**Levels of simplicity in the courtship displays of a lek-mating bird**

**ABSTRACT**

In polygynous lek mating systems, males perform complex courtship displays for females. Studies of vocal displays in birds have linked reproductive success to behavioral complexity. These studies often address a simple level of complexity—repertoire complexity—which focuses on e.g., the number of unique display elements in a display. Here, we address another level—syntax complexity—which focuses on how organisms actually arrange the elements of their display. We analyze 312 displays of lek-mating Golden-winged Manakins, *Masius chrysopterus,* by coding their dances as strings of behavioral elements. We calculate not only classic measures of repertoire complexity, but also multiple measures of syntax complexity using methods from computer and information science (Shannon entropy, LZ77 string compressibility, and Kolmogorov algorithmic complexity). We find that solo male displays (SOLO) were the most complex, while unsuccessful displays performed for a female audience (FEM) were of intermediate complexity, and successful displays ending in copulation (COP) were the simplest in terms of both repertoire and syntax. Using Jaro string distance, a method from record-linkage theory, we find that COP displays were the most uniform and that displays varied more across contexts (SOLO, FEM, COP) than across individuals. In this lek-mating species, females preferred simple displays building to a dramatic high-speed element (audible log-approach dive). Our results raise questions about how aesthetic sexual preferences for complexity vary between song and dance displays. Looking forward, we highlight the need for a third level of complexity—element complexity—that reckons with the qualities of display elements themselves.

Keywords: complexity, repertoire, syntax, courtship, display element, lossless compression, entropy, Jaro string distance, lek, manakin

## **INTRODUCTION**

In mating systems where females can exercise mate choice, considerable attention has focused on the basis for choice and the benefits thereof (Kirkpatrick, 1982). As one basis for female choice, elaborate courtship displays have been a special source of fascination. Courtship displays often entail a sequence of distinct display elements (e.g., wing flutters or hops) that can form a complex sequence. Manakins (Aves, Pipridae) are well known for their extraordinary courtship displays, and ornate and varied plumage; they are the “lekkiest” family in the animal kingdom (McDonald, 2010), with by far the highest percentage of lek-mating species (~84%; the birds of paradise, Paradisaeidae, are a distant second with ~50%). Published studies of manakin courtship behaviour have included high-speed display elements (Fusani, Giordano, Day, & Schlinger, 2007) , acrobatic display elements (Prum & Johnson, 1987; McDonald, 1989), vocalizations (McDonald, 1989) and mechanical sound production (Bostwick & Prum, 2003). Uniquely among lek-mating species, manakins exhibit a spectrum of cooperative courtship behaviour, from solo-displaying males to obligate non-kin cooperation (McDonald & Potts, 1994, Ryder et al., 2008). Here, we address the display sequences of the little-studied (Prum & Johnson, 1987; Snow & Snow, 1992) Golden-winged Manakin (*Masius chrysopterus*), which performs highly acrobatic displays with high-speed elemens. Males occasionally display together, though not in a consistently coordinated manner and never in displays leading to copulations (Prum & Johnson, 1987; pers. obs.). Because *Masius* is currently monotypic, we will refer to it throughout by its genus.

It is important to specify what we mean by complexity (Benedict & Najar, 2019). We distinguish three types of complexity: repertoire complexity, syntax complexity and element complexity. 1) Repertoire complexity – we use this term for analyses that assess the number of distinct elements or syllables that occur in a display. This is often, but not always, the measure of complexity used in analyses of vocal repertoires in birds (REFS). For example, Miles & Fuxjager (2018), for comparative assessment of displays, assessed complexity as the number of display elements or gestures.. Primarily for comparative purposes, we briefly consider this type of complexity in our analysis. 2) Syntax complexity – we use this term for analyses of the arrangement of elements in a display. For example, using letters as abbreviations for distinct display elements (such as bows, *b,* or neck twists, *n*) we can code sequences of elements as strings of letters (e.g., *bbnnrr* vs. *rbnnr*). Our major focus will be on this type of complexity. Our tools for these analyses, described in greater detail below, use sequences (strings) of display elements for ethogrtram networks, information entropy, compressibility (REF), Kolmogorov complexity and Jaro string distance (REF). 3) Element complexity – we use this term for analyses of the internal complexity of elements (e.g., trills vs. pure tones in vocal repertoires). We ignore this last type of complexity and focus on the first two types, particularly syntax (arrangement complexity).

We distinguish four major contexts for the displays: solo male displays, with no male or female audience (SOLO); display bouts in which two or more males display, but with no female audience (MULT); display bouts with an audience of one or more females, regardless of the number of males present (FEM) and those display bouts that resulted in copulations (COP). We assess variation across the contexts with five major quantitative tools: 1) the number of distinct display elements in the courtship displays, and the length of those sequences; 2) ethograms of distinct display elements, analysed as weighted networks (network density and effective degree, a metric based on variance in edge weights; McDonald and Hobson 2018); 3) the information entropy of sequences of display elements; 4) the compression ratio (via lossless compression analysis; Huffman, 1952; Ziv & Lempel, 1977) of the display sequences; 5) the Kolmogorov complexity of the sequences. Finally, we use the Jaro string distance (REF), a measure of similarity between strings of elements, to assess whether males are individually consistent or, alternatively, whether successful displays (regardless of the identity of the performing male) converge on a canonical order of display elements. Low Jaro distances between display sequences means they are more similar to each other, high Jaro distance means they are more dissimilar. Entropy and compression ratio, as well as Jaro string distances, require coding the displays as strings of distinct display elements, which we code as letters (e.g., *b* for *Bows*). We use all these metrics to assess the relative complexity of the displays and the within-individual consistency of males, across the gradient of display contexts, from no female audience (SOLO, MULT) to courtship displays with a female audience, but no copulation (FEM), to successful courtship ending in copulation (COP).

