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

2

3   Motor performance describes the vigour or skill required to perform a particular display. It is a

4   behaviourally salient variable in birdsong and other animal displays, but little is known about

5   within-individual variation in performance over short time scales. The metric "frequency

6   excursion" (FEX) quantifies birdsong performance as cumulative frequency modulation per unit

7   time. We measured FEX in a large sample of recordings from free-living male Adelaide's

8   warblers (*Setophaga adelaidae*). Our objectives were to quantify natural variation in

9   performance, and test the hypotheses that performance (1) improves as a function of recent

10   practice, (2) decreases over consecutive repetitions of a single song-type, (3) improves with rest

11   between songs, (4) varies by singing mode, and (5) changes during vocal interactions with

12   neighbours. We found significant variation in performance among individuals and song-types.

13   Consecutive repetition of a song-type, rest between songs, singing mode, and vocal interaction

14   did not strongly affect performance. Performance consistently increased with song order,

15   however, indicating that males warm up during morning singing. This is the first demonstration

16   of such an effect in a sexual display. The warm up effect may explain the prevalence of intense

17   dawn singing in birds (dawn chorus), if rivals engage in an arms race to warm up.

18

21

22

23   Sexual selection drives the elaboration of sexual ornaments and displays in animals

24   (Bradbury & Vehrencamp 2011). As these traits evolve to become increasingly extreme, costs

25   accumulate and constraints take effect, limiting further elaboration. Motor constraints may be

26   particularly important in limiting the evolution of sexual displays (Podos & Patek 2015), as

27   suggested by systems in which the display performance improves with experience and age.

28   Examples include the cartwheel displays of lance-tailed manakins (*Chiroxiphia lanceolata*;

29   DuVal 2012); and trill performance in swamp sparrows (*Melospiza georgiana*; Ballentine 2009)

30   and banded wrens (*Thryophilus pleurostictus*; Vehrencamp et al. 2013). Although it has not been

31   studied as intensively as year-to-year variation in performance, within-day variation in

32   performance may also provide evidence of performance constraints. In the present study, we

33   examine the factors that influence short term variation in performance in a neotropical songbird,

34   Adelaide's warblers (*Setophaga adelaidae*).

35   Birdsong production requires precise coordination of the intricate avian vocal apparatus,

36   so it is likely that the evolution of birdsong has been affected by motor constraints (Suthers

37   2004). The anti-exhaustion hypothesis proposes that motor fatigue limits birds' ability to

38   repeatedly produce the same song-type, but birds can escape fatigue by switching to a new song-

39   type (Lambrechts & Dhondt 1988). A test of this hypothesis in chaffinches (*Fringilla coelebs*)

40   found no support (Brumm et al. 2009). We propose a novel hypothesis about the effects of motor

41   constraints over short time scales: signal performance could improve over short time scales if

42   animals 'warm up'. Improvement in a motor task due to recent practice, or 'warming up,' is

43   known to affect human athletes and singers (Stewart et al. 2003, Motel et al. 2003, Amir et al.

44   2005, Moorcroft & Kenny 2013). The warm up hypothesis predicts that singing performance

45   improves with recent practice, regardless of song-type. We quantify recent practice with the

46   variable song *order*, which describes the number of songs the bird has already sung on the focal

47   morning. Throughout this report, predictor variables are indicated by the use of italics.

48   Superficially, the warm up hypothesis may seem to be at odds with the anti-exhaustion

49   hypothesis. We suggest that the two hypotheses are not mutually exclusive, however, because

50   birds that switch song-types might warm up, even though they would become exhausted if they

51   were to continue singing the same song-type. Thus, we invoke the anti-exhaustion hypothesis to

52   predict that performance will decrease over consecutive repetitions of a given song-type

53   (measured as the ordinal number of a song in a run of same-type songs, or the *run number*).

54   Alternatively, singing may temporarily exhaust a bird's resources regardless of song-type. This

55   'song-type general exhaustion' hypothesis leads to the prediction that performance will covary

56   positively with the *latency* since the prior song.

57   New World warblers (Family: Parulidae), including Adelaide's warblers, early morning

58   (Type II) singing differs from daytime (Type I) singing during the breeding season (Spector

59   1992, Staicer et al. 1996, Staicer 1996a, Catchpole & Slater 2008, Burt & Vehrencamp 2005).

60   Distinct functions for these two singing modes have been proposed. A study showing that Type I

61   songs have higher performance than Type II songs suggests that Type I singing may have

62   evolved to showcase performance (Beebee 2004). We therefore tested whether *singing mode*

63   affects song performance.

64   Several studies on natural and simulated territorial interactions show that songbirds adjust

65   their performance level when they are vocally interacting ('counter-singing') with a potential

66   rival (Trillo & Vehrencamp 2005, Price et al. 2006, DuBois et al. 2009, Benedict et al. 2012).

67   We tested this hypothesis by estimating the effect of *counter-singing* on song performance.

68   Additionally, performance may be affected by various time-dependent factors, such as the air

69   temperature, the amount of time the bird has been awake, or the rate of social interactions. We

70   included *time of day* in our analysis, to control for these potential influences on performance, and

71   to account for covariance between time and *order*. Table 1 summarises our predictions for each

72   hypothesis.

73   Motor performance in trilled birdsong has traditionally been quantified by measuring the

74   orthogonal distance of a song from the upper-bound regression line of a plot of trill rate and

75   frequency bandwidth (Podos 1997). Using this measure of performance (termed 'vocal

76   deviation', VDev), and its components trill rate and frequency bandwidth, studies have revealed

77   variation in performance among species (Podos 1997, Podos 2001), individuals (Podos 2001,

78   Ballentine et al. 2004), contexts (Trillo & Vehrencamp 2005, Price et al. 2006, DuBois et al.

79   2009, Benedict et al. 2012), and song-types (Trillo & Vehrencamp 2005, Illes et al. 2006,

80   Cramer and Price 2007, Cardoso et al. 2009, Caro et al. 2010, DuBois et al. 2011, Cardoso et al.

81   2012). Evidence is accumulating that variation in these metrics is salient to conspecifics of both

82   sexes (Ballentine et al. 2004, Illes et al. 2006, Caro et al. 2010, DuBois et al. 2011, Moseley et al.

83   2013). Although VDev has proven to be a useful measure of performance for many species of

84   songbirds, it fails to capture potentially important aspects of vocal performance. For example,

85   VDev does not account for adjustments to the vocal apparatus during silent intervals between

86   notes. It also ignores variation in syllable structure other than bandwidth and duration

87   (Geberzahn & Aubin 2014). Because VDev does not account for differences among different

88   syllable types, it is not a suitable metric to compare the performance of songs that contain more

89   than one syllable type (Geberzahn & Aubin 2014, Podos et al. 2016). A new metric of motor

90   performance, frequency excursion (FEX), overcomes these limitations (Podos et al. 2016).

