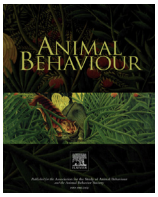
Animal Behaviour 139 (2018) 117–125

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav



Song type matching and vocal performance in territorial signalling by male swamp sparrows

Irene A. Liu*, Jill A. Soha, Stephen Nowicki

Department of Biology, Duke University, Durham, NC, U.S.A.

ARTICLE INFO

Article history:
Received 18 October 2017
Initial acceptance 17 November 2017
Final acceptance 7 February 2018

MS. number: A17-00836R

Keywords:
aggression
audience effect
birdsong
communication network
playback
song type matching
swamp sparrow
vocal performance

In songbird species with repertoires of multiple songs, individuals in territorial interactions can engage in song type matching, in which one bird responds to another using the same song type. Song type matching is thought to be associated with aggressive intent, although empirical support for this hypothesis is mixed. Here we test the alternative hypothesis that males selectively use song type matching, depending on singing ability, to optimize their relative performance in a communication network. We recorded the responses of male swamp sparrows, *Melospiza georgiana*, to playback trials in which they heard stimulus songs of higher or lower vocal performance relative to their own version of those songs. We predicted that, if males use song type matching to influence the perceptions of conspecifics outside the interacting dyad, males would (1) match stimulus songs that they themselves could perform better and (2) respond with a different song type to stimulus songs that they could not perform as well. We found that males song-type matched more often than expected by chance across trials, but contrary to our expectations, they were at least as likely to match to playback of higher-performance songs as to playback of lower-performance songs. As in previous studies, we also found that males sang with higher vocal performance in response to playback than when singing spontaneously, and that they did not preferentially respond with their highest-performance song type as a countersinging strategy. Our results support the idea that in swamp sparrows, song type matching functions primarily within the dyad rather than to broadcast superior performance ability to other conspecifics in the communication network.

Songbirds use song as a signal in territorial interactions. Even when these interactions directly involve only two birds, the songs of each bird may be heard not only by the opponent but also by other conspecifics in the area. Central to the concept of eavesdropping within communication networks is the idea that eavesdroppers can gain more information from hearing an interaction between two males than from hearing each male separately (McGregor & Dabelsteen, 1996). If eavesdropping ultimately affects the fitness of the signallers, then an individual may make signalling decisions that account for both the opponent and a wider audience (Logue & Forstmeier, 2008). One of these signalling decisions, in species with repertoires of multiple song types, is which song to sing next. Does a bird merely select songs from its repertoire at random, or does it choose particular songs, and if the latter, what drives this choice? Here we investigate the interaction of two phenomena that might affect the choice of songs in aggressive vocal interactions: vocal performance and song type matching.

One factor that can influence a male's choice of song is the song type sung by the intruder itself. When two birds share one or more types, song type matching is a possible outcome. First proposed to direct a signal to a particular rival (Brémond, 1968, cited in Armstrong, 1973), song type matching was then described as a signal of threat to new rivals. Male great tits, *Parus major*, were observed to decrease matching after habituation to playback (Krebs, Ashcroft, & Orsdol, 1981) and match more often to unfamiliar males than to neighbours (Falls, Krebs, & McGregor, 1982). These observations have been repeated in other species as well (e.g. song sparrows, *Melospiza melodia*: Beecher, Stoddard, Campbell, & Horning, 1996; western meadowlarks, *Sturnella neglecta*: Falls, 1985). Further studies linking song type matching with conflict escalation between neighbours (Burt, Campbell, & Beecher, 2001) and with indirect measures of aggression (Vehrencamp, 2001) also suggested that matching may function as a signal of threat.

Song type matching does not always predict physical attack, however. Simulations of territorial intrusions using song playback

---

* Correspondence: I. Liu, Department of Biology, University of Maryland, College Park, MD 20742, U.S.A.
E-mail address: irene.a.liu@gmail.com (I. A. Liu).

https://doi.org/10.1016/j.anbehav.2018.03.007
0003-3472/

and a taxidermy mount found that song type matching does not predict whether a male will attack the mount in song sparrows (Searcy, Anderson, & Nowicki, 2006) or in swamp sparrows, *Melospiza georgiana* (Ballentine, Searcy, & Nowicki, 2008). Furthermore, while song type matching was found to predict escalated signalling and subsequent attack in a western population of song sparrows (Akçay, Tom, Campbell, & Beecher, 2013), this result was not replicated in an eastern population (Searcy, DuBois, Rivera-Cáceres, & Nowicki, 2013). These results indicate that we still do not know the extent to which song type matching functions as a signal of aggression or escalation in these species.

Another potential function of song type matching, not mutually exclusive with aggressive signalling, is to facilitate direct comparison of the two singers based on their songs (Logue & Forstmeier, 2008). When two interacting birds song-type match, the acoustic features of those songs can be compared directly. Individual differences in song quality should be particularly easy to detect during the countersinging bouts that commonly occur in territorial disputes, in which the two singers alternate in rapid succession. Such comparison might be useful to the singers themselves, helping them decide how to proceed in the interaction, and also to potential rivals and mates evaluating each singer's competitive ability and mate quality (Bateson & Healy, 2005; McGregor, 1993).

To compare singing ability, listeners need a reliable metric for comparison. One measure of vocal performance that could function as this metric concern the rate and frequency bandwidth of trilled syllables. This type of vocal performance was first studied in swamp sparrows, whose songs consist of a single repeated syllable (Podos, 1997; Fig. 1). Male swamp sparrows face constraints on beak movement that result in a trade-off between the maximum rate at which they can sing these trilled syllables (trill rate) and their frequency bandwidth (Podos & Nowicki, 2004; Podos, 1997). An upper bound regression of frequency bandwidth versus trill rate, based on all of the songs in a population, represents a theoretical performance maximum. Actual performance, which varies across song types and across males, can be assessed relative to this maximum: high-performance songs, and high-performance renditions of a given song type, deviate relatively little from the upper bound, and low-performance songs or renditions deviate further. Female swamp sparrows give more copulation solicitation displays to higher-performance renditions (those with less vocal deviation) of a given song type (Ballentine, Hyman, & Nowicki, 2004), and males generally respond more aggressively to playback of high-performance renditions than to low-performance ones (DuBois, Nowicki, & Searcy, 2011; Moseley, Lahti, & Podos, 2013). Males can modulate their vocal performance to some extent (DuBois, Nowicki, & Searcy, 2009), but for a given song type, within-male variation in performance is generally lower than between-male variation (DuBois et al., 2011). The relative vocal performance of two swamp sparrow males singing the same song type in a territorial interaction might therefore be informative to the interacting males and to conspecifics in the area.

