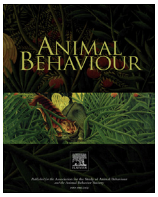
Animal Behaviour 140 (2018) 99–107

Contents lists available at ScienceDirect

Animal Behaviour

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



Animal choreography of song and dance: a case study in the Montezuma oropendola, *Psarocolius montezuma*

Meredith C. Miles, Matthew J. Fuxjager*

Department of Biology, Wake Forest University, Winston-Salem, NC, U.S.A.

ARTICLE INFO

Article history:
Received 10 January 2018
Initial acceptance 26 February 2018
Final acceptance 19 March 2018
Available online 16 May 2018
MS. number: A18-00038R

Keywords:
animal communication
birdsong
choreography
courtship
multimodal signalling

Many multimodal displays incorporate choreography, which occurs when animals modulate how body movements are timed across the display. Choreography typically involves pairing specific gestures with vocalizations. This allows the signaller to effectively produce a display that is more complex than either of its components in isolation. Moreover, some animals appear to use a special case of choreography that can augment vocal performance. Expanding the multimodal framework to incorporate choreography is therefore a necessary step towards understanding how combining two signals into one impacts a display's structure. We explore this in a case study of free-living Montezuma oropendolas, *Psarocolius montezuma*, a polygynous songbird that performs a dramatic song and dance. We found that two elements of this display (bow and wing spread), are each choreographed with the song's loudest note (dB_{max}) and lowest peak frequency (LPF), respectively. This suggests that oropendolas electively time the swing and wing spread gesture with key song elements. Interestingly, there was a correlation between the depth of an individual's swing and LPF, which was not explained by body size or social context. However, social context did predict a difference in vocal performance in terms of frequency modulation. Meanwhile, there was no relationship between wing display performance and dB_{max}. This means that oropendolas choreograph their swing gesture to predict LPF, which might reflect an individual's motor skill or even directly influence vocal performance. Altogether, our data suggest that animals can incorporate phenotypically distinct forms of choreography into their display repertoire, where each instance of choreography serves as an opportunity to generate a novel signal when one did not exist before.

Many animals communicate using complex multimodal displays, which incorporate two or more signalling elements across sensory modalities (Hebets, Vink, Sullivan-Beckers, & Rosenthal, 2013; Rowe, 1999). Each component of a multimodal display can independently influence receiver behaviour (Hebets & Uetz, 1999), but the very act of combining two signals into one can essentially generate a 'third signal' with its own unique structure and function (Taylor & Ryan, 2013; Uetz & Roberts, 2002). The temporal patterning of gestural displays (ritualized body movements) is called choreography, a phenomenon that typically involves performing certain gestures instead of others with a given acoustic signal (Dalziell et al., 2013; Ullrich, Norton, & Scharff, 2016; Williams, 2001). Considering the ongoing challenge presented by understanding how even single signals evolve and function, it is unsurprising that we know relatively little about how choreography works. Nevertheless, expanding the current multimodal framework to consider the role played by choreography is vital to advancing our understanding of how animals communicate.

Some of the most common multimodal displays include both vocalizations (calls or songs) and visual signals in the form of dance, which is at its essence a high-complexity gestural display (Fuxjager et al., 2015; Miles & Fuxjager, 2017; Miles, Cheng, & Fuxjager, 2017; Soma & Garamszegi, 2015). Although both song and dance are phenotypically diverse across animal species, there also appears to be profound variation in the ways that two signals are combined into one integrated display. First, just as human dance is coordinated with music on multiple hierarchical levels (Krumhansl & Schenck, 1997), some species exhibit variation in the fine-scale timing of dance with song (Ullrich et al., 2016), whereas other species elect to perform specific gestures with some songs over others (Dalziell et al., 2013). Although there may be functional differences in choreography at these two scales, they share one important factor: individuals can behaviourally vary the timing or

---

* Correspondence: M. J. Fuxjager, Department of Biology, Wake Forest University, Box 7325 Reynolda Station, Winston-Salem, NC, 27109, U.S.A..
E-mail address: mfoxhunter@gmail.com (M. J. Fuxjager).

https://doi.org/10.1016/j.anbehav.2018.04.006

pattern with which they pair song and dance. In other words, choreography is elective: it is a behavioural trait that is sometimes produced and sometimes not (e.g. singing without dancing or vice versa). This is fundamentally distinct from other multimodal displays constructed from components that are impossible to produce separately, as in some frog species that must inflate their vocal sac (thus producing a visual signal) to call (Starnberger, Preininger, & Hödl, 2014). We therefore make a distinction between such mechanistic dependence and true choreography because the former case does not allow for the signals to be produced independently (i.e. vocalizing without the visual signal is impossible due to physiological constraint).

Choreography can function in more than one way as part of a multimodal display. For example, male lyrebirds, *Melanura novaehollandiae*, have a large vocal and gestural repertoire but choreograph their dance and song with surprising stereotypy (Dalziell et al., 2013), which likely serves to showcase a male's ability to perform a cognitively and/or physically challenging task in this highly polygynous species. Similarly, group-living zebra finches, *Taeniopygia guttata*, preferentially choreograph tap-dancing displays with the introductory portion of their song, possibly leveraging a more complex display to divert female attention towards the signaller (Ullrich et al., 2016). In both of these cases, choreography enhances a display's aesthetic function in mate attraction by increasing its overall complexity, which is known to be important for female choice across a variety of species (Hebets et al., 2013; Miles et al., 2017; Miles & Fuxjager, 2018).

By contrast, gesturing while singing can also influence acoustic output on a physiological basis (Cooper & Goller, 2004). Here, choreography takes on an additional function as a novel mechanism to modulate a target signal, while still enhancing display complexity. If this special case of choreography exists, it should result in a correlation between individual performance of gesture and song. Uncovering such a correlation is insufficient to conclude that gesture is physiologically impacting song performance, however. Instead, both song and gesture performance may be better explained by an external third factor, such as an individual's motor skill, or the innate ability to perform a challenging manoeuvre (Byers, Hebets, & Podos, 2010). Despite the fact that we cannot conclusively determine the function of correlated performance, this special form of choreography stands distinct from its uncorrelated cousin and is thus worth evaluating independently. Therefore, a given song–dance combination should be able to independently manipulate the signal receiver's behaviour, regardless of whether dancing modifies the song.

Theoretically, one species can exhibit multiple forms of choreography within a single display by producing different components of song and dance together. If this is true, then every pair of signals within a species' display repertoire can be combined in a different way, be it across different hierarchical scales or introducing novel ways that song and dance can influence each other. The behavioural and evolutionary ramifications of this possibility are entirely unknown, but one potential outcome is a novel route to increased display complexity without needing the evolutionary ‘innovation’ of a new song type or gesture.

