



## Within-day improvement in a behavioural display: wild birds 'warm up'



Hannes A. Schraft<sup>a, b, \*</sup>, Orlando J. Medina<sup>c</sup>, Jesse McClure<sup>d</sup>, Daniel A. Pereira<sup>e</sup>, David M. Logue<sup>f</sup>

<sup>a</sup> Department of Biology, San Diego State University, San Diego, CA, U.S.A.

<sup>b</sup> Graduate Group in Ecology, Department of Neurobiology, Physiology, and Behavior, University of California, Davis, CA, U.S.A.

<sup>c</sup> Department of Biology, University of Puerto Rico, Mayagüez, Puerto Rico

<sup>d</sup> Program in Bioinformatics & Integrative Biology, University of Massachusetts, Medical School, Worcester, MA, U.S.A.

<sup>e</sup> GEAHNA Group, Biology Department, Universidad del Valle, Cali, Valle del Cauca, Colombia

<sup>f</sup> Department of Psychology, University of Lethbridge, Lethbridge, AB, Canada

### ARTICLE INFO

#### Article history:

Received 12 June 2016

Initial acceptance 26 August 2016

Final acceptance 21 November 2016

MS. number: A16-00613R

#### Keywords:

anti-exhaustion hypothesis  
behavioural display  
birdsong  
dawn chorus  
frequency excursion  
performance

Motor performance describes the vigour or skill required to perform a particular display. It is a behaviourally salient variable in birdsong and other animal displays, but little is known about within-individual variation in performance over short timescales. The metric 'frequency excursion' (FEX) quantifies birdsong performance as cumulative frequency modulation per unit time. We measured FEX in a large sample of recordings from free-living male Adelaide's warblers, *Setophaga adelaidae*. Our objectives were to quantify natural variation in performance and test the hypotheses that performance (1) improves as a function of recent practise, (2) decreases over consecutive repetitions of a single song type, (3) improves with rest between songs, (4) varies by singing mode and (5) changes during vocal interactions with neighbours. We found significant variation in performance among individuals and song types. Consecutive repetition of a song type, rest between songs, singing mode and vocal interaction did not strongly affect performance. Performance consistently increased with song order, however, indicating that males warm up during morning singing. This is the first demonstration of such an effect in a sexual display. The warm-up effect may explain the prevalence of intense dawn singing in birds (dawn chorus), if rivals engage in an arms race to warm up.

© 2017 The Association for the Study of Animal Behaviour. Published by Elsevier Ltd. All rights reserved.

Sexual selection drives the elaboration of sexual ornaments and displays in animals (Bradbury & Vehrencamp, 2011). As these traits evolve to become increasingly extreme, costs accumulate and constraints take effect, limiting further elaboration. Motor constraints may be particularly important in limiting the evolution of sexual displays (Podos & Patek, 2015), as suggested by systems in which the display performance improves with experience and age. Examples include the cartwheel displays of lance-tailed manakins, *Chiroxiphia lanceolata* (DuVal, 2012), and trill performance in swamp sparrows, *Melospiza georgiana* (Ballentine, 2009) and banded wrens, *Thryophilus pleurostictus* (Vehrencamp, Yantachka, Hall, & de Kort, 2013). Although it has not been studied as intensively as year-to-year variation in performance, within-day variation in performance may also provide evidence of performance

constraints. In the present study, we examine the factors that influence short-term variation in performance in a Neotropical songbird, Adelaide's warblers, *Setophaga adelaidae*.

Birdsong production requires precise coordination of the intricate avian vocal apparatus, so it is likely that the evolution of birdsong has been affected by motor constraints (Suthers, 2004). The anti-exhaustion hypothesis proposes that motor fatigue limits birds' ability to repeatedly produce the same song type, but birds can escape fatigue by switching to a new song type (Lambrechts & Dhondt, 1988). A test of this hypothesis in chaffinches, *Fringilla coelebs*, found no support (Brumm, Lachlan, Riebel, & Slater, 2009). We propose a novel hypothesis about the effects of motor constraints over short timescales: signal performance could improve over short timescales if animals 'warm up'. Improvement in a motor task due to recent practice, or 'warming up,' is known to affect human athletes and singers (Amir, Amir, & Michaeli, 2005; Moorcroft & Kenny, 2013; Motel, Fisher, & Leydon, 2003; Stewart, Macaluso, & De Vito, 2003). The warm-up hypothesis predicts

\* Correspondence: H. A. Schraft, Department of Biology, San Diego State University, San Diego, CA 92182, U.S.A.

E-mail address: [haschraft@ucdavis.edu](mailto:haschraft@ucdavis.edu) (H. A. Schraft).

that singing performance improves with recent practice, regardless of song type. We quantify recent practice with the variable song ‘order’, which describes the number of songs the bird has already sung on the focal morning.

Superficially, the warm-up hypothesis may seem to be at odds with the anti-exhaustion hypothesis. We suggest that the two hypotheses are not mutually exclusive, however, because birds that switch song types might warm up, even though they would become exhausted if they were to continue singing the same song type. Thus, we invoke the anti-exhaustion hypothesis to predict that performance will decrease over consecutive repetitions of a given song type (measured as the ordinal number of a song in a run of same-type songs, or the ‘run number’). Alternatively, singing may temporarily exhaust a bird’s resources regardless of song type. This ‘song type general exhaustion’ hypothesis leads to the prediction that performance will covary positively with the latency since the prior song.

Early morning (Type II) singing of New World warblers (family: Parulidae), including Adelaide’s warblers, differs from daytime (Type I) singing during the breeding season (Burt & Vehrencamp, 2005; Catchpole & Slater, 2008; Spector, 1992; Staicer, 1996a; Staicer, Spector, & Horn, 1996). Distinct functions for these two singing modes have been proposed. A study showing that Type I songs have higher performance than Type II songs suggests that Type I singing may have evolved to showcase performance (Beebe, 2004a). We therefore tested whether singing mode affects song performance.

Several studies on natural and simulated territorial interactions show that songbirds adjust their performance level when they are vocally interacting (‘countersinging’) with a potential rival (Benedict, Rose, & Warning, 2012; DuBois, Nowicki, & Searcy, 2009; Price, Earnshaw, & Webster, 2006; Trillo & Vehrencamp, 2005). We tested this hypothesis by estimating the effect of countersinging on song performance. Additionally, performance may be affected by various time-dependent factors, such as the air temperature, the amount of time the bird has been awake, or the rate of social interactions. We included time of day in our analysis to control for these potential influences on performance and to account for covariance between time and order. Table 1 summarizes our predictions for each hypothesis.