91   Frequency excursion attempts to estimate the rate at which the vocal apparatus adjusts by

92 measuring the rate of change in the fundamental frequency of a signal, including changes during

93 silent gaps.

94   Here, we measured FEX in a large sample of songs from free-living male Adelaide's

95 warblers. We then modelled variation in FEX as a function of several variables that may affect

96 performance over short time scales, allowing us to test the hypotheses summarised in Table 1.

97

98 **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 counter-singing	*counter-singing*
Time-dependent factors influence performance	Varied	*Time*

99

100

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 1996, 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 1996). In an unpublished song playback study (Pereira et al. in prep), male Adelaide's warblers type matched low-performance songs (digitally slowed down by 15%) more than controls, and 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 (Sept. 17, 2010) and adhered to the ASAB/ABS Guidelines for the Use of Animals. Birds were captured under DML's federal bird banding permit (#23696). The US 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 one aluminum leg band prior to the onset of the study. At the same time blood samples were taken for another study. We used a hypodermic

123   needle to puncture the left brachial vein, and take a blood sample with a capillary tube. The

124   bleeding was stanched with direct pressure, and standard measurements were taken. No birds

125   were otherwise injured or killed during capture. The birds appeared to forage and sing normally

126   while being recorded, suggesting that our observations did not cause significant distress.

127
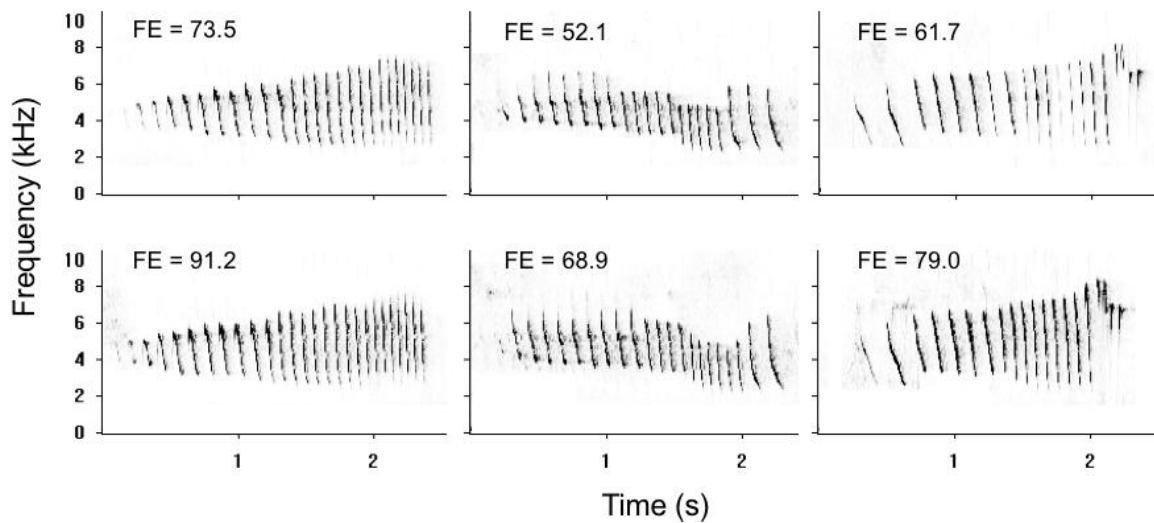
128

129   **Figure 1.** Examples of high and low performance (FEX) songs. Each column shows songs

130   belonging to the same song-type, sung by the same male, but with different FEX. Careful visual

131   inspection shows that higher performance songs tend to have higher trill rate and/or bandwidth.

132

133   *Song recordings*

134   We recorded nine colour-banded males for four days each, averaging $3:30 \pm 0:13$ hours

135   (mean $\pm$ SD) of recording per day. Successive recording sessions of a given male were separated

136   by at least four days except on two occasions when recordings were made on consecutive days

137   because of logistical constraints. Observations started 30 minutes before sunrise – allowing us to

138   capture the start of the dawn chorus and thus the ordinal number of each song – and continued

139   until three hours after sunrise. Although it was too dark to see their coloured leg bands at the

140   beginning of the recording sessions, we are confident we recorded the right individuals because

141   (1) males are highly territorial, (2) males have high fidelity to specific dawn chorus trees, or

142   cluster of trees, and (3) observers followed the birds continuously and confirmed the band

143   combinations once the sun was up. We recorded individuals continuously with a portable solid

144   state audio recorder (Marantz PMD661) and a directional 'shotgun' microphone (Sennheiser

145   ME67). An observer followed each bird through its territory at a distance that did not appear to

146   disturb the individual. We saved recordings as .wav files (sample rate = 44.1kHz, 16 bits).

147   We visualised recordings as spectrograms in Syrinx PC v2.6f sound-analysis software

148   (settings: Blackman window, transform size = 1024 points; John Burt,

149   http://www.syrinxpc.com/). Observers were blind to the identity of focal males. We recorded the

150   occurrence, time, and song-type for each song. Each song recording from a focal male was saved

151   as a separate file and assessed for recording quality. We only used high-quality recordings (high

152   signal-to-noise ratio, minimal overlap with other sounds) for song measurements.

153   In Adelaide's warblers, Type I and Type II songs differ with respect to several variables

154   including time of day, time of year, patterns of song switching (immediate versus eventual

155   variety), fine-scale structure, and social context (Staicer 1996, Staicer 1996b). However, there is

156   no published diagnostic criterion to assign song-types to a singing mode. We therefore relied on

157   median time of delivery to assign a singing mode to each song-type for each male (different

158   individuals can assign a given song-type to different singing modes, Staicer 1996b). We built a

159   histogram of median time of delivery, treating song-type within individual as the sampling unit.

160   The histogram revealed a bimodal distribution, with an antimode shortly after sunrise. We

161   assigned song-types in the left peak of the distribution to Type II and those in the right peak to

162   Type I for each individual. In doing so, we operationally defined Type II songs as those that

163   were usually sung before sunrise, and Type I songs as those that were usually sung after sunrise.