Here we explore this possibility by studying the acrobatic displays performed by male Montezuma oropendolas, *Psarocolius montezuma*. These tropical songbirds compete intensely for mates at their nesting colonies and rely on multimodal displays to mediate both courtship and competition (Webster, 1994a, 1997). The oropendola's complex song (Fig. 1) can last as long as 3 s, combining high-frequency sweeps and a rapid, arythmic series of low-frequency pulses (Price & Lanyon, 2004). While singing, males also perform a gestural display, which starts by slowly leaning forward until the centre of gravity is level with the perch (Jaramillo & Burke, 1999). At this point, the bird rapidly spreads its wings while swinging forward and dangling upside down for up to 1 s. The song itself has multiple vocal elements that should be important for courtship and thus potential targets of choreography. These include the lowest peak frequency (LPF), known to be important in male–male competition (Price & Lanyon, 2004), and a song's maximum amplitude (dB_{max}), the point at which the display is most audible (Janicke, Hahn, Ritz, & Peter, 2008; Ryan, 1988). Either could be accentuated by the high-intensity swinging display. Finally, both the length and frequency modulation ($\Delta F0$) of a song are important in attracting females and signalling aggressive intent (Caro, Sewall, Salvante, & Sockman, 2010; Nelson & Poesel, 2011). These metrics reflect vocal performance over the song's entirety, rather than an isolated time point, and therefore cannot specifically be a target for fine-scale choreography. However, they provide an excellent basis of comparison for guiding inference into the function of choreography.

Here we use video analysis to examine how Montezuma oropendolas choreograph their body and wing displays with different vocal elements (LPF and dB_{max}). We then examine the degree to which gestural and vocal performance are correlated. To distinguish whether correlations between song and gesture are due to physiological interdependence (rather than both modulated due to the influence of some third unknown factor), we also examine the effects of social context and body size on song performance. This is because both factors are well-known modulators of acoustic output in oropendolas and other species (Price, Earnshaw, & Webster, 2006). Considering what is currently known about animal choreography, we operationally defined a gesture as being choreographed with a given song element if the two signals (1) are combined electively and (2) occur together more often than predicted by chance alone. For each gesture choreographed with a different song element, individual gesture and song performance may be correlated or unrelated. When a correlation is present between the two, this suggests that both gesture and song performance are governed by an external factor such as individual motor skill (Byers et al., 2010). Alternatively, the correlation could be due to an intrinsic morphological or physiological link between producing the gesture and song simultaneously (Cooper & Goller, 2004). Here we consider this to be a special case of choreography that is structurally (and perhaps functionally) distinct from the more standard case of choreography with uncorrelated performance. Of course, one animal's display can contain numerous gestural and vocal elements, which makes it possible for multiple choreographic structures to exist within a single display.

METHODS

Ethical Note

This study relied exclusively on the analysis of videos collected from citizens around the world, who posted footage of free-living Montezuma oropendolas display on the Internet. Accordingly, institutional approval for this project was not required.

Study Species and Data Source

To complete this study we relied on audiovisual analysis of displaying birds, an approach that has previously been used to successfully measure both gestural and vocal components of avian displays (Manica, Macedo, Graves, & Podos, 2016; Westneat, Long, Hoese, & Nowicki, 1993). Specifically, we took advantage of publicly archived video recordings (Supplementary Table S1), which are a valuable resource for biologists studying animal behaviour (Corn, Farina, Brash, & Summers, 2016; Nelson & Fijn, 2013; Raine, Pisanski, & Reby, 2017). As such, we gathered video recordings of displaying Montezuma oropendolas from the Macaulay Library (Cornell University, Ithaca, NY, U.S.A.), Handbook of Birds of the World Alive Archive (Lynx Edicións) and YouTube (Alphabet, Inc.). Recordings from natural history libraries are all unedited, and we ensured that all YouTube videos were unmodified with respect to capturing accurate representations of individual songs and dances following previously established criteria (Nelson & Fijn, 2013).

The Montezuma oropendola's song is composed of two major elements: (1) a 'low song', containing a series of abbreviated and arrhythmic pulses between 500 and 900 Hz; and (2) a 'high song', which is a loud train of frequency sweeps with fundamental frequency (F0) roughly between 1000 and 4000 Hz. The song types do not appear to be shared among individuals, although one individual can sing multiple songs. All songs share a gradual increase in amplitude over time, such that the first half of the song is nearly inaudible unless the recorder is at close range (Fig. 1).

Previous studies have detailed the challenges associated with using audiovisual resources from multiple sources, so we adopted conservative quality-control measures. First, we only collected data from recordings in which the focal individual was visible in profile and performed a complete display. To ensure that audio was comparable across files, we only kept recordings with a parent sample rate of 44 100 Hz, and a low level of unfiltered background noise. Because of potential differences in recording distance across files, which makes it impossible to compare low-amplitude song among recordings, we restricted all analyses to the loud song only. To do this, we standardized all recordings to a peak amplitude (dB_{max}) of 0 dB prior to thresholding at −25 dB from the peak, a value which left only the loud song visible. Thresholding in this way allowed us to measure loud song elements, including duration, and compare them across files despite differences in recording equipment and distance (Davidson, Antonova, Dlott, Barber, & Francis, 2017; Podos et al., 2016).

Acoustic Data Collection

We collected all acoustic data in Adobe Audition CC, which allows for side-by-side viewing of high-quality spectrograms with video input. This eliminates the need to perform conversions between frame-based video time codes and audio time. We measured three typical markers of avian song performance: (1) loud song duration ('song length'), (2) lowest peak frequency (LPF) and (3) frequency modulation (ΔF0). After resampling each recording at 44 100 Hz and filtering out all signals −25 dB from the standardized 0 dB threshold, we measured loud song duration as the time elapsed between the start and end of the loud song. Measuring the thresholded loud song (instead of estimating when a song starts and ends based on human observation) also helped to avoid overestimates of song duration due to echoes of the final note. To compute LPF and frequency modulation, we first analysed power spectra generated at 5 ms intervals (Hamming window, fast Fourier transform (FFT) length 4096). For each interval, we recorded which frequency had the greatest amplitude (highest peak) between 500 and 900 Hz ('low song') as well as the peak frequency in the main portion of the song >1000 Hz ('high song'). LPF was then computed as the minimum low-song peak frequency across each 5 ms sample. To calculate average frequency modulation (Δ*F*0) in the upper song, we took the mean change in F0 from one interval to the next and then divided this by 5 to obtain the average change in F0/ms $\left(\Delta F0/_{ms} = \sum(F_n - F_{n+1}) \times \frac{1}{\text{number of intervals}} \times \frac{1}{5\ \text{ms}}\right)$.

To ensure that recording quality would not confound our acoustic analysis, we also measured the average amplitude of background noise in each recording after it was normalized to peak amplitude of 0 dB. Variation in background noise did not affect measurements of loud song length ($R^2 = 0.013$, $P = 0.594$), LPF ($R^2 = 0.041$, $P = 0.338$) or ΔF0 ($R^2 = 0.003$, $P = 0.981$), so we are confident that our measurements were uncompromised by differences in recording quality.