Motor performance in trilled birdsong has traditionally been quantified by measuring the orthogonal distance of a song from the upper-bound regression line of a plot of trill rate and frequency bandwidth (Podos, 1997). Using this measure of performance (termed ‘vocal deviation’, VDev) and its components, trill rate and frequency bandwidth, studies have revealed variation in performance among species (Podos, 1997, 2001), individuals (Ballentine, Hyman, & Nowicki, 2004; Podos, 2001), contexts (Benedict et al., 2012; DuBois et al., 2009; Price et al., 2006; Trillo & Vehrencamp, 2005) and song types (Cardoso, Atwell, Hu, Ketterson, & Price, 2012; Cardoso, Atwell, Ketterson, & Price, 2009; Caro, Sewall, Salvante, & Sockman, 2010; Cramer & Price, 2007; DuBois, Nowicki, & Searcy, 2011; Illes, Hall, & Vehrencamp, 2006; Trillo & Vehrencamp, 2005). Evidence is accumulating that variation in

these metrics is salient to conspecifics of both sexes (Ballentine et al., 2004; Caro et al., 2010; DuBois et al., 2011; Illes et al., 2006; Moseley, Lahti, & Podos, 2013). Although VDev has proven to be a useful measure of performance for many species of songbirds, it fails to capture potentially important aspects of vocal performance. For example, VDev does not account for adjustments to the vocal apparatus during silent intervals between notes. It also ignores variation in syllable structure other than bandwidth and duration (Geberzahn & Aubin, 2014). Because VDev does not account for differences among different syllable types, it is not a suitable metric to compare the performance of songs that contain more than one syllable type (Geberzahn & Aubin, 2014; Podos et al., 2016). A new metric of motor performance, frequency excursion (FEX), overcomes these limitations (Podos et al., 2016). Frequency excursion attempts to estimate the rate at which the vocal apparatus adjusts by measuring the rate of change in the fundamental frequency of a signal, including changes during silent gaps.

Here, we measured FEX in a large sample of songs from free-living male Adelaide’s warblers. We then modelled variation in FEX as a function of several variables that may affect performance over short timescales, allowing us to test the hypotheses summarized in Table 1.

## METHODS

### Study System

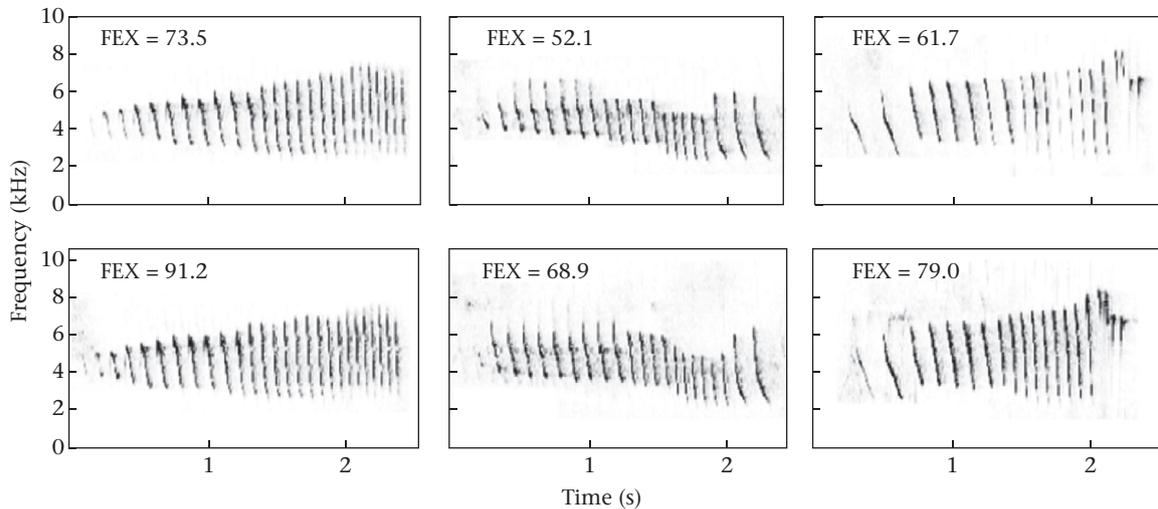
We recorded mated male Adelaide’s warblers at the Cabo Rojo National Wildlife Refuge, Puerto Rico (17°59’N, 67°10’W) during the breeding season between March and June 2012. Adelaide’s warblers are resident wood warblers endemic to Puerto Rico and the neighbouring island of Vieques (Staicer, 1996a; Toms, 2011). Males sing a repertoire of discrete song types (mean = 29.0 song types/male), many of which they share with neighbours (Staicer, 1996b). Songs are frequency-modulated trills, with among-note variation in structure (Fig. 1). Like many other wood warblers, individual Adelaide’s warblers use distinct repertoires for Type I and Type II singing (Staicer, 1996a). In an unpublished song playback study (Pereira, Medina, & Logue, n.d.), male Adelaide’s warblers type-matched low-performance songs (digitally slowed down by 15%) more than controls and type-matched high-performance stimuli (accelerated by 15%) less than controls, indicating that vocal performance is behaviourally salient in this species.

### Ethical Note

This research was approved by the Institutional Animal Care and Use Committee at the University of Puerto Rico at Mayagüez (17 September 2010) and adhered to the ASAB/ABS Guidelines for the use of animals in research. Birds were captured under D.M.L.’s federal bird banding permit (no. 23696). The U.S. Fish and Wildlife Service granted permission to work at the Cabo Rojo Wildlife Refuge (permit 2012-01). Birds were captured in mist nets and fitted with a unique combinations of three coloured leg bands, and

**Table 1**  
Hypotheses and predictions regarding within-day variation in performance of birdsong

Hypothesis	Prediction	Independent variable
Song type specific exhaustion	Performance decreases over consecutive repetitions of a song type	Run number
Song type general exhaustion	Performance increases with latency since the prior song	Latency
Warm up	Performance increases with number of songs sung	Order
Type I singing showcases high performance	Type I songs have higher performance than Type II songs	Singing mode
Vocal interaction	Performance increases when countersinging	Countersinging
Time-dependent factors influence performance	Varied	Time



**Figure 1.** Examples of high- and low-performance ('frequency excursion', FEX) songs. Each column shows songs belonging to the same song type, sung by the same male, but with different FEX. Careful visual inspection shows that higher-performance songs tend to have higher trill rate and/or bandwidth.

one aluminium leg band prior to the onset of the study. At the same time, blood samples were taken for another study. We used a hypodermic needle to puncture the left brachial vein and take a blood sample with a capillary tube. The bleeding was stanching with direct pressure, and standard measurements were taken. No birds were otherwise injured or killed during capture. The birds appeared to forage and sing normally while being recorded, suggesting that our observations did not cause significant distress.