For two networks of similar order (number of nodes) a denser network (one with more edges) is more complex; denser networks will, for example, have more alternative paths connecting disparate nodes. A network with higher effective degree (McDonald and Hobson, 2018). For entropy, higher entropy means greater surprise, and greater complexity. Low entropy denotes low complexity – imagine a string of ones. Lossless compression of sequences of information (Ziv & Lempel, 1977), ranges from complex (low compression; e.g., the high diversity and high surprise characterizing a random string) to simple (high compression; e.g., repeated blocks of the same elements). Jaro string distance assesses intra-individual consistency (a given male's displays resemble each other) versus inter-individual consistency by context (all copulation displays resemble each other, regardless of male or female identity). In summary, we assess the gradient of complexity from male-only displays (SOLO and MULT displays) to displays with a female audience (AUDI) to bouts culminating in a successful copulation (COP bouts), as well as variation in displays across contexts and within and between individual males. We bring our results on the organization of male courtship displays to bear on how they provide novel tools for understanding the basis for female choice.

## **METHODS**

### *Study Site*

We collected data in January and from June to August in 2014, 2015 and 2016, and from September to December in 2017, at the Milpe Bird Sanctuary of the Mindo Cloudforest Foundation, in northwestern Ecuador (~0° 1’ 48” N, 78° 57’ 12” W). Milpe comprises 100 hectares of west slope, Chocó-Andean forest, at an elevation of 1,100 m, and features a network of maintained trails.

### *Study Species*

Male *Masius* use and maintain a small section of fallen mossy logs (~20–60 cm2 surface) on which they perform their courtship displays. The dispersion of display logsconstitutes an exploded lek, a lekking system in which display logs are often in aural, but rarely in visual, contact (Bradbury, 1981). Male *Masius*, like many other manakin species (Ryder & Durães, 2005), undergo a plumage maturation from green (hatch-year) to pre-definitive (second-year) to definitive plumage (third-year) and females copulate only with definitively plumaged males (pers. obs.). Pre-definitive males differ from females and hatch-year males by the addition of golden horn or crest feathers on the forehead and the emergence of waxy, but still greenish, nape feathers. Definitive males differ from pre-definitive males by fully black or yellow wing and body feathers and brownish-orange nape feathers. Female plumage consists of a soft, green wash over the majority of the body, with a crown that is slightly darker than the body feathers, extending to a point down the back of the neck. Darkening of the crown and neck may be less apparent in younger females. Females are best differentiated from green males by their darker nape feathers, as well as by behavioural cues, though distinguishing these features in the field or from video footage can be challenging. Displaying males almost certainly can distinguish females from pre-definitive males by plumage (Morales-Betancourt & Castaño-Villa, 2018; pers. obs.). Many of the females we observed were individually colour-banded and recaptured or resighted at intervals long enough to assign sex with certainty, but for unbanded green birds, we deemed them females based on the behavioural reactions of the males.

### *Field Methods*

We searched for and observed display logs, both in person and with video surveillance, and determined the approximate location of new display logs by conducting regular observations within the lek. We listened for the diagnostic descending whistle and pop of the log-approach display (Prum & Johnson, 1987), detectable from up to 50 meters away. Once a display log was detected, we set up a video camera (Sony Handycam HDR-CX405 or Sony Handycam HDR-CX240, Sony Corp., Tokyo, Japan) three to five m from the display log to record continuously, starting between 7:00 and 9:00 a.m. The cameras were housed in a weatherproof container and externally powered, with either a 12v motorcycle battery or a 10,000 mAh GETIHU power bank (Shenzhen Top Star Industry Co. Ltd., China). Recording ended when the memory card reached capacity, the batteries died, or weather necessitated camera retrieval. We ran 4 to 6 cameras daily, with each camera recording approximately 7 hours of video. We prioritized video surveillance at logs where manakin activity and female visitation were highest, or at newly discovered logs. We timed and counted *nurrt* advertising calls during less active times of the year (Jun to Aug 2014; field notebooks) and at the height of the courtship display season (Nov to Dec 2017; event recorder). We delimited bouts of *nurrts* by applying a broken-stick model (Sibly, Nott & Fletcher, 1990) in the *SiZer* R package (Sonderegger, 2020).

We used mist-nets (6 and 12-m-long, 30mm mesh) to capture and mark individuals for in-person or video identification. All individual *Masius* were banded with a numbered aluminum band and a unique combination of plastic colour bands. Other species were released immediately. All field methods were approved by the University of Wyoming Institutional Animal Use and Care Committee (Protocol# 20160602DM00242-02).

### *Video Coding*

To identify when *Masius* came to the display log so that those clips could be excised from the full video later, we filtered continuous video footage with a movement detection program developed in Python by Liam Taylor. We define a display bout as a sequence of distinct behavioural elements, in which male absence never exceeded 60 seconds. All bouts analysed had at least 20 behavioural elements, lasted at least 60 seconds, and contained at least one of each of the two most important behavioural elements: an audible log-approach display (ALAD) and a Bow (Table 1). We coded bouts using Behavioural Observation Research Interactive Software (BORIS), an open-source event logging tool (Friard & Gamba, 2016). Our dataset included 198 male-only (abbreviated SOLO or MULT) display bouts, 100 bouts with a female audience (abbreviated AUDI) and 14 bouts with at least one copulation (abbreviated COP; Table 1) coded with BORIS. A total of 41 [CHECK] display elements occurred in the raw BORIS data files. The full list of unabbreviated behaviour codings and their condensation into the 21[CHECK] elements used for analyses is given in Table S1. Some elements were subsumed into others and then abbreviated for further analyses (e.g., "Male 1 bow right" and "Male 1 bow left" were combined as "Bow"). "Other display behaviour" (OthM, OthF and OthC) included a slightly different set of elements for the three contexts. A list of descriptions and counts of occurrences of the most important behavioural elements is given in the Results. Voucher clips of the relevant behavioural elements and vocalizations, are archived at the Macaulay Library of Natural Sounds at Cornell University.