The roles of song type matching and relative vocal deviation in territorial signalling thus far have been studied separately. In this study, we examine how the two measures interact. Specifically, we investigate whether the likelihood of song type matching in swamp sparrows is affected by relative vocal performance, using vocal deviation as the performance measure. In a playback experiment with paired trials, we recorded males' responses to stimuli of higher or lower vocal performance relative to their own renditions of the same type. We then tested whether differences in vocal performance between focal males and their perceived rivals predicted the frequency of song type matching. We predicted that males would song-type match more often when they could produce a higher-performance version of that type than the perceived opponent (i.e. in response to low-performance playbacks), because matching in this situation enables the focal male to demonstrate vocal superiority. We also predicted that males would not match when confronted with relatively high-performance song of a given type, because matching in this case would reveal the subjects' own lower performance. In this case, males might be expected to reply with a different song so that listeners cannot as easily compare song quality across the two singers. These predictions are in line with the model of Logue and Forstmeier (2008) simulating the conditions that confer network-dependent advantages to song type matching.

Our experimental design additionally enabled us to assess whether males respond to playback using their highest-performance song type when not song-type matching. DuBois et al. (2009) found that male swamp sparrows do not consistently respond to playback with their highest-performance song type, but song type matching was not considered in that study. We examined whether singing one's highest-performance song type represents an alternative strategy from song type matching and, if so, whether its use depends on the performance level of the playback song. Males might be expected to use such an alternative strategy in high-performance playback trials, in which they could not respond with a higher-performance rendition of the playback song type. Finally, male swamp sparrows can modulate their vocal performance and sing with higher performance when confronted by a perceived intruder (DuBois et al., 2009). We asked whether males modulate their responses to low- and high-performance playback to similar or different degrees.

METHODS

Ethics and Animal Welfare

Research protocols for this study were approved by the Institutional Animal Care and Use Committee of Duke University (A113-05-04) and the Pennsylvania Game Commission (30-2008). All recording and playback procedures were performed on wild, free-living birds. To minimize invasiveness, we limited the duration of recording to that required for full repertoire sampling and tested each male with only one song type (two trials per male). We did not revisit the territories after playback trials were completed.

Field Recordings and Playback Stimuli

The complete song repertoires of 31 territorial free-living male swamp sparrows were recorded from 9 May to 15 June 2008 in Conneaut Marsh, Pennsylvania, U.S.A. (41°35′26″N, 80°15′54″W). Males were recorded between 0530 and 1200 hours Eastern Standard Time (EST) using a Marantz PMD660 digital recorder, a Sony parabolic reflector PBR-330 and a Shure SM57 Dynamic microphone. Most males were already present and singing when we began recording, but when necessary, we played at most two songs outside the territory to determine whether the male was present. We did not record males if they were countersinging with another male. Most males were not colour-banded but were readily identified by their song posts, territorial boundaries and vocal repertoires. Each male's repertoire was recorded within 2 days.

Swamp sparrow songs consist of one multinote syllable repeated in a continuous trill (Fig. 1). Song types are thus identified by the unique sequence of two to five note types that comprise its repeated syllable (Marler & Pickert, 1984). The ranges of vocal performance values differ across song types as well as across males (Ballentine et al., 2004). To calculate the vocal performance of each song type for each male, we first viewed recordings as waveforms in Audacity v.1.2.6 (http://audacity.sourceforge.net) and selected six

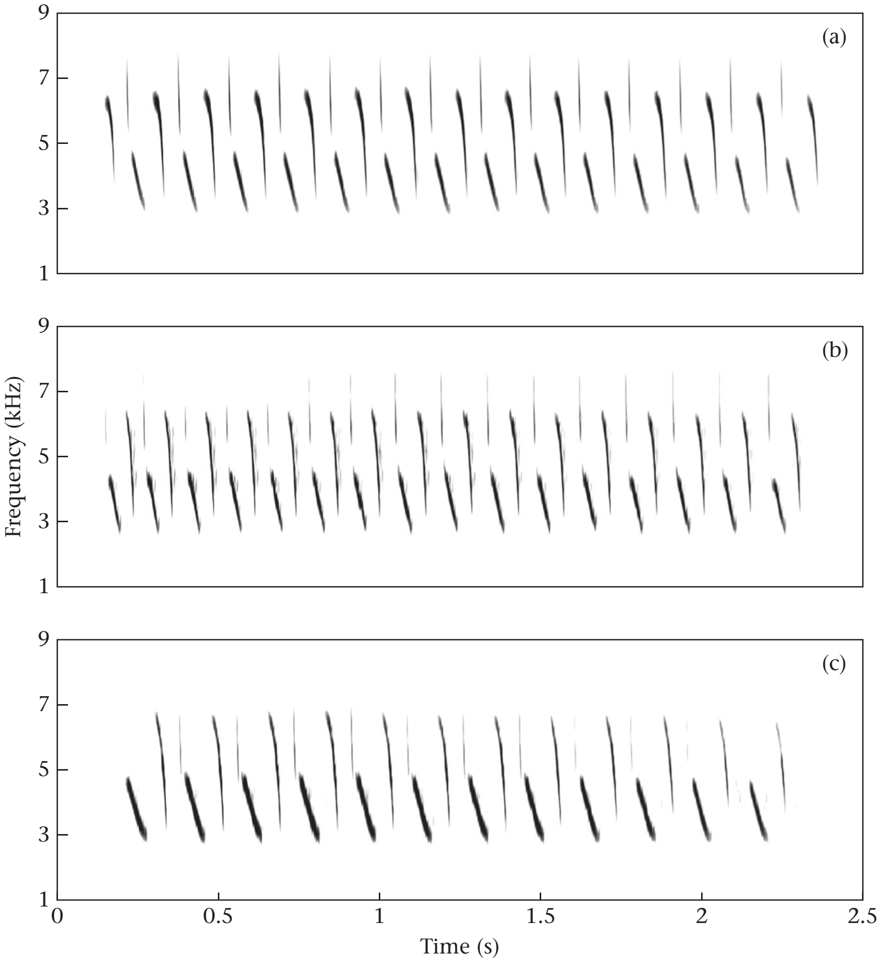


Figure 1. Three exemplars of one swamp sparrow song type. (a) Song recorded from male RWB2. (b) Higher-performance and (c) lower-performance songs of the same type, recorded from two other individuals and used as playback stimuli for male RWB2.