### Song Recordings

We recorded nine colour-banded males for 4 days each, averaging  $3:30 \pm 0:13$  h (mean  $\pm$  SD) of recording per day. Successive recording sessions of a given male were separated by at least 4 days except on two occasions when recordings were made on consecutive days because of logistical constraints. Observations started 30 min before sunrise, allowing us to capture the start of the dawn chorus and thus the ordinal number of each song, and continued until 3 h after sunrise. Although it was too dark to see their coloured leg bands at the beginning of the recording sessions, we are confident we recorded the right individuals because (1) males are highly territorial, (2) males have high fidelity to specific dawn chorus trees, or cluster of trees, and (3) observers followed the birds continuously and confirmed the band combinations once the sun was up. We recorded individuals continuously with a portable solid-state audio recorder (Marantz PMD661) and a directional 'shotgun' microphone (Sennheiser ME67). An observer followed each bird through its territory at a distance that did not appear to disturb the individual. We saved recordings as .wav files (sample rate = 44.1 kHz, 16 bits).

We visualized recordings as spectrograms in Syrinx PC v.2.6f sound analysis software (settings: Blackman window, transform size = 1024 points; J. Burt, <http://www.syrinxpc.com/>). Observers were blind to the identity of focal males. We recorded the occurrence, time and song type for each song. Each song recording from a focal male was saved as a separate file and assessed for recording quality. We only used high-quality recordings (high signal-to-noise ratio, minimal overlap with other sounds) for song measurements.

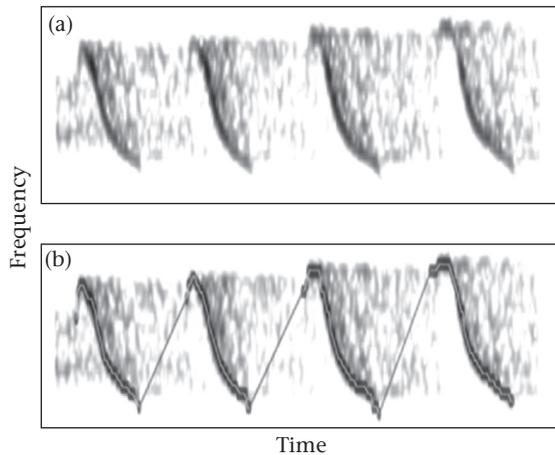
In Adelaide's warblers, Type I and Type II songs differ with respect to several variables including time of day, time of year, patterns of song switching (immediate versus eventual variety), fine-scale structure and social context (Staicer, 1996a, 1996b).

However, there is no published diagnostic criterion to assign song types to a singing mode. We therefore relied on median time of delivery to assign a singing mode to each song type for each male (different individuals can assign a given song type to different singing modes; Staicer, 1996b). We built a histogram of median time of delivery, treating song type within individual as the sampling unit. The histogram revealed a bimodal distribution, with an antimode shortly after sunrise. We assigned song types in the left peak of the distribution to Type II and those in the right peak to Type I for each individual. In doing so, we operationally defined Type II songs as those that were usually sung before sunrise, and Type I songs as those that were usually sung after sunrise. This division is consistent with the literature on wood warbler singing modes (Spector, 1992).

The complete data set contained 9499 song recordings, 2825 of which were of sufficient quality for structural analysis. Prior to analysis, we further reduced the data set by eliminating songs that appeared to be missing sections relative to other songs of the same type, by the same individual,  $N = 42$ , 'double songs' (two songs sung in rapid succession,  $N = 6$ ) and songs belonging to song types with 10 or fewer exemplars ( $N = 50$  songs). The rationale for eliminating rare song types is that the mixed model, which includes song type as a random factor, would have little information with which to estimate their parameters. Finally, we eliminated all first songs of the day (order = 1) because they lacked a latency since prior song ( $N = 10$ ), leaving 2717 songs in the final data set.

### Acoustic Analysis

Frequency excursion was measured with the custom software FEX Calculator (J. McClure, <https://github.com/BehaviorEnterprises/Fex>; Podos et al., 2016; Fig. 2). The software Fourier transforms the audio input creating a matrix of relative amplitude values for each frequency  $\times$  time bin. The highest amplitude point in each time bin (above a selected minimum threshold) is used to create a path through the matrix tracing the peak frequency across time. The total length of this path divided by its duration is the measure of frequency excursion. The amplitude matrix data are also used to create an interactive spectrogram allowing the scorer, when necessary, to remove background noise with an eraser tool that masks the selected frequency  $\times$  time bin, preventing it from being



**Figure 2.** Schematic representation of semiautomated frequency excursion measurement. First, a spectrogram is generated (a), here with a window size of 256 and a bin size of 64 samples, originating from a 44.1 kHz recording. This spectrogram is overlaid with a layer indicating the points with the highest amplitude in each time bin and drawing the frequency excursion path line (b). The length of this connecting line, divided by the time it spans, represents the frequency excursion (FEX). Higher FEX values indicate higher-performance songs.

included in the path without requiring any changes to the original audio signal.

### Statistical Analysis

We fitted a linear mixed model to examine variation in FEX and test for potential effects of warm up, exhaustion, rest, singing mode and countersinging. The following fixed variables were included in the model: time of day, order, latency, run number, singing mode and countersinging (Table 1). Time of day (*s*) represents the time relative to sunrise (positive values are after sunrise). Order is the sequential order of the song for a given male on a given day (range 2–474). For example, order = 1 for the first song that a male sings in the morning, order = 2 for the second song, and so on. Latency (*s*) is the time elapsed since the subject's previous song. This variable was strongly right-skewed, so we log transformed it to prevent the extreme values from exerting excessive leverage in our model. Run number is the song's order within a run of the same song type (range 1–43). Countersinging was scored as '1' if any of the following were true: the song was sung within 1.5 s following a neighbour's song, the song matched the song type sung by a neighbour in the previous 2 s, or the song temporally overlapped a neighbour's song (separating these three kinds of countersinging did not affect the outcome; Pereira et al., n.d.). An alternative way to measure performance changes during countersinging is to ask whether performance changes over the course of a countersinging bout. We defined a bout of countersinging as an uninterrupted string of songs for which at least five of the previous 10 songs were possible responses to a neighbour (i.e. were scored as a '1'). We scored the order of songs within a bout of countersinging to define the variable 'countersinging order', which we substituted for 'countersinging' in a separate model (including countersinging and countersinging order in the same model would be inappropriate given their high degree of correlation). Singing mode separates Type I and Type II singing (Stacer, 1996b; see above). Random variables in the model were ID, day (within ID) and song type (song type is known to affect performance in other species: Cardoso et al., 2012; Cardoso et al., 2009; Cramer & Price, 2007; Trillo & Vehrencamp, 2005).