Table 1. Behaviours used as nodes in the ethogram networks. X denotes behaviours present in that bout type. Silent log-approach displays (SLADs) occurred only when there was no female at the display log. Similarly, males performing in what would eventually become a Cop bout never performed a mixed element, nor was a second male ever present. Note that all three networks had 16 nodes, 11 of which were shared across all three contexts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Node Abbreviation | Full Name | Mal network | Fem network | Cop network |
| ALAD | Audible log-approach display | X | X | X |
| AttC | Attempted copulation |  | X | X |
| B2AL | Bird2 ALAD | X |  |  |
| B2Bw | Bird2 Bow | X |  |  |
| B2Nk | Bird2 neck twist | X | X |  |
| Bow | Side-to-side Bow | X | X | X |
| Cop | Copulation |  |  | X |
| End | Bout end | X | X | X |
| Fff | Female off log |  | X | X |
| Fon | Female on log |  | X | X |
| HafB | Half bow | X | X | X |
| HdBw | Head-down bow | X | X | X |
| Metr | Metronome |  |  | X |
| Mix | Mixed element | X | X |  |
| Neck | Neck twist | X | X | X |
| Oth | Other | X | X | X |
| SLAD | Silent log-approach display | X |  |  |
| Start | Bout start | X | X | X |
| Swtc | Switch | X | X | X |
| Taf | To-and-from log flight | X | X | X |
| Zro | Zero | X | X | X |

### *Ethograms as Networks*

Social network analysis provides a useful quantitative toolkit for evaluating social behaviour (McDonald, 2007; Pinter-Wollman et al., 2014), including applications to analyses of dominance hierarchies (McDonald & Shizuka, 2013) and collective decision-making involving a certain degree of self-organizing behaviour (Couzin, Krause, Franks, & Levin, 2005). Ethograms have a long history as tools for organizing and assessing sequences of behavioural elements (Huxley, 1914; Tinbergen, 1963). We constructed ethograms of the display sequences from the distinct behavioural elements listed in Table 1. Ethograms suffer from the drawback that they are largely descriptive and not particularly amenable to quantitative analysis. Here, we extend the utility of ethograms by treating them as weighted networks, allowing us to quantify variation across display sequences in different contexts (REDO for the two distinct SOLO and MULT types? Mal vs. Fem vs. Cop). We constructed a network for the 198 Mal bouts from the transitions (8,590 edge weights) between 16 distinct behavioural elements (nodes). The Fem network had 16 node types, linked by the 10,600 edge weights in 100 bouts. The Cop network had 16 nodes and 1,893 transitions (edge weights) in the 14 Cop bouts. Eleven of the 16 node types were shared by all three categories of display (Table 1). The "dummy" node types “Start” and “End,” and their concomitant edges, are shown for reference in the network diagrams, but were excluded when computing network metrics. We assessed four node-based metrics: degree, effective degree (McDonald & Hobson, 2018), betweenness, and eigenvector centrality. With the exception of effective degree (*EffDgr*, R script available on request), all were calculated with the *R* package *igraph* 1.2.5 (Csardi & Nepusz, 2006). Effective degree, by analogy to effective number of alleles in population genetics, assesses the number of edges of equal weight that would have the same observed edge-weight diversity as the observed node’s edges (McDonald & Hobson, 2018). If all edges emanating from a node have equal weights, the effective degree equals the simple binary degree. As edge weight variance increases, the effective degree (effective number of edges) decreases. Effective degree is given by the inverse of the summed, squared edge-weight frequencies, *p*i over *k* edges:

We sized nodes in the network diagrams according to eigenvector centrality, because of its familiarity and the integrative nature of eigen-analysis. We compared network density and effective degree across the three types of networks (Mal, Fem, Cop). Because the number of bouts and transitions (edge weights) was much greater for Mal and Fem displays than for Cop displays, we conducted randomizations to verify the comparative metrics. For 10,000 replicates, we therefore randomly selected, with replacement, 1,893 transitions (edge-weights, based on the smaller Cop sample size) from the 145 observed edge types (8,590 edge weights) in the Mal bouts and the 105 observed edge types (10,600 edge weights) in the Fem network, and assessed the distribution of those network metrics in the resampled networks.

As a measure of network complexity the function *all\_simple\_paths()* in the *igraph* 1.2.5 *R* package, calculates the total number of simple paths (no nodes revisited, and self-loops omitted) through the network. Because the metric grows exponentially with the number of edges, it was computationally intractable to calculate the number of simple paths for the 154-edge Mal network. We therefore calculated the number of simple paths for a set of 200 resampled networks, each with 102 edges. That number of edges is twice that for the Cop network (51 edges, ignoring self-loops) and approximately two thirds of the number of edges in the full Mall network. We resampled edges, with replacement, by selecting additional edge weights with probabilities determined by the observed edge weights in the full set of 154 weighted, directed edges. We then repeated the resampling calculation 200 times.

### *Entropy*

Recently, information entropy has been applied to assessing complexity in vocal repertoires (Da Silva, Piqueria, & Vielliard, 2000; McCowan, Doyle, & Hanser, 2002; Freeberg & Lucas, 2012; Palmero, Espelosín, Laiolo, & Illera, 2014). Repertoires with high entropy have a large "surprise" factor, whereby the next syllable type is unpredictable, and the repertoire can be considered highly complex. In contrast, low entropy means low surprise and high levels of repetition. Further, complexity, per se, has been proposed as a fundamental aspect of animal communication (Bradbury & Vehrencamp, 2014; Patricelli & Hebets, 2016). Here, we use entropy to explore variation across display contexts. For each bout type (Mal, Fem, Cop) we calculated the overall frequencies of the elements, *p*i. The maximum, zero-order entropy for bouts (14 possible elements) is given by log2(14) = 3.81 bits. The first-order entropy, in bits, is then given by

(1)

### Because the maximum possible entropy is dependent on the number of elements (here distinct behavioural elements) and because that number varied from 4 to 12 across bouts, we used the adjusted entropies, given by the raw entropy, divided by the maximum possible entropy determined by the number of elements (Vanderbilt, Kelley, & DuVal 2015). [THIS EXPLAINS our 0-1 ENTROPY SCORES] Thus, the entropy scores ranged from 0 to 1. We compared the adjusted entropy scores across the three categories (Mal, Fem, Cop) with ANOVA and Tukey's HSD.

