRUE	TRUE	FALSE	2	0.527	0
3	TRUE	FALSE	TRUE	2	0.787	0.107*
4	TRUE	FALSE	FALSE	2	0.487	0
5	FALSE	TRUE	TRUE	2	0.913	0
6	FALSE	TRUE	FALSE	2	0.567	0
7	FALSE	FALSE	TRUE	2	0.907	0
8	FALSE	FALSE	FALSE	2	0.547	0
9	TRUE	TRUE	TRUE	4	0.780	0.14*
10	TRUE	TRUE	FALSE	4	0.540	0
11	TRUE	FALSE	TRUE	4	0.553	0.073*
12	TRUE	FALSE	FALSE	4	0.553	0
13	FALSE	TRUE	TRUE	4	0.907	0
14	FALSE	TRUE	FALSE	4	0.520	0
15	FALSE	FALSE	TRUE	4	0.933	0
16	FALSE	FALSE	FALSE	4	0.593	0

Table D.2.5: Changing the matrix shape had little effect on dialect establishment. Annotations are the same as in **Tables D.1.1** and **D.2.1**. Dim is the dimensions of the matrix. Females had a uniform template.

Par #	Breed	Tutor	Vertical	Dim	Dialects	Corner	Far Corner
1	TRUE	TRUE	TRUE	4x100	2	0.392	0.37*
2	TRUE	TRUE	TRUE	5x80	2	0.499	0.416*
3	TRUE	TRUE	TRUE	10x40	2	0.500	0.386*
4	TRUE	TRUE	TRUE	20x20	2	0.512	0.416*
5	TRUE	TRUE	FALSE	4x100	2	0.304	0
6	TRUE	TRUE	FALSE	5x80	2	0.348	0.001
7	TRUE	TRUE	FALSE	10x40	2	0.403	0
8	TRUE	TRUE	FALSE	20x20	2	0.451	0
9	TRUE	FALSE	TRUE	4x100	2	0.568	0.157*
10	TRUE	FALSE	TRUE	5x80	2	0.566	0.21*
11	TRUE	FALSE	TRUE	10x40	2	0.558	0.205*
12	TRUE	FALSE	TRUE	20x20	2	0.559	0.178*
13	TRUE	FALSE	FALSE	4x100	2	0.405	0.001
14	TRUE	FALSE	FALSE	5x80	2	0.523	0.001