164   This division is consistent with the literature on wood warbler singing modes (Spector 1992).

165   The complete dataset contained 9499 song recordings, 2825 of which were of sufficient

166   quality for structural analysis. Prior to analysis, we further reduced the dataset by eliminating

167   songs that appeared to be missing sections relative to other songs of the same type, by the same

168   individual, N = 42), 'double songs' (two songs sung in rapid succession; N = 6), and songs

169   belonging to song-types with 10 or fewer exemplars (N = 50 songs). The rationale for

170   eliminating rare song-types is that the mixed model, which includes song-type as a random

171   factor, would have little information with which to estimate their parameters. Finally, we

172   eliminated all first songs of the day (order = 1) because they lacked a *latency* since prior song (N

173   = 10), leaving 2717 songs in the final dataset.

174

175   *Acoustic analysis*

176   Frequency excursion was measured with the custom software FEX Calculator (Jesse

177   McClure, https://github.com/BehaviorEnterprises/Fex; Podos et al. 2016; Fig. 2). The software

178   Fourier transforms the audio input creating a matrix of relative amplitude values for each

179   frequency x time bin. The highest amplitude point in each time bin (above a selected minimum

180   threshold) is used to create a path through the matrix tracing the peak frequency across time. The

181   total length of this path divided by its duration is the measure of frequency excursion. The

182   amplitude matrix data are also used to create an interactive spectrogram allowing the scorer –

183   when necessary – to remove background noise with an eraser tool that masks the selected

184 frequency x time bin, preventing it from being included in the path without requiring any

185 changes to the original audio signal.

186

187
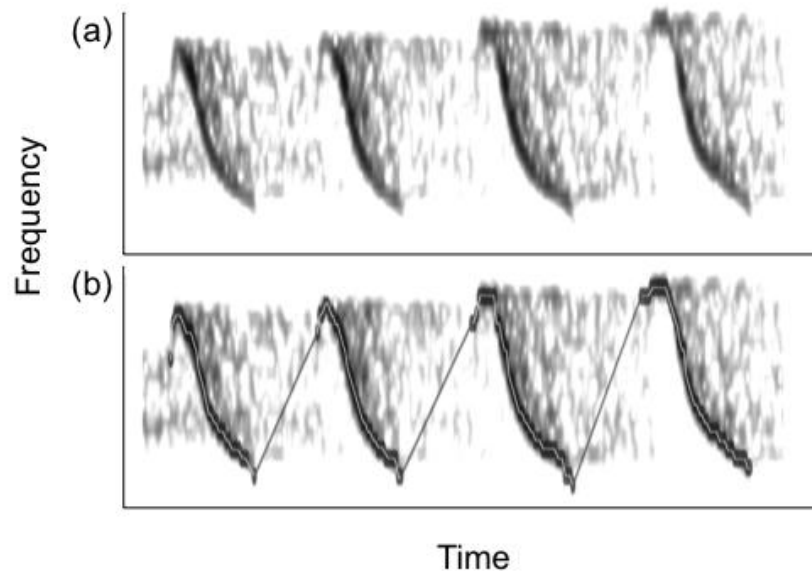
196

197 **Figure 2.** Schematic representation of semi-automated frequency excursion measurement. First a

198 spectrogram is generated (a), here with a window size of 256 and a bin size of 64 samples,

199 originating from a 44.1 kHz recording. This spectrogram is overlaid with a layer indicating the

200 points with the highest amplitude in each time bin and drawing the frequency excursion path line

201 (b). The length of this connecting line, divided by the time it spans, represents the frequency

202 excursion. Higher FEX values indicate higher performance songs.

203

204 *Statistical analysis*

205   We fitted a linear mixed model to examine variation in FEX and test for potential effects

206 of warm up, exhaustion, rest, singing mode, and counter-singing. The following fixed variables

207   were included in the model: *time of day*, *order*, *latency*, *run number*, *singing mode*, and *counter-*

208   *singing* (Table 1). *Time of day* (sec) represents the time relative to sunrise (positive values are

209   after sunrise). *Order* is the sequential order of the song for a given male on a given day (range =

210   2 - 474). For example, *order* = 1 for the first song that a male sings in the morning, *order* = 2 for

211   the second song, and so on. *Latency* (sec) is the time elapsed since the subject's previous song.

212   This variable was strongly right-skewed, so we log transformed it to prevent the extreme values

213   from exerting excessive leverage in our model. *Run number* is the song's order within a run of

214   the same song-type (range = 1 - 43). *Counter-singing* was scored as "1" if any of the following

215   were true: the song was sung within 1.5 sec following a neighbour's song, the song matched the

216   song-type sung by a neighbour in the previous 2 sec, or the song temporally overlapped a

217   neighbour's song (separating these three kinds of counter-singing did not affect the outcome;

218   unpublished data). An alternative way to measure performance changes during counter-singing is

219   to ask whether performance changes over the course of a counter-singing bout. We defined a

220   bout of counter-singing as an uninterrupted string of songs for which at least five of the previous

221   ten songs were possible responses to a neighbour (i.e., were scored as a "1"). We scored the

222   order of songs within a bout of counter-singing to define the variable *counter-singing order*,

223   which we substituted for *counter-singing* in a separate model (including *counter-singing* and

224   *counter-singing order* in the same model would be inappropriate given their high degree of

225   correlation). *Singing mode* separates Type I and Type II singing (Staicer 1996b, see above).

226   Random variables in the model were *ID*, *day (within ID)*, and *song-type* (song-type is known to

227   affect performance in other species Trillo & Vehrencamp 2005, Cramer and Price 2007, Cardoso

228   et al. 2009, Cardoso et al. 2012).

229   Linear mixed models were developed in R v3.2.0 (R Core Team 2014) using the *lme4*

230   package (Bates et al. 2015). Continuous predictors (*time of day*, *order*, *run number*, and *latency*)

231   were mean-centred and standardised prior to analysis by subtracting the mean and dividing by

232   two standard deviations, in order to make effect sizes comparable to those of binary predictors

233   (Gelman 2008). We began with a full model that included all main effects and two-way

234   interactions, and random intercepts for *ID*, *day* (*within ID)*, and *song-type*. We wanted to identify

235   which of these potential explanatory variables were strongly associated with song performance

236   while minimizing the risk of type I errors, so we implemented a conservative model selection

237   procedure. The full model was subjected to the *dredge* function in the R package *MuMIn* (Bartoň

238   2015) to rank all sub-models by AIC (Burnham & Anderson 2002). Variables that were in all

239   models with $\Delta AIC \leq 2$ were included in the reduced model (the full model and a model

240   averaging procedure produced qualitatively similar results, indicating that the results are robust

241   to variation in model selection procedures). We tested for a quadratic effect of *time of day* (as

242   suggested by Fig. 3b), but found no support, so only linear effects were included in the final

243   model. We then added random slopes for all retained fixed variables (Barr et al. 2013), and

244   examined residuals for normality and homoscedasticity. We used the *anova* function to conduct

245   likelihood ratio tests of statistical significance of random and fixed effects. We used the REML

246   method for parameter estimation and the maximum likelihood method for significance tests.