### *Compression Ratio*

Lossless compression of information is an important feature of many digital applications. As with entropy, many compression algorithms use a bit-based approach to compressing sequences of elements, which can then be applied to text, signal waves, pixels or sequences of *Masius* display elements. We term the extent to which a sequence of elements can be compressed its *compression ratio*. We computed compression ratio as the original length of a display string (using the same single-letter alphabetic abbreviations as for entropy), divided by the length of the string after compression by Huffman encoding (Huffman, 1952) and the LZ77 algorithm (Ziv & Lempel, 1977), using the *R* package *brotli* (Ooms & Google Inc., 2018). Huffman encoding (Huffman, 1952) creates optimal prefixes for subsequent compression of strings of elements (here, our *d*i), assigning the smallest bit length to the most common element. LZ77 (Ziv & Lempel, 1977) uses a sliding window, distance-length algorithm to detect and compress recurring motifs in the data stream. An increase in compression ratio, like a decrease in entropy, indicates more repetition of elements and motifs, and thus a decrease in complexity. One potential advantage of compression ratio over entropy is that the latter is computed just from the frequencies of display elements, whereas LZ77 compression also assesses the order of the display elements. Further, LZ77 compression counts repeated motifs, as well as single elements, via its sliding window, distance-length feature, providing an additional layer of pattern detection and compression. We compared the *compression ratio* values across the three categories (Mal, Fem, Cop) with ANOVA and Tukey's HSD.

### *Jaro String Distance*

Matching similar records has long been a goal of computer-assisted methodologies in a field known as record linkage (Fellegi & Sunter, 1969). Jaro (1989) developed a simple but elegant method for matching census records from disparate sources. The algorithm analyses the number of transpositions and mismatches between two strings, and assigns a distance between 0 (no matches) and 1 (complete match). We used Jaro distances to assess similarity of display sequences within and across contexts and within and between males, via the *R* package *stringdist* (van der Loo, 2014). If context-dependence is the major factor explaining behavioural variation, then we expect behaviours in a given context to resemble one another, regardless of male identity. That is Mal Jaro distances **within** a male's bouts should be similar to Mal Jaro distances **between** different males. If, however, individual differences are most important, Within-male Jaro distances, for both Mal and Cop, should be low and Between-male distances should be high, for both Mal and Cop bouts. We created a matrix of Jaro distances between a given string and all other strings to which we were comparing it (e.g., a 198 X 198 matrix between all the Mal strings). We then used the mean of the distance between a string (display sequence) and all others of its context, to assess similarity (low Jaro distances denote high similarity of the compared strings of display elements). Jaro string distances, unlike some alternative algorithms, are little affected by differences in string length (Pearson correlation coefficient for Jaro-distance matrix vs. string-length matrix for the 198X198 Mal display matrix = -0.038). The Expected values, under the hypothesis of individual differences as the driving force in generating variation in displays, were generated by simply labelling the Observed Jaro distance values, such that high Jaro distances were labelled Between and low values were labelled Within. Note that these Expected values are simply a vehicle for visualizing the differences between the predictions of the individual-difference and context-dependence hypotheses, not for any quantitative comparison of Observed vs. Expected. [DOES THIS NEED FURTHER CLARIFICATION?]

### *Randomizations to Assess Effect of Small Sample of Copulation Bouts*

As for the network analyses, we used randomization procedures for the entropy, compression and Jaro analyses, to assess their sensitivity to the large discrepancy in sample size between COP bouts (*n* =14) and Mal (198) and female audience (AUDI, *n* =100) bouts. Across each of 10,000 replicates, we randomly selected (without replacement) 14 strings from the SOLO, MULT and AUDI bouts, and calculated the mean of the ensuing 14 entropy, compression or Jaro metrics. We then compared the distribution of the random replicate metrics to the values calculated using the full sample. All analyses were conducted in Program R (4.0.2) (R Core Team 2020). All R scripts used for the analyses are available, on request, from the corresponding author. [OR AS SUPPLEMENTARY MATERIAL?]

## **RESULTS**

All display bouts occurred in Julian weeks 25 through 3 (24-Jun to 14-Jan), with a pronounced peak from October to December (234 of the 312 display bouts), and a pronounced dip in June, July and August. Male-only displays (SOLO and MULT) were first observed on 24-Jun, female audience bouts (AUDI) were first observed on 15-Jul, and the first copulation was observed on 26-Oct. Males do a low-amplitude "*nurrt*" advertising call from various perches in the vicinity of the display log. (Prum & Johnson, 1989). For Jun-Aug 2014 (low courtship activity), 2,895 mins of observation yielded 2,182 nurrts (0.75 nurrts/min). For Nov-Dec 2017 (peak activity), 1,518 mins of observation yielded 3,311 nurrts (2.2 nurrts/min). Using broken-stick analysis (Sibly, Nott & Fletcher 1990), we defined bouts as sets of nurrts with no gaps > 46 sec. For the 2017 data, 48 nurrts were singletons. Of the 236 bouts with ≥ 2 nurrts, 106 comprised > 10 nurrts. For these 106 longer bouts, the nurrt rate averaged 4.23 nurrts/min (sd = 0.8) within calling bouts, with a median of 4.17 nurrts/min.

Table 2 describes and quantifies 11 major behavioural elements performed by male *Masius* during courtship displays at their logs for our three contexts (Mal, Fem, Cop). The full list of display elements and their abbreviations is in Table S1. When coding videos, we recorded whether the female was above or below the male on the log. Although virtually all displays logs were on slopes, "above," on level terrain would mean further along the log than the incoming trajectory of the swoop down during an ALAD. Overwhelmingly (Pearson's *Χ*2 with Yates' continuity correction = 4140.5, df = 1, p < 0.0001), the female was below the male for Bows (4,540/4,602), whereas she was above the male for Neck twists (553/607). However, most Neck twists (3,159/3,807 = 83%) were performed when the female was not on the log (females not on the log were not counted in the above-below comparison). Thus, males often performed a long series of Neck twists after females left the log following copulations, and, less frequently, when the female was close to, but not on, the log.