Linear mixed models were developed in R v.3.2.0 (R Core Team, 2014) using the 'lme4' package (Bates, Maechler, Bolker, & Walker,

2015). Continuous predictors (time of day, order, run number, latency) were mean-centred and standardized prior to analysis by subtracting the mean and dividing by two standard deviations, in order to make effect sizes comparable to those of binary predictors (Gelman, 2008). We began with a full model that included all main effects and two-way interactions, and random intercepts for ID, day (within ID) and song type. We wanted to identify which of these potential explanatory variables were strongly associated with song performance while minimizing the risk of type I errors, so we implemented a conservative model selection procedure. The full model was subjected to the 'dredge' function in the R package 'MuMIn' (Bartoń, 2015) to rank all submodels by Akaike's Information Criterion, AIC (Burnham & Anderson, 2002). Variables that were in all models with  $\Delta\text{AIC} \leq 2$  were included in the reduced model (the full model and a model-averaging procedure produced qualitatively similar results, indicating that the results were robust to variation in model selection procedures). We tested for a quadratic effect of time of day (see Results, Fig. 3b), but found no support, so only linear effects were included in the final model. We then added random slopes for all retained fixed variables (Barr, Levy, Scheepers, & Tily, 2013) and examined residuals for normality and homoscedasticity. We used the 'anova' function to conduct likelihood ratio tests of statistical significance of random and fixed effects. We used the REML method for parameter estimation and the maximum likelihood method for significance tests.

## RESULTS

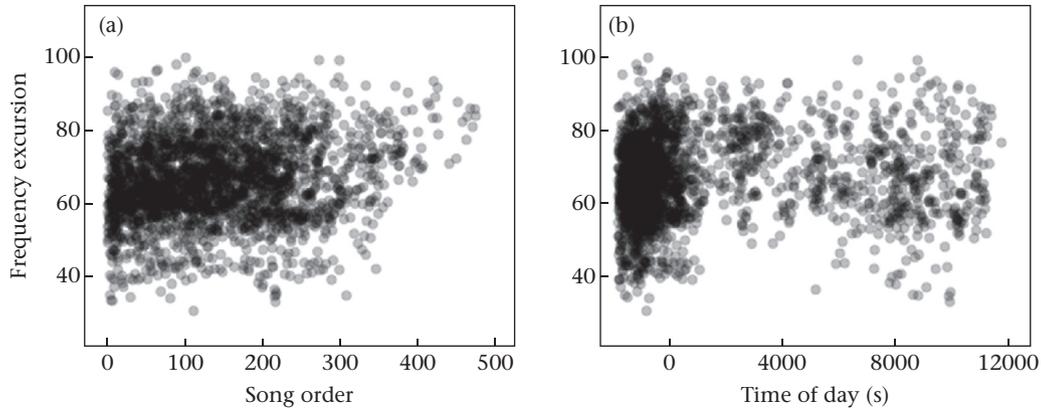
The 2717 songs in the data set included 40 unique song types. Each subject contributed an average  $\pm$  SD of  $301.89 \pm 179.01$  songs, representing  $16.67 \pm 3.16$  song types (Supplementary Table S3). FEX varied among individuals, days and song types (Supplementary Tables S5–S7). Independent variables were intercorrelated, including time of day and order (Pearson's correlation:  $r = 0.68$ ; Supplementary Table S4). No correlations exceeded  $|r| = 0.7$ , limiting the risk of multicollinearity (Dormann et al., 2013).

### Linear Mixed Model

The model selection procedure produced the following model, which we refer to as the 'main model':  $\text{FEX} \sim \text{time} + \text{order} + (1 + \text{time} + \text{order}|\text{ID}) + (1 + \text{time} + \text{order}|\text{ID}:\text{day}) + (1 + \text{time} + \text{order}|\text{type})$ . That is, the main model explained variation in FEX as a function of the fixed effects of time of day and order, with random intercept and slopes for both time of day and order versus ID, day within ID, and song type. The independent variables latency, run number, countersinging, countersinging order and singing mode were not retained by the model selection procedure. FEX increased with order and decreased with time of day (Table 2, Fig. 3). The random variables ID, day within ID and song type were all highly significant ( $P < 0.0001$ ; Supplementary Tables S5–S7).

We found a positive effect of order on performance for all individuals, every day, for almost every song type (Fig. 4a, Supplementary Tables S5–S7). Translating the effects into a biologically relevant scale, we estimate that birds' performance improves by 5.78 FEX points, on average, over the course of 200 songs (subjects sang  $\geq 200$  songs in 27 of 36 observation periods). The among-individual standard deviation in average performance was 4.11 FEX points, so the effect of singing 200 songs was equivalent to 1.41 standard deviations of the among-individual average.

To further address the potential confound between the correlated variables of time of day and order, we conducted a linear regression of time of day versus order (order as the dependent variable) and ran the residuals in our main model. Residual order