to 10 exemplars with a high signal-to-noise ratio. We then used Signal software v.4 (Engineering Design, Berkeley, CA, U.S.A.) to measure (1) frequency bandwidth, by identifying the highest and lowest frequencies at -36 dB relative to the peak amplitude of each song (the same criteria used by Ballentine et al., 2004), and (2) trill rate, by calculating the inverse of the duration of each syllable plus the subsequent intersyllable interval, then averaging across syllables in a song. Finally, we calculated the vocal performance score for each song following the methods of Podos (1997, 2001). We measured the orthogonal deviation of each song from an upper-bound regression of frequency bandwidth versus trill rate, previously derived by Ballentine et al. (2004) from 280 songs recorded from 91 other males in this population. Although this upper-bound regression does not take into account the sampling limitation described by Wilson, Bitton, Podos, and Mennill (2014) and therefore might not accurately represent the true performance limit, it is sufficient for our purposes of comparing relative performance levels among songs within this population. We averaged scores across exemplars to obtain one vocal performance score for each song type in a male's repertoire.

We ranked the song types in each male's repertoire by performance score and chose a song type of intermediate performance to use as the stimulus song type for playback to that male. From an archive of swamp sparrow songs recorded in 2001 from the same population for which the vocal performance scores had already been calculated, we selected exemplars of both higher and lower vocal performance than the subject male's own rendition of that song type to be used as song playback stimuli (Fig. 1), thereby tailoring pairs of playback stimuli to each male. The deviation score used to measure vocal performance is a unitless measure (Podos, 1997); for consistency, we chose stimulus songs that differed equally in deviation relative to the male's own performance score (mean deviation between stimulus song and male's own score ± SD = 7.2 ± 2.7, range 1.31–12.75, $N = 62$). Songs used in high- versus low-performance trials differed significantly in their deviation from the vocal performance boundary (high-performance: mean ± SD = 0.33 ± 4.24; low-performance: 14.60 ± 3.94; t test: $t_{60} = 13.74$, $P < 0.01$).

In five cases, equidistant high- and low-performance song pairs were not available for a male's intermediate-performance song type. Three of these males received high- and low-performance stimulus songs matching their highest-performance song types, and two received stimulus songs matching their lowest-performance song types. Five other males had only two song types in their repertoire (Table 1); three of these received stimulus songs of their higher-performance song type, and two received stimulus songs of their lower-performance song types. To avoid pseudoreplication (Kroodsma, 1989), we used different song exemplars across all trials. Stimulus songs were normalized to the same amplitude using Signal and broadcast through a speaker (Advent Powered Partners) mounted at 1 m above ground level and placed near the centre of the male's territory. Source volume level was adjusted to 85 dB SPL using a BK Precision sound pressure level meter.

Playback Trials

We conducted paired playback trials from 24 May to 23 June 2008 between 0600 and 1030 hours EST. Each male heard two natural versions of the same playback song type: one with a higher performance score than its own rendition of that type (hereafter ‘high-performance trials’) and one with a lower performance score than its own rendition (‘low-performance trials’). The two trials for a given male were presented 48 h apart, with a balanced design to control for order effects. We did not assess pair status or stage in the nesting cycle, both of which can affect responsiveness to playback, but within males, little change in either status was expected between trials. During each trial, we recorded the male's vocal responses to playback. If a male was singing before the trial, we waited until it was silent for at least 30 s before beginning playback. Males heard one song type per trial, repeated once every 10 s (a song rate typical for this species, Mowbray, 1997) for 5 min, yielding 30 repetitions. A male was considered to have song-type matched if its first song in response to playbacks was of the same type as the stimulus song.

We continued recording each male until it had sung all of the song types in its repertoire at least once, except for four males that either stopped singing or flew away before cycling through their entire repertoire after one or both trials. These four males were still included in the analysis of initial responses to playback but not in other analyses requiring the entire repertoire.

We measured vocal performance of up to 10 (mean = 8.4) exemplars of each song type recorded during and after playback, and from these we calculated an average score for each song type in a male's repertoire in each trial. Thus, we had vocal performance scores for songs sung in a neutral context and for those sung in two playback contexts.

Statistical Analysis

Song type matching

To test whether males matched the stimulus song type more often than expected by chance, we ran a chi-square test comparing observed versus expected frequencies of matching across all 62 trials. We calculated the expected frequency using the following equation:

$$\Sigma(n_j/31) \times (1/j) \tag{1}$$

which accounts for variation in repertoire size across males (Table 1). This equation represents the probability that a male with a repertoire of j song types will sing the matching song type in any given playback, with n_j representing the number of males with each repertoire size.

The variation in repertoire size also required us to compute probability values for each possible trial outcome. Using a custom script in R v.3.4.1 (R Core Team., 2017) that incorporated repertoire size data (Table 1), we ran simulations with 100 000 iterations to generate curves showing the probabilities that different numbers of males would respond in a given way (see Fig. 2; details below). For two-tailed analysis with $\alpha = 0.05$, counts of males occurring at cumulative probabilities below 0.025 at either end of the curve were considered to be significantly different from those expected by chance.

Table 1
Distribution of song repertoire sizes of the 31 swamp sparrows included in this study

Repertoire size (j)	Number of males (n_j)
2	5
3	16
4	9
5	1

We first determined the *P* values of various rates of matching within one trial type (i.e. we determined how many males out of 31 must song-type match to constitute a statistically significant proportion; Fig. 2a). Next, we generated three additional curves calculating the *P* values associated with the proportions of males responding in various ways to the paired trials. A male could song-type match in both trials, match in neither trial, or match in one trial and not the other (i.e. match the high- but not the low-performance stimulus song, or match the low- but not the high-performance stimulus song). Matching to the low- but not the high-performance song should occur if males match only when it is advantageous for them to do so, according to the model of Logue and Forstmeier (2008).

