Communication Value of Mistakes in Dark-Eyed Junco Song

André C. Ferreira,[1] **Jonathan W. Atwell,**[2] **Danielle J. Whittaker,**[2,3] **Ellen D. Ketterson,**[2] **and Gonçalo C. Cardoso**[1,*,†]

1. CIBIO—Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Vairão 4485-661, Portugal;   2. Department of Biology, Indiana University, 1001 East Third Street, Bloomington, Indiana 47405;   3. BEACON Center for Study of Evolution in Action, Michigan State University, East Lansing, Michigan 48824

abstract: Sexual signals contain information on individual quality or motivation, and most explanations for their reliability are based on signal costs. A recent suggestion is that signaling mistakes, defined as deviations from typical signal design, provide cues on individual quality, contributing to reliable communication even when signal design is not costly. We describe several atypical song traits in dark-eyed juncos (*Junco hyemalis*) that may be mistakes during song production or development and occur in up to 6% of songs. These putative mistakes were more frequent in an urban versus a wildland population, and individuals differed in their frequency of mistakes. Some atypical signals were more frequent in younger males or were negatively related to paternity success, supporting the hypothesis that fewer mistakes indicate individual quality. We also discuss unexpected results, such as some atypical signals being more frequent in more ornamented males and in songs with lower performance demands. Song consistency (similarity across syllable renditions) was positively related to male age and paternity success; nonetheless, relations with paternity were stronger when looking at the most deviant syllable renditions, suggesting that the perceptual salience of large mistakes may mediate receiver responses to song consistency. Results indicate that signaling mistakes reveal relevant information to play a role in communication.

Keywords: animal communication, signal reliability, signaling mistakes, song performance, song consistency.

Introduction

Reliable communication is challenging to explain when the interests of signalers and receivers differ. For example, weak individuals would benefit from signaling high competitive ability during agonistic signaling, or senders would benefit from signaling with maximum attractiveness during mate attraction. Such conflicts select for receivers that attend only to signaling that reliably indicates phenotypic or genetic traits, or the motivation of the sender (Krakauer and Johnstone 1995; Bradbury and Vehrencamp 1998; Searcy and Nowicki 2005).

* Corresponding author; e-mail: gcardoso@cibio.up.pt.

† ORCIDs: Whittaker, http://orcid.org/0000-0002-3491-5642.

There are several mechanisms that maintain reliable signaling when the interests of signalers and receivers differ, mostly involving signaling costs (Hurd and Enquist 2005): signaling may be costly such that only high-quality individuals benefit from signaling at high intensity, signal design may be mechanistically linked to a physical trait (e.g., body size), or there may be social challenges toward individuals signaling dominance (reviewed in Bradbury and Vehrencamp 1998; Searcy and Nowicki 2005).

A recent suggestion is that signaling also exposes mistakes that reveal information on individual quality or motivation (Cardoso 2013*a*). Putative mistakes are here identified as marked deviations from the typical signal design of the individual or the species. Such atypical signals are rare within populations, suggesting that they are not beneficial signal variants (otherwise they would have become common over evolutionary time) but possibly anomalies during signaling or development of the signal (i.e., mistakes). Many species have stereotyped or repetitive sexual signals (e.g., the repetitive, steady-rate mating calls of anurans or orthopterans, or ritualized courtship movements) that should allow detection of mistakes by receivers. If receivers assess signaling mistakes or other aspects of signaling performance, then simple and stereotyped sexual signals could even evolve as a means for senders to advertise good-quality signals (Cardoso and Hu 2011).

Larger or more frequent mistakes could reveal negative qualities of the sender, such as stress during the development or performance of signals, while the absence of mistakes would reveal positive qualities, such as dominance or neuromuscular ability. Reliable communication could then be maintained due to receivers attending to signaling as an advertiser of resistance to mistakes. This hypothesis conforms to existing theory in that, for mistakes to convey reliable information, senders' resistance to mistakes should have costs (e.g., costly development, reduced sensitivity to fear or stress; Nowicki et al. 2002; see discussion in Cardoso 2013*a*). A novel and attractive aspect of this hypothesis is that costs are not necessarily linked to signal design, and the mecha-

nism could thus work even when the act of signaling has trivial costs.

Here we use birdsong to test whether atypical signals convey information on aspects of individual quality. Previous work has found some song traits that can be interpreted as signaling mistakes and that influence communication (e.g., interrupted songs [Podos 1996; Nowicki et al. 2001], inconsistent syllable form [de Kort et al. 2009]; reviewed in Cardoso 2013a). Here we provide a comprehensive study of atypical traits in the songs of the dark-eyed junco (*Junco hyemalis*). The dark-eyed junco is a socially monogamous passerine with substantial extrapair paternity (Ketterson 1998; Price et al. 2008) whose long-range song (hereafter, simply "song") is used in male-male competition (e.g., territoriality; Titus 1998) and mate attraction (Ketterson et al. 1992; Reichard et al. 2013). In both these contexts, there is conflict of interest between sender and receiver, and we thus expect that singing reveals individual quality. Dark-eyed junco song is very simple: a single syllable repeated several times to form a trill (fig. 1a), and each male has a repertoire of ∼2–6 song or syllable types (Newman et. al. 2008). Song bouts typically repeat a song type many times before switching to another song type (Titus 1998). Although simple, song types differ in aspects of performance (sensu Podos et al. 2009), and higher-performance song types are preferentially used in more aggressive or motivated singing (Cardoso et al. 2009).

We asked whether the occurrence of putative mistakes provides cues of male quality, considering occasional atypical signals that are best interpreted as failure or interference with the neuromotor execution of song (e.g., incomplete syllables, atypically short songs) and more recurrent signals that suggest developmental or learning errors (e.g., species-atypical syllables used as part of a male's repertoire). Our specific objectives are sevenfold: (1) To describe the diversity and frequency of atypical dark-eyed junco songs. (2) To test whether the frequency of these atypical songs differs between ecologically distinct populations, as this may suggest environmental stressors associated with mistakes, or (3) differs among individual males within populations. (4) To test whether the frequency of mistakes indicates age or aspects of the physiological condition of males or (5) predicts their paternity success. (6) To test whether the occurrence of mistakes increases with the vocal demands of different song types (evaluated by differences in song performance), which would suggest that more demanding signals provide better cues of resistance to mistakes. Last, in addition to discrete atypical signals, small and diffuse deviations to song structure are studied under the header of song consistency and often indicate individual traits such as age (reviewed in Sakata and Vehrencamp 2012). In order to assess whether overall song consistency or extreme deviations to song structure (i.e., mistakes) are more

informative, we (7) test whether consistency in junco song (i.e., degree of similarity in syllable rate or frequency within trills) is related to male phenotype or to paternity success and then contrast these results with similar analyses focusing on more extreme deviations (i.e., mistakes) on syllable rate or frequency. These tests provide the first comprehensive analysis of the potential role of signaling mistakes in animal communication.

Methods

Subjects and Categories of Atypical Signals

We used song recordings from two dark-eyed junco populations in Southern California (campus of the University of California at San Diego [UCSD], lat. 32°40′N, long. 117°10′W, and Mount Laguna, lat. 32°52′N, long. 116°25′W), collected during the breeding seasons of 2006 and 2007. Information on the study populations is in Yeh and Price (2004) and that on the field methods is in Cardoso et al. (2007). Here we analyzed only recordings of individually identified males singing spontaneously or occasionally stimulated with a very short song playback. In total, we analyzed 870 recordings from 165 males, identified either by colored leg bands (128 males) or by territory ownership and song repertoire, corresponding to 9,746 songs.

Songs were screened on spectrograms with a resolution of 86 Hz by 1.45 ms in Avisoft SASLab Pro (settings: fast Fourier transform length of 512, hamming window, 100% window duration, and 87.5% overlap, using sound files with a 22,050 Hz sampling rate). We initially screened one song from each of 200 randomly chosen recordings to identify different types of atypical signals. Based on this initial screening, we defined six types of atypical song traits, as described next and illustrated in figure 1. The full set of recordings was then screened for these putative mistakes, counting the total number of songs in each recording, and the number of songs where each atypical trait occurred.

Multitype song (fig. 1*B*): A song composed by two or more different syllable types (up to three in our sample), each making a trill. Such songs were documented in several populations of dark-eyed juncos, always at low frequencies (between 5% and 21% of songs; Konishi 1964; Williams and MacRoberts 1977, 1978; Titus 1998; Newman et al. 2008).

Atypical syllables (figs. 1*C*, A1; fig. A1 available online): In which a song comprises one or more syllables, with a structure much different from the typical dark-eyed junco's syllables—usually much longer, unlike the short syllables typical of trilled song, and more similar to syllables found either in subsong (an early stage of song development; Marler et al. 1962), short-range song (a different class of junco low-amplitude song; Titus 1998), or the more complex songs
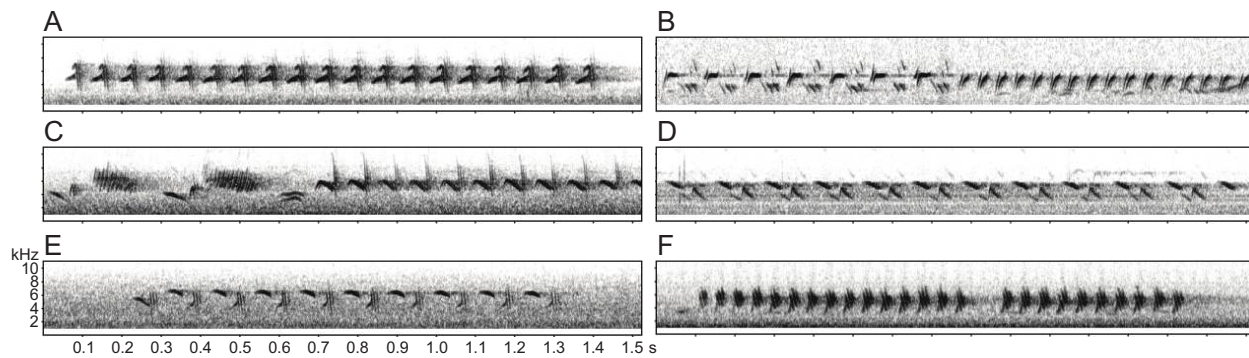
Figure 1: Spectrograms with examples of typical dark-eyed junco song (*A*), multitype song (*B*), atypical syllables (in this case, at the beginning of the song; *C*), incomplete syllable (in this case, at the end of the song; *D*), modified syllable (first syllable in the song; *E*), and break in the song (*F*).

of the closely related yellow-eyed junco (*Junco phaeonotus*; Marler and Isaac 1961; Pieplow and Francis 2011).