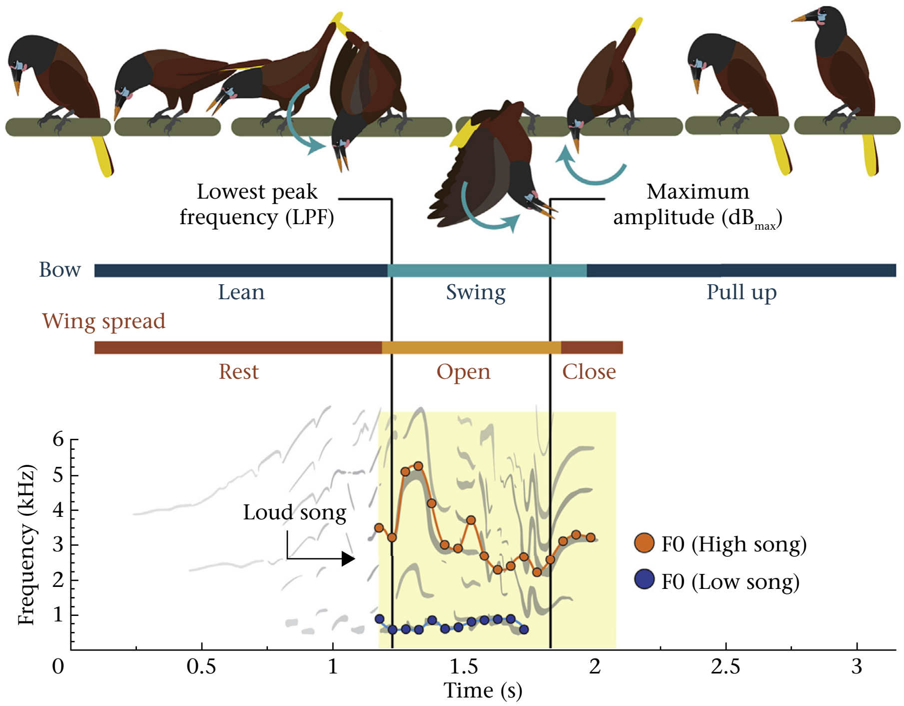


Figure 1. Schematic depicting a representative oropendola display (starting at 33 s in https://www.youtube.com/watch?v=Mz0t05hKVTM) as it unfolds over time. The main graph superimposes acoustic sampling of peak frequencies (via power spectra) taken at 5 ms intervals on a spectrogram of the song. The yellow highlighted area indicates the loud song as thresholded at −25 dB from normalized dB_{max} of 0. Above the graph, the coloured bars correspond to the timing of body display (blue) and wing display (orange) phases.

Testing for Choreography

We determined whether a given pairing of gesture and vocal elements were choreographed by testing whether the two occurred together more often than predicted by chance alone. Specifically, we investigated whether two gestural components (bow display and wing display, see below) were each choreographed with two acoustic elements: the time of lowest peak frequency (LPF) and maximum amplitude (dBmax). As such, we first defined the time of lowest peak frequency as the centre of our 5 ms frequency interval during which the LPF occurred, and defined time of maximum amplitude as the time point identified by Audition's amplitude statistics function as 'peak amplitude' (verified by ensuring that $dBmax = 0$). Using the video files, we defined each display as beginning (i.e. t=0.0 s) when the individual initiates a bow display by leaning forward. Likewise, the display was considered to end after the individual fully returned to the rest position after swinging upside down. We then subdivided this long bow gesture into three distinct phases: (1) Lean, where the bird tilts its head down and stretches the body forward and slightly down; (2) Swing, in which an individual rotates upside down and swings beneath the perch; and (3) Pull Up, when the bird recovers from the inversion and returns to rest. During the bow display, each individual also performs a wing display that consists of a 'rest period' (wings folded along back), as well as an 'opening' and 'closing' phase. To ensure that we consistently characterized gesture phases across videos and individuals, we used straightforward criteria to identify the precise frame that marked a transition between gesture phases (see Supplementary Methods).

After collecting these data on the timing of display elements, we compared when these two acoustic markers occurred with the distribution of display phase times. To do this, we scored which phase of the bow display and wing display contained the time at which LPF and dB_{max} each occurred, effectively generating a contingency table. We then used a chi-square analysis to test the hypothesis that LPF and dB_{max} were timed with the Swing and Open Wings gesture phases more than predicted by chance alone (as weighted by the mean duration of each gesture phase). To correct for multiple hypothesis testing, we applied a Holm–Bonferroni correction to all output P values, and accordingly report these adjusted values.

We also attempted to determine whether each song–dance pair had the potential to exhibit a special case of choreography, wherein a specific gesture performed during the song influences phonation (Cooper & Goller, 2004). Although it is not possible to determine what the mechanism underlying any performance correlation would be without collecting physiological data, we tested whether the oropendola displays conformed to two predictions that are consistent with this unique form of choreography: (1) gesture and vocal performance should be correlated, and (2) this should be independent of other influences such as body size and/or social context. To explore this possibility, we compared variation in our three acoustic variables (song length, LPF and $\Delta F0$) to measurements of gestural intensity for each display. We measured angular distances in each video, an approach used in similar studies because it accounts for differences in distance to the focal individual while still being comparable across recordings (Manica et al., 2016; Westneat et al., 1993).

Because our sample videos were recorded at an unknown distance from the focal individual, we decided to measure gesture intensity as maximum angular displacement of the body relative to the perch (bow display) or the wings relative to the body (wing display; Supplementary Fig. S1). This is an effective approach because most of the oropendola display's movement consists of simple rotation, which can be easily measured as an angular distance from a reference point. We therefore defined body displacement as the angle formed between the horizontal perch and the bird's eye, such that birds performing a deep wing display will have a higher bow angle than individuals only dipping slightly down. Similarly, we measured wing spread as the angle formed between the body line and the chord of the wing (which is anchored by tip of the longest flight feather and the wrist joint).

We took these measurements digitally using the 'angle between two lines' tool in the program MB-Ruler Pro (http://markus-bader.de). The tool calculates high-precision digital angles as a screen overlay, which allowed us to superimpose the reference lines on each video file at the frame in which the bird was at its deepest bow or widest wing position (Supplementary Fig. S1). To ensure that data collection was unbiased, we randomized the file order and measured display angles blind to any acoustic features of the display (i.e. with audio off and spectrogram disabled). Finally, to confirm that our digital measurements were accurate, we measured each video in triplicate and characterized the average error associated with this approach ($CV_{bow} = 0.0137$; $CV_{wing} = 0.0203$).

To statistically test for an association between gesture and vocal performance, we used linear mixed models (LMMs) run in R version 3.4.3 (R Foundation for Statistical Computing, Vienna, Austria), which allowed us to include videos in which one individual performed multiple displays by including individual identity as a random factor. The fixed factors were bow depth and wing angle, and we ran three separate LMMs on each of our three acoustic variables. To ensure we produced no spurious results due to multiple testing, we only report P values that were Holm-adjusted for multiple testing.