Table 2. A list of the common Masius courtship display behaviours, their abbreviations, descriptions and how many times they were performed in the 312 coded display bouts.

|  |  |  |  |
| --- | --- | --- | --- |
| Behaviour | Abbreviation | Description | *n* |
| Audible log- approach display | ALAD | The most dramatic element in the display repertoire. Males swoop down from in or near the canopy to land with a snap on the display log. Males will perform this alone, with other males and when a female is present. Copulations always occur immediately following an ALAD. | 814 |
| Side-to-side bow | Bow | The most frequent display element. A male moves from one edge of the display log to the other bowing his head almost to the surface at each edge. If a female is present, he is virtually always (4,540/4,602) upslope of the female. | 6,658 |
| Copulation | Cop | Copulations generally occur after a long series of Bows and an ALAD. The male lands on the display log and then mounts the female from downslope on the log. If a second copulation occurs during the same bout, the number of Bows proceeding the ALAD is greatly reduced. | 24 |
| Head down bow | HdBw | The male stands in one spot and bows repeatedly, such that his cheek is almost touching the surface of the log; almost always performed after an ALAD. | 2,069 |
| Metronome | Metr | Most commonly performed by solo males soon after a female leaves the log. The male sways back and forth in a rhythmic manner evocative of a metronome, while perched near, but not on, the log. | 8 |
| Mixed element | Mix | Most often performed during display sequences with no female present. It represents a mixture of two other elements, such as a Bow and a Neck twist. | 202 |
| Silent log- approach display | SLAD | Silent log-approach displays are mostly performed when no female is present. The head down bow performed after an ALAD is often omitted after an SLAD, suggesting that it might function mostly to practice approach trajectories. | 408 |
| Neck twist | Nck | Most frequently occurs when a female is nearby, but not on the log. If the female is off the log, he often does a string of Neck twists until she descends to the log. If the female is present, he is virtually always (553/607) downslope of the female. | 4,590 |
| Switch | Swtc | Quick rotation of the body ~45 degrees back and forth, without moving up or down the display log. Females often perform switches after an ALAD, while up-slope of the male. Males were not observed performing switches in the presence of a female. | 912 |
| To and from log flights | Taf | Flights of variable duration and distance from the log to the surrounding vegetation and back. Taf are frequently performed immediately after a female leaves the display log. | 935 |
| Zero | Zro | A period in which the male is present on the log, but inactive for >5 seconds | 1,466 |

Table 3 presents rates (per min) for Bows, Neck twists, and ALADs in the three contexts (Mal, Fem, Cop). The rate for Bows (Mal 3.4 ± 3.1, Cop 16.6 ± 4.1) was far lower in Male-only display bouts than in Copulation bouts and lower, though not as dramatically, for Neck twists (Mal 6.4 ± 8.2; Cop 8.7 ± 6.4). Interestingly, males actually performed ALADs at a higher rate during Mal bouts (1.22 ± 0.5) than during Cop bouts (0.5 ± 0.2), likely because, during copulation bouts, ALADs almost always preceded a copulation or attempted copulation, and were therefore less likely to be repeated (X of Y Cop bouts had > 1 copulation). During copulation bouts, males performed many more Bows before the first copulation (54 ± 16.9) compared to before the second copulation (5.1 ± 13.3). The converse was true for Neck twists, with 21.6 ± 39.6 before the first copulation and 35.4 ± 28.3 thereafter. All but two of the 14 bouts leading to a copulation followed a consistent pattern of 1 to 17 behaviours (5.1 ± 4.9) followed by a long string of Bows (range 25 to 76), then an ALAD and the copulation. The two unusual bouts differed in that the female left the log one or more times, after 20 or more male display elements, before settling on the log for the copulation. The two males with multiple copulations (ID980 n = 3 range 51 to 67; ID296 n = 10, range = 25 to 76) did not differ in the number of Bows they performed prior to the first copulation (Welch's two-sample t-test: df = 8.16 effect size = 0.37 P = 0.44).

Table 3. Rates of Bow, Neck and ALAD display for Masius. During Cop bouts, mean Bow and Neck rates were higher, but had lower maxima, than during Mal and Fem bouts.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Bows/minute | | | | Neck twists/minute | | | | ALADs/minute | | | |
|  | Mean | s.d. | Min | Max | Mean | s.d. | Min | Max | Mean | s.d. | Min | Max |
| Mal | 3.35 | 3.08 | 0.2 | 22.72 | 6.43 | 8.23 | 0.18 | 30.65 | 1.22 | 0.46 | 0.39 | 2.8 |
| Fem | 10.96 | 6.21 | 0.39 | 26.71 | 8.69 | 6.09 | 0.16 | 32.22 | 0.53 | 0.2 | 0.18 | 1.15 |
| Cop | 16.57 | 4.05 | 7.97 | 22.36 | 8.71 | 6.35 | 1.33 | 27.05 | 0.49 | 0.16 | 0.22 | 0.80 |

Table 4 shows the first and last record (capture or sighting) for each of the 12 colour-banded males in the study, and the number of Mal, Fem and Cop bouts in which each participated. We observed 12 individually colour-banded females copulate with the three colour-banded copulator males. Some of those females had histories of visitation at multiple display logs. In addition, three other colour-banded females were observed watching displays, but never copulated during our video monitoring.