247

248

249   **Results**

250   The 2717 songs in the dataset included 40 unique song-types. Each subject contributed an

251   average of 301.89 ± 179.01 songs, representing 16.67 ± 3.16 song-types (Table S3

252   Supplementary Material). FEX varied among individuals, days, and song-types (Tables S5-S7

253   Supplementary Material). Independent variables were intercorrelated, including *time of day* and

254   *order* (Pearson's correlation: r = 0.68, Table S4 Supplementary Material). No correlations

255   exceeded |r| = 0.7, limiting the risk of multicollinearity (Dormann et al. 2013).

256

257   *Linear mixed model*

258   The model selection procedure produced the following model, which we refer to as the

259   'main model':

260   FEX ~ time + order + (1 + time + order|ID) + (1 + time + order|ID:day) + (1 + time + order|type)

261   That is, the main model explained variation in FEX as a function of the fixed effects of *time of*

262   *day* and *order*, with random intercept and slopes for both *time of day* and *order* versus *ID*, *day*

263   *within ID*, and *song-type*. The independent variables *latency*, *run number*, *counter-singing*,

264   *counter-singing order*, and *singing mode* were not retained by the model selection procedure.

265   FEX increased with *order* and decreased with *time of day* (Table 2, Fig. 3). The random

266   variables *ID*, *day within ID*, and *song-type* were all highly significant (p < 0.0001, Tables S5-S7

267   Supplementary Materials).

268   We found a positive effect of *order* on performance for all individuals, every day, for

269   almost every song-type (Fig. 4a, Tables S5-S7 Supplementary Materials). Translating the effects

270   into a biologically relevant scale, we estimate that birds' performance improves by 5.78 FEX

271   points, on average, over the course of 200 songs (subjects sang ≥ 200 songs in 27 of 36

272   observation periods). The among-individual standard deviation in average performance is 4.11

273   FEX points, so the effect of singing 200 songs is equivalent to 1.41 standard deviations of the

274   among-individual average.

275   To further address the potential confound between the correlated variables *time of day*

276   and *order*, we conducted a linear regression of *time of day* versus *order* (*order* as the dependent

277   variable) and ran the residuals in our main model. *Residual order* was a significant positive

278   predictor of FEX (effect size = 4.69, χ^2 = 7.34, df = 1, p = 0.0067).

279

280   **Table 2.** Results of a mixed effects model for the dependent variable 'frequency excursion'.

281   Covariate estimates refer to scaled data.

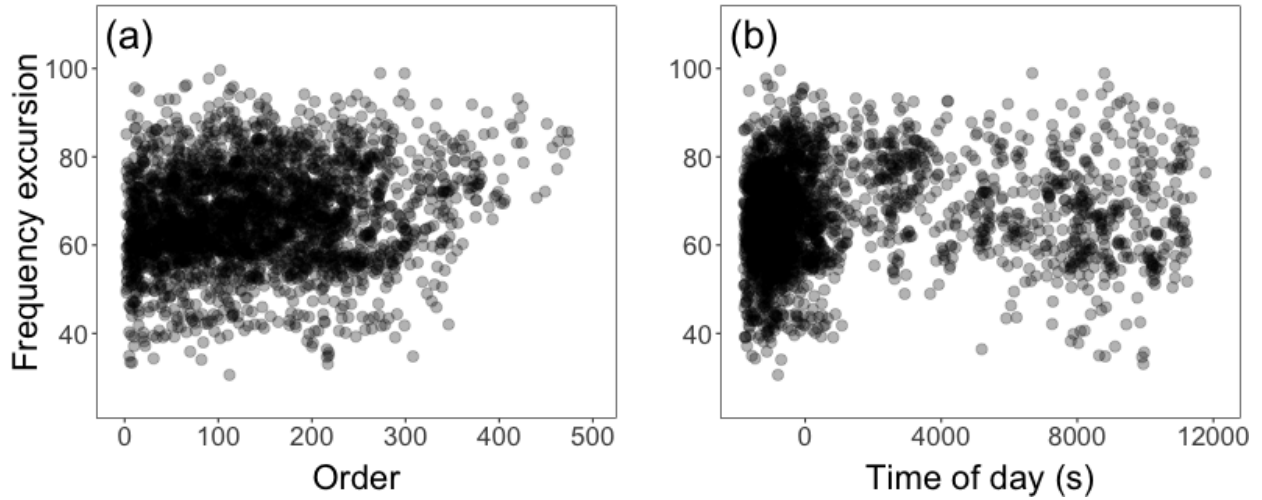| Term	Estimate	χ^2	df	p
Intercept	66.48			
Order	5.43	11.39	1	0.0007
Time of day	-2.99	2.45	1	0.12

288   Random intercepts: (σ_{ID} = 4.04, $\sigma_{Day\ w/in\ ID}$ = 3.03, $\sigma_{Song-type}$ = 9.89, all p < 0.0001; $\sigma_{Residual}$ = 5.06)

289   Random slopes (Order): (σ_{ID} = 1.89, $\sigma_{Day\ w/in\ ID}$ = 3.53, $\sigma_{Song-type}$ = 3.31)

290   Random slopes (Time of day): (σ_{ID} = 2.57, $\sigma_{Day\ w/in\ ID}$ = 5.02, $\sigma_{Song-type}$ = 6.58)
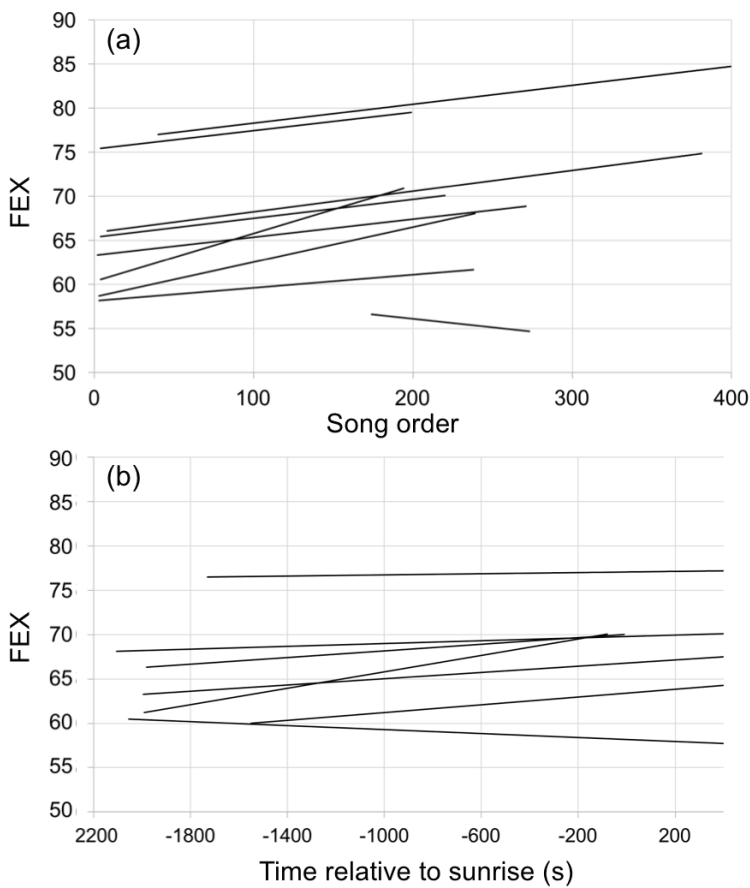
291