Effects of Body Size and Social Context

Body size is a major factor that can enhance or limit vocal performance, so we developed methodology to characterize relative body size across unknown individuals. Our approach is based on previous studies that show that a known-size anchor point in digital photographs allows one to accurately estimate body size (Deakos, 2010; Yoshihara, 1997). Because we analysed digitally archived video from multiple sources, we aimed to find an 'anchor measurement' within Montezuma oropendolas, in the form of a morphological variable that changes minimally across individuals. To do this, we first took basic morphological measurements (wing chord, tail length, tarsus length, culmen depth and length of orange bill marking) on adult male Montezuma oropendola specimens ($N = 26$) housed at the Smithsonian Museum of Natural Science (USNM; Supplementary Table S2). We then ran a principal component analysis (PCA) to characterize how morphological variables were related, and found that our first principal component (PC1) explained 88.5% of variation in the data set (Supplementary Table S3). Surprisingly, tarsus length was the most uniform of all the measurements ($CV = 0.046$) and was thus uncorrelated with PC1 scores ($R^2 = 0.027$, $P = 0.652$), while wing chord was most strongly correlated with PC1 scores ($R^2 = 0.97$, $P < 0.0001$; Supplementary Fig. S2a). We therefore used each individual's tarsus length as a relatively invariable index measurement, and characterized relative body size based on wing chord. Consistent with the use of an index, our actual measurements of wing chord were highly correlated with our relative measurements in tarsus units (i.e. $\frac{\text{wing chord}}{\text{tarsus length}}$; $R^2 = 0.632$, $P < 0.0001$), while tarsus length itself is uncorrelated with wing length ($R^2 = 0.018$, $P = 0.514$; Supplementary Fig. S2b). Because the oropendola bow display involves unfolding the wings to clearly expose the manus joint and the primary feather tips, wing chord was easily measurable in all of our display videos. Similarly, our videos portrayed individuals at a lateral angle, and oropendolas have little tarsal feathering, which allowed us to visually distinguish the joints that mark the extent of the tarsometatarsus bone.

Because this is a novel way to quantify relative body size from unknown individuals, we performed a set of verification exercises to ensure that wing chord and tarsus length could be accurately measured at a variety of distances (more details in Supplementary Methods). Briefly, we digitally photographed a random sample of specimens ($N = 8$) in profile at distances between 1 m and 10 m. Using the MB-Ruler distance tool, we then measured wing chord and tarsus length in arbitrary pixel units. We then normalized both the digital and the real-life specimen wing chord measurements to tarsus units by dividing wing chord by tarsus, and compared these to one another. This exercise supports the idea that relativizing wing chord to tarsus is an effective way to estimate body size, as digital measurements had low measurement error (mean CV = 0.016), were statistically insignificant from their real-life counterparts ($t = 0.774$, $P = 0.464$; Supplementary Fig. S3a) and were highly correlated with real-life data collected from the specimens themselves ($R^2 = 0.623$, $P < 0.0001$; Supplementary Fig. S3b). We therefore used the MB-Ruler distance tool to obtain morphological measurements from each video, taken in triplicate at the frame in which either tarsus length (CV = 0.062) or wing chord (CV = 0.071) were most easily visible. To see whether relative body size influenced song performance, we ran LMMs to test for the effect of body size (fixed) on each acoustic variable, with individual identity as a random factor.

Finally, we also examined whether social context affects choreography, as the social environment has potent effects on behavioural output (Price et al., 2006; Schuppe, Sanin, & Fuxjager, 2016). In particular, Montezuma oropendolas are known to produce songs with lower LPF values when overlapping their song with a neighbour (Price et al., 2006). Previous studies also suggest that oropendolas exhibit maximum aggression when they are at the nesting colony rather than away from it (Webster, 1994a). Because oropendolas display both at and away from the colony, this allowed us to categorize videos based on where the birds appeared to be displaying. This was easy to distinguish visually and aurally, because oropendolas nest in high densities (up to 45 nests started in a single colony) accompanied by near-constant activity by both males and females (Webster, 1994b). We therefore considered videos to be recording birds at a colony if they exhibited the following characteristics: (1) the presence of one or more woven nests, where the males display; and (2) the presence of multiple individuals in the recording, either other males displaying or calling females and/or nestlings. Videos that met both of these criteria were considered to be ‘at the colony’ and those that did not were considered to be ‘away from the colony’. Notably, there was no video that met one criterion but not the other, including no cases in which it was difficult to tell whether there were background vocalizations in videos with nests present. This ensured that poor audio quality did not impact our ability to partition videos by location relative to the nesting colony, which was also supported by statistical similarity in background noise across recordings from the two locations ($t = 0.847$, $P = 0.410$). We tested for an effect of social context in this way by performing a *t* test on acoustic variables in songs performed at and away from the colony. Therefore, a significant difference in song performance would suggest that any change in vocalizations could be attributed to social context rather than gesture intensity.

RESULTS

We first tested whether Montezuma oropendolas choreograph their displays by timing different gestures with specific vocal components. However, for a temporal link between gesture and song to be choreographed, it must elective (i.e. an individual must be physically capable of singing without performing the gestural display). Whereas most of our videos featured individuals producing the full display, we did find instances in which birds only sang without the bow or wing gesture. This allowed us to rule out the possibility that song production is impossible without gesturing (or vice versa).

We next tested whether the high-intensity Swing and Open Wing phases of the dance coincided with the song's LPF and dB$_{max}$ more often than predicted by chance alone (Fig. 2). Indeed, individuals were significantly more likely to sing their lowest note during the Swing ($\chi^2_2 = 8.068$, $P = 0.0045$) and Open Wing ($\chi^2_2 = 9.624$, $P = 0.0019$) phases of the display routine, compared to the Lean and Pull Up phases (Fig. 2). We found a similar case for the timing of dB$_{max}$, which also occurred more often during the Swing ($\chi^2_2 = 28.76$, $P<0.0001$) and Open Wing phases ($\chi^2_2 = 30.37$, $P < 0.0001$), relative to the other phases. In other words, the oropendola's display exhibits choreography of the bow display with LPF and choreography of the wing display with dB$_{max}$.

To investigate the relationship between the performance of choreographed gesture and song, we then tested for a correlation between the two. We found that the depth of an individual's bow display predicted LPF ($F_{1,20.9} = 11.431$, $P = 0.0084$), where birds that produced lower swings also vocalized at a lower frequency (Fig. 3a). Wing spread angle did not predict LPF alone ($F_{1,22.6} = 0.305$, $P = 0.586$; Fig. 3b) or through an interaction with bow depth ($F_{1,22.1} = 1.091$, $P = 0.308$). Of course, this correlation can only reflect a direct relationship between song and dance if it is not also explained by other factors such as social context and body size. We therefore compared our same indices of display performance among individuals displaying at the nesting colony (high aggression social context) or away from the colony (low aggression social context; Fig. 4). We found no difference between display locations for either LPF ($t_{22} = 0.297$, $P = 0.769$) or bow depth ($t_{22} = 0.863$, $P = 0.397$), which suggests that the effect of bow depth on LPF

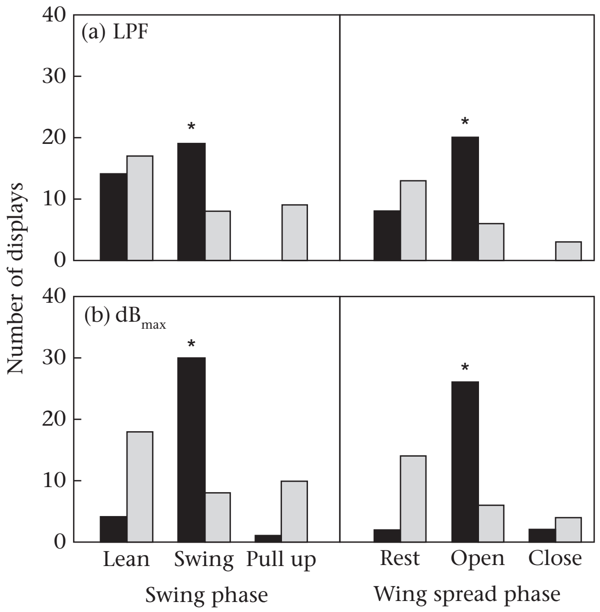


Figure 2. Observed (black bars) and expected (grey bars) number of times that the (a) lowest peak frequency (LPF) or (b) greatest amplitude (dB$_{max}$) occurred during different gestural display phases. Asterisks indicate phases during which the observed counts were significantly greater ($P < 0.05$ after Bonferroni correction) than expected counts as weighted by mean duration of different display phases.