Table 4. Identities of males performing courtship displays. Note that 10 of the 12 banded males displayed for at least one female (i.e., participated in ≥ 1 Fem bout). Only three banded males copulated. In the MaleID column, Unb denotes bouts performed by an unbanded male. Span is the number of days over which they were observed.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| MaleID | Total bouts | Mal | Fem | Cop | Date banded | Date last seen | Span (days) |
| 112 | 20 | 8 | 12 | 0 | 23-Sep-17 | 20-Dec-17 | 89 |
| 113 | 5 | 4 | 1 | 0 | 17-Sep-17 | 20-Nov-17 | 65 |
| 291 | 1 | 1 | 0 | 0 | 26-Jul-15 | 8-Aug-16 | 380 |
| 296 | 153 | 99 | 44 | 10 | 10-Jul-15 | 20-Dec-17 | 895 |
| 299 | 9 | 6 | 3 | 0 | 4-Jul-15 | 30-Nov-17 | 881 |
| 940 | 7 | 2 | 4 | 1 | 26-May-13 | 16-Jan-15 | 601 |
| 948 | 2 | 1 | 1 | 0 | 10-Jun-13 | 13-Sep-17 | 1557 |
| 965 | 20 | 15 | 5 | 0 | 3-Jul-14 | 3-Oct-17 | 1189 |
| 976 | 2 | 1 | 1 | 0 | 22-Jul-15 | 11-Jan-17 | 905 |
| 978 | 3 | 1 | 2 | 0 | 23-Jul-14 | 21-Aug-15 | 395 |
| 980 | 78 | 53 | 22 | 3 | 24-Jul-14 | 20-Dec-17 | 1246 |
| 982 | 1 | 1 | 0 | 0 | 24-Jul-14 | 17-Jan-16 | 543 |
| Unbanded | 11 | 6 | 5 | 0 | NA | NA | NA |

Time to first copulation in the 14 Cop bouts was 129.1 ± 41.3 secs, with a range from 70.4 to 212.7. The mean number of copulations in a Cop bout was 1.7 ± 0.8 (range, 1 to 4). Total bout length in Cop bouts averaged 270.0 secs (± 140; range, 134.7 to 690.3). The number of times that females left the display log (abbreviated *Fff* as a network node) was higher and more variable in Fem bouts (2.6 ± 2.8; range, 1 to 23; c.o.v. 1.08) than in Cop bouts (2.1 ± 1.1; range, 1 to 5; c.o.v. 0.53).

Ethogram networks for each of the three contexts (Mal, Fem, Cop) are shown in Figure 1 (with nodes representing the behavioural elements listed in Table 1). A clear progression is apparent, from dense, with high effective degree for the Mal network to sparse (low density), with low effective degree for the Cop network. None of the 10,000 randomly subsampled male-only networks had an effective degree as low as that of the copulation network (mean = 24.5 ± 0.83 for the Mal random replicates vs. 2.8 for the Cop network).

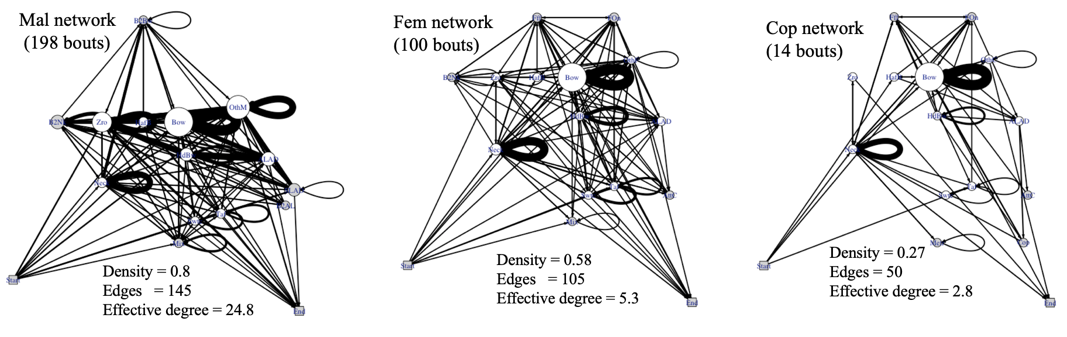


Figure 1. Masius display ethograms visualized as networks in three contexts (Mal, Fem, Cop). The nodes represent distinct behavioural elements (e.g., ALADs and Bows). The directed edges represent transitions between elements, with weights representing how often those transitions occurred. For example, Bows were often repeated in a long string, creating a self-loop (edge emanating from, and returning to, a node) with heavy weight. Size of node is proportional to eigenvector centrality. Start and End (lower left and lower right) are dummy nodes, to indicate the general direction of flow through the ethogram/network. The major difference across contexts is increasing simplicity – from dense, with high effective degreein male-only displays (Mal) to sparse (low density), with low effective degree,in displays leading to copulations (Cop). Effective degree is based on variance in edge weights. If all edges from a node have equal weight, effective degree equals the binary degree (number of distinct edges). As variance in edge weight increases, effective degree declines relative to binary degree. For the Cop network (effective degree 2.8), self-loops on Bow and Neck twist are, effectively, the only ones that really matter. In contrast, edge weight is much more evenly distributed in the Mal network (24.8 effective edges).

To assess the relative complexity of the Mal and Cop networks, we enumerated the number of simple paths through the Cop network. For the Mal network, because the enumeration was computationally intractable for the full network, we computed the number of simple paths for 200 subsampled Mal networks with a tractable number of edges (102), as described in Methods. The Cop network had 1,530 simple paths. The 200 resampled Mal networks, each with 102 of the 154 possible edges, had more than two orders of magnitude more simple paths (mean = 649,610; sd = 751,673; range 88,850 to 4,702,072).

Figure 2 summarizes the information entropy analysis. The 198 Mal bouts had high adjusted entropy (diversity, randomness or surprise), often approaching the maximum possible value of 1.0. In contrast, the 14 Cop bouts had low entropy. The adjusted entropies differed among the three contexts (Mal, Fem, Cop) under ANOVA followed by Tukey's HSD (P < 0.001). Higher-order entropies (Fig. S1) continue to decline beyond second-order for Mal bouts more than they do for Fem or Cop bouts. Those differences across higher-order entropies suggest that, in Fem and Cop bouts, the current element depends only on the preceding element. In contrast, Mal bouts elements depend, in complex ways, on earlier elements.