15	TRUE	FALSE	FALSE	10x40	2	0.431	0
16	TRUE	FALSE	FALSE	20x20	2	0.524	0
17	FALSE	TRUE	TRUE	4x100	2	0.635	0.001
18	FALSE	TRUE	TRUE	5x80	2	0.636	0
19	FALSE	TRUE	TRUE	10x40	2	0.628	0.001
20	FALSE	TRUE	TRUE	20x20	2	0.637	0.001
21	FALSE	TRUE	FALSE	4x100	2	0.422	0
22	FALSE	TRUE	FALSE	5x80	2	0.448	0
23	FALSE	TRUE	FALSE	10x40	2	0.491	0
24	FALSE	TRUE	FALSE	20x20	2	0.446	0
25	FALSE	FALSE	TRUE	4x100	2	0.685	0
26	FALSE	FALSE	TRUE	5x80	2	0.676	0.003
27	FALSE	FALSE	TRUE	10x40	2	0.699	0.001
28	FALSE	FALSE	TRUE	20x20	2	0.679	0
29	FALSE	FALSE	FALSE	4x100	2	0.527	0
30	FALSE	FALSE	FALSE	5x80	2	0.480	0
31	FALSE	FALSE	FALSE	10x40	2	0.444	0
32	FALSE	FALSE	FALSE	20x20	2	0.497	0
33	TRUE	TRUE	TRUE	4x100	4	0.416	0.164*
34	TRUE	TRUE	TRUE	5x80	4	0.403	0.213*
35	TRUE	TRUE	TRUE	10x40	4	0.434	0.295*
36	TRUE	TRUE	TRUE	20x20	4	0.484	0.306*
37	TRUE	TRUE	FALSE	4x100	4	0.277	0
38	TRUE	TRUE	FALSE	5x80	4	0.320	0
39	TRUE	TRUE	FALSE	10x40	4	0.402	0
40	TRUE	TRUE	FALSE	20x20	4	0.366	0
41	TRUE	FALSE	TRUE	4x100	4	0.328	0.178*
42	TRUE	FALSE	TRUE	5x80	4	0.305	0.149*
43	TRUE	FALSE	TRUE	10x40	4	0.312	0.192*
44	TRUE	FALSE	TRUE	20x20	4	0.313	0.185*
45	TRUE	FALSE	FALSE	4x100	4	0.379	0
46	TRUE	FALSE	FALSE	5x80	4	0.382	0
47	TRUE	FALSE	FALSE	10x40	4	0.339	0
48	TRUE	FALSE	FALSE	20x20	4	0.387	0
49	FALSE	TRUE	TRUE	4x100	4	0.611	0.003
50	FALSE	TRUE	TRUE	5x80	4	0.606	0.004
51	FALSE	TRUE	TRUE	10x40	4	0.593	0.004
52	FALSE	TRUE	TRUE	20x20	4	0.608	0
53	FALSE	TRUE	FALSE	4x100	4	0.280	0
54	FALSE	TRUE	FALSE	5x80	4	0.361	0
55	FALSE	TRUE	FALSE	10x40	4	0.418	0
56	FALSE	TRUE	FALSE	20x20	4	0.474	0
57	FALSE	FALSE	TRUE	4x100	4	0.660	0.004
58	FALSE	FALSE	TRUE	5x80	4	0.664	0.001
59	FALSE	FALSE	TRUE	10x40	4	0.646	0.006

60	FALSE	FALSE	TRUE	20x20	4	0.644	0
61	FALSE	FALSE	FALSE	4x100	4	0.538	0
62	FALSE	FALSE	FALSE	5x80	4	0.471	0
63	FALSE	FALSE	FALSE	10x40	4	0.449	0
64	FALSE	FALSE	FALSE	20x20	4	0.566	0

APPENDIX E

Quantifying syllable repertoires

E.1 Quantification

For **Chapter 2**, we determined estimates for *Carpodacus mexicanus* (**Table E.1.1**) and *Wilsonia canadensis* (**Table E.1.2**) using recordings obtained from xeno-canto[111]. Average species syllable repertoire size could not be estimated for *Saxicola caprata* or *Phylloscopus trochiloides* (**Figures E.1.1** and **E.1.2**), so these species were not included in the species syllable repertoire dataset. In the case of *Saxicola caprata*, only a small number of short recordings were available, and many new syllables were still being discovered by the end of the recording. We searched the Macaulay Library at the Cornell Lab of Ornithology's recording database[284] for additional songs, but they were not longer than the recordings on xeno-canto.org. *Phylloscopus trochiloides* is a proposed ring species with very different syllable and song types between regions[157]. We did know the region that each measurement was associated with (India, Siberia, or Kyrgyzstan), but only one or two usable recordings were available for each region, and new syllables were still being discovered at the end of these short recordings. We searched for additional recordings in the Macaulay Library database, but did not find longer recordings in the geographic regions of interest.

Table E.1.1: Syllable repertoire estimates for House Finch (*Carpodacus mexicanus*). (top table) XC ID is the xeno-canto ID number for each recording. The first half of the table shows when a percentage of the discovered unique syllables was found in regards to the number of bouts analyzed, while the second half shows the percentage of unique syllables discovered over time in seconds. Analyzed recordings were all at least two minutes long. In most cases, the half of the syllable repertoire was revealed in the first few bouts with the majority revealed before bout 12. These estimates were similar to those in Tracy and Baker 1999 [77] (bottom table), so we felt confident that they had defined "syllable types" per bird similarly to how we defined "unique syllables" per bird, and we combined the two datasets to get a final mean of 41.3 unique syllables per bird.

