ORIGINAL PAPER

A proxy of social mate choice in prairie warblers is correlated with consistent, rapid, low-pitched singing

Bruce E. Byers[1] · Michael E. Akresh[2] · David I. King[3]

Abstract In songbirds, female mate choice may be influenced by how well a male performs his songs. Performing songs well may be especially difficult if it requires maximizing multiple aspects of performance simultaneously. We therefore hypothesized that, in a population of prairie warblers, the males most attractive to females would be those with superior performance in more than one facet of singing. We tested this prediction by comparing different aspects of song performance, as well as different combinations of these aspects, to determine which were the best predictors of first-egg date, which we took to be a proxy for social mate choice. We found that first-egg date was best predicted by a combination of song performance traits that included consistent performance, rapid rate, and low pitch. Female preference for males capable of physically challenging song performance may have contributed to the evolution of acoustically complex vocalizations in oscine songbirds, because if complex sounds are more difficult to perform, they may be favored by selection for signal reliability.

Communicated by M. Leonard

✉ Bruce E. Byers
bbyers@bio.umass.edu

[1]   Department of Biology, University of Massachusetts Amherst, 221 Morrill Science Center, Amherst, MA 01003, USA

[2]   Department of Environmental Conservation, University of Massachusetts Amherst, 204 Holdsworth Hall, Amherst, MA 01003, USA

[3]   U.S. Forest Service Northern Research Station, University of Massachusetts Amherst, 201 Holdsworth Hall, Amherst, MA 01003, USA

Introduction

Why are the songs of many oscine songbird species so much more acoustically complex than the vocal signals found in other animal taxa (Kroodsma 2005)? A possible explanation is that, if female listeners evaluate potential mates on the basis of how well their songs are performed, sexual selection might favor songs that are increasingly difficult to perform, and one way in which a vocal signal can become more difficult to perform is through increased acoustic complexity (Suthers and Goller 1997). In this view, signals that are more difficult to perform may arise as a result of fitness benefits that accrue to females that evolve a mating preference for good performance of physically challenging signals that reliably reveal the signaler's quality (Zahavi and Zahavi 1997; Byers et al. 2010). Alternatively, increased performance difficulty may arise as a side effect of an arbitrary female preference for signal features that happen to make the signal more challenging to perform (Prum 2010, 2012).

Whether or not song performance is an indicator of male quality, empirical evidence suggests that variation in how songs are performed is salient in songbird communication (see references below). Many different aspects of song performance have been identified as potentially having signal value, reflecting the diversity of song form and singing behavior among songbird species. Despite this variety, however, it is possible to identify general categories of song performance indicators. In particular, the song performance traits that have

been implicated in communication can be viewed as falling into five categories: output and rate, consistency, frequency (pitch), amplitude, and song structure.

Traits in each of these song performance categories have been linked to aspects of mate choice in some songbird species. That is, proxies of mate choice have been found to be associated with 1) the related traits of song output (the total amount of singing by a bird) and singing rate (a bird's song output per unit time while singing) (e.g., Alatalo et al. 1990; Houtman 1992; Chiver et al. 2008); 2) consistent performance, i.e., maintaining consistency between renditions of a song (reviewed in Sakata and Vehrencamp 2012); 3) especially high- or low-frequency songs (reviewed in Cardoso 2012); and 4) song amplitude (e. g., Ritschard et al. 2010) or the degree to which amplitude is sustained across a difficult-to-perform song component (Forstmeier et al. 2002; Poesel et al. 2012).

The fifth performance category includes aspects of performance that are inherent in the structure of individual songs, especially in the form and arrangement of the elements that make up a song. For example, in species whose songs contain rapid trills, individuals may differ in "vocal deviation" (Podos 1997), a measure of the degree to which a trill simultaneously maximizes frequency bandwidth and the rate at which the trill's component elements are repeated. Some studies have found this measure to be correlated with proxies of mate choice (reviewed in Wilson et al. 2014). The components of vocal deviation (bandwidth and element repetition rate) may also be independently salient in signaling (e.g., Christensen et al. 2006; Linhart et al. 2013). Other structural features that have been found to be associated with mate choice include song duration (e.g., Kempenaers et al. 1997; Nolan and Hill 2004) and number of switches between different element types within a song (e.g., Forstmeier and Leisler 2004).

Working hypotheses

Given the evidence that various categories of song performance may have a signaling function in at least some songbird species, we hypothesized that female social mate choice might be influenced by a combination of performance traits from different categories, much as extra-pair mate choice appeared to be in an earlier study of the chestnut-sided warbler (*Setophaga pensylvanica*; Byers 2007). We broke this general hypothesis into a set of alternative working hypotheses (Table 1). Our broadest hypothesis was that females would gain the maximum possible information about male attractiveness by attending to all categories of performance simultaneously (although we were not able to include song amplitude in our hypotheses, due to the difficulty of accurately measuring this feature from field recordings). Alternatively, females might not have the capacity to simultaneously assess all aspects of performance and would be influenced by a subset

Table 1 Nested alternative hypotheses about the aspects of prairie warbler song performance that best predict first-egg date

| First-egg date best predicted by the following: |
| --- |
| 1. Rate, consistency, frequency, structure |
| 2. Rate, frequency, structure |
| 3. Rate, consistency, structure |
| 4. Consistency, frequency, structure |
| 5. Rate, consistency, frequency |
| 6. Rate, structure |
| 7. Frequency, structure |
| 8. Consistency, structure |
| 9. Rate, consistency |
| 10. Rate, frequency |
| 11. Consistency, frequency |
| 12. Rate only |
| 13. Consistency only |
| 14. Structure only |
| 15. Frequency only |

only. We therefore included all possible performance category subsets in our group of alternative hypotheses, including the possibility that females use only a single type of song performance to evaluate males.

We tested our alternative working hypotheses in a population of prairie warblers (*Setophaga discolor*), a species we chose in part because of its phylogenetic and ecological similarity to the previously studied chestnut-sided warbler. Our alternative hypotheses were expressed as a set of 15 candidate models in which male song features were modeled as predictors of the date on which a male's social mate laid her first egg.