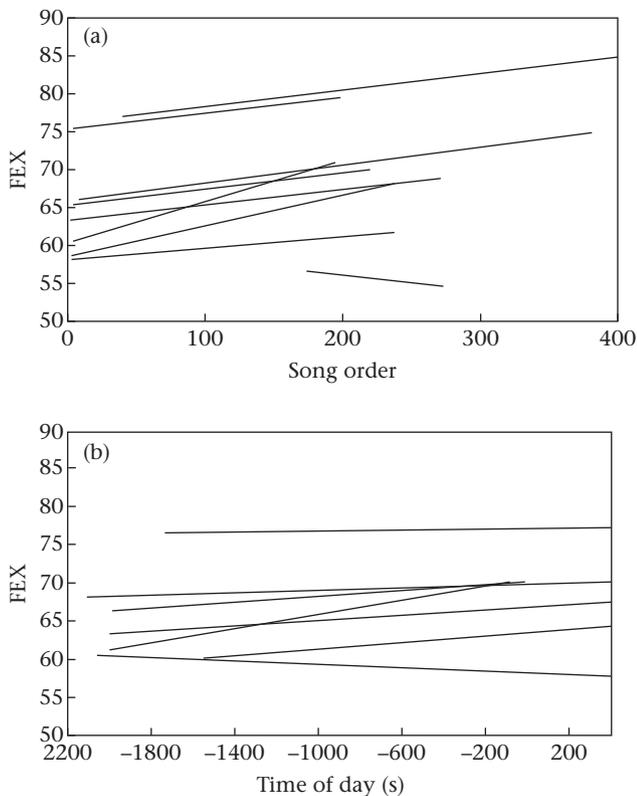


**Figure 3.** Frequency excursion (FEX) as a function of (a) song order ( $r = 0.16$ ) and (b) time relative to sunrise ( $r = 0.10$ ). Points are semitransparent, so darker regions indicate overlapping data points.

**Table 2**  
Results of a mixed effects model for the dependent variable ‘frequency excursion’

Term	Estimate	$\chi^2_1$	$P$
Intercept	66.48		
Order	5.43	11.39	0.0007
Time of day	-2.99	2.45	0.12

Covariate estimates refer to scaled data. Random intercepts:  $\sigma_{ID} = 4.04$ ,  $\sigma_{Day\ w/in\ ID} = 3.03$ ,  $\sigma_{Song\ type} = 9.89$ , all  $P < 0.0001$ ;  $\sigma_{Residual} = 5.06$ . Random slopes: order:  $\sigma_{ID} = 1.89$ ,  $\sigma_{Day\ w/in\ ID} = 3.53$ ,  $\sigma_{Song\ type} = 3.31$ ; time of day:  $\sigma_{ID} = 2.57$ ,  $\sigma_{Day\ w/in\ ID} = 5.02$ ,  $\sigma_{Song\ type} = 6.58$ .



**Figure 4.** Regression lines showing the relationship between vocal performance (frequency excursion, FEX), (a) song order and (b) time relative to sunrise for the most common song type from each of the nine subjects.

was a significant positive predictor of FEX (effect size = 4.69,  $\chi^2_1 = 7.34$ ,  $P = 0.0067$ ).

FEX relies on a linear frequency scale, but production and perception of sound frequency is better modelled on a log scale (Cardoso, 2013). If song duration covaries with order, and if long songs tend to include more high fundamental frequencies than do short songs, our model would overestimate the salience of covariation between FEX and order to signal receivers. To test this possibility, we used linear regression to model FEX as a function of song duration (ms) and average mid-frequency (the midpoint of the average maximum frequency and average minimum frequency) and used the residuals as the dependent variable in our main model. The model with residual FEX estimated similar effects of time of day (-2.31) and order (4.69) as did our main model. As in the main model, the effect of order was statistically significant ( $\chi^2_1 = 7.34$ ,  $P = 0.007$ ).

We were interested in the specific structural changes responsible for the effects of time of day and order on FEX, so we ran variants of the main model with each of the following dependent variables: trill rate (number of notes/duration), frequency bandwidth (Hz, averaged over all notes), number of notes, duration (s), minimum frequency (Hz, averaged over all notes), maximum frequency (Hz, averaged over all notes) and vocal deviation (Supplementary Table S1, see Supplementary Material for methodological details). The only metrics that showed a statistically significant relationship with order were trill rate and number of notes, both of which had positive coefficients (Table 3). We then regressed FEX on trill rate and ran the residuals in the main model to determine whether order affected FEX after accounting for the covariation between FEX and trill rate. The effect of order on the residuals of FEX was positive and significant (effect size = 3.78,  $\chi^2_1 = 6.39$ ,  $P = 0.011$ ).

**DISCUSSION**

We found strong evidence that song motor performance (FEX) improves over the course of morning singing in male Adelaide’s warblers (Table 2, Fig. 3). This effect was not explained by the time of day, which we would expect if factors like air temperature, amount of time that a bird has been awake, or increase in social interactions caused the improvement. Rather, the cumulative number of songs that a bird had sung during the morning explained the observed improvement in performance on a given day. We therefore conclude that Adelaide’s warblers warm up during morning singing. The magnitude of the order effect (+1.41 SD of among-individual variation over 200 songs) is likely to be

**Table 3**  
Estimates for the effects time of day and order on seven song structure metrics

Song metric	Fixed effect	
	Time	Order
Average frequency bandwidth (kHz)	−0.02	−0.06
Trill rate (Hz)	0.03	0.72*
Number of notes	−1.02	3.11**
Duration (ms)	−56.30	93.66
Average minimum frequency (kHz)	−0.07	0.10
Average maximum frequency (kHz)	−0.09	0.07
Vocal deviation	0.02	−0.03

\* $P < 0.01$ ; \*\* $P < 0.001$ .

biologically relevant if receivers compare performance among males. Controlling for song duration and mid-frequency had little effect on the relationship between order and performance, indicating that receivers are likely to perceive the warm-up effect in spite of the nonlinearity of sound frequency production and perception (Cardoso, 2013). As birds warm up they might choose to sing higher-performance song types or they might continue to sing the same song types, but with higher performance. Our analyses do not address the first possibility, but the performance of almost all song types (39/40) improved with song order (Supplementary Table S7).

This is the first report of increased song performance as a function of recent practice, but other studies have considered changes in performance over longer timescales. Banded wrens and swamp sparrows show an increase in performance (VDev) with age, and banded wrens increase their trill rates over the course of a single season (Ballentine, 2009; Vehrencamp et al., 2013). Similarly, frequency bandwidth increases with age in Java sparrows, *Lonchura oryzivora* (Ota & Soma, 2014). One study, however, found that song performance decreases between successive years in white-crowned sparrows, *Zonotrichia leucophrys* (Poesel & Nelson, 2015). We conclude that changes in performance can take place over a range of timescales, from minutes to years.

The performance metric frequency excursion is a function of trill rate, frequency bandwidth and other structural variables. We tested several such variables in an attempt to better understand the positive influence of song order on FEX (Table 3). Trill rate and especially the number of notes in the song, were positively influenced by song order. We found a positive effect of order when controlling for the effects of trill rate on FEX, which means that order affects additional components of FEX beyond just trill rate. Candidate components include frequency modulation within notes and frequency jumps between notes (Geberzahn & Aubin, 2014). The effects of song order on song duration, average frequency bandwidth and average minimum and maximum frequencies were not significant. As Adelaide's warblers warm up, they add more notes and sing faster, but they do not greatly alter the frequency characteristics of their songs. The performance metric VDev trended in the expected direction (lower VDev corresponds to higher performance). Its small effect size is probably attributable to the fact that frequency bandwidth influences VDev much more strongly than does trill rate in this population (see Supplementary Material).