Incomplete syllable (fig. 1*D*): One or more notes of a syllable are missing from the last or first syllable in the song. Since, by definition, this type of atypical signal cannot occur in syllables with a single note, analyses of incomplete syllables use only those song types with more than one note per syllable. Thus, as noted throughout the "Results" section, sample sizes are smaller in analyses of incomplete syllables than of other atypical traits.

Modified syllable (fig. 1*E*): When the first or the last syllables in the trill have a different shape from the others but still share clear similarities.

Break (fig. 1*F*): An interval within a song that is much longer than the regular intervals between the other syllables in the trill. Intervals that appeared atypically long on spectrograms were measured and scored as a break if they were more than twice the average duration of the other intersyllable intervals in that song.

Alternate: Cases where a different song type appears within a song bout without being repeated (i.e., A, A, A, B, A, A, A . . .), as opposed to the typical behavior of repeating the same song type several times before switching to another and then repeating it several times also (i.e., A, A, A, B, B, B . . .).

In addition to the above six categories of discretely identifiable atypical signals, we also looked for extreme cases in three quantitative traits: (1) song duration, (2) changes in syllable rate within a song, and (3) changes in syllable peak frequency within a song. Unlike traits that vary discretely, for which atypical signals may be identified with a simple screening, deviations in these quantitatively varying traits are best identified by reference to frequency distributions (histograms) of signal measurements for the entire pool of song recordings (Cardoso 2013*a*). The clearest cases are when histograms show disjunct distributions or multimodal distribu-

tions and a quantitative threshold can be drawn to define atypical signals apart from the main distribution of trait values. We generated distribution histograms for song duration and for changes in syllable rate or peak frequency within a song using earlier measurements of up to five songs per song type per recording (methods in Cardoso et al. 2007). Syllable rate was quantified as the inverse of the onset-to-onset duration (i.e., duration from the beginning of a syllable to the next, identical syllable in the trill). Peak frequency was quantified as the logarithm of the frequency of highest amplitude, because log-transformed values reflect frequency changes more faithfully to the mechanisms of sound production and perception (Cardoso 2013*b*). We computed deviations in syllable rate or in peak frequency by subtracting from each syllable measurement the mean value for all syllables of the same song type in its recording, thus standardizing for differences among song types and among recordings, and then drew the distribution histogram of these deviations (i.e., absolute value of the differences) using the entire data set. Extreme cases in song duration (see "Results" for identification of extreme cases) were afterward screened for in the entire set of recordings and used in the several analyses described below. Extreme cases in deviations of syllable rate or peak frequency (see "Results") were screened for in the data set of quantitative syllable measurements noted above and used in the analyses of song consistency in the last section of the "Methods."

Population and Individual Difference in the Occurrence of Atypical Signals

Habitat differences between the UCSD campus and Mount Laguna (urbanized habitat vs. mountain forest) are associated with divergence in song frequency (Slabbekoorn et al. 2007; Cardoso and Atwell 2011) and in other sexual signals (ornamental white in the tail and black in the hood; Yeh

2004; Price et al. 2008; Atwell et al. 2014). It is thus plausible that these populations also differ in the occurrence of mistakes. We calculated the proportion of songs that had an atypical trait for each male and compared populations using a nonparametric Mann-Whitney U test for each type of atypical trait. We used nonparametric tests because proportions of atypical songs per male were strongly nonnormally distributed. On the better-sampled population (UCSD, not Mount Laguna because of lower sample sizes and, as shown later, rarer atypical song traits), we also used pair-wise nonparametric Spearman correlations (r_s) on these proportions of songs with atypical traits per male to assess whether different types of atypical traits were strongly related and, thus, redundant as possible communication cues.

Establishing whether the occurrence of atypical traits differs among individuals is a precondition for later tests that relate these song traits to phenotypes. Therefore, also on the better-sampled population, we tested whether males differ in the occurrence of atypical song traits using nonparametric Kruskal-Wallis tests comparing the proportion of songs with atypical traits across males, without making assumptions about the distribution of those proportions (Zar 1996). Here we use the different recordings of the same male as replicates, in order to assess within-male variation, and each individual male defines a group, to test the significance of differences in the frequency of atypical songs among males (e.g., Stöwe et al 2006; Jones et al. 2011).

Relation with Age, Male Phenotype, and Survival

Male juncos were assigned to one of two age classes—year after fledging (second year) and after second year—based on plumage and on banding history (details in Cardoso et al. 2012). Birds for which we could not assign an age class confidently were not included in these analyses. For males recorded in 2006 as second-year individuals but recorded again in 2007 (10 males), we excluded the recordings from 2007.

For the recorded males that were captured and color-banded (85 at UCSD and 38 at Mount Laguna), we quantified (1) body condition as the residual mass on a measure of body size (first principal component of a principal component analysis on tarsus length, head breadth, and length of head plus bill), population of origin, and capture date (to correct for population differences and seasonal and year effects; see details in Cardoso et al. 2012), and measured (2) hematocrit level (also corrected for population differences and seasonal and year effects), (3) ectoparasite load (load of feather mites in the tail, corrected for population differences and seasonal and time-of-day effects), (4) maximum testosterone level (after stimulation with injected gonadotropin-releasing hormone; marginal mean values for each male as summarized across repeated measurements and corrected for date, body mass, handing time, and capture method), and (5) the size of the white patch on the tail and of (6) the homogeneous black on the hood (both of which are ornamental phenotypes).

The above data are described in Cardoso et al. (2012) or Atwell et al. (2014), to which we refer for detailed methods. In addition to body condition (evaluated as the residual mass), we considered hematocrit, because high hematocrit levels normally indicate better body condition on avian species (Brown 1996; Fair et al. 2007), and ectoparasite load, as a general indication of susceptibility to infestation. We considered testosterone level because, in juncos, it mediates a trade-off between investment in sexual signals and mating effort versus in parental care and survival (Ketterson et al. 1992; Enstrom et al. 1997; Schoech et al. 1998; McGlothlin et al. 2008a, 2008b, 2010).

For UCSD, where breeding juncos are nearly a closed population (Yeh and Price 2004), we assessed whether the males that were individually marked survived to the next season. Survival of males recorded in 2006 was determined by censuses during fieldwork in 2007 (February–July), and survival of males recorded in 2007 was determined by a thorough census conducted for this purpose during May 2008. For males recorded both in 2006 and 2007, we assigned the mean of these two survival scores.

Also using data from UCSD, the better-sampled population, we assessed whether some of the above phenotypes (age, condition, and survival) were strongly correlated so as to justify excluding redundant variables. The strongest Pearson correlation coefficient (r_p) was between the size of black on the hood and age ($r_p = 0.46$, $P < .001$, $N = 84$ males), but the correlation coefficient (0.46) was not strong enough to qualify one of these phenotypic traits as redundant or to cause multicollinearity problems. Absolute values of all other correlation coefficients were lower than 0.22 ($P > .04$; N varies between 73 and 87 males depending on the pair of phenotypic traits).

Because atypical signals occur rarely, count data on their occurrence across males had a large number of zeros and were often overdispersed (the variance largely exceeding the mean). We chose to analyze these overdispersed data with generalized linear models (GLM), modeling the dependent variable (counts of atypical songs per male) with a negative binomial distribution and log link function (e.g., Boveng et al. 2003), which allows different values for the mean and the variance and is thus appropriate to deal with overdispersion (O'Hara and Kotze 2010). Because we have various candidate phenotypes that may influence the occurrence of atypical song traits, we fit models using Akaike's information criterion (AIC). We computed AICc (AIC corrected for small samples size) for all possible combinations of the eight predictors (age, body condition, hematocrit, parasites, testosterone, size of white patch on the tail, size

of black on the hood, and survival), always including the number of songs screened for each male as a fixed covariate in the models to control for the obvious effect of sample size on the number of atypical traits counted. The model with the lowest AICc is the best fit, but support for models within an interval of 2 AICc is not statistically distinguishable (Burnham and Anderson 2002). For every type of atypical song trait tested, there were other models less than 2 AICc from the best one, and we calculated model-averaged standardized partial coefficients ($\hat{\bar{\beta}}_{st}$) among these to obtain estimates of the effect of each predictor taking into account model selection uncertainty. If some GLM models did not converge (i.e., estimation of coefficients did not stabilize [Hilbe 2015], likely due to too many factors for the sample size), they were not included in model averaging. Model averaging is done by computing the Akaike weights (the relative estimated probability of a model being the best) for each model and then averaging the estimated standardized partial coefficients across all models within 2 AICc from the best, weighted by their Akaike weights (Symonds and Moussalli 2011). All predictors were rescaled before analysis (mean = 0, SD = 1) so that the predictor partial coefficients are comparable. The relative importance of each predictor can also be measured by summing the Akaike weights of the models in which a given predictor was selected. Predictors included in all models within 2 AIC from the best have a relative importance value of 1, and predictors that are rarely included have low relative importance. This can be interpreted as the probability of a predictor to be in the true best model (Symonds and Moussalli 2011). We report the above two statistics, $\hat{\bar{\beta}}_{st}$ and relative importance. For guidance to readers accustomed to more conventional statistical approaches, we also report the simple standardized partial coefficients (β_{st}) and the *P* values of predictors using the most complete model (i.e., the one with most predictors) within the 2 AICc interval from the best-fit one.

Relation with Mating and Paternity Success

For a subset of males (38 in UCSD, 19 in Mount Laguna) we have information on mating and paternity success, based on nest monitoring and paternity assignment through genotyping of nestlings. We refer to Atwell et al. (2014) for details on field and molecular methods characterizing paternity success. Here we used the following measures of mating and paternity success: the total number of mates (including extrapair mates), the number of extrapair young gained (genetic offspring of a male in nests outside the social pair), the number of young lost to cuckolding (nestlings in own nest fathered by other males), and the total number of young per male (genetic offspring within-pair and extrapair). For males recorded in a single year (2006 or 2007), only the measures

of reproductive mating and paternity success relative to that breeding season were used, and for males recorded in both years, these measures were averaged across the two breeding seasons.