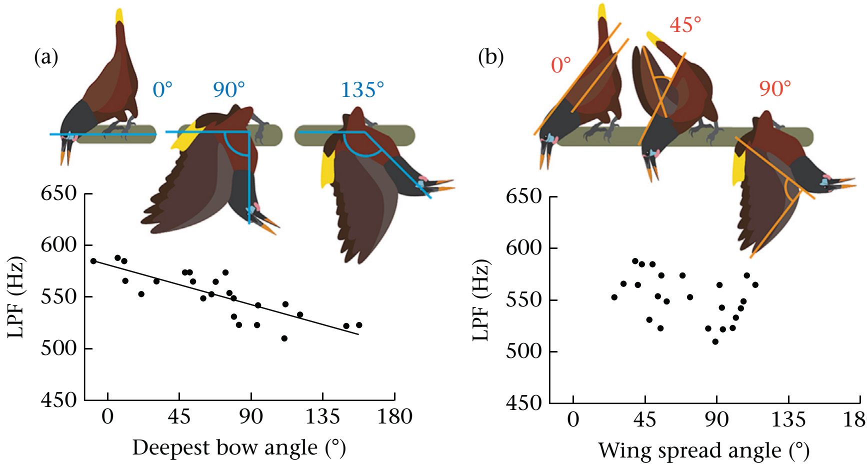


Figure 3. Lowest peak frequency (LPF) of Montezuma oropendola song as it varies with (a) the farthest extent of the bow gesture or (b) the widest angle of the wing spread gesture. Solid line indicates a statistically significant ($P = 0.0084$ after correction for multiple testing) predictive relationship of bow angle on LPF.

occurred independently of social context. Meanwhile, social context did appear to influence other aspects of display performance; at the colony, individuals performed wider wing spread displays ($t_{22} = 2.79$, $P = 0.021$) and longer songs ($t_{22} = 2.77$, $P = 0.027$) with higher frequency modulation ($t_{22} = 2.86$, $P = 0.027$) than those displaying elsewhere.

Relative body size also did not appear to play any role in altering song performance, as it did not explain variation in either LPF ($F_{1,13.9} = 0.113$, $P = 0.742$), song length ($F_{1,15.4} = 0.021$, $P = 0.889$) or $\Delta F0$. There also appeared to be no interaction between bow and relative body size that predicted LPF ($F_{1,15.4} = 0.113$, $P = 0.322$), song length ($F_{1,16.9} = 0.010$, $P = 0.920$) or $\Delta F0$ ($F_{1,17} = 0.995$, $P = 0.332$). This is consistent with a predictive relationship between bow depth and LPF that occurs independently of body size.

DISCUSSION

Using the Montezuma oropendola as a model, we show how two different gestures are each choreographed differently as part of a complex multimodal display used in both courtship and competition. First, we establish that combining gesture with vocal elements is elective, as song and dance can be performed independently. Next, we demonstrate that the bird's upside-down swing display and wing spread are each timed so that the song's LPF and dB$_{max}$ occur during the most dynamic phases of each gesture (Swing or Open Wings, respectively). This suggests that birds intentionally combine these two elements of their display repertoire, ultimately generating a display aesthetic that cannot be produced by either signal alone.

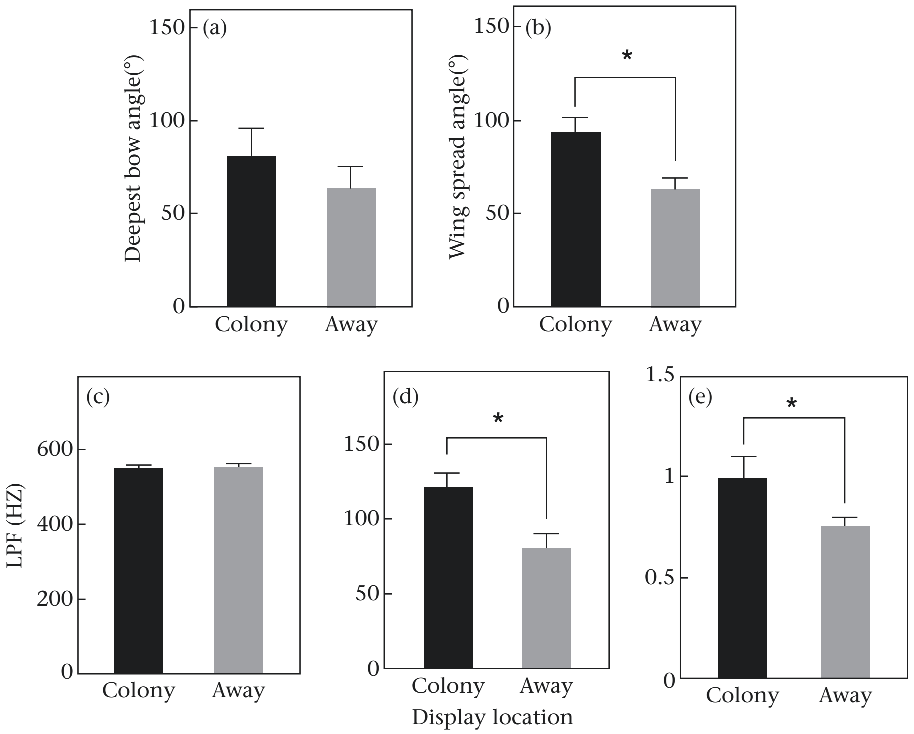


Figure 4. Acoustic variables as they vary with social context based on display location: (a) lowest peak frequency, (b) frequency modulation and (c) loud song duration. Statistically significant differences are marked with an asterisk.

We also found a strong correlation between two aspects of the display: LPF and bow depth. This means that individuals that swing lower under their perch while displaying also sing with a lower LPF. This finding supports the idea that these traits are functionally linked, where gesture may influence vocal output. However, an alternate causal explanation is that both of these traits covary with display motivation. If this were the case, we would expect that birds displaying in the high-stakes social context of the nesting colony would produce a lower LPF and deeper bow than those displaying away. However, because we found no difference in LPF and bow depth across social contexts (but did find that birds displaying at the colony sang longer songs with greater $\Delta F0$), we can rule out this possibility. Our results also do not support the alternative explanation that individual morphology influences LPF or swing depth, because we found no relationship between these variables and relative body size.