XC ID	bouts analyzed	50% rep	90% rep	full rep	Length (s)	50% rep	90% rep	full rep	syl rep
XC320727	17	1	7	11	124	4	48	73	29
XC268463	27	2	14	20	260	10	82	178	31
XC268458	24	2	6	14	190	9	36	95	33
XC353018	40	8	23	32	136	30	57	102	34
XC219331	20	3	11	12	167	15	69	78	35
XC179232	25	1	5	6	197	5	34	44	37
XC268457	18	2	11	11	139	9	82	84	44
XC268465	45	3	6	10	222	13	23	41	44
XC73494	19	3	10	11	178	12	71	74	49
XC268464	49	3	11	25	300	10	45	89	52

Average: 38.8 Unique syllables

Note: bout length varies widely within and between individuals

Syllable Repertoires from Tracy and Baker 1999 [77]:

31	30	31	72	64
45	38	41	34	35
58	43	40		

Average: 43.23 Unique syllables

Combined Average: 41.3 Unique syllables

Table E.1.2: Syllable repertoire estimates for Canada Warbler (*Wilsonia canadensis*). XC ID is the xeno-canto ID number for each recording. The first half of the table shows when a percentage of the discovered unique syllables was found in regards to the number of bouts analyzed, while the second half shows the percentage of unique syllables discovered over time in seconds. Analyzed recordings were all at least two minutes long. In most cases, half of the syllable repertoire was revealed in the first few bouts. The majority of the syllable repertoire was usually revealed by the fifth bout and well before the end of the recordings. The final mean used in this study was 17.43 unique syllables per bird.

XC ID	bouts analyzed	50% rep	90% rep	full rep	Length (s)	50% rep	90% rep	full rep	syl rep
XC51468	13	1	1	1	168	5	5	5	9
XC189300	27	1	3	3	324	2	27	27	11
XC179679	29	3	9	13	225	11	44	66	18
XC371402	22	1	5	10	174	3	34	72	14
XC294137	14	2	2	10	155	18	18	79	19
XC189302	39	2	4	9	239	8	19	36	25
XC370937	38	1	3	6	260	4	13	41	26

Average: 17.43 Unique syllables

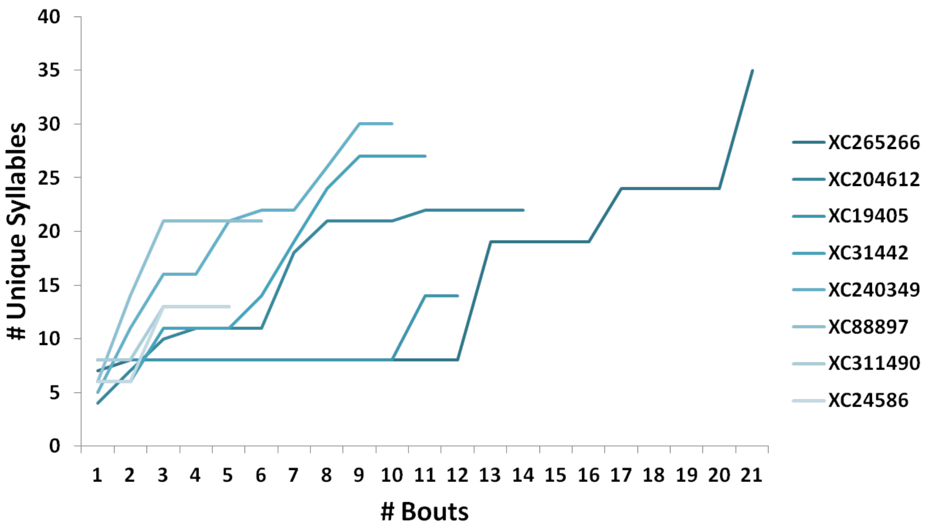


Figure E.1.1: Syllable repertoire estimates of Pied Bush Chat (*Saxicola caprata*) for rarefaction curves. Recordings ranged from only 40 seconds to 2.5 minutes; however, even in the longer recordings, new syllables were still being encountered in the last 30 seconds of recording. Based on this evidence, we were not convinced that we had complete repertoire data for individuals of this species and thus did not include it in the syllable repertoire size dataset.