292

293

Figure 3. Frequency excursion as a function of (a) song order (r = 0.16) and (b) time relative to

sunrise (r = 0.10). Points are semi-transparent, so darker regions indicate overlapping data points.

296

297   **Figure 4.** Regression lines showing the relationship between vocal performance (FEX), (a) song

298   order and (b) time relative to sunrise for the most common song-type from each of the nine

299   subjects.

300

301   FEX relies on a linear frequency scale, but production and perception of sound frequency

302   is better modelled on a log scale (Cardoso 2013). If song duration covaries with *order*, and if

303   long songs tend to include more high fundamental frequencies than do short songs, our model

304   would overestimate the salience of covariation between FEX and *order* to signal receivers. To

305   test this possibility, we used linear regression to model FEX as a function of song *duration* (ms)

306   and *average mid-frequency* (the midpoint of the average maximum frequency and average

307   minimum frequency), and used the residuals as the dependent variable in our main model. The

308   model with residual FEX estimated similar effects of *time of day* (-2.31) and *order* (4.69) as did

309   our main model. As in the main model, the effect of *order* was statistically significant ($\chi^2 = 7.34$,

310   df = 1, p = 0.007)

311   We were interested in the specific structural changes responsible for the effects of *time of*

312   *day* and *order* on FEX, so we ran variants of the main model with each of the following

313   dependent variables: trill rate (number of notes / duration), frequency bandwidth (Hz, averaged

314   over all notes), number of notes, duration (s), minimum frequency (Hz, averaged over all notes),

315   maximum frequency (Hz, averaged over all notes), and vocal deviation (Table S1, see

316   Supplementary Material for methodological details). The only metrics that showed a statistically

317   significant relationship with *order* were trill rate and number of notes, both of which had positive

318   coefficients (Table 3). We then regressed FEX on trill rate and ran the residuals in the main

319   model to determine whether *order* affected FEX after accounting for the covariation between

320   FEX and trill rate. The effect of *order* on the residuals of FEX was positive and significant

321   (effect size = 3.78, χ^2 = 6.39, df = 1, p = 0.011).

322

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

fixed effect	average frequency bandwidth (kHz)	trill rate (Hz)	number of notes	duration (ms)	average minimum frequency (kHz)	average maximum frequency (kHz)	vocal deviation
time	-0.02	0.03	-1.02	-56.30	-0.07	-0.09	0.02
order	-0.06	0.72*	3.11**	93.66	0.10	0.07	-0.03

324   * p < 0.01; **p < 0.001

325

326   **Discussion**

327   We found strong evidence that song motor performance (FEX) improves over the course

328   of morning singing in male Adelaide's warblers (Table 2, Fig. 3). This effect was not explained

329   by the time of day, which we would expect if factors like air temperature, amount of time that a

330   bird has been awake, or increase in social interactions caused the improvement. Rather, the

331   cumulative number of songs that a bird had sung during the morning explained the observed

332   improvement in performance on a given day. We therefore conclude that Adelaide's warblers

333   warm up during morning singing. The magnitude of the *order* effect (+ 1.41 SD of among-

334   individual variation over 200 songs) is likely to be biologically relevant if receivers compare

335   performance among males. Controlling for song duration and mid-frequency had little effect on

336   the relationship between order and performance, indicating that receivers are likely to perceive

337 the warm up effect in spite of the non-linearity of sound frequency production and perception

338 (Cardoso 2013). As birds warm up they might choose to sing higher performance song-types or

339 they might continue to sing the same song-types, but with higher performance. Our analyses do

340 not address the first possibility, but the performance of almost all song-types (39/40) improved

341 with song order (Table S7 Supplementary Materials).

342   This is the first report of increased song performance as a function of recent practice, but

343 other studies have considered changes in performance over longer time scales. Banded wrens

344 (*Thryophilus pleurostictus*) and swamp sparrows (*Melospiza georgiana*) show an increase in

345 performance (VDev) with age, and banded wrens increase their trill rates over the course of a

346 single season (Ballentine 2009, Vehrencamp et al. 2013). Similarly, frequency bandwidth

347 increases with age in Java sparrows (*Lonchura oryzivora*; Ota & Soma 2014). One study,

348 however, found that song performance decreases between successive years in white-crowned

349 sparrows (*Zonotrichia leucophrys*, Poesel & Nelson 2015). We conclude that changes in

350 performance can take place over a range of time scales, from minutes to years.

351   The performance metric frequency excursion is a function of trill rate, frequency

352 bandwidth, and other structural variables. We tested several such variables in an attempt to better

353 understand the positive influence of song *order* on FEX (Table 3). Trill rate and especially the

354 number of notes in the song, were positively influenced by song *order*. We found a positive

355 effect of *order* when controlling for the effects of trill rate on FEX, which means that *order*

356 affects additional components of FEX beyond just trill rate. Candidate components include

357 frequency modulation within notes and frequency jumps between notes (Geberzahn & Aubin

358 2014). The effects of song *order* on song duration, average frequency bandwidth, and average

359 minimum and maximum frequencies were not significant. As Adelaide's warblers warm up, they

360   add more notes and sing faster, but they do not greatly alter the frequency characteristics of their

361   songs. The performance metric VDev trended in the expected direction (lower VDev

362   corresponds to higher performance). Its small effect size is probably attributable to the fact that

363   frequency bandwidth influences VDev much more strongly than does trill rate in this population

364   (see Supplementary Material).