When a multimodal display like that of the Montezuma oropendola consists of multiple elements in both song and dance, choreography can introduce a novel way to enhance display complexity by behaviourally modifying how gesture is temporally mapped onto song. This is primarily illustrated by the fact that individuals display so that the Swing Phase of the bow display overlaps with LPF and that the Open Wings phase of the wing spread occurs with dB_{max}. These new combinations of sound and movement likely forge a novel facet to the display's gestalt. Other studies similarly suggest that the effectiveness of such an integrated signal often supersedes the effectiveness of either signalling component presented alone (Hebets et al., 2016; Patricelli & Hebets, 2016). If receivers therefore evaluate how different signals are intertwined to generate novel display elements, then the process of choreography must also transmit valuable information. We suspect that this is the case in Montezuma oropendolas, since LPF plays a fundamental role in mediating male–male competition and is one of the main aspects of the bird's song that is choreographed (Price et al., 2006).

How might choreography augment a display's function in courtship and competition? Studies in other species point to several nonmutually exclusive possibilities. First, choreographing redundant signals may help displaying males draw female attention to their performance in a crowded or noisy environment (Ullrich et al., 2016). Considering that much of a Montezuma oropendola's breeding season takes place at a cacophonous breeding colony, observability may be an important driver of choreography in this species as well. This notion is further supported by the fact that oropendolas specifically time their loudest vocalizations to coincide with the display's visual climax, when a male is dangling upside down with wings spread.

A second compelling possibility is that fine-scale choreography is an inherently challenging motor task, which offers the opportunity for males to showcase subtle differences among individuals (Byers et al., 2010). Although no studies to date explicitly address choreography on this scale, previous work in lyrebirds has found that choreographing different elements of this species' vast song and dance repertoire functions as a signalling challenge. The same may be true for blue-black grassquits, *Volatinia jacarina*, which are thought to exhibit motor skill by singing while performing a jump display (Manica et al., 2016). There is a correlation between song length and jump height during this display too, suggesting that choreography may also play a role. Considering the Montezuma oropendola's highly polygynous mating system and extreme display complexity, this is a potential driver in this species as well.

Alternatively, differences in choreography may facilitate individual recognition. This occurs in doves that display by cooing and bowing, with little structural variation in these signals among males. However, each individual exhibits a unique rhythmic signature in the way coos and bows are choreographed, which facilitates individual recognition (Fusani, Hutchison, & Hutchison, 1997). However, this is unlikely to explain the Montezuma oropendola's choreography of LPF and dB_{max} with the bow and wing gestures, because we found strong similarities in choreography across individuals.

Beyond the potential for choreography to act as a signal in its own right, introducing gesture into an acoustic display has another surprising consequence: the combination of certain gestures with complex song can be a novel way to modify the physiological underpinnings of vocal performance (Cooper & Goller, 2004). Our data are consistent with the predictions associated with this phenomenon, first by exhibiting a negative correlation between bow depth and LPF. This means that individuals that are more apt to dangle upside down during the display also sing lower LPFs, which again are known to mediate male–male competition (Hall, Kingma, & Peters, 2013; Price et al., 2006). The same does not hold true with respect to wing display, so multiple forms of choreography are being integrated into one display.

The proximate reason for the gesture–song correlation may lie in motor mechanisms that underlie both gesture and vocalization. Although no study has examined the motor physiology of oropendola displays, the bird's upside-down swing gesture undoubtedly involves dynamic movement of the wings, legs and torso. Because the avian respiratory system circulates air by modulating pressure in air sacs positioned throughout the body, local muscle contraction unrelated to respiration can ultimately affect how air moves through the body, including the vocal apparatus (Boggs, Jenkins, & Dial, 1997; Mackelprang & Goller, 2013). Indeed, work in the brown-headed cowbird, *Molothrus ater*, shows that wing raising reduces respiratory effort just before phonation, potentially making it easier to produce challenging song elements (Cooper & Goller, 2004). This is because the wing raise secondarily shortens the abdominal muscles as if they were contracting for respiration, but without the muscles actually being activated. In other words, the cowbird's gestural display also serves a secondary respiratory function important to vocalization. Although this study illustrates the potential for gesture to influence phonation, the cowbird's wing display is not known to result in any measurable change to song performance itself. How, then, might gesture's influence on respiratory airflow result in lower-frequency notes in birds that display more intensely? Low-frequency notes in songbirds are produced in the pulse tone register, wherein phonation frequency is modulated by changing the rate at which rapid pulses of air are allowed to vibrate the respiratory membrane (Goller & Larsen, 2002; Jensen, Cooper, Larsen, & Goller, 2007). The Montezuma oropendola's low-frequency pulses are likely produced in this register as well, and larger gesture-induced shifts in respiratory pressure dynamics may make it easier to generate the fine-scale pulse control required to sing at a lower frequency.

However, an underlying mechanistic link between gesture and song is not the only possible explanation for a correlation between these two display traits. Another explanation for our results could simply be that an individual's ability to perform challenging motor tasks influences both swing depth and LPF. This may be the case because sexual selection for male motor skill appears to drive the evolution of numerous displays, including song and dance (Barske, Schlinger, Wikelski, & Fusani, 2011; Byers et al., 2010; Manica et al., 2016; Schuppe & Fuxjager, 2017). Because performance of both song and dance ultimately relies on a similar suite of neuromotor mechanisms (e.g. ‘superfast’ muscle contraction; Elemans, Spierts, Müller, van Leeuwen, & Goller, 2004; Fuxjager, Goller, Dirkse, Sanin, & Garcia, 2016), individual differences in motor skill could underlie both LPF and swing performance.

We also assess two major alternative explanations, and our data support neither of them. The first possibility is that body size influences the performance of both song and dance. This is because body size constrains low-frequency sound generation, where larger individuals are able to vocalize at lower frequencies (Charlton & Reby, 2016; Riede & Goller, 2014). This pattern is even seen among oropendolas and their close relatives, wherein large species like the Montezuma oropendola are capable of singing the lowest notes (Price et al., 2006). Similarly, although little is known about the relationship between body size and upside-down display performance, it may be that larger individuals also tend to swing lower. However, our data do not support this alternative because we found no evidence that individual differences in body size predicted LPF. Therefore, body size does not explain the functional correlation we found between LPF and swing depth.

A second factor that influences both gestural and vocal display performance is aggressive motivation, which shifts according to social context. Indeed, animals tend to produce high-quality displays more often in a social context where there are positive consequences to displaying well (e.g. increased copulations or decreased intruder threat; Mager, Walcott, & Piper, 2012; Schuppe & Fuxjager, 2017; Schuppe et al., 2016). Again, however, we found that social context failed to explain individual differences in LPF. Instead, individuals displaying in the high aggression nesting colony (Webster, 1994a) tended to sing longer songs with greater frequency modulation. Interestingly, previous work in the Montezuma oropendola has shown that males do modulate their LPF during direct vocal contests when songs of two males overlap (Price et al., 2006). Our results do not contradict this finding, considering that we did not include overlapped songs in our analyses; instead this indicates that there may be multiple 'tiers' of social display context to which animals respond.