As do the songs of many other wood-warbler species, prairie warbler songs fall into two categories (Fig. 1), dubbed "first category" and "second category" (Spector 1992). Songs from different categories tend to be used in different contexts in both prairie warblers (Nolan 1978; Houlihan 2000) and other warbler species (Highsmith 1989; Kroodsma et al. 1989; Spector 1991; Byers 1996; Staicer 1996; Staicer et al. 2006; Bolsinger 2000). First category songs predominate in singing by unmated males, singing near females, and undisturbed singing after dawn. Second category songs predominate in singing near territorial boundaries or other males, during aggressive interactions between males, and at dawn. These differences suggest a function in mate attraction for first category songs and a function in male-male aggression for second category songs. Because of the differences between the two categories of prairie warbler songs, we treat the categories separately when assessing evidence related to our hypotheses. In addition, we also treat separately songs uttered during the dawn bout, a period of highly distinctive singing behavior that begins before dawn and consists of 30–45 min of rapidly

Fig. 1 Spectrograms (512-point FFT) of prairie warbler songs, showing two examples of first category songs (*top row*) and two examples of second category songs (*bottom row*)
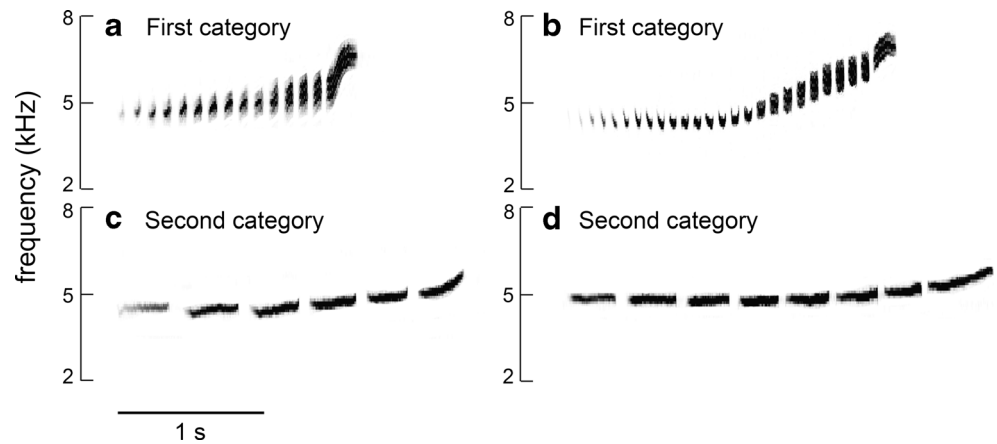
repeated second category songs interspersed with characteristic chips.

Mate choice proxy

We interpreted first-egg date as a proxy for social mate choice, as has been done in a number of field studies of songbird song function (e.g., Krebs et al. 1978; Poesel et al. 2001; Reid et al. 2004). Such proxies are necessary because it is not realistically possible to directly assess and measure female mating decisions in wild, free-living birds. The validity of our proxy depends on the assumption, first articulated by Darwin (1871, p. 254), that more-preferred males are selected as mates earlier than are less-preferred males. Validity further depends on the assumption that first-egg date reflects mating order in prairie warblers. Evidence that it does comes from Nolan's comprehensive investigation of the breeding biology of an Indiana prairie warbler population (Nolan 1978). Nolan found that 1) earlier-forming pairs tended to begin nest-building earlier, and 2) the time elapsed from the onset of nest-building to first-egg date was fairly constant among pairs (coefficient of variation 0.3). Thus, the order of first-egg dates generally matched the order of nest initiation and, therefore, the order of pair formation.

Even if first-egg date does reflect mating order as we expect, the link between mating order and mate choice might be compromised if breeding females tend to return to the same area in subsequent years and mate again with their prior mate and/or settle on the territory on which they had nested the prior year. The settlement order of females that behave in this fashion might not reflect free choice from among all available males. However, our banding records suggest that such female fidelity is not common in our study population. Of 38 females banded in the breeding seasons immediately prior to those in our study period, 10 (26 %) returned to the study area, and only 5 of these returning females (13 % of banded females) settled on the same territory as in the prior year. Similarly, Nolan (1978) found a 19 % return rate among female prairie warblers, and about a third of returning females (7 % of all breeding females) nested on the same territory in consecutive years. Thus, it seems likely that only a small proportion of mate choice decisions in our study population could potentially be driven by mate or site fidelity.

Methods

Study site

We conducted our study in 2009 and 2010 at the Montague Plains Wildlife Management Area in Franklin County, Massachusetts (N 42° 34′, W 72° 31′). The management area is a 600-ha mosaic of hardwood forest, power line corridors, stands of pitch pine (*Pinus rigida*), and stands of scrub oak (*Quercus ilicifolia*). Prairie warblers nested primarily in the shrubby habitat beneath power lines and in the pitch pine and scrub oak stands. We captured most of the male prairie warblers in these habitats (and many females) and banded each one with a USGS aluminum band and a unique combination of color bands (see Akresh et al. 2015 for additional information about the study site and field methods).

Mate choice

We monitored prairie warblers in the study area each year, beginning with the first arrival of birds on breeding territories in late April. Frequent surveys of singing males during the first few weeks of the breeding season allowed us to establish the arrival date of almost all of the territorial males in the study area. Once a male established a territory, we visited it every 2–4 days for detailed observation. This ongoing observation, including plotting of GPS coordinates of bird locations, allowed us to map territory boundaries for most males. We also conducted systematic nest searches during these visits and monitored the status of active nests (with each nest monitored until the nestlings fledged or the contents of the nest

disappeared). We were thus able to identify the date on which each male's mate laid her first egg. We found most nests while they were under construction, but some were discovered only after eggs had been laid; for these nests, we estimated first-egg date based on hatch date. In some cases, we could not be certain that the first nest we found on the territory was the first nest built there rather than a re-nest after a nest failure. We excluded these cases from further analysis.

Song sample

We recorded singing prairie warblers on most days from 4 May to 27 June 2009 and every other day from 6 May to 28 June 2010. All songs were recorded with a Nagra LB solid state recorder (48 kHZ sampling rate, 16-bit sample depth) and a Sennheiser MKH62 microphone mounted in a Telinga Pro parabolic reflector. We began recording before dawn each day, arriving on a territory before the resident male had begun singing. After the initial focal male began singing, we recorded a 5-min sample, moved to a neighboring bird to record another sample, and so on, cycling among ten or so birds in the vicinity, recording samples of 3–10 min duration from each bird, with multiple samples from each individual over the course of the morning. We noted the GPS coordinates of each song sample. We stopped recording at around 0900 hours, when singing became sporadic. By systematically rotating the starting location of each day's recording session, we were able to obtain multiple samples, spread over different times of day and over a range of dates, from each of 61 birds in 2009 and 21 birds in 2010 (when we focused on a smaller area). Our sampling regime was designed to record each bird's singing in a variety of contexts, in hope of capturing the full range of variation in each bird's songs. The identity of each sampled singer was determined by observation of color bands or, for instances in which bands could not be been, by the bird's location and the form of its song (most birds used distinctive songs that differed from those of their neighbors).