We used GLMs similar to the ones in the previous section to test for a relation between counts of atypical song traits per male (dependent variable) and measures of mating and paternity success (predictors). As before, the number of songs screened was fixed in the models to control for its effect on the number of atypical traits counted. Additional male phenotypes that the previous analyses had shown to be related to the occurrence of atypical song traits were also included in the models so as to test for a relation between the occurrence of atypical traits and mating or paternity success while controlling for the effect of the relevant confounding phenotypic traits. Unlike in the previous section, the predictors here are correlated by definition, because they all use the same paternity data. Therefore, we did not use model averaging (which would dilute the importance of each predictor amid that of correlated measures) but instead used AICc to select the best model, which selects the most relevant reproductive variables to the exclusion of correlated but less well-fitted alternatives.

Note that the statistical models in this section reverse the direction of the causal effect: we are interested in whether signaling mistakes affect paternity, but we modeled counts of atypical song traits as the dependent variable instead. This is necessary to account for the overdispersion of the occurrence of atypical traits, and it is also convenient to sort among the alternative aspects of mating behavior and to suggest which is the most relevant.

Relation with Vocal Demands of Song

Metrics for different aspects of song performance attempt to compare vocal demands among different songs and are often based on trade-offs between acoustic traits suggestive of motor or physiological constraints (reviewed in Podos et al. 2009). Here we used three such metrics of performance for junco song. Below is a brief rationale for each, and we refer to Cardoso et al. (2009) for detailed explanations and methods on their measurement.

Vocal deviation: Calculated as the distance to an upper limit of the relation between frequency bandwidth and repetition rate of syllables. In juncos, as in a number of other species with trilled songs (Wilson et al. 2014), there is a negative upper boundary for the relation on how fast birds can repeat a syllable and how widely they can modulate its sound frequency (Cardoso et al. 2007). Vocal deviation is calculated as the distance from this upper boundary, such that lower vocal deviations (i.e., songs that approach the upper boundary), indicate higher performance (Podos 2001).

Residual intervals: Quantifies the duration of intervals between syllables in relation to the airflow demands of different syllable types. In some species, longer syllables require longer intervals for minibreaths (e.g., Hartley and Suthers 1989), but in juncos, the covariation between the duration of syllables and of the intervening intervals is only apparent when controlling for syllable traits whose phonation may affect airflow (Cardoso et al. 2007). Therefore, we use the residuals from the multiple regression of duration of intervals on syllable traits in Cardoso et al. (2007); this is an improved way to assess airflow performance relative to measures based solely on the duration of syllables and/or intervals. Lower residual intervals indicate higher performance, since the bird uses smaller intervals than predicted for recovery of air volume.

Predicted amplitude: Uses the trade-off between sound amplitude and syllable complexity, as evaluated from changes in relative amplitude when juncos switch song types within a song bout (i.e., when singing with identical motivation), and is calculated as the predicted value from the multiple regression of relative sound amplitude on syllable traits in Cardoso et al. (2007). The amplitude of birdsong is very labile (Anderson et al. 2008), such that any two song types can be sung at identical amplitudes. The objective of computing predicted amplitude is not to predict actual amplitudes of the different song types but instead to infer vocal demands when juncos sing different song types at the same, high amplitude. The assumption is that song traits associated with lower amplitudes should be more demanding to sing at a given, high amplitude. Thus, a lower predicted amplitude indicates a more demanding, higher-performance song type.

All these metrics of performance differ strongly among song types and vary only slightly within renditions of the same song type (Cardoso et al. 2009). Therefore, we tested whether more demanding (i.e., higher-performance) song types increase the occurrence of atypical traits using within-individual comparisons, with the statistical units being the different song types on the repertoire of each male. The data on mean song performance per song type/male came from Cardoso et al. (2012), to which we added values of predicted amplitude not used in that earlier study. Here we analyzed only atypical signals that can be interpreted as motor failure (incomplete syllables, modified syllables, very short songs) and therefore could be affected by the motor demands of song. For each type of atypical song trait, we ran a generalized linear mixed model (GLMM) with a negative binomial distribution and log link function, with the count of atypical traits in each song type/male as the dependent variable, the number of songs screened and the average song performance of each song type/male as independent variables, and male identity as a random effect to control for differences between individuals in the occurrence of atypical song

traits. We discarded individuals with only one song type recorded or with zero atypical traits for all their song types, because these individuals are uninformative in these within-repertoire comparisons (i.e., they have no within-repertoire variation in the occurrence of atypical song traits). Therefore, sample sizes differ in the analyses of different types of atypical traits.

Consistency versus Extreme Deviations in Syllable Rate and Frequency

With the same measurements of syllable rate and peak frequency used to identify putative mistakes (i.e., extreme syllable deviations in those traits; see first section of "Methods"), we also calculated coefficients of variation in syllable rate and in syllable peak frequency for each song recording. This is a standard metric of song consistency (i.e., the ability to replicate acoustic features in songs; Sakata and Vehrencamp 2012) independent of scale (e.g., independent of differences in syllable rate or frequency among song types). We calculated coefficients of variation using all syllables of the same type in each recording and then averaged across syllable types for those recordings with more than one syllable type. Next, we averaged across the recordings of each male to obtain a single score of song consistency per male.

We tested whether song consistency is related to male age, condition, or mating or paternity success, using statistical approaches similar to those described in the previous sections of the "Methods," with two differences: GLMs here used normal distribution and identity link function because the coefficient of variation in peak frequency and in syllable rate (the latter after $\log(x)$ transformation) were approximately normally distributed across males (Kolmogorov-Smirnov tests: $Z = 0.92$, $N = 71$, $P = .36$ and $Z = 1.17$, $N = 71$, $P = .13$, respectively), and the number of songs screened per male was not used in the models because here the dependent variable is an average (rather than a count). To evaluate whether the signaling value of song consistency was driven by large and discrete deviations in syllable rate or frequency (i.e., putative mistakes) rather than small and diffuse deviations, we replicated these tests instead using counts of extreme deviations, as identified from the analysis of distribution histograms; since these are count data, we used the exact same statistical approaches described in previous sections of the "Methods."

GLMs, model averaging, and GLMMs were performed in R, version 2.15.3 (R Development Core Team 2013), with the packages MASS (Venables and Ripley 2002), MuMIn (Bartoń 2013), and glmmADMB (Fournier et al. 2012). All other tests were performed in SPSS, version 21 (SPSS IBM, New York). All data sets used in this article are deposited in the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.c6344 (Ferreira et al. 2016).

Results

Frequency of Occurrence of Atypical Signals

The different types of discretely identified atypical signals (putative mistakes) ranged from <2% of songs with breaks to 5% of songs with incomplete syllables (fig. 2*A*). The histogram of song duration did not show a bimodal or disjunct distribution but did show that songs of up to 1 s in duration are infrequent and that at approximately 1 s, the frequency of occurrence increases abruptly (fig. 2*B*). We therefore used 1 s as the threshold to screen for atypically short songs, suggestive of premature interruption in singing, which totaled 6% of all songs (fig. 2*A*).

The right tail of the distribution of song duration is gradual and does not suggest a threshold for atypically long songs (fig. 2*B*). Furthermore, unlike very short songs, long songs in juncos likely reflect highly motivated singing (Cardoso et al. 2009) rather than a motor disruption. Thus, we did not further consider long songs as possible mistakes. Histograms of deviations in syllable rate or peak frequency also showed gradual changes in the distributions, not suggestive of a threshold for atypically large deviations (fig. 2*C*, 2*D*).

We therefore did not define discrete atypical song traits based on deviations in syllable rate or peak frequency. Nonetheless, in order to assess whether the signaling value of song consistency may be driven by extreme deviations (see "Consistency versus Extreme Deviations in Syllable Rate and Frequency"), we defined extreme deviations using as an arbitrary threshold the 5% most extreme cases, which approximates the frequencies of other atypical traits studied (fig. 2*A*).

Population and Individual Difference in the Occurrence of Atypical Signals

Males in UCSD sang atypical syllables more often than in Mount Laguna (Mann-Whitney U test, $Z = 2.46$, $P = .01$; table 1; fig. 3) and tended to sing atypically short songs more often ($Z = -1.80$, $P = .07$; table 1). Occurrence of atypical traits in the remaining categories did not differ significantly between populations (table 1). Among males from UCSD, individuals differed significantly in the frequency of occurrence of all types of atypical traits (Kruskal-Wallis tests $\chi^2 > 156$, $P < .01$; table 1).
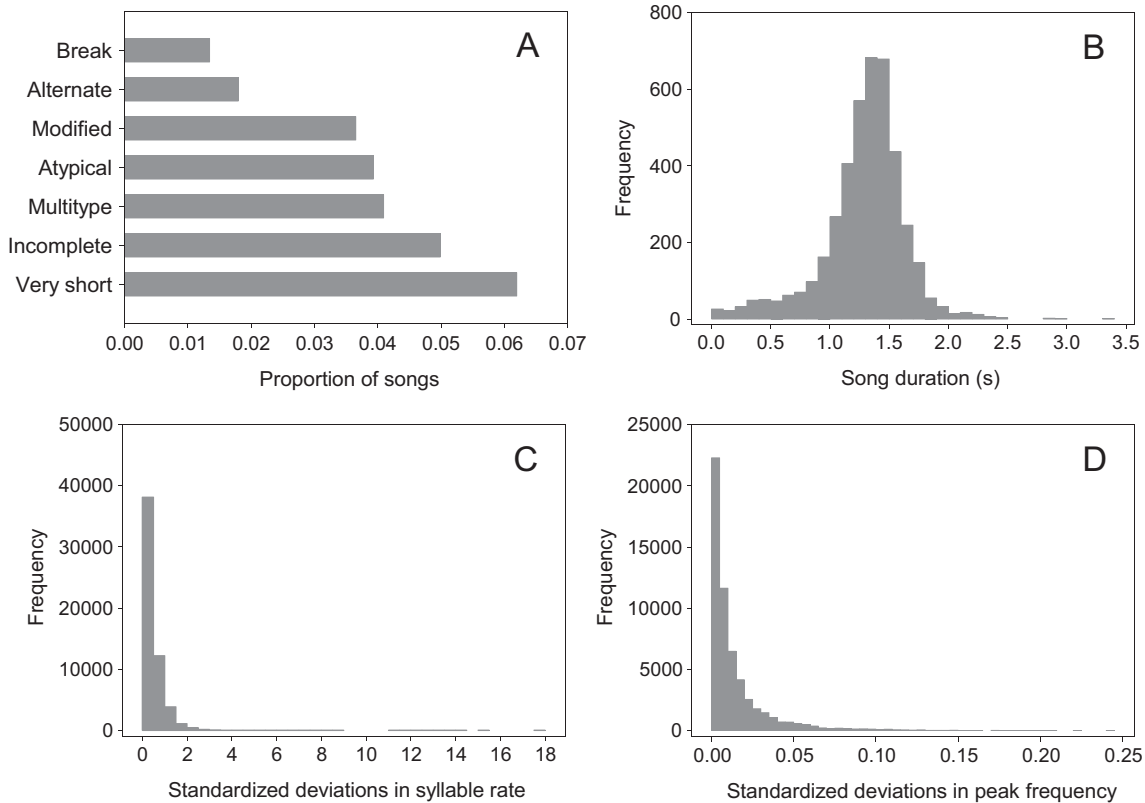