Conclusions

Here we outlined a testable framework designed to characterize how animals choreograph song and dance into an integrated display. We then applied this framework to a case study, revealing that the Montezuma oropendola's multimodal display is structurally defined by pairing both wing and body movement with specific vocal elements. More significantly, the choreography of this species' swing display with low-frequency vocalization is also marked by correlated performance of both signals, which is not explained by social context or body size. This suggests that the performance correlation is better explained by an intrinsic physiological link between gesture and vocalization, or otherwise by another factor such as motor skill influencing both signals. Because only one set of choreographed signals exhibits this relationship, it suggests that multiple choreographic structures (and perhaps functions as well) can exist in a single multimodal display, and potentially serve as a mechanism to enhance display complexity without introducing a new signal. The inherent complexity of multimodal signals makes understanding their function and evolution challenging, but integrating choreography into the existing framework will only serve to deepen our understanding of animal communication.

Acknowledgments

This study was supported by the National Science Foundation (IOS-1655730) to M.J.F. and Wake Forest University intramural funds. We thank Christopher Milensky (USNM) for coordinating our museum work, as well as the many citizen scientists who posted videos of displaying oropendolas on the Internet.

Supplementary Material

Supplementary data related to this article can be found at https://doi.org/10.1016/j.anbehav.2018.04.006.

References

Barske, J., Schlinger, B. A., Wikelski, M., & Fusani, L. (2011). Female choice for male motor skills. *Proceedings of the Royal Society B: Biological Sciences, 278*(1724), 3523–3528. https://doi.org/10.1098/rspb.2011.0382.

Boggs, D. F., Jenkins, F. A., & Dial, K. (1997). The effects of the wingbeat cycle on respiration in black-billed magpies (*Pica pica*). *Journal of Experimental Biology, 200*, 1403–1412.

Byers, J., Hebets, E., & Podos, J. (2010). Female mate choice based upon male motor performance. *Animal Behaviour, 79*, 771–778. https://doi.org/10.1016/j.anbehav.2010.01.009.

Caro, S. P., Sewall, K. B., Salvante, K. G., & Sockman, K. W. (2010). Female Lincoln's sparrows modulate their behavior in response to variation in male song quality. *Behavioral Ecology, 21*(3), 562–569. https://doi.org/10.1093/beheco/arq022.

Charlton, B. D., & Reby, D. (2016). The evolution of acoustic size exaggeration in terrestrial mammals. *Nature Communications, 7*, 12739. https://doi.org/10.1038/ncomms12739.

Cooper, B. G., & Goller, F. (2004). Multimodal signals: Enhancement and constraint of song motor patterns by visual display. *Science, 303*(5657), 544–546. https://doi.org/10.1126/science.1091099.

Corn, K. A., Farina, S. C., Brash, J., & Summers, A. P. (2016). Modelling tooth–prey interactions in sharks: The importance of dynamic testing. *Royal Society Open Science, 3*(8), 160141. https://doi.org/10.1098/rsos.160141.

Dalziell, A. H., Peters, R. A., Cockburn, A., Dorland, A. D., Maisey, A. C., & Magrath, R. D. (2013). Dance choreography is coordinated with song repertoire in a complex avian display. *Current Biology, 23*(12), 1132–1135. https://doi.org/10.1016/j.cub.2013.05.018.

Davidson, B. M., Antonova, G., Dlott, H., Barber, J. R., & Francis, C. D. (2017). Natural and anthropogenic sounds reduce song performance: Insights from two emberizid species. *Behavioral Ecology, 84*(4), E1–E9. https://doi.org/10.1093/beheco/arx036.

Deakos, M. (2010). Paired-laser photogrammetry as a simple and accurate system for measuring the body size of free-ranging manta rays *Manta alfredi*. *Aquatic Biology, 10*(1), 1–10. https://doi.org/10.3354/ab00258.

Elemans, C. P. H., Spierts, I. L. Y., Müller, U. K., van Leeuwen, J. L., & Goller, F. (2004). Bird song: Superfast muscles control dove's trill. *Nature, 431*(7005). https://doi.org/10.1038/431146a, 146–146.

Fusani, L., Hutchison, R. E., & Hutchison, J. B. (1997). Vocal–postural co-ordination of a sexually dimorphic display in a monomorphic species: The barbary dove. *Behaviour, 134*(5/6), 321–335. https://doi.org/10.1163/156853997X00566.

Fuxjager, M. J., Eaton, J., Lindsay, W. R., Salwiczek, L. H., Rensel, M. A., Barske, J., et al. (2015). Evolutionary patterns of adaptive acrobatics and physical performance predict expression profiles of androgen receptor - but not oestrogen receptor - in the forelimb musculature. *Functional Ecology, 29*(9), 1197–1208. https://doi.org/10.1111/1365-2435.12438.

Fuxjager, M. J., Goller, F., Dirkse, A., Sanin, G. D., & Garcia, S. (2016). Select forelimb muscles have evolved superfast contractile speed to support acrobatic social displays. *eLife, 5*. https://doi.org/10.7554/eLife.13544.

Goller, F., & Larsen, O. N. (2002). New perspectives on mechanisms of sound generation in songbirds. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology, 188*(11–12), 841–850. https://doi.org/10.1007/s00359-002-0350-6.

Hall, M. L., Kingma, S. A., & Peters, A. (2013). Male songbird indicates body size with low-pitched advertising songs. *PLoS One, 8*(2), e56717. https://doi.org/10.1371/journal.pone.0056717.

Hebets, E. A., Barron, A. B., Balakrishnan, C. N., Hauber, M. E., Mason, P. H., & Hoke, K. L. (2016). A systems approach to animal communication. *Proceedings of the Royal Society B: Biological Sciences, 283*(1826), 20152889. https://doi.org/10.1098/rspb.2015.2889.

Hebets, E. A., & Uetz, G. W. (1999). Female responses to isolated signals from multimodal male courtship displays in the wolf spider genus *Schizocosa* (Araneae: Lycosidae). *Animal Behaviour, 57*, 865–872. https://doi.org/10.1006/anbe.1998.1048.

Hebets, E. A., Vink, C. J., Sullivan-Beckers, L., & Rosenthal, M. F. (2013). The dominance of seismic signaling and selection for signal complexity in *Schizocosa* multimodal courtship displays. *Behavioral Ecology and Sociobiology, 67*(9), 1483–1498. https://doi.org/10.1007/s00265-013-1519-4.

Janicke, T., Hahn, S., Ritz, M. S., & Peter, H. U. (2008). Vocal performance reflects individual quality in a nonpasserine. *Animal Behaviour, 75*, 91–98. https://doi.org/10.1016/j.anbehav.2007.04.007.

Jaramillo, A., & Burke, P. (1999). *New World blackbirds: The icterids*. Princeton, NJ: Princeton University Press.

Jensen, K. K., Cooper, B. G., Larsen, O. N., & Goller, F. (2007). Songbirds use pulse tone register in two voices to generate low-frequency sound. *Proceedings of the Royal Society B: Biological Sciences, 274*(1626), 2703–2710. https://doi.org/10.1098/rspb.2007.0781.