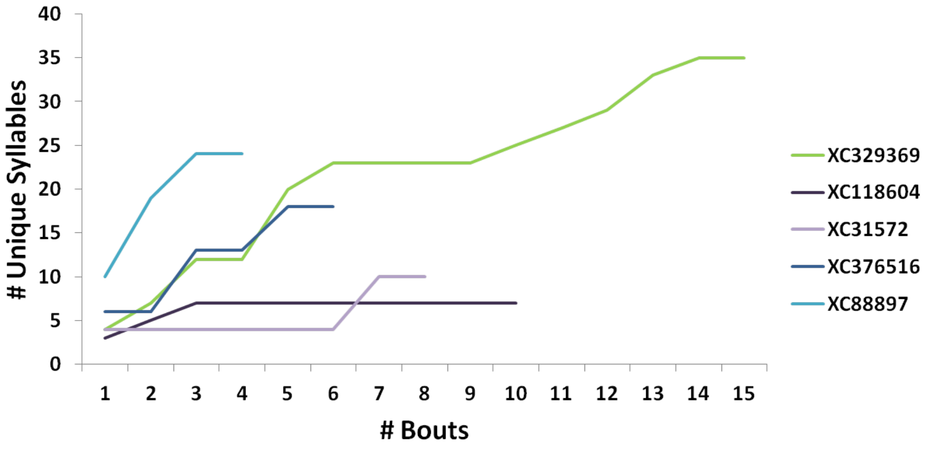


Figure E.1.2: Syllable repertoire estimates of Greenish Warbler (*Phylloscopus trochiloides*) for rarefaction curves. Green line is the recording for Kyrgyzstan, blue lines are Russian recordings, and purple lines are Indian recordings. Curves suggest that the recordings were not long enough to capture the full repertoire. While the longest recording (Kyrgyzstan) was ~2.5 minutes, the remaining recordings were only 40-70 seconds, and the intervals between songs in these recordings was often between 5-8 seconds. Based on this evidence, we were not convinced that we had complete repertoire data for individuals of this species and thus did not include it in the syllable repertoire size dataset.

For **Chapter 3**, we estimated the repertoires of three additional species: for *Philesturnus rufusater*, Jenkins[188] coded songs into different types and gave the song repertoire types of each male studied in Table 2 of that paper. Note that information from one male was missing from this table. We deduced the repertoire of the missing male by first looking at Table 1 from that paper (Jenkins' Table 1), which showed that 16 males had a repertoire size of one song. Only 15 of the males in Jenkins' Table 2 had a repertoire size of one song. Thus, the missing male had a repertoire size of one song. We then compared the bands of males present in Jenkins' Table 2 to the territory map in Figure 7 of that paper, and A_RW was the only male missing from Jenkins's Table 2. A_RW was located in the DC region of Jenkins's Figure 7, so we assigned that as his repertoire. Jenkins notes that neighboring males share song types, so the only other song A_RW could have known instead of DC was ZZ, which has the same number of unique syllables as DC. We divided each song type into unique syllables and counted the syllable repertoire for each male (**Table E.1.3**). For *Geospiza fortis* **Table E.1.4** and *G. scandens* **Table E.1.5**, we used recordings from the Macaulay Library to estimate the syllable repertoire size and syllables per song. For *G. fortis*, we also used the sonogram examples present in Grant and Grant[27].