365   The mechanisms underlying the warm up effect are not known. In humans, warming up

366   has an especially pronounced effect on sprint performance (Yaicharoen et al. 2012) and maximal

367   muscle performance (Wittekind et al. 2012) but no detectable effect on sub-maximal muscle

368   performance (McCrary et al. 2015). Warm up improves perceived tone quality and may

369   influence vibrato rate in human singing (Amir et al. 2005, Moorcroft & Kenny 2012, Moorcroft

370   & Kenny 2013). It remains to be determined whether similar physiological mechanisms underlie

371   human and avian vocal warm up.

372   The adaptive significance of vocal performance in Adelaide's warbler is not known.

373   Here, however, we speculate on how warming up and female choice for high performance might

374   interact to affect the evolution of singing behaviour. Many birds, including Adelaide's warblers,

375   participate in the 'dawn chorus,' during which song rates are elevated. There are several

376   proposed explanations for the dawn chorus, including the hypothesis that males are advertising to

377   females who compare potential mating partners at dawn (reviewed in Catchpole & Slater 2008

378   pp. 128-135). According to the warm up hypothesis, males that begin singing at a high rate early

379   in the morning would sing with higher performance than males that begin later or sing at a lower

380   rate, all else equal. If females choose males by comparing their singing performances, selection

381   would favour earlier and more intense dawn singing, spurring an evolutionary arms race for ever

382   earlier and more frequent dawn singing. This hypothesis emphasises the importance of

383   constraints in shaping the evolution of behavioural phenomena.

384   Song performance decreased with *time of day*, suggesting that, when accounting for

385   *order*, performance decreases toward the later hours of the morning. The fixed variable *time of*

386   *day* was not statistically significant in the final model because there was a lot of variation in the

387   slopes of *time of day* versus FEX among levels of ID, day, and song-type (Tables S5-S7).

388   Nevertheless, the AIC-based model selection procedure (conducted before we added random

389   slopes) included *time of day* in all of the best models, suggesting that it is important for

390   explaining variation in FEX. Unmeasured biotic or abiotic factors that correlate with time (air

391   temperature, amount of time the bird has been awake, rate of social interactions) could affect

392   performance, or *time of day* may capture the effect of singing mode better than the *mode*

393   variable. Finally, it is possible that the recordists tended to be farther away from the birds later in

394   the day, perhaps because the birds fly longer distances in the late morning. If this were the case,

395   later recordings might tend to miss high frequency components of songs which attenuate with

396   distance more severely than do low frequency components. The negative relationship between

397   time and maximum frequency would seem to support this hypothesis, but that effect is small and

398   non-significant (Table 3).

399   The positively correlated variables *order* and *time of day* (r = 0.68) exert opposite effects

400   on FEX in our model, raising the question of whether *order* has a positive influence on song

401   performance that is independent of *time of day*. Critically, *order* and *time* were not perfectly

402   correlated (r < 1.0) because birds did not sing at perfectly regular temporal intervals. This fact

403   allowed us to separate out variation in *order* that cannot be attributed to variation in *time* and

404   show that this variable (*residual order*), which was statistically independent of *time of day*,

405   explained substantial variation in FEX. We conclude that the positive effect of *order* on FEX is

406   not a collinearity artefact.

407   Song performance did not change in a consistent direction over repeated renditions of the

408   same song-type, failing to support a key prediction of the song-type specific exhaustion

409   hypothesis (Lambrechts & Dhondt 1988). Birds may switch to a different song-type before

410   exhaustion occurs, but our data do not allow us to test that idea. To our knowledge, this is the

411   first test of the anti-exhaustion hypothesis that quantifies the effect of song-type repetition on

412   song performance. A previous test of this hypothesis also failed to find support; singing rate and

413   song-type switching rate in chaffinches (*Fringilla coelebs*) are not different between males with

414   only one song-type and males with a repertoire of several song-types (Brumm et al. 2009). The

415   time elapsed since a subject's previous song (*latency*) did not affect FEX, failing to support the

416   song-type general exhaustion hypothesis.

417   We also found no support for the hypothesis that songs differ in performance based on

418   singing mode. Staicer (1996) found that Type I and Type II songs were structurally distinct in

419   Adelaide's warblers, but she did not measure performance. In the congeneric American yellow

420   warbler's (*S. petechia*) Type I songs are higher performance (VDev) than Type II songs (Beebee

421   2004, Beebee 2004b), however in the light of the current study this result could have been

422   attributable to warm up effects. Our finding comes with the caveat that there is no definitive

423   method for discriminating singing modes in our focal species, so our simple classification

424   scheme may not have accurately separated the two modes. We are investigating this topic

425   further. We also failed to find an effect of vocal interaction on song performance. We interpret

426   that finding to mean that Adelaide's warblers do not significantly alter their vocal performance in

427   a consistent direction during counter-singing. Playback experiments will be required to

428   determine whether performance is affected by specific social contexts (e.g., the presence of a

429   male or female intruder).

430   Performance varied significantly among individuals, song-types, and days. Variation in

431   performance among song-types was much greater than variation in performance among

432   individuals, highlighting the importance of controlling for song-type in performance analyses

433   (Cardoso et al. 2009). The evolutionary maintenance of song-types with low FEX scores could

434   be explained in several ways. First, there are aspects of performance that FEX does not capture,

435   such as song rate (Lambrechts & Dhondt 1987), amplitude (Ritschard et al. 2010), or consistency

436   (Byers 2007). Some low FEX song-types may be high-performance with respect to one or more

437   unmeasured metrics. Second, some functions of song (e.g., advertisement of territory occupancy)

438   might not require maximal performance, so selection may favour the use of low-performance

439   songs, perhaps because they are less taxing. Third, some birds could be sufficiently constrained,

440   at least some of the time, that they are incapable of producing high performance song-types.

441   These birds might make the best of a bad situation by resorting to low performance song-types.

442   Fourth, low-performance songs might function as a first stage in escalating to higher-

443   performance songs during aggressive interactions. Individuals' performance differed between

444   recording days. This effect could be attributable to breeding phase, a long-term practice effect, or

445   idiosyncratic variation in, for example, nutritional stress or temperature (Ballentine 2009,

446   Vehrencamp et al. 2013).

447   We encourage further tests for warm up effects in birdsong and other animal displays.

448   Given the strength of the effect in this study, efforts to accurately measure and compare song

449   performance should account for possible order effects. This requires continuous observation,

450   beginning with the first display of the day. Extending observation periods later in the day or over

451   a longer segment of the year would reveal whether the effect of order tapers off or reverses later

452   in the day, and if its strength is affected by breeding phase.

References

Amir, O., Amir, N., and Michaeli, O. (2005). Evaluating the influence of warmup on singing

    voice quality using acoustic measures. J. Voice 19, 252–260.