Figure 2: *A*, Proportion of songs analyzed ($N = 9{,}746$) containing each of six categories of mistakes. *B–D*, Frequency distributions for song duration and for the degree of deviations in syllable rate and syllable peak frequency within trills.

Table 1: Differences between junco populations at University of California at San Diego (UCSD) and Mount Laguna, and among individuals in each population

Characteristic	Differences between populations: Mann-Whitney test, Z (P)	Differences among individuals at UCSD: Kruskal-Wallis test, χ^2 (P)
Multitype songs	−1.63 (.10)	219.54 (<.001)
Atypical syllables	−2.46 (.01)	178.68 (<.001)
Incomplete syllables	−1.49 (.14)	158.20 (.001)
Very short songs	−1.80 (.07)	156.10 (.005)
Break	−.66 (.51)	258.86 (<.001)
Alternate	−.26 (.80)	272.36 (<.001)
Modified syllables	−1.15 (.25)	181.75 (<.001)

Note: Data are from 115 males at UCSD and 50 at Mount Laguna (except for incomplete syllables: 107 and 45 males, respectively).

In the remaining analyses, we used only data from UCSD because some atypical song traits were less frequent in Mount Laguna and the sample size was much smaller (33 males for which we also had information on age and physical traits, compared to 72 males in UCSD), resulting in low statistical power. Also because of reduced statistical power, here we do not analyze further the atypical trait with the lowest frequency of occurrence (breaks in song; fig. 2A). The proportions of songs with incomplete or with modified syllables were correlated across UCSD males ($r_s = 0.48$, $P < .001$), suggesting perhaps some redundancy among these types of atypical song traits. But all other pair-wise correlations were lower (multitype songs vs. atypical syllables: $r_s = 0.29$, $P < .001$; multitype songs vs. modified syllables: $r_s = 0.30$, $P = .001$; modified syllables vs. atypical syllables: $r_s = 0.19$, $P = .04$; modified syllables vs. short songs: $r_s = 0.21$, $P = .02$; all other: $|r_s| < 0.15$, $P > .13$, $N = 107$ males for correlations involving incomplete syllables, $N = 115$ for the remaining), indicating that different atypical traits may have distinct communication value. We thus looked at the different atypical song traits separately in the following analyses.

Relation with Age, Male Phenotypes, and Survival

Here we used UCSD males with information for all predictors (72 males), since AICc values are comparable only when models have the same sample size. Table 2 shows the results of AICc-based model averaging for tests relating age and physical traits of males to the occurrence of atypical song traits and the conventional partial regression coefficients of predictors for the most complete model within the best-supported models (i.e., within 2 AICc from the best).

Compared to second-year males, older males (greater than second year) were less likely to sing atypical syllables

($\hat{\bar{\beta}}_{st} = -1.99$, relative importance $= 1$; fig. 4A) or multitype songs ($\hat{\bar{\beta}}_{st} = -1.17$, relative importance $= 1$; fig. 4B) and alternated less often a song type without repeating it ($\hat{\bar{\beta}}_{st} = -0.90$, relative importance $= 0.74$; fig. 4C; table 2). Males with larger ornamental white patches on the tail sang very short songs more often ($\hat{\bar{\beta}}_{st} = 0.42$, relative importance $= 1$) and alternated a song type without repetition more often than less ornamented males ($\hat{\bar{\beta}}_{st} = 0.99$, relative importance $= 0.72$; table 2). Males with higher hematocrit levels also sang more often very short songs ($\hat{\bar{\beta}}_{st} = 0.31$, relative importance $= 1$). As expected, the number of songs recorded per male (fixed a priori in all models) was strongly related to atypical trait counts in most cases (table 2).

The remaining male phenotypes had low relative importance (<0.66). Some of them had significant effects under conventional statistical testing (using only the most complete model; table 2), suggesting that the occurrence of atypical song traits is influenced by yet other male phenotypes. But since they had low relative importance in model averaging, we do not draw strong conclusions regarding them.

Relation with Mating and Paternity Success

Here we used the subset of males from UCSD for which we have information on paternity (up to 38 males; mean songs screened per male: 74 ± 62 SD). Males that either sang multitype songs more often, alternated a song type without repetition more often, or sang modified syllables more often had less total paternity (i.e., within- plus extrapair paternity) than did males singing with fewer atypical traits (standardized partial coefficient, $\beta_{st} = -2.90$, $P = .001$; $\beta_{st} = -1.41$, $P = .05$; and $\beta_{st} = -1.27$, $P < .01$, respectively; table 3; fig. 5). Occurrence of the other types of atypical song traits was not significantly related to any measure of mating or
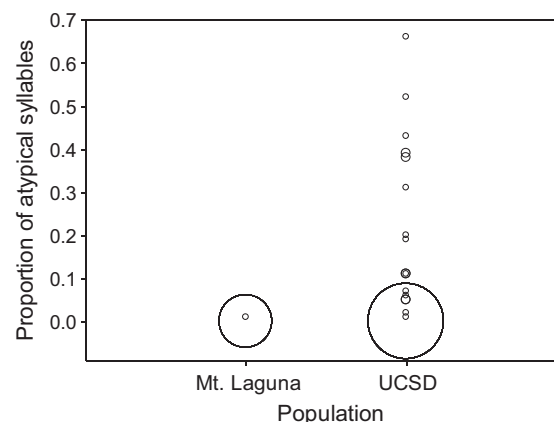

Figure 3: Differences between populations in the proportion of songs with atypical syllables. Here and in figures 4–6, the area of the circles is proportional to the square root of the number of males.

Table 2: Relationship between the occurrence of mistakes and male traits

Mistake, predictors	Model averaging of the best-supported models: $\hat{\beta}_{st}$ (SE), relative importance	Most complete GLM among best-supported models: β_{st} (P)
Atypical syllables (5):		
Age	−1.99 (.82), 1	−1.41 (.004)
Black hood	1.32 (.65), .47	NA
White patch	−1.18 (.50), .19	−1.18 (.02)
Parasites	.981 (.49), .37	1.08 (.03)
Testosterone	.49 (.46), .13	NA
Songs screened	1.87 (.51), NA	2.11 (<.001)
Multitype songs (4):		
Age	−1.17 (.43), 1	−1.14 (.007)
Testosterone	.90 (.43), .25	.90 (.04)
Hematocrit	−.69 (.48), .19	NA
Body size	.55 (.44), .19	NA
Songs screened	.54 (.43), NA	.40 (.32)
Alternation (9):		
Age	−.90 (.44), .74	−1.06 (.01)
White patch	.99 (.41), .72	.85 (.03)
Survival	.80 (.42), .66	.70 (.06)
Body size	.70 (.45), .30	.64 (.15)
Songs screened	.89 (.41), NA	.67 (.04)
Modified syllables (17):		
Age	−.47 (.31), .56	−.51 (.10)
White patch	.24 (.30), .04	NA
Survival	−.41 (.29), .39	−.39 (.15)
Black hood	−.78 (.33), .90	−.60 (.07)
Hematocrit	−.39 (.30), .48	−.40 (.18)
Parasites	−.38 (.27), .38	−.34 (.20)
Songs screened	.85 (.27), NA	.94 (<.001)
Incomplete syllables (13):		
White patch	.15(22), .09	NA
Survival	−.18 (.22), .10	NA
Black hood	.29 (.22), .25	.24 (.27)
Body size	.34 (.23), .28	.29 (.19)
Hematocrit	.29 (.22), .14	NA
Songs screened	1.03 (.22), NA	1.09 (<.001)
Very short songs (3):		
White patch	.42 (.16), 1	.49 (.002)
Hematocrit	.31 (.16), 1	.33 (.04)
Black hood	−.19 (.16), .28	−.19 (.24)
Testosterone	.13 (.15), .21	NA
Songs screened	.59 (.15), NA	.56 (<.001)

Note: Model averaging was based on models within the 2Δ corrected Akaike information criterion (AICc) interval from the best model (number indicated in parentheses beside each type of mistake). Shown are predictors in at least one model within the 2ΔAICc interval from the best model or, in the last column, predictors in the most complete model within this 2ΔAICc interval. Number of songs screened was fixed in all models. GLM = generalized linear models; NA = not applicable.

paternity success, and as expected, the number of songs screened per male was, in most cases, positively related to atypical trait counts (table 3). For atypical syllables, statistical models failed to converge, likely due to low frequency of atypical syllables in this sample.