The mechanisms underlying the warm-up effect are not known. In humans, warming up has an especially pronounced effect on sprint performance (Yaicharoen, Wallman, Bishop, & Morton, 2012) and maximal muscle performance (Wittekind, Cooper, Elwell, Leung, & Beneke, 2012) but no detectable effect on submaximal muscle performance (McCrary, Halaki, Sorkin, & Ackermann, 2015). Warm up improves perceived tone quality and may influence vibrato rate in human singing (Amir et al., 2005; Moorcroft &

Kenny, 2012, 2013). It remains to be determined whether similar physiological mechanisms underlie human and avian vocal warm up.

The adaptive significance of vocal performance in Adelaide's warbler is not known. Here, however, we speculate on how warming up and female choice for high performance might interact to affect the evolution of singing behaviour. Many birds, including Adelaide's warblers, participate in the 'dawn chorus', during which song rates are elevated. There are several proposed explanations for the dawn chorus, including the hypothesis that males are advertising to females who compare potential mating partners at dawn (reviewed in Catchpole & Slater, 2008, pp. 128–135). According to the warm-up hypothesis, males that begin singing at a high rate early in the morning would sing with higher performance than males that begin later or sing at a lower rate, all else equal. If females choose males by comparing their singing performances, selection would favour earlier and more intense dawn singing, spurring an evolutionary arms race for ever earlier and more frequent dawn singing. This hypothesis emphasizes the importance of constraints in shaping the evolution of behavioural phenomena.

Song performance decreased with time of day, suggesting that, when accounting for order, performance decreases toward the later hours of the morning. The fixed variable of time of day was not statistically significant in the final model because there was a lot of variation in the slopes of time of day versus FEX among levels of ID, day and song type (Supplementary Tables S5–S7). Nevertheless, the AIC-based model selection procedure (conducted before we added random slopes) included time of day in all of the best models, suggesting that it is important for explaining variation in FEX. Unmeasured biotic or abiotic factors that correlate with time (air temperature, amount of time the bird has been awake, rate of social interactions) could affect performance, or time of day may capture the effect of singing mode better than the mode variable. Finally, it is possible that the recordists tended to be farther away from the birds later in the day, perhaps because the birds fly longer distances in the late morning. If this were the case, later recordings might tend to miss high-frequency components of songs, which attenuate with distance more severely than do low-frequency components. The negative relationship between time and maximum frequency would seem to support this hypothesis, but that effect was small and nonsignificant (Table 3).

The positively correlated variables order and time of day ( $r = 0.68$ ) exerted opposite effects on FEX in our model, raising the question of whether order has a positive influence on song performance that is independent of time of day. Critically, order and time were not perfectly correlated ( $r < 1.0$ ) because birds did not sing at perfectly regular temporal intervals. This fact allowed us to separate out variation in song order that could not be attributed to variation in time and to show that this variable ('residual order'), which was statistically independent of time of day, explained substantial variation in FEX. We conclude that the positive effect of order on FEX was not a collinearity artefact.

Song performance did not change in a consistent direction over repeated renditions of the same song type, failing to support a key prediction of the song-type specific exhaustion hypothesis (Lambrechts & Dhondt, 1988). Birds may switch to a different song type before exhaustion occurs, but our data did not allow us to test that idea. To our knowledge, this is the first test of the anti-exhaustion hypothesis that quantifies the effect of song type repetition on song performance. A previous test of this hypothesis also failed to find support: singing rate and song type switching rate in chaffinches did not differ between males with only one song type and males with a repertoire of several song types (Brumm et al., 2009). The time elapsed since a subject's previous

song (latency) did not affect FEX, failing to support the song-type general exhaustion hypothesis.

We also found no support for the hypothesis that songs differ in performance based on singing mode. Staicer (1996a) found that Type I and Type II songs were structurally distinct in Adelaide's warblers, but she did not measure performance. In the congeneric American yellow warbler, *Setophaga petechia*, Type I songs are higher performance (VDev) than Type II songs (Beebee, 2004a, 2004b); however, in the light of the current study, this result could have been attributable to warm-up effects. Our finding comes with the caveat that there is no definitive method for discriminating singing modes in our focal species, so our simple classification scheme may not have accurately separated the two modes. We are investigating this topic further. We also failed to find an effect of vocal interaction on song performance. We interpret that finding to mean that Adelaide's warblers do not significantly alter their vocal performance in a consistent direction during counter-singing. Playback experiments will be required to determine whether performance is affected by specific social contexts (e.g. the presence of a male or female intruder).

Performance varied significantly among individuals, song types and days. Variation in performance among song types was much greater than variation in performance among individuals, highlighting the importance of controlling for song type in performance analyses (Cardoso et al., 2009). The evolutionary maintenance of song types with low FEX scores could be explained in several ways. First, there are aspects of performance that FEX does not capture, such as song rate (Lambrechts & Dhondt, 1987), amplitude (Ritschard, Riebel, & Brumm, 2010), or consistency (Byers, 2007). Some low-FEX song types may be high performance with respect to one or more unmeasured metrics. Second, some functions of song (e.g. advertisement of territory occupancy) might not require maximal performance, so selection may favour the use of low-performance songs, perhaps because they are less taxing. Third, some birds could be sufficiently constrained, at least some of the time, that they are incapable of producing high-performance song types. These birds might make the best of a bad situation by resorting to low-performance song types. Fourth, low-performance songs might function as a first stage in escalating to higher-performance songs during aggressive interactions. Individuals' performance differed between recording days. This effect could be attributable to breeding phase, a long-term practice effect, or idiosyncratic variation in, for example, nutritional stress or temperature (Ballentine, 2009; Vehrencamp et al., 2013).

We encourage further tests for warm-up effects in birdsong and other animal displays. Given the strength of the effect in this study, efforts to accurately measure and compare song performance should account for possible order effects. This requires continuous observation, beginning with the first display of the day. Extending observation periods later in the day or over a longer segment of the year would reveal whether the effect of order tapers off or reverses later in the day, and whether its strength is affected by breeding phase.