    (doi: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*. Anim. Behav. *77*, 973–978.

    (doi:10.1016/j.anbehav.2008.12.027)

Ballentine, B., Hyman, J., and Nowicki, S. (2004). Vocal performance influences female

    response to male bird song: an experimental test. Behav. Ecol. *15*, 163–168.

    (doi:10.1093/beheco/arg090)

Ballentine, B., Searcy, W., and Nowicki, S. (2008). Reliable aggressive signalling in swamp

    sparrows. Anim. Behav. *75*, 693–703. (doi:10.1016/j.anbehav.2007.07.025)

Barr, D.J., Levy, R., Scheepers, C., and Tily, H.J. (2013). Random effects structure for

    confirmatory hypothesis testing: Keep it maximal. J. Mem. Lang. *68*, 255–278.

    (doi:10.1016/j.jml.2012.11.001)

Bartoň, K. (2015). MuMIn: Mul4--- Model Inference. R package version 1.15.1.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models

    Using lme4. J. Stat. Softw. *67*, 1–48. (doi:10.18637/jss.v067.i01.)

Beebee, M. (2004). Variation in vocal performance in the songs of a wood-warbler: Evidence for

    the function of distinct singing modes. Ethology *110*, 531–542. (doi:10.1111/j.1439-

    0310.2004.00994.x)

Beebee, M. (2004b). The functions of multiple singing modes: Experimental tests in yellow

    warblers, *Dendroica petechia*. Anim. Behav. *67*, 1089–1097.

476   (doi:10.1016/j.anbehav.2003.05.016)

477   Benedict, L., Rose, A., and Warning, N. (2012). Canyon wrens alter their songs in response to

478   territorial challenges. Anim. Behav. *84*, 1463–1467. (doi:10.1016/j.anbehav.2012.09.017)

479   Brumm, H., Lachlan, R.F., Riebel, K., and Slater, P.J.B. (2009). On the function of song type

480   repertoires: testing the 'antiexhaustion hypothesis' in chaffinches. Anim. Behav. *77*, 37-42.

481   (doi: 10.1016/j.anbehav.2008.09.009)

482   Burnham, K., and Anderson, D. (2002). Model Selection and Multimodel Inference: A Practical

483   Information-Theoretic Approach. (New York: Springer).

484   Burt, J. Syrinx-PC. http://www.syrinxpc.com/

485   Burt, J., and Vehrencamp, S.L. (2005). Dawn chorus as an interactive communication network.

486   In Animal Communication Networks, P. McGregor, ed. (Cambridge: Cambridge

487   University Press), pp. 320–343.

488   Byers, B.E. (2007). Extrapair paternity in chestnut-sided warblers is correlated with consistent

489   vocal performance. Behav. Ecol. *18*, 130–136. (doi:10.1093/beheco/arl058)

490   Byers, J., Hebets, E., and Podos, J. (2010). Female mate choice based upon male motor

491   performance. Anim. Behav. *79*, 771–778. (doi:10.1016/j.anbehav.2010.01.009)

492   Cardoso, G., Atwell, J., Ketterson, E., and Price, T. (2009). Song types, song performance, and

493   the use of repertoires in dark-eyed juncos (*Junco hyemalis*). Behav. Ecol. 20, 901–907.

494   (doi:10.1093/beheco/arp079)

495   Cardoso, G., Atwell, J., Hu, Y., Ketterson, E., and Price, T. (2012). No Correlation Between

496   Three Selected Trade-Offs in Birdsong Performance and Male Quality for a Species With

497   Song Repertoires. Ethology *118*, 584–593. (doi:10.1111/J.1439-0310.2012.02047.X)

498   Cardoso, G.C. (2013). Using frequency ratios to study vocal communication. Anim. Behav. *85*,

499   1529-1532. (doi: 10.1016/j.anbehav.2013.03.044)

500 Caro, S., Sewall, K., Salvante, K., and Sockman, K. (2010). Female Lincoln's sparrows

501   modulate their behavior in response to variation in male song quality. Behav. Ecol. *21*,

502   562–569. (doi:10.1093/beheco/arq022)

503 Catchpole, C., and Slater, P. (2008). Bird Song: Biological Themes and Variations. (Cambridge:

504   Cambridge University Press).

505 Cramer, E., and Price, J. (2007). Red-winged blackbirds *Ageliaus phoeniceus* respond differently

506   to song types with different performance levels. J. Avian Biol. *38*, 122–127.

507   (doi:10.1111/j.2006.0908-8857.03839.x)

508 Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G.,

509   Gruber, B., Lafourcade, B., Leitão, P.J., and Münkemüller, T. (2013). Collinearity: a

510   review of methods to deal with it and a simulation study evaluating their performance.

511   Ecography *36*, 27–46. (doi:10.1111/j.1600-0587.2012.07348.x)

512 DuBois, A., Nowicki, S., and Searcy, W. (2009). Swamp sparrows modulate vocal performance

513   in an aggressive context. Biol. Lett. *5*, 163–165. (doi:10.1098/rsbl.2008.0626)

514 DuBois, A., Nowicki, S., and Searcy, W. (2011). Discrimination of vocal performance by male

515   swamp sparrows. Behav. Ecol. Sociobiol. *65*, 717–726. (doi:10.1007/s00265-010-1073-2)

516 DuVal, E.H. (2012). Variation in annual and lifetime reproductive success of lance-tailed

517   manakins: experience as an alpha offsets senescence in siring success. Proc. R. Soc. B Biol.

518   Sci. *279*, 1551–1559. (doi:10.1098/rspb.2011.1840)

519 Geberzahn, N., and Aubin, T. (2014). Assessing vocal performance in complex birdsong: a novel

520   approach. BMC Biol. *12*, 58. (doi:10.1186/s12915-014-0058-4)

521   Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. Stat. Med.

522   *27,* 2865-2873. (doi: 10.1002/sim.3107)

523   Illes, A., Hall, M., and Vehrencamp, S.L. (2006). Vocal performance influences male receiver

524   response in the banded wren. Proc. R. Soc. B Biol. Sci. *273*, 1907–1912.

525   (doi:10.1098/rspb.2006.3535)

526   Lambrechts, M, and Dhondt, A. (1987). Differences in singing performance between male great

527   tits. Ardea *75*, 43–53.

528   Lambrechts, M., and Dhondt, A. (1988). The anti-exhaustion hypothesis: a new hypothesis to

529   explain song performance and song switching in the great tit. Anim. Behav. *36*, 327–334.