Table E.1.3: Syllable repertoire for *Philesturnus rufusater*. Adapted from Jenkins 1977[188] table 2 with new syllable repertoire and syllables per song data included. Syl Rep is the sum of distinct syllables within each song, while Corrected is the number of unique syllables across the full repertoire (note that SR and VPH shared one syllable). The mean of Corrected was used in our paper. Syllables per song was calculated following the definition from our main methods. Asterisk (*) marks the individual that was missing from table 2 in the original publication.

Song Type	SR	VPH	CC	PH	KS	DC	ZZ	SE			
Syl\Song	3	2	3	3	2	1	1	4	Syl Rep	Corrected	Syll\Song
_A	1	1							5	4	2.5
♂Fern	1	1							5	4	2.5
♂Knob	1	1							5	4	2.5
BR_A			1						3	3	3
_WA			1						3	3	3
A_RB			1						3	3	3
AW_			1						3	3	3
AG_			1						3	3	3
A_RY			1						3	3	3
♂YW_A			1						3	3	3
AT_			1	1					6	6	3
_YA				1					3	3	3
GW_A				1					3	3	3
A_RG				1			1		4	4	2
RA_Y				1					3	3	3
YR_AG				1	1				5	5	2.5
AR_					1	1			3	3	1.5
Y_AR					1	1			3	3	1.5
A_WR						1			1	1	1
A_GW						1			1	1	1
_A							1		1	1	1
♂H. Gully							1	1	5	5	2.5
G_RA							1	1	5	5	2.5
AY_GR							1	1	5	5	2.5
_AY							1	1	5	5	2.5
B_WA								1	4	4	4
Y_AY								1	4	4	4
A_RW*						1			1	1	1
								Mean:	3.5	3.393	2.5

Table E.1.4: Syllable repertoire of *Geospiza fortis* Recordings from the Macaulay Library (top section of the table) and sonograms published in Grant and Grant[27] (bottom section of the table) were examined. The first column gives the Macaulay Library recording ID or the Grants' sonogram ID. The mean number of syllables we counted is at the bottom of the column 3. Because this species has a song repertoire of one, the value for syllables per song was the same as the syllable repertoire.

Geospiza fortis		
ML	Recordist	Syllables
86782	Robert I. Bowman	1
86729	Robert I. Bowman	1
86728	Robert I. Bowman	2
86727	Robert I. Bowman	1
86726	Robert I. Bowman	1
86724	Robert I. Bowman	1
86723	Robert I. Bowman	1
86719	Robert I. Bowman	2
86718	Robert I. Bowman	1
86717	Robert I. Bowman	1
86716	Robert I. Bowman	2
86714	Robert I. Bowman	1
82869	Robert I. Bowman	2
82865	Robert I. Bowman	1
82863	Robert I. Bowman	1
82597	Robert I. Bowman	2
82595	Robert I. Bowman	1
82576	Robert I. Bowman	1
82575	Robert I. Bowman	1
82574	Robert I. Bowman	1
46372	Margery R. Plymire	1

Grant and Grant (1996)	Recordist	Syllables
2666	Grant and Grant	1
4446	Grant and Grant	1
10826	Grant and Grant	1
4339	Grant and Grant	1
5555	Grant and Grant	1
5921	Grant and Grant	1
13901	Grant and Grant	1
17835	Grant and Grant	1
3612	Grant and Grant	1
16805	Grant and Grant	1
17103	Grant and Grant	1
5505	Grant and Grant	1
17796	Grant and Grant	1