Song measurements

From our collection of recorded songs, we selected a subset to be measured. This subset was drawn from samples that contained a continuous bout of singing with high-quality recordings of at least five consecutive songs. From this group of song bouts, we excluded those from individuals from whom we had recorded bouts on fewer than four different days and those from individuals for whom we did not have complete nesting data. For five birds that were recorded in both years of the study, we also excluded recordings from 1 year, chosen at random. The resulting pool included 647 bouts, containing a

total of 7958 songs (mean songs/bout 12.3, standard deviation 5.3, range 5–43) from 51 birds (mean bouts/bird 12.7, standard deviation 3.1, range 5–21).

Our sampling regime ensured that samples for almost all birds included multiple bouts from each stage of a bird's nesting cycle. Four stages were delineated: courtship and nesting, laying and incubation, nestlings, and fledglings. We could not reliably determine the onset of courtship because the cryptic behavior of females prevented reliable assessment of female arrival dates, so all samples prior to first-egg date were classed together. Fourteen birds were not recorded during the nestling phase, because their nests failed before hatching (12) or their eggs did not hatch until after the end of the recording period (2). Despite the absence of nestling-stage recordings for some birds, a chi-square test of independence suggests that the proportion of samples in each nesting stage was similar across birds (χ^2=97.6, df=100, p=0.55). The fledgling stage was not included in the chi-square analysis due to infrequent occurrence. Only eight birds were recorded during the fledgling phase; the remaining birds fledged young after the end of the recording period, or failed to fledge young.

For each song in the pool, we used onscreen cursors to make measurements on spectrograms (for time measurements) and power spectra (for frequency measurements) generated in Raven 1.3 (Bioacoustics Research 2011; 128 point FFT for time measurements, 1024 point FFT for frequency measurements). The features measured were chosen to encompass four categories of song performance that have been identified in the literature as having potential signal value: singing rate, performance consistency, frequency, and song structure. To assess these features, we measured song duration, intersong interval, frequency of peak power, lowest frequency (the value 30 dB below peak), highest frequency (30 dB above peak), and frequency bandwidth. In addition, we counted the number of elements in each song; element count was divided by song duration to produce an element rate for each song. Each measured variable was averaged across songs in a bout to produce a single value for each bout. In addition, to assess consistency of song performance within a song bout, the coefficient of variation for each measured variable was calculated for each bout.

For subsequent analysis, we divided the per-bout values into three categories, to ensure that between-individual comparisons were based on comparable songs. The three categories were songs produced during the dawn bout (these were exclusively second category songs); first category songs produced after the dawn bout; and second category songs produced after the dawn bout. In each category, we calculated the median value of each bird's per-bout means for each variable; these median values (i.e., a single value of each variable for each individual) formed the basis of all subsequent analyses of song features.

Model specification

To assess whether aspects of male singing predict a proxy of female mate choice preference, we fit linear mixed effects models with random intercepts and first-egg date (measured as the number of days after May 1) as the dependent variable. The models were fitted with the R package lme4 (Bates et al. 2013; R Core Team 2014), using a maximum likelihood estimation. We fit three separate sets of models: for dawn songs, daytime first category songs, and daytime second category songs. Sample sizes varied among the three sets, because some birds in our analysis sample were not recorded extensively during the dawn bout and because post-dawn singing was more likely to consist of first category songs than second category songs. Thus, the sample sizes (number of birds) for the dawn ($n=27$) and daytime second category ($n=28$) models were smaller than for the daytime first category ($n=51$) models.

Before fitting models, we screened potential predictor variables for intercorrelation. We calculated correlation coefficients for all possible pairwise combinations of our song variables. We considered a value of $r=0.7$ or higher to be an indication that retaining both variables would risk multicollinearity in a model. We calculated correlation matrices separately for each of our three song categories, but results were very similar across categories. Within the song variables, highest frequency was strongly correlated with frequency bandwidth; we were interested in bandwidth as a song structure measure, so we retained it and excluded highest frequency from our models. All of our coefficient of variation measures were strongly intercorrelated, even though they measured performance consistency of diverse frequency and temporal features. We therefore used principle component analysis to reduce this suite of variables to a single variable that measured overall consistency of performance (Table 2). Because the scores for this composite variable represented levels of coefficient of variation (which is inversely related to consistency), we reversed their signs so that they could be interpreted as measures of consistency (i.e., so that higher values signified greater consistency).

Our screen for correlations also included non-song variables that might be associated with first-egg date: male arrival date, male age (second year versus after second year), and male tarsus length and wing length (indices of body size). All four non-song variables were strongly correlated with multiple song variables, making it impossible to include both song and non-song variables in a given model. The correlations also make it very challenging, in our observational study, to disentangle the causal effects of the two classes of variables. In addition, the two classes of variables seem conceptually distinct; the song variables represent signal features whereas the non-song variables represent aspects of male phenotype that might be revealed by signaling. Overall, it seemed to us

Table 2 Results of a principal component analysis of variables related to song performance consistency

	PC1
Eigenvalue	6.95
% variance explained	89.9
Component loadings	
CV intersong interval	0.669
CV lowest frequency	0.967
CV highest frequency	0.988
CV peak frequency	0.946
CV bandwidth	0.923
CV duration	0.974
CV number of elements	0.966
CV element rate	0.984

Only the single factor with an eigenvalue greater than 1 is shown. Factor loadings are based on varimax rotation

prudent to proceed by analyzing the two classes of variables separately. We therefore limit our current analysis to the effects of variation in song performance and plan to address the potential association between song traits and non-song traits in a future analysis.

For each song category, we fit a set of candidate models that represented our multiple working hypotheses. One of these hypotheses was that first-egg date would be best predicted by a model in which the fixed factors included singing rate, performance consistency, song frequency, and song structure. The remaining hypotheses were represented by models containing the 14 possible subsets of those four predictor categories (see Table 1).