530   (doi:10.1016/S0003-3472(88)80002-2)

531   Mappes, J., Alatalo, R., Kotiaho, J., and Parri, S. (1996). Viability Costs of Condition-Dependent

532   Sexual Male Display in a Drumming Wolf Spider. Proc. R. Soc. B Biol. Sci. *263*, 785–789.

533   McClure, J. (2015). FEX Calculator. https://github.com/BehaviorEnterprises/Fex

534   McCrary, J.M., Halaki, M., Sorkin, E., and Ackermann, B.J. (2015). Acute Warm-up Effects in

535   Submaximal Athletes: An EMG Study of Skilled Violinists. Med. Sci. Sports Exerc. *48*,

536   307–315. (doi:10.1249/MSS.0000000000000765)

537   Moorcroft, L., and Kenny, D. (2012). Vocal warm-up produces acoustic change in singers'

538   vibrato rate. J. Voice *26*, 667.e13–667.e18. (doi:10.1016/j.jvoice.2011.10.007)

539   Moorcroft, L., and Kenny, D. (2013). Singer and listener perception of vocal warm-up. J. Voice

540   *27*, 258.e1–258.e13. (doi:10.1016/j.jvoice.2012.12.001)

541   Moseley, D., Lahti, D., and Podos, J. (2013). Responses to song playback vary with the vocal

542   performance of both signal senders and receivers. Proc. R. Soc. B Biol. Sci. *281*,

543   20131401. (doi:10.1098/rspb.2013.1401)

544   Motel, T., Fisher, K., and Leydon, C. (2003). Vocal warm-up increases phonation threshold

545   pressure in soprano singers at high pitch. J. Voice *17*, 160–167. (doi:10.1016/S0892-

546   1997(03)00003-1)

547   Ota, N., and Soma, M. (2014). Age-dependent song changes in a closed-ended vocal learner:

548   elevation of song performance after song crystallization. J. Avian Biol. *45*, 566–573.

549   (doi:10.1111/jav.00383)

550   Poesel, A., and Nelson, D.A. (2015). Changes in performance of shared and unshared songs

551   within and between years in the white-crowned sparrow. Ethol. *121*, 1-11.

552   (doi:10.1111/eth.12399)

553   Podos, J. (1997). A performance constraint of the evolution of trilled vocalizations in a songbird

554   family (Passeriformes: Emberizidae). Evolution *51*, 537–551. (doi:10.2307/2411126)

555   Podos, J. (2001). Correlated evolution of morphology and vocal signal structure in Darwin's

556   finches. Nature *409*, 185–188. (doi:10.1038/35051570)

557   Podos, J., Moseley, D., Goodwin, S., McClure, J., Taft, B., Strauss, A., Rega-Brodsky, C., and

558   Lahti, D. (2016). A fine-scale, broadly-applicable index of vocal performance: frequency

559   excursion. Anim. Behav. *116*, 203–213. (doi: 10.1016/j.anbehav.2016.03.036)

560   Podos, J., and Patek, S. (2015). Acoustic signal evolution: Biomechanics, size, and performance.

561   In Animal Signaling and Function: An Integrative Approach, D. Irshick, M. Briffa, and J.

562   Podos, eds. (New York: John Wiley & Sons), pp. 175-204.

563   Price, J., Earnshaw, S., and Webster, M. (2006). Montezuma oropendolas modify a component

564   of song constrained by body size during vocal contests. Anim. Behav. *71*, 799–807.

565   (doi:10.1016/j.anbehav.2005.05.025)

566   R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria:

567   R Foundation for Statistical Computing. http:// www.R-project.org/

568 Ritschard, M., Riebel, K., and Brumm, H. (2010). Female zebra finches prefer high-amplitude

569   song. Anim. Behav. *79*, 877–883. (doi:10.1016/j.anbehav.2009.12.038)

570 Spector, D. (1992). Wood-warbler song systems: a review of paruline singing behaviors. In

571   Current Ornithology, D. Power, ed. (New York: Plenum), pp. 199–238.

572 Staicer, C. (1996). Acoustical Features of Song Categories of the Adelaide's Warbler (*Dendroica*

573   *adelaidae*). Am. Nat. 113, 771–783.

574 Staicer, C. (1996b). Honest advertisement of pairing status: evidence from a tropical resident

575   wood-warbler. Anim. Behav. *51*, 375–390. (doi:10.1006/anbe.1996.0036)

576 Staicer, C., Spector, D., and Horn, A. (1996). The dawn chorus and other diel patterns in acoustic

577   signaling. In Ecology and Evolution of Acoustic Communication in Birds, D.E.

578   Kroodsma, and E.H. Miller, eds. (Ithaca: Comstock Publishing), pp. 426–453.

579 Stewart, D., Macaluso, A., and De Vito, G. (2003). The effect of an active warm-up on surface

580   EMG and muscle performance in healthy humans. Eur. J. Appl. Physiol. *89*, 509–513.

581   (doi:10.1007/s00421-003-0798-2)

582 Suthers, R. (2004). How birds sing and why it matters. In Nature's Music: The Science of

583   Birdsong, P. Marler, H. Slabbekoorn, eds. (New York: Elsevier), pp. 272–295.

584 Toms, J. (2011). Non-breeding competition between migrant American Redstarts (*Setophaga*

585   *ruticilla*) and resident Adelaide's Warblers (*Dendroica adelaidae*) in the Guánica

586   Biosphere Reserve, southwest Puerto Rico. Dissertation, University of Missouri.

587 Trillo, P.A., and Vehrencamp, S.L. (2005). Song types and their structural features are associated

588   with specific contexts in the banded wren. Anim. Behav. *70*, 921–935.

589   (doi:10.1016/j.anbehav.2005.02.004)

590   Vehrencamp, S.L., Yantachka, J., Hall, M., and de Kort, S. (2013). Trill performance

591   components vary with age, season, and motivation in the banded wren. Behav. Ecol.

592   Sociobiol. *67*, 409–419. (doi:10.1007/s00265-012-1461-x)

593   Wittekind, A., Cooper, C.E., Elwell, C.E., Leung, T.S., and Beneke, R. (2012). Warm-up effects

594   on muscle oxygenation, metabolism and sprint cycling performance. Eur. J. Appl.

595   Physiol. *112*, 3129–3139. (doi:10.1007/s00421-011-2262-z)

596   Yaicharoen, P., Wallman, K., Bishop, D., and Morton, A. (2012). The effect of warm up on

597   single and intermittent-sprint performance. J. Sports Sci. *30*, 833–840.

598   (doi:10.1080/02640414.2012.675084)

Highlighted supplementary materials

Non-highlighted supplementary materials