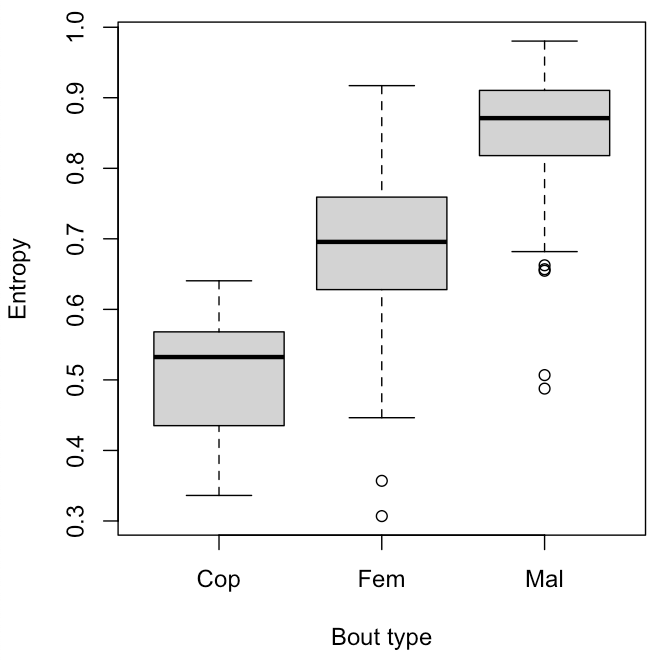


Figure 2. Entropy (relative to the maximum possible) for 14 Cop bouts, 100 Fem bouts and 198 Mal bouts. In contrast to the avian song repertoire literature, more successful (Cop) displays were simpler (had lower entropy) than were less successful (Mal and Fem) display sequences. We coded each display bout as a string of characters (*a* to *n*) representing the sequence of 14 discrete behavioural elements in Table 1, and then computed its information entropy (in bits). Mal strings had high entropy (diversity, surprise or randomness; mean 0.86; range, 0.49 to 0.98), often approaching the theoretical maximum of 1.0. In contrast, Cop strings had low entropy (mean 0.51; range, 0.34 to 0.64), indicating considerable redundancy (repetition of elements such as Bow. See self-loops in Fig. 1). Fem bouts (mean 0.68, range 0.31 to 0.92) had variable entropy, overlapping with both Mal and Cop bouts.

Figure 3 depicts compression ratio of the display sequences across the three contexts (Mal, Fem, Cop). Cop bouts had far higher compression ratios than did Mal bouts. The compression ratios differed among the three contexts (Mal, Fem, Cop) under ANOVA followed by Tukey's HSD (P < 0.001). Figure 4 shows the compression ratio of all bouts performed by our most prolific male (M296; Table 4), by sample day. M296 performed many Mal, Fem and Cop displays, but always had Mal bouts with low compression ratio, even after multiple copulation bouts.

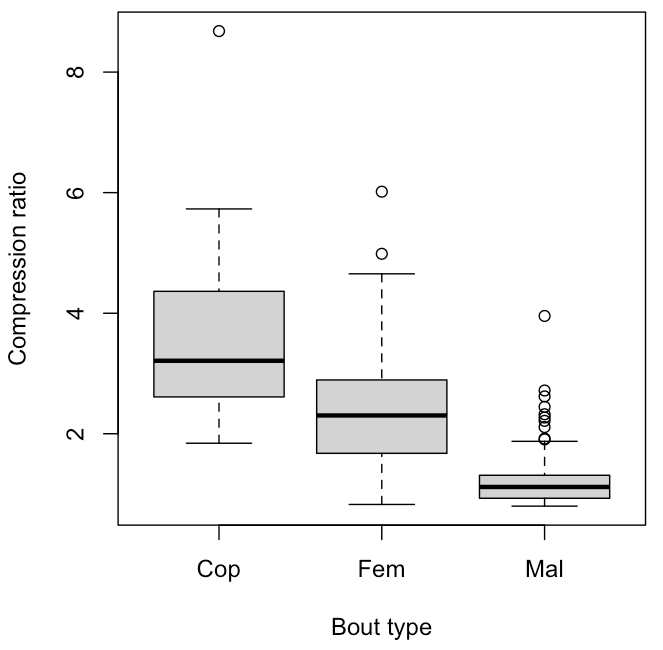


Figure 3. Compression ratio of *Masius* displays (strings of behavioural elements) for Cop, Fem and Mal bouts. Lossless compression was conducted with Huffman encoding and the LZ77 algorithm. Compression ratio is the ratio of the length of the uncompressed string to the compressed string. For example, the maximum compression ratio (8.68) was for a Cop sequence (upper left). Mal sequences had significantly lower compression ratios (range 0.8 to 3.96; mean 1.2) than did Cop sequences (range 1.8 to 8.68; mean 3.7). Mal sequences were often a complex jumble of disparate behavioural elements. Cop sequences, in contrast, generally had had oft-repeated elements, especially Bows, as well as repeated motifs of varying length. Fem sequences (mean 2.4) were intermediate. The repeated motifs detected (and compressed) by the length-distance, sliding window feature of the LZ77 compression algorithm are one major difference from the entropy analyses (Fig. 2), which depend only on relative frequencies of elements.