To test whether a significant number of males matched in one trial but not the other, we performed a binomial simulation using the following equation:

$$(1/j) \times ((j-1)/j) \tag{2}$$

with j again representing the number of songs in a male's repertoire. In particular, using equation (2), we ran a second simulation with 100 000 iterations to generate a probability curve representing the expected proportion of males matching in one trial only (Fig. 2b). Finally, we ran additional simulations calculating the probabilities that a male matched in both trials (Fig. 2c),

$$(1/j) \times (1/j) \tag{3}$$

or in neither trial (Fig. 2d),

$$((j-1)/j) \times ((j-1)/j) \tag{4}$$

We followed this with a chi-square test to test whether any one outcome occurred with significantly greater frequency than the others (Table 2).

Effect of playback on vocal performance

We asked whether males sing at different vocal performance levels (1) when they are singing in neutral (non-playback) versus aggressive (playback) contexts, and (2) when they are singing in response to high- versus low-vocal-performance playback trials. To compare performance levels across contexts over a male's entire repertoire, we ran linear mixed-effects models testing whether recording context (playback versus non-playback) or trial type (low- versus high-performance trials) predicted vocal performance. As fixed effects, we entered the recording context or trial type, and as random effects, we entered song type nested within male. Song types with missing values were excluded. Full models with the fixed effect of interest were compared against a null model using an analysis of deviance. Analyses were performed using the R package lme4 v.1.1.7 (Bates, Mächler, Bolker, & Walker, 2015).

To examine the extent to which males modify the performance of individual song types in response to playback, we compared performance across contexts. We chose two categories of song types for this analysis: the ‘stimulus-matching song type’ (the type used in playbacks) and the ‘first response song type’ sung to playback. In trials where males song-type matched, these two types are the same. Of the four males that did not cycle through their

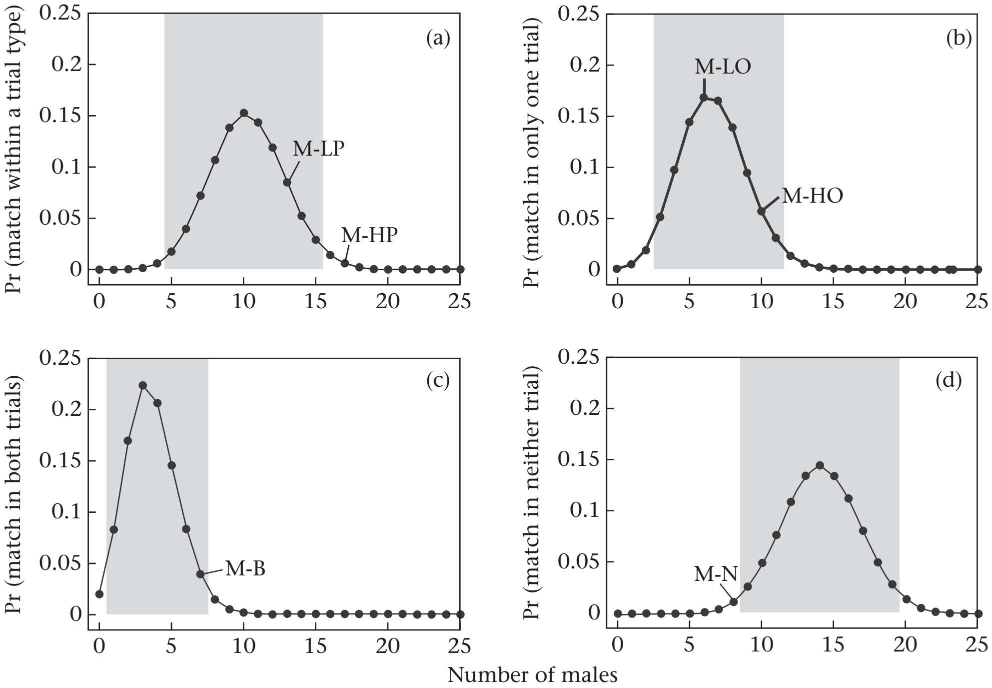


Figure 2. Probability curves based on simulations with 100 000 iterations of song type matching responses by 31 male swamp sparrows. Shaded areas indicate the range of values (numbers of males responding in a given manner) that are expected by chance, based on a two-tailed cumulative α level of 0.05. (a) Probability that a given number of males will match during playbacks of a given type (either high- or low-performance). Actual numbers of males that matched in high- and low-performance trials are indicated by the points labelled M-HP and M-LP, respectively. (b) Probability that a given number of males will song-type match in one trial and not the other. The point labelled M-HO indicates the actual number of males that matched only in the high-performance trial and not in the low-performance trial; the point labelled M-LO indicates the number of males matching only in the low-performance and not in the high-performance trial. (c) Probability that a given number of males will song-type match in both trials, with the label M-B indicating the actual result. (d) Probability that a given number of males will song-type match in neither trial, with the label M-N indicating the actual result.

complete repertoires in one or both trials, three did not have recordings of the relevant song type in both contexts and were excluded from analysis. We used Wilcoxon signed-ranks tests to compare the mean performance scores within males of the ‘stimulus-matching song type’ in the non-playback context versus in low-performance trials, and in the non-playback context versus in high-performance trials. We did the same comparisons with the performance scores of the ‘first response song type’.

For each male, we also assessed whether playback altered the ranking, by performance score, of song types in the repertoire. Specifically, we identified the song type with the highest mean performance score in response to playback and asked whether this song type (hereafter the ‘evoked’ highest-performance song) was the same as in the non-playback condition (the ‘baseline’ highest-performance song). We did this separately for low-performance and high-performance trials.

Song type choice in response to playback

Instead of song-type matching, males might respond to playbacks with their highest-performance song type. To test this prediction, we ran a chi-square test comparing the observed and expected proportions of trials in which males sang their highest-performance song type, then used the probability curve derived from equation (1) above (Fig. 2a) to assess significance of the observed proportion, as this analysis too concerns the selection of one particular song from each male's repertoire. We did this twice, once for the baseline highest-performance song type and once for the evoked highest-performance type. We additionally used chi-square tests to test whether males responded with their baseline highest-performance song type more often to one type of playback trial (e.g. high-performance playback song) than the other.

To test whether males respond with their highest-performance song type when not song-type matching, we ran additional simulations as described above using equation (1) but with repertoire size reduced by one for each bird to exclude the stimulus-matching song type. These simulations included only those birds that did not song-type match and excluded birds with only two songs in their full repertoire, as well as the three birds that were tested with their highest-performance song type (for these birds, song type matching is the same as responding with the baseline highest-performance song type). Simulations were run separately for high-performance trials ($N = 10$ males) and low-performance trials ($N = 15$ males), first for baseline and then for evoked highest-performance song types.