In the models, singing rate was represented by mean intersong interval (shorter intervals = faster singing rate). Performance consistency was represented by our composite consistency variable. Song frequency was represented in the models by two variables—lowest frequency and peak frequency—because both the extreme frequency and peak frequency of songs have been important predictors in prior studies (e.g., Byers 2007). Song structure was represented by three variables: song duration, element rate (elements per second within a song), and frequency bandwidth. In prairie warbler songs, with their steady increase in pitch as a song proceeds (see Fig. 1), frequency bandwidth represents the increase from the lowest frequency of the first song element to the highest frequency of the final element.

In addition to the fixed factors described above, all of our models also included a term for habitat type. Some of our subjects nested in pitch pine ($N=19$ birds), some in scrub oak ($N=13$), and some in power line corridors ($N=19$). Exploratory plots of our data suggested that mean first-egg date differed in different habitats, such that the dates within a habitat type may be correlated. We therefore included habitat type in all models as a random factor. In addition, because the

timing of mating and nesting might vary from year to year, we included year as a fixed factor in our models.

None of our candidate models included measures of song elaboration (i.e., measures of the variety or diversity of sounds produced, such as song repertoire size or element repertoire size). We elected to focus on song performance and exclude elaboration because prairie warbler songs and singing have characteristics that suggest that a signal function of variation in elaboration is unlikely. For example, song repertoire size varied hardly at all among the males in our study population; we recorded two song types from more than 90 % of males, and the remaining birds used just one additional song type. In addition, the elements that compose a prairie warbler song cannot, for the most part, be said to fall into discrete types; each successive element is slightly different than the preceding one (see Fig. 1). Thus, the elements in most songs constitute a continuum of similar forms and cannot be readily separated into distinct types; the concept of element repertoire size does not seem to be meaningful in prairie warblers. Nonetheless, we cannot definitely rule out the possibility that prairie warblers perceive variation in song elaboration that is not apparent to human observers and that our models would be improved had we devised and included a measure of song elaboration.

Model selection

To evaluate our three sets of candidate models (i.e., working hypotheses), we assessed the strength of evidence in support of each model in a set using the method and rationale outlined by Burnham and Anderson (2002). Our comparisons were based on values of second-order Akaike's information criterion (AIC_c), calculated for each fitted model using the R package AICcmodavg (Mazerolle 2013). (AIC_c is an adjusted version of AIC that performs better than AIC when the ratio of sample size to the number of parameters estimated is less than about 40, as was the case in our analysis). Models with lower values of AIC_c are better supported by the data than are models with higher values.

To keep our final model set relatively small, we employed a two-stage, hierarchical selection process (Dinsmore and Dinsmore 2007; Shake et al. 2011). Prior to the main model selection process that compared our working hypotheses, we compared model variants within the hypotheses that included song frequency or structure. Each of these two predictors included two or three potential component variables and, thus, could be represented in models by multiple variables or just one. We therefore compared AIC_c for all possible variable combinations within a hypothesis (results summarized in Online Resource 1) and chose the combination with the lowest AIC_c to represent that hypothesis in the main model selection comparison. We performed this procedure separately for each song category.

Once we had finalized our set of candidate models, we fit each one with and without the inclusion of year as a predictor. In each case, we found that a model with year excluded had a lower AIC_c than did the corresponding model with year included. These findings suggest that year did not have an important effect on first-egg date in our sample, a conclusion also supported by the observation that median first-egg date (26 May) was the same in both years of the study. We therefore subsequently excluded year from all models.

For the main model selection comparison, we calculated several additional statistics to quantify the evidence in support of each model in a set of candidates. We calculated the AIC_c difference (Δ_i), i.e., the difference between a model's AIC_c value and the lowest AIC_c value in the set of candidates. Values of Δ_i allow ranking of the candidate models; smaller values reflect a higher likelihood that a model is the best model, given the data. We used Δ_i values to calculate Akaike weights (w_i) for each model. The Akaike weights of the models in a candidate set sum to 1; the statistic provides a measure of the probability that a model is the best approximating model in a candidate set. We also calculated the ratio of each model's Akaike weight to the weight of the best model in a candidate set. This ratio, known as the evidence ratio, provides another tool for assessing the relative likelihood of each model in a set to be the best approximating model.

We also assessed the relative importance of each predictor that appeared in the model set. To do so, for each predictor, we summed Akaike weights across the subset of candidate models that contained the predictor (each predictor appeared in eight different models). Higher sums indicate a more important predictor. In addition, because our model selection procedure assessed the relative fit of alternative models but did not reveal information about their absolute fit, we estimated goodness-of-fit for each of our best models by calculating $R^2_{GLMM(m)}$ (marginal R^2) and $R^2_{GLMM(c)}$ (conditional R^2) using the method developed by Nakagawa and Schielzeth (2013). Marginal R^2 estimates the proportion of variance explained by the fixed effects; conditional R^2 estimates the proportion of variance explained by the fixed and random effects together. R^2 values were estimated with the R package MuMIn (Bartoń 2013).

Because our model selection procedure did not reveal a single unequivocally best model in any of our candidate sets (see Results), we derived parameter estimates by averaging across the most plausible models in each set. We calculated these weighted-average parameter estimates as "natural" averages (see Burnham and Anderson 2002, p. 152). For each model-averaged parameter estimate, we also estimated unconditional confidence intervals, which account for both uncertainty associated with parameter estimation and uncertainty associated with model selection. Model-averaged parameter estimates and unconditional confidence intervals were estimated with AICcmodavg.

Repeatability

To assess whether song traits were persistent characteristics of individuals or were more plastic, we assessed within-year repeatability for each song variable, i.e., the proportion the variation in our measurements that is attributable to variation among individuals. We estimated repeatability as the intraclass correlation coefficient. Intraclass correlation coefficients and confidence intervals were calculated with the R package ICC (Wolak et al. 2012), which estimates these quantities using variance components from a one-way ANOVA.

Results

Our assessment of empirical support for our candidate models suggests that, for first category songs, first-egg date is best predicted by a model that includes singing rate, song performance consistency, and song frequency (Table 3). Performance consistency is also in important predictor of first-egg date for second category songs, along with song structure (Table 4).