Krumhansl, C. L., & Schenck, D. L. (1997). Can dance reflect the structural and expressive qualities of music? A perceptual experiment on Balanchine's choreography of Mozart's Divertimento No. 15. *Musicae Scientiae, 1*(1), 63–85. https://doi.org/10.1177/102986499700100105.

Mackelprang, R., & Goller, F. (2013). Ventilation patterns of the songbird lung/air sac system during different behaviors. *Journal of Experimental Biology, 216*(19), 3611–3619. https://doi.org/10.1242/jeb.087197.

Mager, J. N., III, Walcott, C., & Piper, W. H. (2012). Male common loons signal greater aggressive motivation by lengthening territorial yodels. *Wilson Journal of Ornithology, 124*(1), 73–80. https://doi.org/10.1676/11-024.1.

Manica, L. T., Macedo, R. H., Graves, J. A., & Podos, J. (2016). Vigor and skill in the acrobatic mating displays of a Neotropical songbird. *Behavioral Ecology, 28*(1), 164–173. https://doi.org/10.1093/beheco/arw143.

Miles, M. C., Cheng, S., & Fuxjager, M. J. (2017). Biogeography predicts macroevolutionary patterning of gestural display complexity in a passerine family. *Evolution, 71*(5), 1406–1416. https://doi.org/10.1111/evo.13213.

Miles, M. C., & Fuxjager, M. J. (2018). Synergistic selection regimens drive the evolution of display complexity in birds of paradise. *Journal of Animal Ecology*. https://doi.org/10.1111/1365-2656.12824. Advance online publication.

Nelson, X. J., & Fijn, N. (2013). The use of visual media as a tool for investigating animal behaviour. *Animal Behaviour, 85*, 525–536. https://doi.org/10.1016/j.anbehav.2012.12.009.

Nelson, D. A., & Poesel, A. (2011). Song length variation serves multiple functions in the white-crowned sparrow. *Behavioral Ecology and Sociobiology, 65*(5), 1103–1111. https://doi.org/10.1007/s00265-010-1120-z.

Patricelli, G. L., & Hebets, E. A. (2016). New dimensions in animal communication: The case for complexity. *Current Opinion in Behavioral Sciences, 12*, 80–89. https://doi.org/10.1016/j.cobeha.2016.09.011.

Podos, J., Moseley, D. L., Goodwin, S. E., McClure, J., Taft, B. N., Strauss, A. V. H., et al. (2016). A fine-scale, broadly applicable index of vocal performance: Frequency excursion. *Animal Behaviour, 116*, 203–212. https://doi.org/10.1016/j.anbehav.2016.03.036.

Price, J. J., Earnshaw, S. M., & Webster, M. S. (2006). Montezuma oropendolas modify a component of song constrained by body size during vocal contests. *Animal Behaviour, 71*, 799–807. https://doi.org/10.1016/j.anbehav.2005.05.025.

Price, J. J., & Lanyon, S. M. (2004). Patterns of song evolution and sexual selection in the oropendolas and caciques. *Behavioral Ecology, 15*(3), 485–497. https://doi.org/10.1093/beheco/arh040.

Raine, J., Pisanski, K., & Reby, D. (2017). Tennis grunts communicate acoustic cues to sex and contest outcome. *Animal Behaviour, 130*, 47–55. https://doi.org/10.1016/j.anbehav.2017.06.022.

Riede, T., & Goller, F. (2014). Morphological basis for the evolution of acoustic diversity in oscine songbirds. *Proceedings of the Royal Society B: Biological Sciences, 281*(1779). https://doi.org/10.1098/rspb.2013.2306, 20132306–20132306.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour, 58*, 921–931. https://doi.org/10.1111/jce.13019.

Ryan, M. J. (1988). Energy, calling, and selection. *American Zoologist, 28*(3), 885–898. https://doi.org/10.1093/icb/28.3.885.

Schuppe, E. R., & Fuxjager, M. J. (2017). Motor skill underlies the effectiveness of a physical, whole-body territorial display. *Functional Ecology, 32*(2), 450–460. https://doi.org/10.1111/1365-2435.13010.

Schuppe, E. R., Sanin, G. D., & Fuxjager, M. J. (2016). The social context of a territorial dispute differentially influences the way individuals in breeding pairs coordinate their aggressive tactics. *Behavioral Ecology and Sociobiology, 70*(5), 673–682. https://doi.org/10.1007/s00265-016-2088-0.

Soma, M., & Garamszegi, L. Z. (2015). Evolution of courtship display in estrildid finches: Dance in relation to female song and plumage ornamentation. *Frontiers in Ecology and Evolution, 3*(4), 1–11. https://doi.org/10.3389/fevo.2015.00004.

Starnberger, I., Preininger, D., & Hödl, W. (2014). The anuran vocal sac: A tool for multimodal signalling. *Animal Behaviour, 97*, 281–288. https://doi.org/10.1016/j.anbehav.2014.07.027.

Taylor, R. C., & Ryan, M. J. (2013). Interactions of multisensory components perceptually rescue Tungara frog mating signals. *Science, 341*(6143), 273–274. https://doi.org/10.1126/science.1237113.

Uetz, G. W., & Roberts, J. A. (2002). Multisensory cues and multimodal communication in spiders: Insights from video/audio playback studies. *Brain, Behavior and Evolution, 59*(4), 222–230. https://doi.org/10.1159/000064909.

Ullrich, R., Norton, P., & Scharff, C. (2016). Waltzing *Taeniopygia*: Integration of courtship song and dance in the domesticated Australian zebra finch. *Animal Behaviour, 112*, 285–300. https://doi.org/10.1016/j.anbehav.2015.11.012.

Webster, M. S. (1994a). Female-defence polygyny in a Neotropical bird, the Montezuma oropendola. *Animal Behaviour, 48*, 779–794. https://doi.org/10.1006/anbe.1994.1302.

Webster, M. S. (1994b). The spatial and temporal distribution of breeding female Montezuma oropendolas: Effects on male mating strategies. *Condor, 96*(3), 722–733. https://doi.org/10.2307/1369475.

Webster, M. S. (1997). Extreme sexual size dimorphism, sexual selection, and the foraging ecology of Montezuma oropendolas. *Auk, 114*(4), 570–580. https://doi.org/10.2307/4089276.

Westneat, M. W., Long, J. H., Jr., Hoese, W., & Nowicki, S. (1993). Kinematics of birdsong: Functional correlation of cranial movements and acoustic features in sparrows. *Journal of Experimental Biology, 182*(1), 147–171.

Williams, H. (2001). Choreography of song, dance and beak movements in the zebra finch (*Taeniopygia guttata*). *Journal of Experimental Biology, 204*, 3497–3506. Retrieved from https://www.researchgate.net/profile/Heather_Williams5/publication/11646573_Choreography_of_song_dance_and_beak_movements_in_the_Zebra_Finch_Taeniopygia_guttata/links/00b4952febf0b7653b000000/Choreography-of-song-dance-and-beak-movements-in-the-Zebra-Finc.

Yoshihara, K. (1997). A fish body length measuring method using an underwater video camera in combination with laser discharge equipment. *Fisheries Science, 63*(5), 676–680. https://doi.org/10.2331/fishsci.63.676.