Table 2
Distribution of subject responses to paired playback trials

	STM in high-VP trial?	
	Yes	No
STM in low-VP trial?		
Yes	7	6
No	10	8

Each male could song-type match (STM) or not, with a stimulus song of higher or lower vocal performance (VP). The value on the top right (6) indicates the predicted response (i.e. that males would song-type match with a low-performance song but not with a high-performance song).

Performance advantage in playback trials

Lastly, we tested whether the playback-induced increase in performance score, combined with choice of first response song, affected the degree of difference in performance between the first response song of each subject and the playback song (i.e. the putative performance advantage or disadvantage of the subject males). First we calculated the difference between the mean performance of the first response song type and the performance score of the playback song, whether or not these were of the same type. We then averaged this difference across males. To assess significance, we compared this average to the distribution of results obtained from simulations (coded in Python 2.7) using 10 000 random selections of performance differences between subject songs and playback songs, across all birds. These simulations were run separately for high-performance and low-performance trials, incorporating the observed playback-induced performance scores of all songs in each bird's repertoire for the respective trial type.

RESULTS

Overall Song-matching Rate

Across the 31 males, the mean repertoire size was 3.2 ± 0.75 song types, and these songs represented 18 different types across the population. Using equation (1), we determined the expected song type matching frequency to be 0.332, a proportion approximately reflecting the average male repertoire of three song types (and thus a 1 in 3 chance of matching at random). Males were expected to match randomly in 20 or 21 trials (0.332 × 62 trials). Instead, they matched in 29 of 62 trials (0.47), a significantly higher proportion than expected (chi-square test: $\chi^2_1 = 4.55$, $P = 0.03$).

Matching to High- and Low-performance Songs

The probability curve from equation (1) indicates that within a trial type, song type matching in ≤ 5 or ≥ 16 of the 31 males would represent significant deviation from chance (Fig. 2a). In response to high-performance songs, 17 males matched ($P < 0.01$), while in response to low-performance songs, 13 males matched ($P = 0.09$). Male swamp sparrows therefore engaged in song type matching significantly more often than expected by chance in high-performance trials, but only at chance levels in low-performance trials. In a direct comparison of these outcomes, the frequency of matching did not differ significantly between the two types of trials (chi-square test: $\chi^2_1 = 1.04$, $P = 0.31$).

We next examined responses to the paired trials, summarized in Table 2. The probability curve from equation (2) indicates that ≤ 2 or ≥ 12 of the 31 males would have had to song-type match in one trial and not the other for proportions to be significantly different from chance (Fig. 2b). Six males matched to the low- but not the high-performance song ($P = 0.17$). Ten males behaved opposite to our prediction by matching to high- but not low-performance song ($P = 0.06$). The probability curves from equations (3) and (4) indicate that matching in both trials by either zero or ≥ 8 birds (Fig. 2c), and matching in neither trial by ≤ 8 or ≥ 20 birds (Fig. 2d), would represent significant deviations from chance. Seven males matched to both stimulus songs and eight males matched to neither ($P = 0.04$ and $P = 0.01$, respectively). Only the last of these four outcomes differed significantly from chance, and comparing them directly, we found that none was significantly more prevalent than the others (chi-square test: $\chi^2_3 = 1.9$, $P = 0.59$).

Vocal Performance in Different Contexts

The linear mixed-effect models showed that, across song types, males sang with significantly higher vocal performance (closer to the performance optimum) in response to playback than in a non-playback context. In response to playback, the average (±SE) vocal deviation was closer to the performance optimum by 1.32 ± 0.19 (analysis of deviance, chi-square test: $\chi^2_1 = 25.1$, $P < 0.01$; Fig. 3a). Thus, we rejected the null hypothesis that vocal performance did not differ between playback and non-playback recording contexts.

Relative to the non-playback context, males sang their ‘stimulus-matching song type’ with significantly higher vocal performance in low-performance playback trials (Wilcoxon signed-ranks tests: $V = 358$, $N = 31$, $P = 0.03$; Fig. 3b). Males also increased their vocal performance in high-performance playback trials, although not significantly ($V = 280$, $N = 28$, $P = 0.08$; Fig. 3b). Likewise, relative to the non-playback context, males sang their ‘first response song type’ with significantly higher performance in playback trials of low-performance song ($V = 351$, $N = 31$, $P = 0.04$; Fig. 3c) but not in playback trials of high-performance song ($V = 316$, $N = 31$, $P = 0.18$; Fig. 3d).

These results might suggest that males increased their vocal performance more in low-performance playback trials than in high-performance trials. However, direct comparison revealed that within playbacks, average vocal performance did not differ between high- and low-performance trials across all song types (analysis of deviance, chi-square test: $\chi^2_1 = 0.14$, $P = 0.71$) or for the ‘stimulus-matching song type’, whenever it was sung (Wilcoxon matched-pairs test: $V = 144.5$, $N = 28$, $P = 0.19$; Fig. 3b). Vocal performance of the ‘first response song type’ sung in response to playback could be compared across trial types only for the 15 birds that sang the same song type first in both trials. For these 15 birds, vocal performance of this song type also did not differ between high- and low-performance trials ($V = 316$, $N = 15$, $P = 0.19$).

Within males, the increase in performance in response to playback varied across song types such that for 11 of the 31 birds, the ranking of song types by performance differed between the two contexts (in non-playback broadcast song versus in response to at least one of the playback trials). For eight birds, in at least one of the playback trials the evoked highest-performance song type differed from the baseline highest-performance type.

Song Type Choice in Response to Playback

Regardless of matching, males responded with their baseline highest-performance song type in 23 of 62 trials (23/62 = 0.37). This proportion was not significantly different from chance (chi-square test: $\chi^2_1 = 0.27$, $P = 0.60$). During high-performance trials, 10 of 31 males responded with their baseline highest-performance song ($P = 0.15$; Fig. 2a), and in low-performance trials, 13 males did so ($P = 0.09$). Males did not respond with their baseline highest-performance song more often in one trial type than in the other ($\chi^2_1 = 0.28$, $P = 0.60$). Among males that did not song-type match and had two or more additional song types in their repertoires, in high-performance trials 4 of 10 responded with their baseline highest-performance song ($P = 0.25$), and in low-performance trials 9 of 15 did so ($P = 0.12$). Neither of these proportions differed significantly from chance.