First category songs

For first category songs, two models had reasonably high likelihoods of being the best approximating model in the candidate set. A model containing singing rate, song performance consistency, and song frequency as fixed effects proved most likely to be the best model, and the same model but with consistency excluded was about two thirds as likely. These two models together had a combined probability of about 0.85 of being the best model. All other candidate models were at least ten times more poorly supported by the data than was the top-ranked model (see Table 3). The summed Akaike weights for all candidate models suggest that singing rate

Table 3 Model selection statistics for a set of candidate models with predictor variables measured from prairie warbler first category songs

Candidate models	k	Δ_i	w_i	ER
Rate + consist + freq	6	0	0.527	1.0
Rate + freq	5	0.95	0.327	1.6
Rate + consist + freq + struc	7	4.90	0.045	11.6
Rate + perf + freq	6	5.54	0.033	16.0
Freq	4	6.74	0.018	29.1
Consist + freq	5	7.14	0.015	35.5

First-egg date is the dependent variable. Statistics shown include the number of parameters estimated (k), AIC$_c$ difference (Δ_i), Akaike weight (w_i), and evidence ratio (ER). Only models with evidence ratios lower than 40 are shown. For this set of models, lowest AIC$_c$=260.45

Consist consistency, *freq* frequency, *perf* performance, *struc* structure

Table 4 Model selection statistics for a set of candidate models with predictor variables measured from prairie warbler second category songs uttered after the dawn bout

Candidate models	k	Δ_i	w_i	ER
Consist + structure	6	0	0.533	1.0
Consist	4	2.01	0.195	2.7
Structure	5	3.10	0.113	4.7
Rate	4	5.91	0.028	19.2
Rate + consist	5	6.07	0.026	26.1
Rate + consist + structure	7	6.52	0.020	27.8

First-egg date is the dependent variable. Statistics shown include number of parameters estimated (k), AIC$_c$ difference (Δ_i), Akaike weight (w_i), and evidence ratio (ER). Only models with evidence ratios lower than 40 are shown. For this set of models, lowest AIC$_c$=173.83

Consist consistency

and song frequency are the most important predictors, that performance consistency is somewhat less important, and that song structure is not an important predictor (Table 5). Marginal R^2 for the model containing singing rate, frequency, and consistency was 0.37, and the conditional R^2 was 0.52.

The signs of estimated coefficients averaged over the two best-supported models indicate that the social mates of males with faster singing rates (smaller intersong intervals), more consistent singing, and lower-frequency songs laid eggs earlier (Table 6, Fig. 2a). The 95 % confidence intervals of the estimated coefficients suggest that the precision of the estimates is not especially high, but that the estimated coefficients would be expected to retain their signs in a large majority of replicate samples (i.e., we are more confident in the direction of the effects than in our estimates of their magnitude). Overall, the data for first category songs seem to provide strong empirical support for the hypothesis that some variation in first-egg date is predicted by a combination of singing rate, song frequency, and song performance consistency.

Second category songs, daytime singing

For second category songs sung after the dawn bout, one model in the candidate set had reasonably strong support,

Table 5 Summed Akaike weights for all candidate models containing each variable type

	Daytime first category	Daytime second category
Singing rate	0.956	0.093
Consistency	0.604	0.821
Structure	0.095	0.713
Frequency	0.974	0.084

Sums are shown for two sets of models: first category songs and second category songs uttered after the dawn bout

Table 6 Fixed effect estimates averaged across the most plausible models from each of two sets of candidate models

Variable	Coefficient	Unconditional 95 % CI	
		Lower	Upper
Daytime first category songs (*N*=51 birds)			
Intersong interval (s)	0.72	0.28	1.16
Consistency	−2.50	−4.96	−0.05
Lowest frequency (kHz)	1.38	0.58	2.19
Daytime second category songs (*N*=28 birds)			
Consistency	−2.64	−5.03	−0.25
Duration (s)	−4.54	−8.30	−0.79
Element rate (elem/s)	3.18	0.48	5.87

Two models were averaged for first category songs and three models for second category songs. First-egg date was the dependent variable. Each model also estimated variance components for a random effect (habitat type) not shown here

and two models had weaker but non-negligible support (see Table 4). A model containing song performance consistency and song structure (duration of songs and elements) had a 0.53 probability of being the best model, and two subsets of that model, consistency only and structure only, were about one third and one fifth as well supported, respectively. The combined probability of the three top-ranked models was 0.84. All other candidate models had only weak or negligible support (evidence ratios of 19 or higher). Akaike weights summed across all candidate models indicated that consistency and structure are important predictors, but singing rate and song frequency are not (see Table 5). Marginal R^2 for the model containing consistency and song structure was 0.34, and the conditional R^2 was 0.57.

As with first category songs, the signs of the model-averaged estimated coefficients (in this case averaged over three models) indicate that more consistent singing of daytime second category songs was associated with earlier first-egg dates (see Table 6, Fig. 2b). In addition, earlier first-egg dates were associated with longer songs and lower element rates (i.e., longer elements). Confidence intervals for the estimated coefficients again indicate limited estimation precision, but suggest that most replicate samples would result in estimated coefficients of unchanged sign. Overall, the data for daytime second category songs provide the most support for the hypothesis that first-egg date is best predicted by a combination of consistency and structure, but models that included only one of those variables were also, though considerably more weakly, supported.

Second category songs, dawn bout

Model ranks for second category songs uttered during the dawn bout were similar to those for second category songs
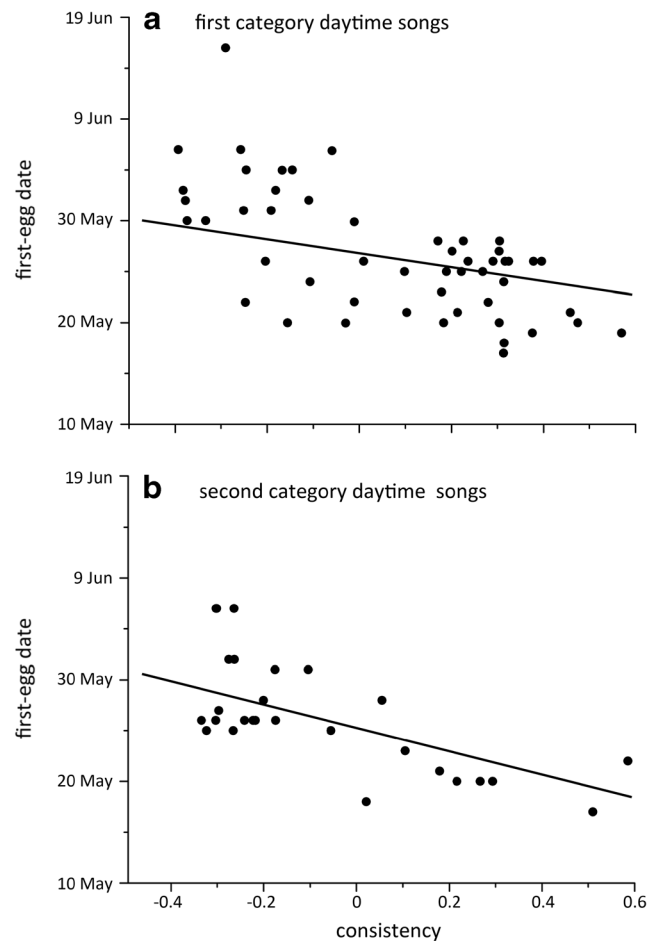