Atypical Signals and Song Performance

Table 4 shows results from GLMMs relating atypical trait counts with the performance of song types within male repertoires. Contrary to the prediction that vocally demanding songs would cause more frequent mistakes, incomplete syl-
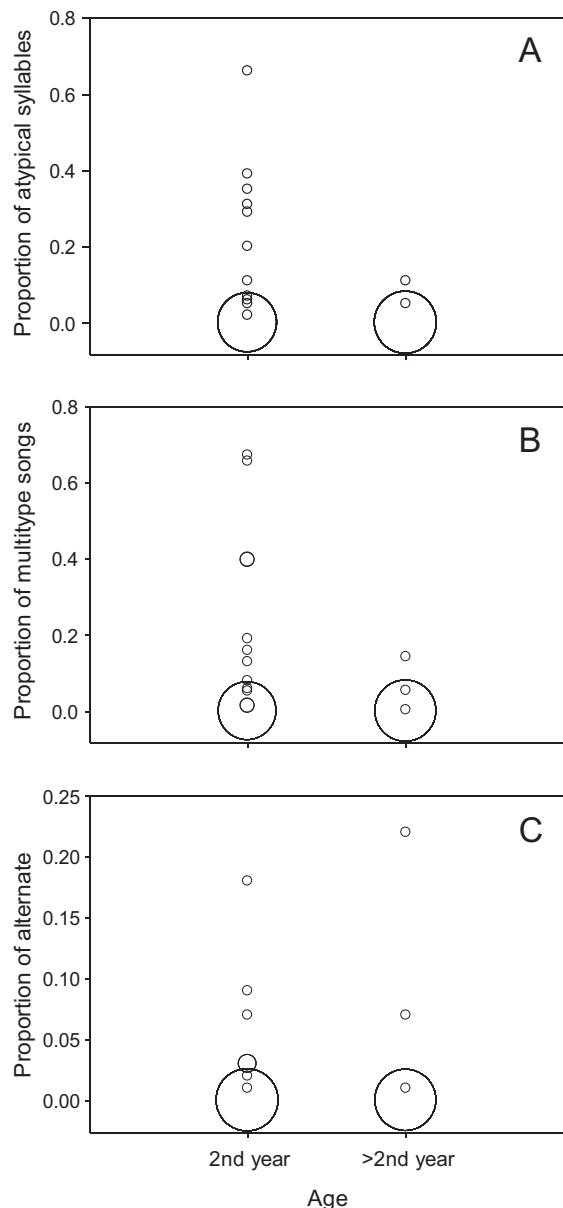
Figure 4: Proportion of songs with atypical syllables (*A*), multitype songs (*B*), and song bouts alternating song types without repeating them (*C*) in young (second-year) versus older (greater-than-second-year) males.

lables were more frequent in song types with high vocal deviation (i.e., with lower performance demands; $\beta_{st} = 0.80$, $P = .004$), and very short songs were more frequent in song types with higher residual intervals (i.e., with lower performance demands; $\beta_{st} = 0.29$, $P = .02$; table 4). Therefore, less demanding singing, as evaluated by vocal deviation and residual intervals, was associated with the occurrence of more atypical songs.

Consistency versus Extreme Deviations in Syllable Rate and Frequency

Consistency in syllable rate was related to male age, with older males being more consistent (i.e., having lower within-song-type coefficients of variation: $\bar{\bar{\beta}}_{st} = -0.30$, relative importance $= 1$; table 5; fig. 6*A*). We did not find evidence for consistency in syllable rate affecting survival, mating, or paternity success (table 5).

We did not find strong predictors of consistency in syllable frequency (table 5). However, in the subset of males with mating and paternity data, males with more consistent peak frequency of syllables (i.e., lower coefficients of variation) had significantly more total paternity ($\beta_{st} = -0.46$, $P = .006$; table 5; fig. 6*B*).

Similar analyses focusing instead on extreme deviations (5% most extreme deviations in syllable rate or peak frequency) showed stronger associations with paternity success. As had been the case for some other atypical traits, extreme deviations in syllable rate were more frequent in males with larger tail white patches ($\bar{\bar{\beta}}_{st} = 0.37$, relative importance $= 1$). Accounting for this effect, extreme deviations in syllable rate or in peak frequency were both strongly related to lower gains of extrapair paternity ($\beta_{st} = -0.69$, $P < .01$ and $\beta_{st} = -0.79$, $P < .01$, respectively; table 5; fig. 6*C*, 6*D*).

Discussion

We found several types of atypical song traits in dark-eyed juncos, ranging in frequency from <1% to 6% of songs

Table 3: Best generalized linear model (i.e., model with lowest corrected Akaike information criterion) relating the occurrence of each type of mistake with components of reproductive success

Mistake, predictors	β_{st} (SE)	P
Multitype songs:		
Total young	−2.90 (.90)	.001
Testosterone	2.81 (.754)	<.001
Songs screened	1.00 (.49)	.04
Alternation:		
Total young	−1.41 (.71)	.05
Songs screened	2.05 (.46)	<.001
Modified syllables:		
Total young	−1.27 (.37)	<.001
Songs screened	.81 (.28)	.004
Incomplete syllables:		
Total young	.48 (.29)	.10
Songs screened	1.20 (.28)	<.001
Very short songs:		
No. mates	−.20 (.18)	.27
White patch	.31 (.17)	.07
Songs screened	.93 (.18)	<.001

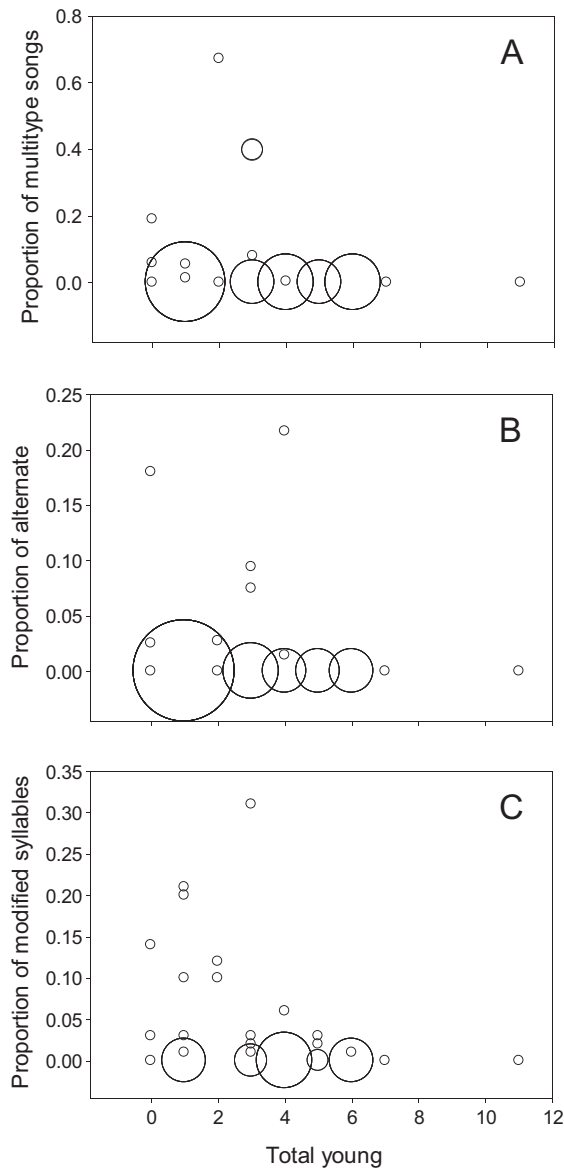Note: The number of songs screened was fixed in all models.

Figure 5: Proportion of multitype songs (*A*), alternation between song types without repeating them (*B*), and songs with modified syllables (*C*), in relation to total young (within-pair plus extrapair young) fathered by each male.

(fig. 2*A*). Some of these atypical traits could be mistakes that occur during song production, and others suggest atypical song development. The occurrence of different types of putative mistakes was generally weakly or not correlated across individuals, suggesting that different atypical song traits may have distinct underpinnings and communication value. The frequency of most atypical song traits differed between individuals, and one type (atypical syllables) was more frequent in the urbanized habitat of UCSD compared to a native forested habitat in the same region (fig. 3; table 1). Some atyp-

ical song traits were less frequent in older males and more frequent in more ornamented males (fig. 4; table 2), showing that these atypical traits can be informative regarding aspects of male quality. Also, some atypical traits were negatively related to the genetic paternity success of males (fig. 5; table 3), supporting interpreting them as song mistakes and suggesting that females assess and respond negatively to these mistakes in male song; this requires direct, experimental confirmation. Contrary to expectations, we found that males sang fewer atypical songs when using more vocally demanding song types, as evaluated by metrics of song performance (table 4). Similarly to studies on other species (reviewed in Sakata and Vehrencamp 2012), we found higher song consistency (similarity across renditions of the same syllable) in older males (fig. 5*A*; table 5); this result was not due to extreme and infrequent deviations to syllable structure (i.e., discrete mistakes) but to a continuum of smaller and larger deviations. We also found that higher song consistency was related to paternity success but that extreme deviations to syllable structure show even stronger relations with extrapair paternity (fig. 6; table 5). We discuss these findings in turn.

Atypical syllables were more common in the urbanized habitat of UCSD compared to the native habitat in the same region, and there was also a trend for UCSD males to sing more atypically short songs. This is similar to the finding that urban great tits (*Parus major*) more often sing phrases with an atypical number of syllables (Slabbekoorn and Boer-Visser 2006). A possible explanation is that the urban environment

Table 4: Results from a generalized linear mixed model relating the occurrence of mistakes (only those mistakes that may reflect motor failure) with song performance of the different song types

Mistake, predictors	β_{st} (SE)	P
Modified syllables:		
Predicted amplitude	.48 (.30)	.11
Vocal deviation	.16 (.26)	.55
Interval residuals	.39 (.30)	.20
Songs screened	.91 (.26)	<.001
Incomplete syllables:		
Predicted amplitude	.018 (.21)	.93
Vocal deviation	.80 (.28)	.004
Interval residuals	−.23 (.33)	.49
Songs screened	1.15 (.30)	<.001
Very short songs:		
Predicted amplitude	−.23 (.13)	.08
Vocal deviation	.09 (.12)	.47
Interval residuals	.29 (.12)	.02
Songs screened	.69 (.13)	<.001

Note: Sample sizes for the analysis of modified syllables: 135 combinations of song types × male, from a total of 35 males; incomplete syllables: 74 combinations of song types × male, from a total of 28 males; very short songs: 240 combinations of song types × male, from a total of 81 males.