Results were similar after accounting for the playback-induced change in performance ranking of song types. In both high- and low-performance trials, 9 of 31 males responded with their evoked highest-performance song type ($P = 0.14$ in both cases; Fig. 2a). Among males that did not song-type match and had at least three song types in their repertoires, in high-performance trials 3 of 10 responded with their evoked highest-performance song ($P = 0.18$), and in low-performance trials 5 of 15 did so ($P = 0.12$).

Performance Advantage in Playback Trials

In high-performance trials, subjects sang their first response song with an average vocal performance 4.8 units lower than the playback song, whereas in low-performance trials, subjects sang their first response song 10.7 units higher. This observed advantage of subjects in low-performance trials (10.7 units) was significantly higher than that expected by chance ($P = 0.01$, based on 10 000-

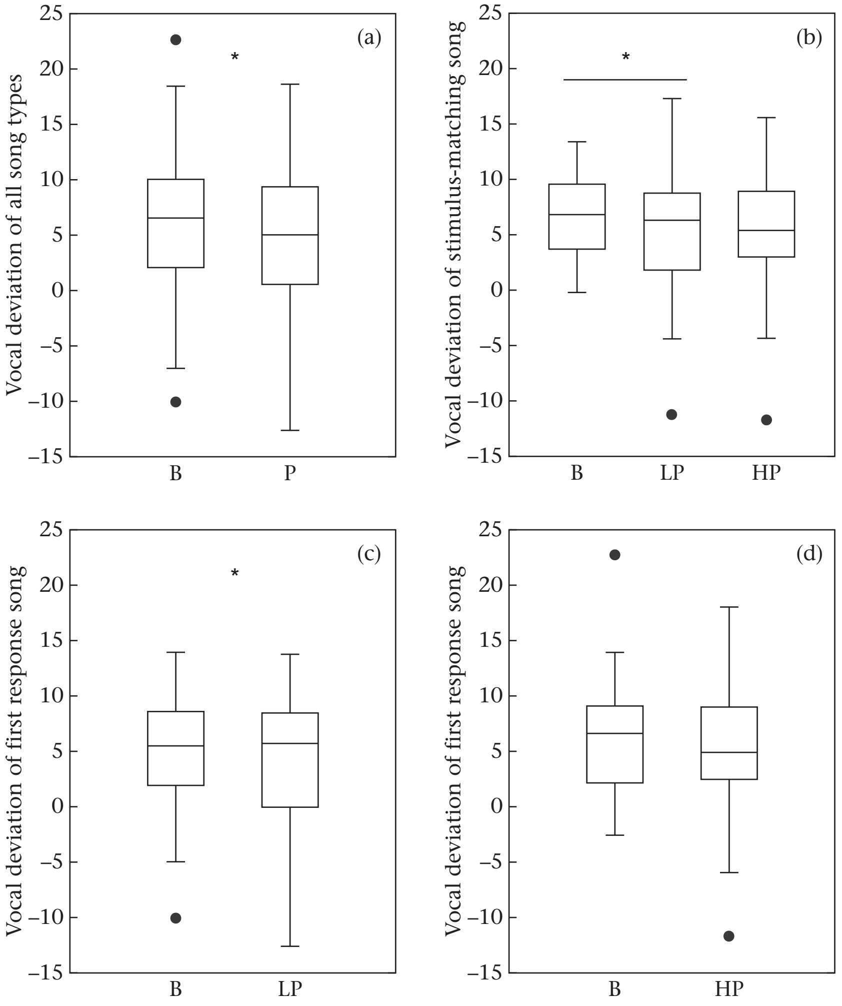


Figure 3. Summary graphs of vocal performance, measured as deviation from the performance boundary, of subject song types in different conditions. Horizontal line within each box indicates median across subjects, upper and lower edges of boxes represent first and third quartiles, whiskers include values falling within 1.5 times the interquartile range, and dots indicate outliers. Asterisks indicate significant differences ($P < 0.05$) based on a linear mixed-effects model (panel a) or repeated-measures (Wilcoxon) tests (panels b-d). (a) Vocal deviation of all song types in the neutral condition (B, broadcast song; $N = 99$) and in response to playback (P; $N = 185$). (b) Vocal deviation of the ‘stimulus-matching song type’ (shared across trials) in the neutral condition and in response to low-performance (LP) playbacks ($N = 31$) and high-performance (HP) playbacks ($N = 28$). (c) Vocal deviation of the LP-trial ‘first response song type’ in neutral conditions and in response to low-performance playback ($N = 31$), and (d) vocal deviation of the HP-trial ‘first response song type’ in neutral conditions and in response to high-performance playback ($N = 31$).

iteration simulations), whereas the observed disadvantage in high-performance trials (4.8 units) was no different from that expected by chance ($P = 0.42$).

DISCUSSION

The male swamp sparrows in this study responded to playback by song type matching more often than expected by chance. Our primary question was whether the likelihood of song type matching depends on the relative performance differences between the playback song and the subject's own performance of the same song type, and we found that it did not. Higher-performance playback elicited matching more often than expected by chance while lower-performance playback did not, but in a direct comparison, the rate of matching in these two trial types did not differ significantly. That males song-type match at least as often to higher-performance songs as to lower-performance ones refutes the prediction that male swamp sparrows selectively song-type match when it is advantageous for them to do so from a network perspective (i.e. when it can help males demonstrate superior singing ability to eavesdroppers; Logue & Forstmeier, 2008).

Our results further refute the broader hypothesis that male swamp sparrows benefit from conveying their relative vocal performance ability to eavesdroppers. Among the males we tested, those that did not song-type match did not sing their highest-performance songs in response to playback as an alternative strategy. Similar to DuBois et al. (2009), we found that male swamp sparrows do not preferentially respond to playback with their baseline highest-performance song. Here we observed this to be true even among the subset of males that did not song-type match. We also found that although playback can alter the performance score ranking of the song types in a male's repertoire, males did not respond with the resulting ‘evoked’ highest-performance song type as an alternative strategy to song type matching.