Fig. 2 Scatterplots of observed values of song performance consistency versus first-egg date for **a** first category songs and **b** daytime second category songs overlaid with partial plots of the conditional effect of consistency, as estimated by linear mixed effects models. Consistency is the only predictor variable that appeared in the best models for both song categories

sung later in the day. However, the absolute fit of even the best-supported model was poor (marginal R^2=0.06, conditional R^2=0.17). In addition, the 95 % confidence intervals of the estimated coefficients were much wider than for the other models and included zero. The comparatively poor fit and imprecise estimation of the dawn bout models suggest that the association between singing and first-egg date was weaker for dawn singing than for daytime singing, or perhaps that, for the dawn bout, the set of candidate models did not include a good approximation of the "true" model. In any case, our results for dawn songs seem inadequate to support inferences about the role of the dawn bout in social mate choice, so we do not consider them further.

Repeatability

Repeatability estimates revealed that, for the song variables included in the best-supported models for daytime singing, except for singing rate, the majority of observed variation

was due to variation among individuals (all $R > 0.6$; Table 7). Thus, comparatively little variation was due to variation across different song bouts within individuals (which includes both individual plasticity and measurement error). For singing rate, however, repeatability was considerably lower than it was for the other variables ($R = 0.18$). For all variables, repeatability was higher for dawn songs than for daytime songs (see Table 7).

Discussion

Our analysis of prairie warbler singing showed that first-egg date, serving as a proxy for female social mate choice, was best predicted by models that included multiple types of song performance indicator. In particular, the best-performing models for first category songs included singing rate, performance consistency, and song frequency, and for second category songs included both consistency and aspects of song structure. This finding is consistent with the hypothesis that female prairie warblers prefer to mate with males that can maximize multiple dimensions of song performance simultaneously and is also consistent with the hypothesis that such males gain an advantage in contests for territories that attract females. Perhaps listening prairie warblers are sufficiently impressed by the neuromuscular skill required for consistent singing only if consistency is achieved while simultaneously meeting the demands of singing rapidly at low pitch (when

Table 7 Repeatability, estimated as the intraclass correlation coefficient (R), of song variables retained in the models that best predicted first-egg date for daytime songs, with values for dawn song shown for comparison

Variable	R	95 % CI	
		Lower	Upper
Day 1 songs ($N=55$ birds, mean $k=6.7$ samples/bird)			
Intersong interval (s)	0.18	0.09	0.27
Consistency	0.61	0.45	0.76
Lowest frequency (kHz)	0.68	0.53	0.83
Element rate (elem/s)	0.82	0.74	0.90
Duration	0.68	0.54	0.82
Dawn songs ($N=28$ birds, mean $k=7.0$ samples/bird)			
Intersong interval (s)	0.60	0.41	0.78
Consistency	0.72	0.55	0.89
Lowest frequency (Hz)	0.82	0.72	0.92
Element rate (elem/s)	0.92	0.91	0.93
Duration (s)	0.79	0.67	0.91

N for repeatability is slightly larger than for the main models due to inclusion of birds for which we had adequate song samples but no nesting data

singing first category songs) or singing longer songs with longer elements (when singing second category songs).

Although our correlative models do not allow us to reliably distinguish effects attributable to female preference for particular males from those attributable to the outcome of male-male competition, we speculate that a distinction between these two causal mechanisms may help explain why our results differed somewhat between first and second category songs. As described earlier, observational and experimental evidence suggests that the first category songs of wood warblers function in mate attraction, whereas second category songs are associated with aggression (see references in the "Introduction"). Thus, good performance of first category songs may appeal directly to females, while good performance of second category songs attracts females indirectly, by assisting in acquisition of a superior territory. The particular aspects of performance that are most important may differ between the two communicative functions. However, determining if and how variation in song performance influences the behavior of male and female receivers will require further investigation.

Constancy of trait expression

A listener can reliably discriminate among males on the basis of song performance traits, and the traits can respond to the resulting selection, only if the traits are constant features of individual males (Boake 1989). This condition appears to be met for most of the performance traits identified as important in our models, almost all of which had relatively high repeatability. Singing rate of daytime songs, however, had low repeatability, a not unexpected finding given that singing rate in many species is known to vary with context, for example, with breeding stage or time of day (e.g., Pinxten and Eens 1998; Amrhein et al. 2004; Foote and Barber 2009). However, our finding that song rate is much more repeatable when observations are restricted to a particular context (i.e., the dawn bout; see Table 7) suggests the possibility that singers differ reliably in singing rate if the listener is able to adjust for context. Such an effect would help explain why our model for first category songs detected an effect of differences among males in average singing rate, despite the large amount of variability around the average.

Signal honesty

As Prum (2010, 2012) points out, robust theory (Lande 1981; Kirkpatrick 1982) suggests that female preference for vigorous, skillful performances can in principle evolve even if the performance reveals no information about the performer's likelihood of providing direct or indirect benefits to females. Nonetheless, investigations of female preferences for performance tend to be informed by the proposition that the signal value of performance arises from its potential to accurately

expose differences in the quality of signalers, by revealing differences in their ability to perform physically challenging motor tasks. Are the aspects of song performance that predict first-egg date demanding enough to reveal underlying differences among singers?

Because song production in oscines can require extraordinarily precise, high-speed neuromuscular coordination (Suthers 2004; Elemans et al. 2008; Suthers et al. 2012), it is intuitively plausible that performance of many aspects of songs would be constrained by the difficulty of the underlying motor tasks. Empirical evidence in support of this intuition, however, is relatively sparse. Among the aspects of performance that our models suggest are correlated with first-egg date, most of the relevant published evidence is related to singing rate and, to a lesser extent, singing consistency.