Table 5: Relationship between song consistency (coefficients of variation in syllable rate or peak frequency; top half of the table) or the occurrence of extreme deviation in syllable rate or peak frequency (lower half) and either male traits (left columns) or reproductive output (right columns)

Variable, predictors	Relations with male traits		Relations with reproductive output	
	Model averaging of the best-supported models: $\hat{\beta}_{st}$ (SE), relative importance	Most complete model among best-supported models: β_{st} (P)	Predictor	Best model: β_{st} (P)
Song consistency:				
CV syllable rate (9):				
Age	−.30 (.11), 1	−.30 (.009)	No. mates	−.05 (.79)
White patch	.21 (.12), .85	.21 (.06)	Age	−.27 (.14)
Survival	.16 (.11), .34	.14 (.22)		
Body size	.09 (.12), .08	NA		
Hematocrit	−.11 (.12), .10	NA		
Parasites	.19 (.11), .65	.17 (.14)		
Testosterone	.08 (.12), .07	NA		
CV peak frequency (9):				
White patch	.11 (.12), .11	NA	Total young	−.46 (.006)
Survival	−.05 (.12), .07	NA		
Testosterone	.14 (.12), .23	.14 (.25)		
Black hood	−.11 (.12), .18	NA		
Hematocrit	.15 (.12), .31	.14 (.24)		
Parasites	.05 (.12), .07	NA		
Extreme deviations:				
Syllable rate (3):				
White patch	.37 (.14), 1	.38 (.003)	Extrapair young	−.69 (<.001)
Black hood	.14 (.14), .16	NA	White patch	.61 (.001)
Body size	−.12 (.14), .14	NA	Syllables screened	1.07 (<.001)
Hematocrit	−.17 (.14), .20	−.17 (.20)		
Testosterone	−.14 (.14), .16	NA		
Syllables screened	.80 (.14), NA	.75 (<.001)		
Peak frequency (3):				
Age	−.31 (.16), .49	−.31 (.07)	Extrapair young	−.79 (<.001)
White patch	.10 (.16), .14	.10 (.54)	Syllables screened	.92 (<.001)
Black hood	−.19 (.13), .22	NA		
Parasites	.09 (.16), .14	NA		
Syllables screened	.43 (.17), NA	.38 (.02)		

Note: Relations with male traits used model averaging (of models within the 2Δ corrected Akaike information criterion [AICc] interval from the best model; number indicated in parentheses beside each type of mistake), and relations with reproductive success used the best model (AICc criterion). The number of syllables screened was fixed in the models analyzing the occurrence of extreme deviations. NA = not applicable.

has detrimental effects on song production or development. One could hypothesize that distraction by human activities interferes with singing or that anthropogenic noise or other stressors interfere with song learning. Juvenile juncos deprived from contact with adult songs develop atypical songs (Marler et al. 1962), and if anthropogenic noise hinders juvenile acoustic experience, then it might similarly facilitate the development of atypical syllables. The higher incidence of atypical songs in cities is an intriguing phenomenon yet to be further investigated.

Of the male phenotypes studied, age, the ornamental white tail patch, and hematocrit level were the most related to atypical song traits. Additional phenotypes might also influence the occurrence of mistakes, but we do not draw strong conclusions from these additional traits because their relative importance was low on model averaging (table 2). Older males sang three types of atypical songs less frequently (atypical syllables, multitype songs, and alternating a song type without repetition), which could be due to either improved singing with age and/or increased survival of better singers. In either case, these atypical songs give an indication of age, which is an important aspect of male quality because it may indicate parental ability or investment (reviewed by Forslund and Pärt 1995 and Bradley and Saffran 2014; in the junco: Reed et al. 2006). The relation between atypical songs and age agrees with the view that these are singing or develop-
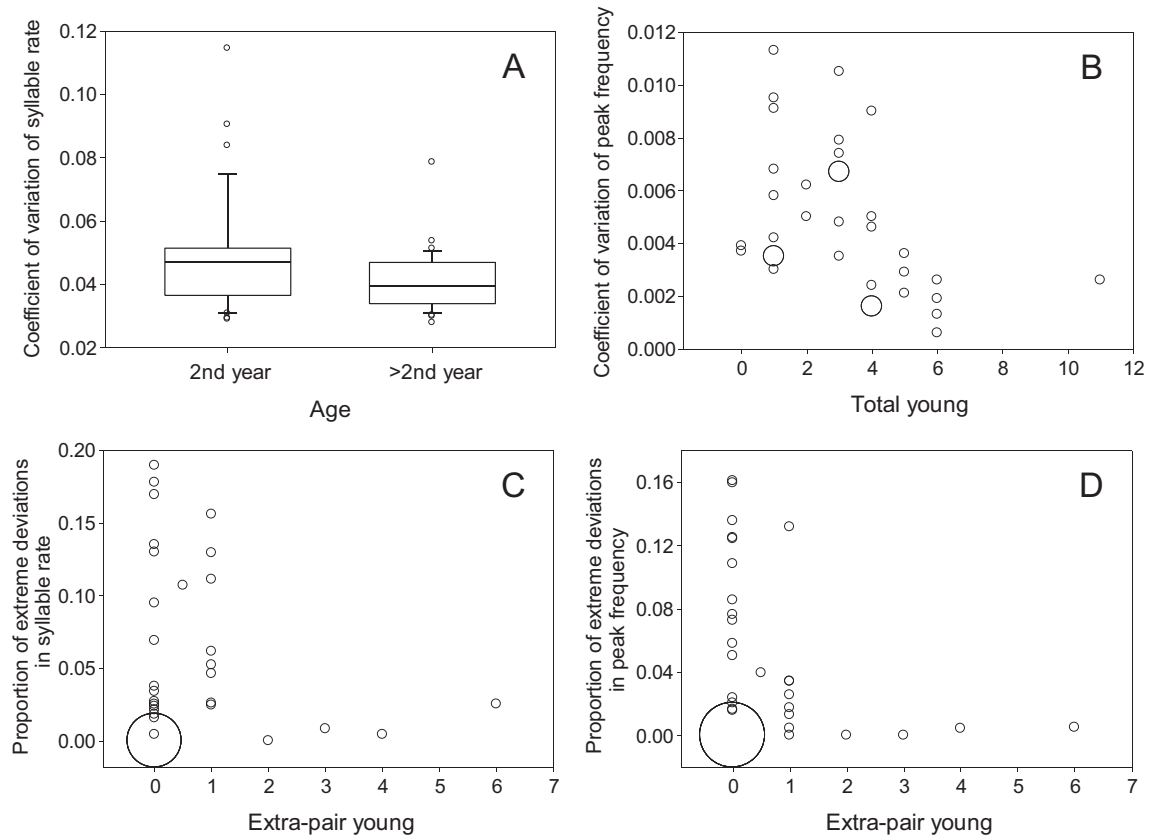
Figure 6: Consistency in syllable rate in relation to male age (*A*) and consistency in peak frequency in relation to total paternity (within-pair plus extrapair young; *B*). Proportion of songs with extreme deviations in syllable rate (*C*) and proportion of songs with extreme deviations in peak frequency (*D*) in relation to the number of extrapair young gained.

mental mistakes and that mistakes reflect negatively on individual quality.

Unexpectedly, males with more ornamental white in the tail and higher hematocrit (both suggestive of high quality) sang two types of atypical songs more often (alternating a song type without repetition and/or very short songs). The white tail patch is preferred by females and indicates dominance in juncos (Holberton et al. 1989; Hill et al. 1999; McGlothlin et al. 2005; but see Price et al. 2008), and therefore it was unexpected that those males should produce more atypical songs. Similarly to our result, however, a previous study also showed that junco males with more tail white have smaller song repertoires (Newman et al. 2008). An ad hoc hypothesis is that circulating testosterone causes a compromise between ornamentation versus song development. It is known that testosterone can depress immune function, increasing the likelihood of infection and stress (Folstad and Karter 1992; Lindstrom et al. 2001; Mougeot et al. 2006; in juncos: Casto et al. 2001; Greives et al. 2006) and that stress during development interferes with song learning (Buchanan et al. 1999; Nowicki et al. 2002). In the case of juncos, in-

creased testosterone in juveniles reduces adult song repertoires and output (Titus et al. 1997). Also, the size of the white patch is related to testosterone (McGlothlin et al. 2008*b*), testosterone levels in juncos are repeatable (Jawor et al. 2006; Atwell et al. 2014), and studies of other vertebrates indicate substantial heritability of circulating testosterone levels (King et al. 2004; van Oers et al. 2011). Therefore, we hypothesize that the increased frequency of atypical songs in juncos with more tail white might be due to detrimental effects of testosterone on song development (Titus et al. 1997). An alternative hypothesis could be that alternating song types is in fact a high-quality signal that emphasizes repertoire size, but we disfavor this for two reasons. First, alternating songs should not be costlier than repeating songs, and if those were good-quality signals, it would be paradoxical that they remain at low frequencies rather than having been selected and becoming widely used. Second, we found a negative relation of paternity success with alternating song types, and with multitype songs and modified syllables as well (table 3). These relations with paternity success are consistent with the view that these are singing or developmental mistakes and

low-quality signals, and that receivers (mating females or competing males) assess song mistakes.

The hypothesis that information in mistakes maintains honest communication does not require a costly signal design in order to function, but it was suggested that more challenging signals could increase the occurrence of mistakes, therefore making them more informative (Cardoso 2013*a*). Contrary to this suggestion, we found that higher-performance song types, evaluated by two metrics of song performance (vocal deviation and residual intervals) were associated with fewer atypical songs (incomplete syllables and a very short song, respectively). It is unclear why higher-performance songs would per se decrease the occurrence of these putative mistakes, and this contrasts with the finding, in a different species, that higher performance in vocal deviation decreases song consistency (Vehrencamp et al. 2013). One possibility is that our unexpected result is mediated by motivation. Juncos singing with increased motivation (longer songs, countersinging against neighbors) preferentially use song types with higher performance on the above two metrics of performance (Cardoso et al. 2009), and increased motivation might, in turn, be associated with fewer mistakes. For example, it was shown, in a different species, that more motivated singing (courtship singing as opposed to singing alone) increases the degree of neuromotor control over song, thus decreasing variability (Kao et al. 2005; Sakata et al. 2008). Therefore, it is possible that increased motivation causes an indirect relation between these aspects of song performance and the low occurrence of atypical songs. This is an ad hoc hypothesis motivated by our unexpected finding. It requires future research manipulating motivation to test whether more motivated singing buffers against the occurrence of mistakes or, on the contrary, causes stress that may enhance mistakes (Cardoso 2013*a*).