## Acknowledgments

Paloma Sánchez-Jaureguí, Fabio L. Tarazona, Jorge Illanas, Krystal Medina and Alfredo Lamela helped score data and provided valuable insight during discussions. Bailey Parker, Tony Shlakoff, Casandra Logue, Joshua Báez, Ruben Irizarry and Alicia García processed song recordings. We are grateful to John Vokey and Tyler Bonnell for their advice on statistical analyses. We also thank Jean-Baptiste Leca, Shannon Whelan, the University of Lethbridge Bird-song Reading Group, Jeff Podos and two anonymous referees for insightful comments on the manuscript. Work was funded in part

by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-06553) to D.M.L.

## Supplementary Material

Supplementary material associated with this article is available, in the online version, at <http://dx.doi.org/10.1016/j.anbehav.2016.12.026>.

## References

- Amir, O., Amir, N., & Michaeli, O. (2005). Evaluating the influence of warmup on singing voice quality using acoustic measures. *Journal of Voice*, *19*, 252–260. <http://dx.doi.org/10.1016/j.jvoice.2004.02.008>.
- 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. <http://dx.doi.org/10.1016/j.anbehav.2008.12.027>.
- Ballentine, B., Hyman, J., & Nowicki, S. (2004). Vocal performance influences female response to male bird song: An experimental test. *Behavioral Ecology*, *15*, 163–168. <http://dx.doi.org/10.1093/beheco/arg090>.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. <http://dx.doi.org/10.1016/j.jml.2012.11.001>.
- Barton, K. (2015). *MuMIn: Mul4 – Model inference* (R package version 1.15.1). Vienna, Austria: R Foundation for Statistical Computing.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. <http://dx.doi.org/10.18637/jss.v067.i01>.
- Beebee, M. (2004a). Variation in vocal performance in the songs of a wood-warbler: Evidence for the function of distinct singing modes. *Ethology*, *110*, 531–542. <http://dx.doi.org/10.1111/j.1439-0310.2004.00994.x>.
- Beebee, M. (2004b). The functions of multiple singing modes: Experimental tests in yellow warblers, *Dendroica petechia*. *Animal Behaviour*, *67*, 1089–1097. <http://dx.doi.org/10.1016/j.anbehav.2003.05.016>.
- Benedict, L., Rose, A., & Warning, N. (2012). Canyon wrens alter their songs in response to territorial challenges. *Animal Behaviour*, *84*, 1463–1467. <http://dx.doi.org/10.1016/j.anbehav.2012.09.017>.
- Bradbury, J. W., & Vehrencamp, S. L. (2011). *Principles of animal communication*. Sunderland, MA: Sinauer.
- Brumm, H., Lachlan, R. F., Riebel, K., & Slater, P. J. B. (2009). On the function of song type repertoires: Testing the 'antiexhaustion hypothesis' in chaffinches. *Animal Behaviour*, *77*, 37–42. <http://dx.doi.org/10.1016/j.anbehav.2008.09.009>.
- Burnham, K., & Anderson, D. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer.
- Burt, J., & Vehrencamp, S. L. (2005). Dawn chorus as an interactive communication network. In P. K. McGregor (Ed.), *Animal communication networks* (pp. 320–343). Cambridge, U.K.: Cambridge University Press.
- Byers, B. E. (2007). Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. *Behavioral Ecology*, *18*, 130–136. <http://dx.doi.org/10.1093/beheco/arl058>.
- Cardoso, G. C. (2013). Using frequency ratios to study vocal communication. *Animal Behaviour*, *85*, 1529–1532. <http://dx.doi.org/10.1016/j.anbehav.2013.03.044>.
- Cardoso, G., Atwell, J., Hu, Y., Ketterson, E., & Price, T. (2012). No correlation between three selected trade-offs in birdsong performance and male quality for a species with song repertoires. *Ethology*, *118*, 584–593. <http://dx.doi.org/10.1111/j.1439-0310.2012.02047.x>.
- Cardoso, G., Atwell, J., Ketterson, E., & Price, T. (2009). Song types, song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). *Behavioral Ecology*, *20*, 901–907. <http://dx.doi.org/10.1093/beheco/arp079>.
- Caro, S., Sewall, K., Salvante, K., & Sockman, K. (2010). Female Lincoln's sparrows modulate their behavior in response to variation in male song quality. *Behavioral Ecology*, *21*, 562–569. <http://dx.doi.org/10.1093/beheco/araq022>.
- Catchpole, C., & Slater, P. (2008). *Bird song: Biological themes and variations*. Cambridge, U.K.: Cambridge University Press.
- Cramer, E., & Price, J. (2007). Red-winged blackbirds *Agelaius phoeniceus* respond differently to song types with different performance levels. *Journal of Avian Biology*, *38*, 122–127. <http://dx.doi.org/10.1111/j.2006.0908-8857.03839.x>.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*, 27–46. <http://dx.doi.org/10.1111/j.1600-0587.2012.07348.x>.
- DuBois, A., Nowicki, S., & Searcy, W. (2009). Swamp sparrows modulate vocal performance in an aggressive context. *Biology Letters*, *5*, 163–165. <http://dx.doi.org/10.1098/rsbl.2008.0626>.
- DuBois, A., Nowicki, S., & Searcy, W. (2011). Discrimination of vocal performance by male swamp sparrows. *Behavioral Ecology and Sociobiology*, *65*, 717–726. <http://dx.doi.org/10.1007/s00265-010-1073-2>.
- DuVal, E. H. (2012). Variation in annual and lifetime reproductive success of lance-tailed manakins: Experience as an alpha offsets senescence in siring success. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1551–1559. <http://dx.doi.org/10.1098/rspb.2011.1840>.