With respect to the question of whether rapid or abundant singing is constrained by physiological limits, the overall evidence is ambiguous. Laboratory studies of oxygen consumption by singing birds have generally shown only a small metabolic cost of singing (summarized in Ward et al. 2004). However, other evidence suggests that the metabolic cost of sustained singing under natural conditions may be higher (Ward and Slater 2005). For example, a field study of great reed warblers (*Acrocephalus arundinaceus*) found that high singing rates were associated with substantially elevated metabolic rates (Hasselquist and Bensch 2008). In addition, there is evidence that free-living birds with higher singing rates lose mass more quickly or gain it more slowly (Thomas 2002; Thomas et al. 2003); that birds experimentally induced to mount a costly immune response reduce their singing rate (Owen-Ashley et al. 2006; Munoz et al. 2010); and that restricting energy intake reduces singing rate (e.g., Ritschard and Brumm 2012) whereas supplementing food increases it (e.g., Berg et al. 2005; Grava et al. 2009). These findings suggest that song output and singing rate can in at least some circumstances depend on the vigor and condition of the singer, but a conclusive demonstration that singing rapidly is physically demanding remains elusive.

Fewer studies have examined the hypothesis that it is physically challenging to sing songs that do not vary in form, duration, pitch, or timing (cadence) across renditions, such that a singer's degree of consistency in singing might reveal its physical capabilities. Some evidence in support of this proposition comes from studies of singing by birds whose condition had been experimentally worsened. For example, the singing consistency of male white-crowned sparrows (*Zonotrichia leucophrys*) decreased after the birds were infected with apicomplexan parasites (Gilman et al. 2007), and singing consistency was likewise reduced in male canaries (*Serinus canaria*) subjected to infestation by ticks (Müller et al. 2013). A few studies have found that singing consistency increases with age (Botero et al. 2009; de Kort et al. 2009;

Rivera-Gutierrez et al. 2010), suggesting that female listeners could evaluate consistency to choose older males.

If the song performance traits that we identified as influencing social mate choice are in fact honest indicators of male traits that benefit their mates, then song characteristics that increase performance difficulty would enhance the ability of females to distinguish males with the best performance ability. The benefits of this increased ability to discriminate could result in selection favoring acoustically complex signals. Such selection may be one of the factors responsible for the structural complexity of prairie warbler songs and, by extension, those of other oscine species.

References

Akresh ME, King DI, Brooks RT (2015) Demographic response of a shrubland bird to habitat creation, succession, and disturbance in a dynamic landscape. Forest Ecol Manag 336:72–80

Alatalo RV, Glynn C, Lundberg A (1990) Singing rate and female attraction in the pied flycatcher: an experiment. Anim Behav 39:601–603

Amrhein V, Kunc HP, Naguib M (2004) Seasonal patterns of singing activity vary with time of day in the nightingale (*Luscinia megarhynchos*). Auk 121:110–117

Bartoń K (2013) MuMIn: Multi-model inference. R package version 1.9.13, http://CRAN.R-project.org/package=MuMIn

Bates D, Maechler M, Bolker B, Walker S (2013) lme4: linear mixed-effects models using Eigen and S4. R package version 1.1-6, http://CRAN.R-project.org/package=lme4

Berg ML, Beintema NH, Welbergen JA, Komdeur J (2005) Singing as a handicap: the effects of food availability and weather on song output in the Australian reed warbler *Acrocephalus australis*. J Avian Biol 36:102–109

Bioacoustics Research Program (2011) Raven Pro: interactive sound analysis software (version 1.4), http://www.birds.cornell.edu/raven

Boake CRB (1989) Repeatability—its role in evolutionary studies of mating-behavior. Evol Ecol 3:173–182

Bolsinger JS (2000) Use of two song categories by golden-cheeked warblers. Condor 102:539–552

Botero CA, Rossman RJ, Caro LM, Stenzler LM, Lovette IJ, de Kort SR, Vehrencamp SL (2009) Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. Anim Behav 77:701–706

Burnham KP, Anderson DR (2002) Model selection and multimodel inference. Springer, New York

Byers BE (1996) Messages encoded in the songs of chestnut-sided warblers. Anim Behav 52:691–705

Byers BE (2007) Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. Behav Ecol 18:130–136

Byers J, Hebets E, Podos J (2010) Female mate choice based upon male motor performance. Anim Behav 79:771–778

Cardoso GC (2012) Paradoxical calls: the opposite signaling role of sound frequency across bird species. Behav Ecol 23:237–241

Chiver I, Stutchbury BJM, Morton ES (2008) Do male plumage and song characteristics influence female off-territory forays and paternity in the hooded warbler? Behav Ecol Sociobiol 62:1981–1990

Christensen R, Kleindorfer S, Robertson J (2006) Song is a reliable signal of bill morphology in Darwin's small tree finch *Camarhynchus parvulus*, and vocal performance predicts male pairing success. J Avian Biol 37:617–624

Darwin C (1871) The descent of man and selection in relation to sex, 2nd edn. Murray, London

de Kort SR, Eldermire ERB, Valderrama S, Botero CA, Vehrencamp SL (2009) Trill consistency is an age-related assessment signal in banded wrens. Proc R Soc Lond B 276:2315–2321

Dinsmore SJ, Dinsmore JJ (2007) Modeling avian nest survival in program MARK. Stud Avian Biol 34:73–83

Elemans CPH, Mead AF, Rome LC, Goller F (2008) Superfast vocal muscles control song production in songbirds. PLoS One 3, e2581

Foote JR, Barber CA (2009) Paired male song sparrows sing more when their mate is incubating. Wilson J Ornithol 121:819–822

Forstmeier W, Leisler B (2004) Repertoire size, sexual selection, and offspring viability in the great reed warbler: changing patterns in space and time. Behav Ecol 15:555–563

Forstmeier W, Kempenaers B, Meyer A, Leisler B (2002) A novel song parameter correlates with extra-pair paternity and reflects male longevity. Proc R Soc Lond B 269:1479–1485

Gilman S, Blumstein DT, Foufopoulos J (2007) The effect of hemosporidian infections on white-crowned sparrow singing behavior. Ethology 113:437–445

Grava T, Grava A, Otter KA (2009) Supplemental feeding and dawn singing in black-capped chickadees. Condor 111:560–564

Hasselquist D, Bensch S (2008) Daily energy expenditure of singing great reed warblers *Acrocephalus arundinaceus*. J Avian Biol 39:384–388

Highsmith RT (1989) The singing behavior of golden-winged warblers. Wilson Bull 101:36–50