Finally, older males sang with higher consistency, and more consistent singers had higher paternity success. Studies in other species similarly found song consistency to improve with age, perhaps due to practice, to indicate male quality, and to be preferred by females (Byers 2007; Gilman et al. 2007; Botero et al. 2009; de Kort et al. 2009; Rivera-Gutierrez et al. 2010; Wegrzyn et al. 2010; Taff et al. 2012; Cramer 2013; Vehrencamp et al. 2013). The novelty of our analysis was testing whether the signaling value of consistency is driven by small deviations ubiquitous in song (Kao and Brainard 2006) or by larger deviations that may be superimposed on those (e.g., Lambrechts 1997). The magnitudes of deviations in syllable rate or peak frequency were continuously distributed in junco song, not showing clear indications for atypically large deviations. But using the 5% most extreme deviations (an arbitrary threshold, chosen for being similar to the frequencies of atypical songs), we found stronger relations with paternity success: extreme deviations in syllable rate and peak frequency were negatively related to extrapair paternity. Extreme deviations to syllable structure should be perceptually salient, and these relations with paternity success suggest that receivers (females, in this case) may respond particularly to them.

To conclude, this is the first comprehensive study of atypical communication signals, which can be viewed as developmental or signaling mistakes, and it showed or suggested a wealth of phenomena largely supportive of the hypothesis that mistakes play a role in animal communication. As had been hypothesized (Cardoso 2013*a*), some types of atypical signals were negatively related to aspects of male quality and negatively related to paternity success. We also found unexpected results, such as atypical signals whose occurrence increased with male ornamentation or decreased with aspects of song performance. These suggest influences on the occurrence of atypical signals that were not anticipated, and we proposed tentative explanations. Additionally, our results support the view that receiver responses to song consistency may be mediated by the perceptual salience of extreme deviations in song, rather than by smaller and ubiquitous deviations in song structure. We thus encourage researchers to explore and refine the study of signaling mistakes; not only of mistakes in behavioral signals but also in ornamentation, where atypical phenotypes may have major effects on receiver response (e.g., Dakin and Montgomerie 2011). Studying mistakes may be especially insightful for understanding sexual signals that do not easily fit explanations based on high costs of signal design, as is the case for the very simple dark-eyed junco songs. We look forward to learning how fruitful this approach will prove to be in other systems.
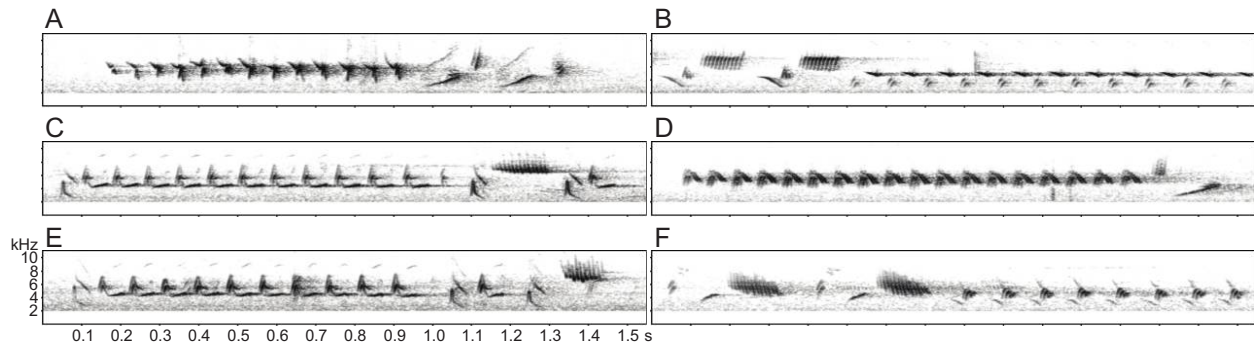
Acknowledgments

APPENDIX

Figure A1: Spectrograms of dark-eyed junco long-range songs that include atypical syllables at the onset (*B, F*) or the end (*A, C–E*) of the song. Note how these atypical syllables are easily distinguishable from the regular trilled syllable in the songs.

Literature Cited

Anderson, R. C., W. A. Searcy, S. Peters, and S. Nowicki. 2008. Soft song in song sparrows: acoustic structure and implications for signal function. Ethology 114:662–676.

Atwell, J. W., G. C. Cardoso, D. J. Whittaker, T. D. Price, and E. D. Ketterson. 2014. Hormonal, behavioral, and life-history traits exhibit correlated shifts in relation to population establishment in a novel environment. American Naturalist 184:E147–E160.

Bartoń, K. 2013. Package "MuMIn": multi-model inference. R package. Version 1.9.5. http://CRAN.R-project.org/package=MuMIn. Accessed February 05, 2015.

Botero, C. A., R. J. Rossman, L. M. Caro, L. M. Stenzler, I. J. Lovette, S. R. de Kort, and S. L. Vehrencamp. 2009. Syllable type consistency is related to age, social status and reproductive success in the tropical mockingbird. Animal Behaviour 77:701–706.

Boveng, P. L., J. L. Bengtson, D. E. Withrow, J. C. Cesarone, M. A. Simpkins, K. J. Frost, and J. J. Burns. 2003. The abundance of harbor seals in the gulf of Alaska. Marine Mammal Science 19:111–127.

Bradbury, J. W., and S. L. Vehrencamp. 1998. Principles of animal communication. Sinauer, Sunderland, MA.

Bradley, R. J., and R. J. Saffran. 2014. Conceptual revision and synthesis of proximate factors associated with age-related improvement in reproduction. Ethology 120:411–426.

Brown, M. E. 1996. Assessing body condition in birds. Current Ornithology 13:67–135.

Buchanan, K. L., C. K. Catchpole, J. W. Lewis, and A. Lodge. 1999. Song as an indicator of parasitism in the sedge warbler. Animal Behaviour 57:307–314.

Burnham, K. P., and D. P. Anderson. 2002. Model selection and multimodel inference. Springer, New York.

Byers, B. E. 2007. Extrapair paternity in chestnut-sided warblers is correlated with consistent vocal performance. Behavioral Ecology 18:130–136.

Cardoso, G. C. 2013*a*. Sexual signals as advertisers of resistance to mistakes. Ethology 119:1035–1043.

———. 2013*b*. Using frequency ratios to study vocal communication. Animal Behaviour 85:1529–1532.

Cardoso, G. C., and J. W. Atwell. 2011. On the relation between loudness and the increased song frequency of urban birds. Animal Behaviour 82:831–836.

Cardoso, G. C., J. W. Atwell, Y. Hu, E. D. Ketterson, and T. D. Price. 2012. No correlation between three selected trade-offs in birdsong performance and male quality for a species with song repertoires. Ethology 118:584–593.

Cardoso, G. C., J. W Atwell, E. D. Ketterson, and T. D. Price. 2007. Inferring performance in the songs of dark-eyed juncos (*Junco hyemalis*). Behavioral Ecology 18:1051–1057.

———. 2009. Song types, song performance, and the use of repertoires in dark-eyed juncos (*Junco hyemalis*). Behavioral Ecology 20:901–907.

Cardoso, G. C., and Y. Hu. 2011. Birdsong performance and the evolution of simple (rather than elaborate) sexual signals. American Naturalist 178:679–686.

Casto, J. M., V. Nolan Jr., ,and E. D. Ketterson. 2001. Steroid hormones and immune function: experimental studies in wild and captive dark-eyed juncos (*Junco hyemalis*). American Naturalist 157:408–420.

Cramer, E. R. 2013. Vocal deviation and trill consistency do not affect male response to playback in house wrens. Behavioral Ecology 24:412–420.

Dakin, R., and R. Montgomerie. 2011. Peahens prefer peacocks displaying more eyespots, but rarely. Animal Behaviour 82:21–28.

de Kort, S. R., E. R. B. Eldermire, S. Valderrama, C. A. Botero, and S. L. Vehrencamp. 2009. Trill consistency is an age-related assessment signal in banded wrens. Proceedings of the Royal Society B: Biological Sciences 276:2315–2321.

Enstrom, D. A., E. D. Ketterson, and V. Nolan Jr. 1997. Testosterone and mate choice in the dark-eyed junco. Animal Behaviour 54:1135–1146.

Fair, J., S. Whitaker, and B. Pearson. 2007. Sources of variation in haematocrit in birds. Ibis 149:535–552.

Ferreira, A. C., J. W. Atwell, D. J. Whittaker, E. D. Ketterson, and G. C. Cardoso. 2016. Data from: Communication value of mistakes in dark-eyed junco song. American Naturalist, Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.c6344.

Folstad, I., and A. J. Karter. 1992. Parasites, bright males, and the immunocompetence handicap. American Naturalist 139:603–622.

Forslund, P., and T. Pärt. 1995. Age and reproduction in birds—hypotheses and tests. Trends in Ecology and Evolution 10:374–378.

Fournier, D. A., H. J. Skaug, J. Ancheta, J. Ianelli, A. Magnusson, M. Maunder, A. Nielsen, and J. Sibert. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterised complex non-linear models. Optimisation Methods and Software 27:233–249.

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

Greives, T. J., J. W. McGlothlin, J. M. Jawor, G. E. Demas, and E. D. Ketterson. 2006. Testosterone and innate immune function inversely covary in a wild population of breeding dark-eyed juncos (*Junco hyemalis*). Functional Ecology 20:812–818.

Hartley, R. S., and R. A. Suthers. 1989. Airflow and pressure during canary song: direct evidence for mini-breaths. Journal of Comparative Physiology A 165:15–26.

Hilbe, J. M. 2015. Practical guide to logistic regression. CRC, Boca Raton, FL.

Hill, J. A., D. A. Enstrom., E. D. Ketterson, V. Nolan Jr., and C. Ziegenfus. 1999. Mate choice based on static versus dynamic secondary sexual traits in the dark-eyed junco. Behavioral Ecology 10:91–96.