The playback-induced increase in vocal performance that we observed replicates previous results in this species (DuBois et al., 2009). In the current study, this increase was significant in response to low- but not high-performance stimuli for the two categories of song type we assessed (the ‘stimulus-matching song type' and the 'first response song type'). We emphasize that in a direct comparison between the high- and low-performance trials, the increase in vocal performance did not differ significantly. Trial type, however, did affect the significance of the difference in performance score between the first response song type and the playback song. In low-performance trials, subjects outperformed the playback song significantly more than expected by chance. In high-performance trials, the first response song types were lower in performance score than the playback songs, but only to an extent expected by chance. The performance difference between the first response song and the playback song depended on both the identity of the response song (with its natural range of performance scores) and the extent to which the tested male modulated his performance of that song type. Our results therefore suggest that males might use modulation of vocal performance, in combination with song type choice—which can include song type matching—to emphasize their apparent performance advantage when responding to rivals singing lower-performance songs but not to minimize their apparent disadvantage when responding to rivals singing higher-performance songs.

Whether or not eavesdropping swamp sparrows actually use information gained from vocal exchanges between territorial males remains unknown. Previous studies have found that swamp sparrow males (DuBois et al., 2011; Moseley et al., 2013) and females (Ballentine et al., 2004) respond differently to higher- versus lower-performance versions of a given song type, at least when the difference in performance is large enough to reflect between-individual rather than within-individual variation (DuBois et al., 2011). These results suggest swamp sparrows of both sexes can detect individual differences in vocal performance and should therefore be able to identify the higher-performance singer when two males use the same song type during an interaction. However, no studies have yet explicitly assessed whether eavesdroppers in this species use information gained from the vocal exchanges between two territorial males, as has been described in species such as great tits (Peake, Terry, McGregor, & Dabelsteen, 2001), black-capped chickadees, *Poecile atricapillus* (Mennill, Ratcliffe, & Boag, 2002) and nightingales, *Luscinia megarhynchos* (Naguib & Todt, 1997). In this study, we examined only whether males respond to territorial intrusion in a way that would be adaptive if their relative vocal performance influences the behaviour of eavesdroppers.

High-performance playbacks elicited more song type matching than expected by chance and low-performance trials did not, whereas low-performance playbacks elicited greater increase in the subjects' vocal performance than expected by chance and high-performance trials did not. Again, direct comparisons across trial types indicated that these differences were subtle (not statistically significant). In our experiment, we presented songs near the centre of each male's territory. Songs played from this location can elicit high responses in general, and discrimination between such stimuli might be obscured (Stoddard, Beecher, Horning, & Campbell, 1991). It is possible that we would have seen more pronounced differences in the responses to the two trial types had we conducted playbacks closer to the territory edge. Our playbacks were also noninteractive, and interactive playback might reveal greater variation in responses to low- and high-performance stimuli. Regardless, previous findings indicate that male swamp sparrows respond with more flights, songs and soft songs to high-performance stimuli than to low-performance stimuli (DuBois et al., 2011). Given this more aggressive response to higher-performance songs, the high rate of song type matching that we observed in response to high-performance songs is consistent with the idea that song type matching in swamp sparrows, while not a direct predictor of attack, may still be a component of the aggressive response directed at an opponent.

A possible explanation for the distinct patterns of response to higher- and lower-performance stimuli in our study is that subject males perceived the higher-performance stimuli as coming from older males and lower-performance stimuli as coming from first-year males. Indeed, Ballentine (2009) found that vocal performance increases from the first year to the second in male swamp sparrows. Because we did not record the ages of our test subjects, a future hypothesis to test is whether older males might represent a greater threat when intruding on a territory and therefore elicit a stronger response. Another possibility is that regardless of age, increased vocal performance and song type matching represent progressively higher stages of escalation in territorial interactions. Our results are consistent with the idea that if an intruder sings low-performance song, a territorial male can escalate by responding with relatively higher-performance song, and if the intruder begins with (or escalates to) higher-performance song, the territory holder can further escalate by song type matching. Such use of song types for hierarchical escalation and de-escalation could help explain the presence of lower-performance song types in the repertoire (Podos, 2017). Although Ballentine et al. (2008) found that song type matching does not predict attack in swamp sparrows, it remains possible that song type matching signals a lower level of aggressive escalation in this species (Searcy & Beecher, 2009).

Whether or not increased vocal performance and song type matching represent stages of aggressive escalation in male swamp sparrows, some variation across individuals in signalling strategy is to be expected. Individual variation in behavioural tendencies (behavioural syndromes) is well documented, including in songbirds (Dingemanse et al., 2012; Minderman, Reid, Evans, & Whittingham, 2009). In song sparrows, males consistently differ in how aggressively they defend their territories (Akçay et al., 2009; Hyman, Hughes, Searcy, & Nowicki, 2004; Nowicki, Searcy, Krueger, & Hughes, 2002). Likewise, in our study, a significant number of swamp sparrow males song-type matched in neither of the two playback trials, suggesting that some males of this species might tend to avoid matching. Moseley et al. (2013) similarly found that some males retreated from high-performance playback song while others responded aggressively.

Such variation in behaviour might reflect not only distinct 'personalities' but also differences in the singer's age or experience. In banded wrens, *Thryothorus pleurostictus*, older males song-type match more than do younger males (Vehrencamp, Hall, Bohman, Depeine, & Dalziell, 2007), whereas the reverse is true in nightingales (Kiefer, Scharff, & Kipper, 2011). At present, we do not know how age and/or experience affect the use of increased vocal performance and song type matching by the swamp sparrows in our study. Nevertheless, our study represents a step forward in understanding how these two components of territorial signalling may interact.

Acknowledgments

Barbara Ballentine provided songs and vocal performance data from other swamp sparrows in our study population. Robert Lachlan and Pete Wyckoff wrote simulations and helped with statistical analyses. We thank Susan Peters, Adrienne DuBois, Kim Rosvall, Rindy Anderson and Kendra Sewall for helpful discussion throughout the study. We also thank two anonymous reviewers for their helpful comments on the manuscript. This work was supported by the Arthur and Barbara Pape Endowment of the Pymatuning Laboratory of Ecology, the Wilson Ornithological Society, a Sigma Xi Grant-in-Aid of Research and a National Science Foundation Graduate Research Fellowship.

References

Akçay, Ç., Tom, M. E., Campbell, S. E., & Beecher, M. D. (2013). Song type matching is an honest early threat signal in a hierarchical animal communication system. *Proceedings of the Royal Society B: Biological Sciences, 280*(1756), 20122517. https://doi.org/10.1098/rspb.2012.2517.