Houlihan PW (2000) The singing behavior of prairie warblers (*Dendroica discolor*). Dissertation, University of Massachusetts

Houtman AM (1992) Female zebra finches choose extra-pair copulations with genetically attractive males. Proc R Soc Lond B 249:3–6

Kempenaers B, Verheyren GR, Dhondt AA (1997) Extrapair paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring quality. Behav Ecol 8:481–492

Kirkpatrick M (1982) Sexual selection and the evolution of female choice. Evolution 36:1–12

Krebs J, Ashcroft R, Webber M (1978) Song repertoires and territory defence in the great tit. Nature 271:539–542

Kroodsma DE (2005) The singing life of birds. Houghton Mifflin, Boston

Kroodsma DE, Bereson RC, Byers BE, Minear E (1989) Use of song types by the chestnut-sided warbler: evidence for both intra- and inter-sexual functions. Can J Zool 67:447–456

Lande R (1981) Models of speciation by sexual selection on polygenic traits. Proc Natl Acad Sci U S A 78:3721–3725

Linhart P, Jaska P, Petruskova T, Petrusek A, Fuchs R (2013) Being angry, singing fast? Signalling of aggressive motivation by syllable rate in a songbird with slow song. Behav Process 100:139–145

Mazerolle MJ (2013) AICcmodavg: model selection and multimodel inference based on (Q)AIC(c). R package version 1.35, http://CRAN.R-project.org/package=AICcmodavg

Müller W, Heylen D, Eens M, Rivera-Gutierrez HF, Groothuis TGG (2013) An experimental study on the causal relationships between (ecto-)parasites, testosterone and sexual signalling. Behav Ecol Sociobiol 67:1791–1798

Munoz NE, Blumstein DT, Foufopoulos J (2010) Immune system activation affects song and territorial defense. Behav Ecol 21:788–793

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R^2 from generalized linear mixed-effects models. Methods Ecol Evol 4:133–142

Nolan V (1978) The ecology and behavior of the prairie warbler, *Dendroica discolor*. American Ornithologists' Union, Tampa

Nolan PM, Hill GE (2004) Female choice for song characteristics in the house finch. Anim Behav 67:403–410

Owen-Ashley NT, Turner M, Hahn TP, Wingfield JC (2006) Hormonal, behavioral, and thermoregulatory responses to bacterial lipopolysaccharide in captive and free-living white-crowned sparrows (*Zonotrichia leucophrys gambelii*). Horm Behav 49:15–29

Pinxten R, Eens M (1998) Male starlings sing most in the late morning, following egg-laying: a strategy to protect their paternity? Behaviour 135:1197–1211

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

Poesel A, Foerster K, Kempenaers B (2001) The dawn song of the blue tit *Parus caeruleus* and its role in sexual selection. Ethology 107:521–531

Poesel A, Nelson DA, Gibbs HL (2012) Song sharing correlates with social but not extrapair mating success in the white-crowned sparrow. Behav Ecol 23:627–634

Prum RO (2010) The Lande-Kirkpatrick mechanism is the null model of evolution by intersexual selection: implications for meaning, honesty, and design in intersexual signals. Evolution 64:3085–3100

Prum RO (2012) Aesthetic evolution by mate choice: Darwin's really dangerous idea. Philos T Roy Soc B 367:2253–2265

R Core Team (2014) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/

Reid JM, Arcese P, Cassidy A, Hiebert SM, Smith JNM, Stoddard PK, Marr AB, Keller LF (2004) Song repertoire size predicts initial mating success in male song sparrows, *Melospiza melodia*. Anim Behav 68:1055–1063

Ritschard M, Brumm H (2012) Zebra finch song reflects current food availability. Evol Ecol 26:801–812

Ritschard M, Riebel K, Brumm H (2010) Female zebra finches prefer high-amplitude song. Anim Behav 79:877–883

Rivera-Gutierrez HF, Pinxten R, Eens M (2010) Multiple signals for multiple messages: great tit, *Parus major*, song signals age and survival. Anim Behav 80:451–459

Sakata JT, Vehrencamp SL (2012) Integrating perspectives on vocal performance and consistency. J Exp Biol 215:201–209

Shake CS, Moorman CE, Burchell MR (2011) Cropland edge, forest succession, and landscape affect shrubland bird nest predation. J Wildlife Manage 75:825–835

Spector DA (1991) The singing behavior of yellow warblers. Behaviour 117:29–52

Spector DA (1992) Wood-warbler song systems: a review of paruline singing behaviors. Curr Ornithol 9:199–238

Staicer CA (1996) Honest advertisement of pairing status: evidence from a tropical resident wood-warbler. Anim Behav 51:375–390

Staicer CA, Ingalls V, Sherry TW (2006) Singing behavior varies with breeding status of American redstarts (*Setophaga ruticilla*). Wilson J Ornithol 118:439–451

Suthers RA (2004) How birds sing and why it matters. In: Marler P, Slabbekoorn H (eds) Nature's music: the science of birdsong. Elsevier Academic Press, San Diego

Suthers RA, Goller F (1997) Motor correlates of vocal diversity in songbirds. Curr Ornithol 14:235–288

Suthers RA, Vallet E, Kreutzer M (2012) Bilateral coordination and the motor basis of female preference for sexual signals in canary song. J Exp Biol 215:2950–2959

Thomas RJ (2002) The costs of singing in nightingales. Anim Behav 63: 959–966

Thomas RJ, Cuthill IC, Goldsmith AR, Cosgrove DF, Lidgate HC, Proctor SLB (2003) The trade-off between singing and mass gain in a daytime-singing bird, the European robin. Behaviour 140:387–404

Ward S, Slater PJB (2005) Raised thermoregulatory costs at exposed song posts increase the energetic cost of singing for willow warblers *Phylloscopus trochilus*. J Avian Biol 36:280–286

Ward S, Lampe HM, Slater PJB (2004) Singing is not energetically demanding for pied flycatchers, *Ficedula hypoleuca*. Behav Ecol 15: 477–484

Wilson DR, Bitton P-P, Podos J, Mennill DJ (2014) Uneven sampling and the analysis of vocal performance constraints. Am Nat 183:214–228

Wolak ME, Fairbairn DJ, Paulsen YR (2012) Guidelines for estimating repeatability. Methods Ecol Evol 3:129–137

Zahavi A, Zahavi A (1997) The handicap principle: a missing piece of Darwin's puzzle. Oxford University Press, New York