5578	Grant and Grant	1
15236	Grant and Grant	1
15359	Grant and Grant	1
14963	Grant and Grant	1
10228	Grant and Grant	1
4946	Grant and Grant	1
4913	Grant and Grant	2
714	Grant and Grant	1
16167	Grant and Grant	2
5110	Grant and Grant	2
14687	Grant and Grant	1
5275	Grant and Grant	1
15514	Grant and Grant	1
A	Grant and Grant	1
10081	Grant and Grant	1
2639	Grant and Grant	1
10211	Grant and Grant	1
10550	Grant and Grant	2
14720	Grant and Grant	1
4620	Grant and Grant	2
B	Grant and Grant	2
	Mean:	1.20

Table E.1.5: Syllable repertoire of *Geospiza scadens*. Recordings from the Macaulay Library were examined. The first column gives the Macaulay Library recording ID. The mean number of syllables we counted is at the bottom of the column 3. Because this species has a song repertoire of one, the value for syllables per song was the same as the syllable repertoire.

Geospiza scandens		
ML	Recordist	Syllables
133749351	Eric DeFonso	1
46235	Robert I. Bowman	1
46234	Robert I. Bowman	2
46233	Robert I. Bowman	2
46228	Robert I. Bowman	2
46224	Robert I. Bowman	2
46222	Robert I. Bowman	2
46220	Robert I. Bowman	1
49219	Robert I. Bowman	2
46218	Robert I. Bowman	1
46217	Robert I. Bowman	1
	Mean:	1.54

E.2 Xeno-canto file citations

Wilsonia canadensis

Andrew Spencer, XC51468. Accessible at www.xeno-canto.org/51468.

Martin St-Michel, XC294137. Accessible at www.xeno-canto.org/294137.

Matt Wistrand, XC370937. Accessible at www.xeno-canto.org/370937.

Matt Wistrand, XC371402. Accessible at www.xeno-canto.org/371402.

Paul J. Hurtado, XC179679. Accessible at www.xeno-canto.org/179679.

Richard E. Webster, XC189300. Accessible at www.xeno-canto.org/189300.

Richard E. Webster, XC189302. Accessible at www.xeno-canto.org/189302.

Carpodacus mexicanus

Ian Cruickshank, XC73494. Accessible at www.xeno-canto.org/73494.

Nick Komar, XC320727. Accessible at www.xeno-canto.org/320727.

Paul Marvin, XC219331. Accessible at www.xeno-canto.org/219331.

Richard E. Webster, XC268457. Accessible at www.xeno-canto.org/268457.

Richard E. Webster, XC268458. Accessible at www.xeno-canto.org/268458.

Richard E. Webster, XC268463. Accessible at www.xeno-canto.org/268463.

Richard E. Webster, XC268464. Accessible at www.xeno-canto.org/268464.

Richard E. Webster, XC268465. Accessible at www.xeno-canto.org/268465.

Richard E. Webster, XC353018. Accessible at www.xeno-canto.org/353018.

Thomas G. Graves, XC179232. Accessible at www.xeno-canto.org/179232.

Saxicola caprata

Allen T. Chartier, XC31442. Accessible at www.xeno-canto.org/31442.

chiefredearth, XC265266. Accessible at www.xeno-canto.org/265266.

David Edwards, XC24586. Accessible at www.xeno-canto.org/24586.

Frank Lambert, XC88897. Accessible at www.xeno-canto.org/88897.

Mike Nelson, XC204612. Accessible at www.xeno-canto.org/204612.

Mike Nelson, XC240349. Accessible at www.xeno-canto.org/240349.

Peter Boesman, XC311490. Accessible at www.xeno-canto.org/311490.

Stuart Fisher, XC19405. Accessible at www.xeno-canto.org/19405.

Phylloscopus trochiloides

Allen T. Chartier, XC31572. Accessible at www.xeno-canto.org/31572.

Frank Lambert, XC88897. Accessible at www.xeno-canto.org/88897.

Lars Edenius, XC376516. Accessible at www.xeno-canto.org/376516.

Manuel Schweizer, XC329369. Accessible at www.xeno-canto.org/329369.

Timo Janhonen, XC118604. Accessible at www.xeno-canto.org/118604.