Holberton, R. L., K. P. Able, and J. C. Wingfield. 1989. Status signalling in dark-eyed juncos, *Junco hyemalis*: plumage manipulations and hormonal correlates of dominance. Animal Behaviour 37:681–689.

Hurd, P. L., and M. Enquist. 2005. A strategic taxonomy of biological communication. Animal Behaviour 70:1155–1170.

Jawor, J. M., J. W. McGlothlin, J. M. Casto, T. J. Greives, E. A. Snajdr, G. E. Bentley, and E. D. Ketterson. 2006. Seasonal and individual variation in response to GnRH challenge in male dark-eyed juncos (*Junco hyemalis*). General and Comparative Endocrinology 149:182–189.

Jones, T. C., T. S. Akoury, C. K. Hauser, and D. Moore. 2011. Evidence of circadian rhythm in antipredator behaviour in the orb-weaving spider *Larinioides cornutus*. Animal Behaviour 82:549–555.

Kao, M. H. 2006. Lesions of an avian basal ganglia circuit prevent context-dependent changes to song variability. Journal of Neurophysiology 96:1441–1455.

Kao, M. H., A. J. Doupe, and M. S. Brainard. 2005. Contributions of an avian basal ganglia–forebrain circuit to real-time modulation of son. Nature 433:638–643.

Ketterson, E. D., V. Nolan Jr., L. Wolf, and C. Ziegenfus. 1992. Testosterone and avian life histories: effects of experimentally elevated testosterone on behavior and correlates of fitness in the dark-eyed junco (*Junco hyemalis*). American Naturalist 140:980–999.

Ketterson, E. D., P. G. Parker, S. A. Raouf, V. Nolan Jr., C. Ziegenfus, and C. R. Chandler. 1998. The relative impact of extra-pair fertilizations on variation in male and female reproductive success in dark-eyed juncos (*Junco hyemalis*). Ornithological Monographs 49:81–101.

King, R. B., J. H. Cline, and C. J. Hubbard. 2004. Heritable variation in testosterone levels in male garter snakes (*Thamnophis sirtalis*). Journal of Zoology 264:143–147.

Konishi, M. 1964. Song variation in a population of Oregon juncos. Condor 66:423–436.

Krakauer, D. C., and R. A. Johnstone. 1995. The evolution of exploitation and honesty in animal communication: a model using artificial neural networks. Philosophical Transactions of the Royal Society B: Biological Sciences 348:355–361.

Lambrechts, M. M. 1997. Song frequency plasticity and composition of phrase versions in great tits *Parus major*. Ardea 85:99–109.

Lindstrom, K. M., D. Krakower, J. O. Lundstrom, and B. Silverin. 2001. The effects of testosterone on a viral infection in greenfinches (*Carduelis chloris*): an experimental test of the immunocompetence-handicap hypothesis. Proceedings of the Royal Society B: Biological Sciences 268:207–211.

Marler, P., and D. Isaac. 1961. Song variation in a population of Mexican juncos. Wilson Bulletin 73:193–206.

Marler, P., M. Tamura, and M. Kreith. 1962. Song development in hand-raised Oregon juncos. Auk 79:12–30.

McGlothlin, J. W., J. M. Jawor, T. J. Greives, J. M. Casto, J. L. Phillips, and E. D. Ketterson. 2008a. Hormones and honest signals: males with larger ornaments elevate testosterone more when challenged. Journal of Evolutionary Biology 21:39–48.

McGlothlin, J. W., J. M. Jawor, and E. D. Ketterson. 2008b. Natural variation in a testosterone-mediated trade-off between mating effort and parental effort. American Naturalist 170:864–875.

McGlothlin, J. W., P. G. Parker, V. Nolan Jr., and E. D. Ketterson. 2005. Correlational selection leads to genetic integration of body size and an attractive plumage trait in dark-eyed juncos. Evolution 59:658–671.

McGlothlin, J. W., D. J. Whittaker, S. E. Schrock, N. M. Gerlach, J. M. Jawor, E. A. Snajdr, and E. D. Ketterson. 2010. Natural selection on testosterone production in a wild songbird population. American Naturalist 175:687–701.

Mougeot, F., S. M. Redpath, and S. B. Piertney. 2006. Elevated spring testosterone increases parasite intensity in male red grouse. Behavioral Ecology 17:117–125.

Newman, M. M., P. J. Yeh, and T. D. Price. 2008. Song variation in a recently founded population of the dark-eyed junco (*Junco hyemalis*). Ethology 114:164–173.

Nowicki, S., W. A. Searcy, M. Hughes, and J. Podos. 2001. The evolution of bird song: male and female response to song innovation in swamp sparrows. Animal Behaviour 62:1189–1195.

Nowicki, S., W. A. Searcy, and S. Peters. 2002. Brain development, song learning and mate choice in birds: a review and experimental test of the "nutritional stress hypothesis." Journal of Comparative Physiology A 188:1003–1014.

O'Hara, R. B., and D. J. Kotze. 2010. Do not log-transform count data. Methods in Ecology and Evolution 1:118–122.

Pieplow, N., and C. Francis. 2011. Song differences among subspecies of yellow-eyed juncos (*Junco phaeonotus*). Wilson Journal of Ornithology 123:464–471.

Podos, J. 1996. Motor constraints on vocal development in a songbird. Animal Behaviour 51:1061–1070.

———. 2001. Correlated evolution of morphology and vocal signal structure in Darwin's finches. Nature 409:185–188.

Podos, J., D. C. Lahti, and D. L. Moseley. 2009. Vocal performance and sensorimotor learning in songbirds. Advances in the Study of Behavior 40:159–195.

Price, T. D., P. J. Yeh, and B. Harr. 2008. Phenotypic plasticity and the evolution of a socially selected trait following colonization of a novel environment. American Naturalist 172(suppl.):S49–S62.

R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Reed, W. L., M. E. Clark, P. G. Parker, S. A. Raouf, N. Arguedas, D. S. Monk, E. Snajdr, V. Nolan Jr., and E. D. Ketterson. Physiological effects on demography: a long-term experimental study of testosterone's effects on fitness. American Naturalist 167:667–683.

Reichard, D. G., R. J. Rice, E. M. Schultz, and S. E. Schrock. 2013. Low-amplitude songs produced by male dark-eyed juncos (*Junco hyemalis*) differ when sung during intra- and inter-sexual interactions. Behaviour 150:1183–1202.

Rivera-Gutierrez, H. F., R. Pinxten, and M. Eens. 2010. Multiple signals for multiple messages: great tit, *Parus major*, song signals age and survival. Animal Behaviour 80:451–459.

Sakata, J. T., C. M. Hampton, and M. S. Brainard. 2008. Social modulation of sequence and syllable variability in adult birdsong. Journal of Neurophysiology 99:1700–1711.

Sakata, J. T., and S. L. Vehrencamp. 2012. Integrating perspectives on vocal performance and consistency. Journal of Experimental Biology 215:201–209.

Schoech, S. J., E. D. Ketterson, V. Nolan Jr., P. J. Sharp, and J. D. Buntin. 1998. The effect of exogenous testosterone on parental behavior, plasma prolactin, and prolactin binding sites in dark-eyed juncos. Hormones and Behavior 34:1–10.

Searcy, W. A., and S. Nowicki. 2005. The evolution of animal communication. Princeton University Press, Princeton, NJ.

Slabbekoorn, H., and A. den Boer-Visser. 2006. Cities change the songs of birds. Current Biology 16:2326–2331.

Slabbekoorn, H., P. Yeh, and K. Hunt. 2007. Sound transmission and song divergence: a comparison of urban and forest acoustics. Condor 109:67–78.

Stöwe, M., T. Bugnyar, M.-C. Loretto, C. Schloegl, F. Range, and K. Kotrschal. 2006. Novel object exploration in ravens (*Corvus corax*): effects of social relationships. Behavioural Processes 73:68–75.

Symonds, M. R. E., and A. Moussalli. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behavioural Ecology and Sociobiology 65:13–21.

Taff, C. C., D. Steinberger, C. Clark, K. Belinsky, H. Sacks, C. R. Freeman-Gallant, P. O. Dunn, and L. A. Whittingham. 2012. Multimodal sexual selection in a warbler: plumage and song are related to different fitness components. Animal Behaviour 84:813–821.

Titus, R. C. 1998 Short-range and long-range songs: use of two acoustically distinct song classes by dark-eyed juncos. Auk 115:386–393.

Titus, R. C., E. D. Ketterson, and V. Nolan Jr. 1997. High testosterone prior to song crystallization inhibits singing behavior in captive yearling dark-eyed juncos (*Junco hyemalis*). Hormones and Behavior 32:133–140.

van Oers, K., K. L. Buchanan, T. E. Thomas, and P. J. Drent. 2011. Correlated response to selection of testosterone levels and immunocompetence in lines selected for avian personality. Animal Behaviour 81:1055–1061.

Vehrencamp, S. L., J. Yantachka, M. Hall, and S. R. de Kort. 2013. Trill performance components vary with age, season, and motivation in the banded wren. Behavioural Ecology and Sociobiology 67:409–419.

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. 4th ed. Springer, New York.

Węgrzyn, E., K. Leniowski, and T. S. Osiejuk. 2010. Whistle duration and consistency reflect philopatry and harem size in great reed warblers. Animal Behaviour 79:1363–1372.

Williams, L., and M. H. MacRoberts. 1977. Individual variation in songs of dark-eyed juncos. Condor 79:106–112.

———. 1978. Song variation in dark-eyed juncos in Nova Scotia. Condor 80:237–240.

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

Yeh, P. J. 2004. Rapid evolution of a sexually selected trait following population establishment in a novel habitat. Evolution 58:166–174.

Yeh, P. J., and T. D. Price. 2004. Adaptive phenotypic plasticity and the successful colonization of a novel environment. American Naturalist 164:531–542.

Zar, J. H. 1996. Biostatistical analysis. 3rd ed. Prentice Hall, Upper Saddle River, NJ.

Associate Editor: Elizabeth Adkins-Regan
Editor: Yannis Michalakis

A dark-eyed junco singing. Photograph by David Levinson.