- Geberzahn, N., & Aubin, T. (2014). Assessing vocal performance in complex bird-song: A novel approach. *BMC Biology*, *12*, 58. <http://dx.doi.org/10.1186/s12915-014-0058-4>.
- Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine*, *27*, 2865–2873. <http://dx.doi.org/10.1002/sim.3107>.
- Illes, A., Hall, M., & Vehrencamp, S. L. (2006). Vocal performance influences male receiver response in the banded wren. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1907–1912. <http://dx.doi.org/10.1098/rspb.2006.3535>.
- Lambrechts, M., & Dhondt, A. (1987). Differences in singing performance between male great tits. *Ardea*, *75*, 43–53.
- Lambrechts, M., & Dhondt, A. (1988). The anti-exhaustion hypothesis: A new hypothesis to explain song performance and song switching in the great tit. *Animal Behaviour*, *36*, 327–334. [http://dx.doi.org/10.1016/S0003-3472\(88\)80002-2](http://dx.doi.org/10.1016/S0003-3472(88)80002-2).
- McCrary, J. M., Halaki, M., Sorkin, E., & Ackermann, B. J. (2015). Acute warm-up effects in submaximal athletes: An EMG study of skilled violinists. *Medicine & Science in Sports & Exercise*, *48*, 307–315. <http://dx.doi.org/10.1249/MSS.00000000000000765>.
- Moorcroft, L., & Kenny, D. (2012). Vocal warm-up produces acoustic change in singers' vibrato rate. *Journal of Voice*, *26*, 667.e13–667.e18. <http://dx.doi.org/10.1016/j.jvoice.2011.10.007>.
- Moorcroft, L., & Kenny, D. (2013). Singer and listener perception of vocal warm-up. *Journal of Voice*, *27*, 258.e1–258.e13. <http://dx.doi.org/10.1016/j.jvoice.2012.12.001>.
- Moseley, D., Lahti, D., & 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*, *281*, 20131401. <http://dx.doi.org/10.1098/rspb.2013.1401>.
- Motel, T., Fisher, K., & Leydon, C. (2003). Vocal warm-up increases phonation threshold pressure in soprano singers at high pitch. *Journal of Voice*, *17*, 160–167. [http://dx.doi.org/10.1016/S0892-1997\(03\)00003-1](http://dx.doi.org/10.1016/S0892-1997(03)00003-1).
- Ota, N., & Soma, M. (2014). Age-dependent song changes in a closed-ended vocal learner: Elevation of song performance after song crystallization. *Journal of Avian Biology*, *45*, 566–573. <http://dx.doi.org/10.1111/jav.00383>.
- Pereira, D.A., Medina, O., & Logue, D.M. (n.d.). Experimental evidence that a songbird matches song-types to manage alignability, not to signal aggressive intent. Manuscript in preparation.
- Podos, J. (1997). A performance constraint of the evolution of trilled vocalizations in a songbird family (Passeriformes: Emberizidae). *Evolution*, *51*, 537–551. <http://dx.doi.org/10.2307/2411126>.
- Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's finches. *Nature*, *409*, 185–188. <http://dx.doi.org/10.1038/35051570>.
- Podos, J., Moseley, D., Goodwin, S., McClure, J., Taft, B., Strauss, A., et al. (2016). A fine-scale, broadly applicable index of vocal performance: Frequency excursion. *Animal Behaviour*, *116*, 203–213. <http://dx.doi.org/10.1016/j.anbehav.2016.03.036>.
- Podos, J., & Patek, S. (2015). Acoustic signal evolution: Biomechanics, size, and performance. In D. Irshick, M. Briffa, & J. Podos (Eds.), *Animal signaling and function: An integrative approach* (pp. 175–204). New York, NY: John Wiley.
- Poesel, A., & Nelson, D. A. (2015). Changes in performance of shared and unshared songs within and between years in the white-crowned sparrow. *Ethology*, *121*, 1–11. <http://dx.doi.org/10.1111/eth.12399>.
- Price, J., Earnshaw, S., & Webster, M. (2006). Montezuma oropendolas modify a component of song constrained by body size during vocal contests. *Animal Behaviour*, *71*, 799–807. <http://dx.doi.org/10.1016/j.anbehav.2005.05.025>.
- R Core Team. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. [www.R-project.org/](http://www.R-project.org/).
- Ritschard, M., Riebel, K., & Brumm, H. (2010). Female zebra finches prefer high-amplitude song. *Animal Behaviour*, *79*, 877–883. <http://dx.doi.org/10.1016/j.anbehav.2009.12.038>.
- Spector, D. (1992). Wood-warbler song systems: A review of paruline singing behaviors. In D. Power (Ed.), *Current ornithology* (Vol. 9, pp. 199–238). New York, NY: Plenum.
- Staicer, C. (1996a). Acoustical features of song categories of the Adelaide's warbler (*Dendroica adelaidae*). *American Naturalist*, *113*, 771–783.
- Staicer, C. (1996b). Honest advertisement of pairing status: Evidence from a tropical resident wood-warbler. *Animal Behaviour*, *51*, 375–390. <http://dx.doi.org/10.1006/anbe.1996.0036>.
- Staicer, C., Spector, D., & Horn, A. (1996). The dawn chorus and other diel patterns in acoustic signaling. In D. E. Kroodsma, & E. H. Miller (Eds.), *Ecology and evolution of acoustic communication in birds* (pp. 426–453). Ithaca, NY: Comstock.
- Stewart, D., Macaluso, A., & De Vito, G. (2003). The effect of an active warm-up on surface EMG and muscle performance in healthy humans. *European Journal of Applied Physiology*, *89*, 509–513. <http://dx.doi.org/10.1007/s00421-003-0798-2>.
- Suthers, R. (2004). How birds sing and why it matters. In P. Marler, & H. Slabbekoorn (Eds.), *Nature's music: The science of birdsong* (pp. 272–295). New York, NY: Elsevier.
- Toms, J. (2011). *Non-breeding competition between migrant American redstarts (Setophaga ruticilla) and resident Adelaide's warblers (Dendroica adelaidae) in the Guánica Biosphere Reserve, southwest Puerto Rico* (Ph.D. Dissertation). Columbia, MO: University of Missouri.
- Trillo, P. A., & Vehrencamp, S. L. (2005). Song types and their structural features are associated with specific contexts in the banded wren. *Animal Behaviour*, *70*, 921–935. <http://dx.doi.org/10.1016/j.anbehav.2005.02.004>.
- Vehrencamp, S. L., Yantachka, J., Hall, M., & de Kort, S. (2013). Trill performance components vary with age, season, and motivation in the banded wren. *Behavioral Ecology and Sociobiology*, *67*, 409–419. <http://dx.doi.org/10.1007/s00265-012-1461-x>.
- Wittekind, A., Cooper, C. E., Elwell, C. E., Leung, T. S., & Beneke, R. (2012). Warm-up effects on muscle oxygenation, metabolism and sprint cycling performance. *European Journal of Applied Physiology*, *112*, 3129–3139. <http://dx.doi.org/10.1007/s00421-011-2262-z>.
- Yaicharoen, P., Wallman, K., Bishop, D., & Morton, A. (2012). The effect of warm up on single and intermittent-sprint performance. *Journal of Sports Sciences*, *30*, 833–840. <http://dx.doi.org/10.1080/02640414.2012.675084>.