Akçay, Ç., Wood, W. E., Searcy, W. A., Templeton, C. N., Campbell, S. E., & Beecher, M. D. (2009). Good neighbour, bad neighbour: Song sparrows retaliate against aggressive rivals. *Animal Behaviour, 78*, 97–102.

Armstrong, E. A. (1973). *A study of bird song*. New York, NY: Dover.

Ballentine, B. (2009). The ability to perform physically challenging songs predicts age and size in male swamp sparrows, Melospiza georgiana. *Animal Behaviour, 77*, 973–978.

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

Ballentine, B., Searcy, W. A., & Nowicki, S. (2008). Reliable aggressive signalling in swamp sparrows. *Animal Behaviour, 75*, 693–703.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 1*(1), 1–48.

Bateson, M., & Healy, S. D. (2005). Comparative evaluation and its implications for mate choice. *Trends in Ecology & Evolution, 20*(12), 659–664.

Beecher, M. D., Stoddard, P. K., Campbell, E. S., & Horning, C. L. (1996). Repertoire matching between neighbouring song sparrows. *Animal Behaviour, 51*, 917–923.

Brémond, J. C. (1968). Recherches sur la sémantique et les éléments vecteurs d'information dans les signaux acoustiques du Rouge-gorge (*Erithacus rubecula* L.). *La Terre et la Vie, 2*, 109–220.

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

Dingemanse, N. J., Bouwman, K. M., van de Pol, M., van Overveld, T., Patrick, S. C., Matthysen, E., et al. (2012). Variation in personality and behavioural plasticity across four populations of the great tit *Parus major*. *Journal of Animal Ecology, 81*(1), 116–126.

DuBois, A. L., Nowicki, S., & Searcy, W. A. (2009). Swamp sparrows modulate vocal performance in an aggressive context. *Biology Letters, 5*(2), 163–165.

DuBois, A. L., Nowicki, S., & Searcy, W. A. (2011). Discrimination of vocal performance by male swamp sparrows. *Behavioral Ecology and Sociobiology, 65*(4), 717–726.

Falls, J. B. (1985). Song matching in western meadowlarks. *Canadian Journal of Zoology, 63*(11), 2520–2524.

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

Hyman, J., Hughes, M., Searcy, W. A., & Nowicki, S. (2004). Individual variation in the strength of territory defense in male song sparrows: Correlates of age, territory tenure, and neighbor aggressiveness. *Behaviour, 141*(1), 15–27.

Kiefer, S., Scharff, C., & Kipper, S. (2011). Does age matter in song bird vocal interactions? Results from interactive playback experiments. *Frontiers in Zoology, 8*(1), 29.

Krebs, J. R., Ashcroft, R., & Orsdol, K. V. (1981). Song matching in the great tit *Parus major* L. *Animal Behaviour, 29*, 918–923.

Kroodsma, D. E. (1989). Suggested experimental designs for song playbacks. *Animal Behaviour, 37*, 600–609.

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

Marler, P., & Pickert, R. (1984). Species-universal microstructure in the learned song of the swamp sparrow (*Melospiza georgiana*). *Animal Behaviour, 32*, 673–689.

McGregor, P. K. (1993). Signalling in territorial systems: A context for individual identification, ranging and eavesdropping. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 340*(1292), 237–244.

McGregor, P. K., & Dabelsteen, T. (1996). Communication networks. In D. E. Kroodsma, & E. H. Miller (Eds.), *Ecology and evolution of acoustic communication in birds* (pp. 409–425). Ithaca, NY: Cornell University Press.

Mennill, D. J., Ratcliffe, L. M., & Boag, P. T. (2002). Female eavesdropping on male song contests in songbirds. *Science, 296*(5569), 873.

Minderman, J., Reid, J. M., Evans, P. G. H., & Whittingham, M. J. (2009). Personality traits in wild starlings: Exploration behavior and environmental sensitivity. *Behavioral Ecology, 20*(4), 830–837.

Moseley, D. L., Lahti, D. C., & Podos, J. (2013). Responses to song playback vary with the vocal performance of both signal senders and receivers. *Proceedings of the Royal Society B: Biological Sciences, 280*(1768), 20131401. https://doi.org/10.1098/rspb.2013.1401.

Mowbray, T. B. (1997). Swamp sparrow (*Melospiza georgiana*). In P. G. Rodewald (Ed.), *The Birds of North America Online*. Ithaca, NY: Cornell Lab of Ornithology Version 2.0. https://doi.org/10.2173/bna.279.

Naguib, M., & Todt, D. (1997). Effects of dyadic vocal interactions on other conspecific receivers in nightingales. *Animal Behaviour, 54*, 1535–1544.

Nowicki, S., Searcy, W. A., Krueger, T., & Hughes, M. (2002). Individual variation in response to simulated territorial challenge among territory-holding song sparrows. *Journal of Avian Biology, 33*(3), 253–259.

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

Podos, J. (1997). A performance constraint on the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution, 51*(2), 537–551.

Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature, 409*(6817), 185–188.

Podos, J. (2017). Birdsong performance studies: Reports of their death have been greatly exaggerated. *Animal Behaviour, 125*, e17–e24.

Podos, J., & Nowicki, S. (2004). Performance limits on birdsong. In P. Marler, & H. Slabbekoorn (Eds.), *Nature's music: The science of birdsong* (pp. 318–342). New York, NY: Elsevier Academic Press.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Searcy, W. A., Anderson, R. C., & Nowicki, S. (2006). Bird song as a signal of aggressive intent. *Behavioral Ecology and Sociobiology, 60*(2), 234–241.

Searcy, W. A., & Beecher, M. D. (2009). Song as an aggressive signal in songbirds. *Animal Behaviour, 78*, 1281–1292.

Searcy, W. A., DuBois, A. L., Rivera-Cáceres, K., & Nowicki, S. (2013). A test of a hierarchical signalling model in song sparrows. *Animal Behaviour, 86*, 309–315.

Stoddard, P. K., Beecher, M. D., Horning, C. L., & Campbell, S. E. (1991). Recognition of individual neighbors by song in the song sparrow, a species with song repertoires. *Behavioral Ecology and Sociobiology, 29*(3), 211–215.

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

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

Wilson, D. R., Bitton, P.-P., Podos, J., & Mennill, D. J. (2014). Uneven sampling and the analysis of vocal performance constraints. *American Naturalist, 183*(2), 214–228.