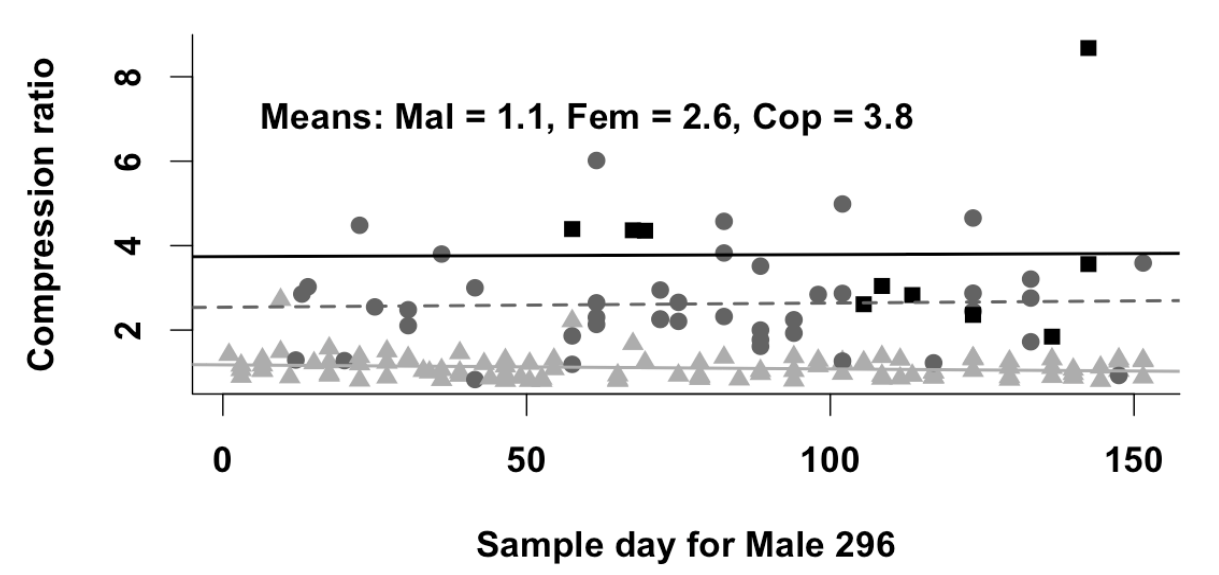


Figure 4. Compression ratio of the Mal (light gray triangles), Fem (dark gray circles) and Cop (black squares) bouts by Male 296, across his 152-day span of appearing in video clips. Note that, despite Cop bouts beginning as early as sample day 57, his later Mal bouts continued to have low (~1.1) compression ratios. The slope of compression ratio as a function of sample day was flat (- 0.001). That pattern runs counter to the hypothesis that males might perform more organized Mal displays after performing one or more successful, organized (high compression ratio) Cop displays, which would produce a positive slope. Instead, male display organization depends largely on the context (Mal vs. Fem vs. Cop) –always simple and with high compression ratio in displays containing copulations, always complex (diverse) and with low compression ratios in Mal displays.

Figure 5 shows expected Jaro string distances (left four bars) under the hypothesis that individual differences between different males ("personalities") drive variation in display sequences. Low Jaro distances mean the displays compared are similar. The four gray and black bars on the right show the observed Jaro distances, which support the hypothesis that display sequences according to context (Mal, Fem, Cop), and do not support the individual-difference hypothesis depicted by the four leftmost expected bars. The pooled Within- and Between Mal Jaro distances differed significantly from the pooled Within- and Between Cop Jaro distances (*t-test*: *t* = 17.24, *df* = 91.2 P < 0.0001). To examine the effect of the small sample size (43) of Cop Jaro distance pairs between males, we randomly selected sets of 43 distance pairs and computed their mean Jaro distances from one another for 10,000 replicates. The absolute differences between the means of the resampled distributions and the observed means were < 0.0002, and within 0.1% of each other. None of the 10,000 random Mal-Within replicates (mean = 0.39; 95% CI = 0.36 to 0.43) overlapped the random Cop-Within replicates (mean = 0.25, 95% CI = 0.22 to 0.28). Likewise, no Mal-Between values (mean = 0.42; 95% CI = 0.39 to 0.45) overlapped with the Cop-Between values (mean = 0.23; 95% CI = 0.21 to 0.26). Differences between the Jaro distances across the three contexts (Mal, Fem, Cop) are shown in Fig. S2.

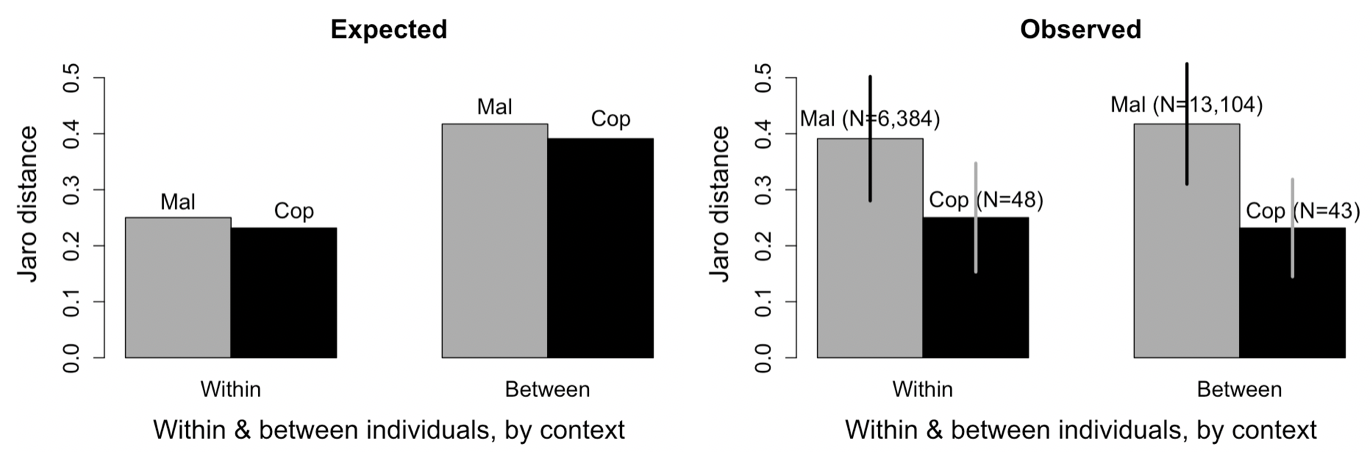


Figure 5. Expected (left four bars) Jaro string distance outcomes under the hypothesis that individual differences ("personalities") drive variation in display behaviour more than do contexts (Mal, Fem, Cop). If "personalities" mattered most, we would expect a male's own displays to be similar to each other (short Within Jaro distances; Mal-Within Cop-Within), regardless of context (Mal, Cop). We would expect a male's displays to be dissimilar to those of other males (high Mal-Between and Cop-Between). Instead (Observed, right four bars), a male's own Mal displays were almost as different from each other (Mal-Within = 0.39) as from those of other males (Mal-Between = 0.42). Cop displays were always very similar (short Jaro distances) whether Within or Between. In fact, males' Cop displays were slightly more similar (Cop-Between = 0.23) to those of other males, than to their own (Cop-Within = 0.25). Thus, variation in display is largely a matter of context; Mal Within resembles Mal Between and both differ greatly from Cop Between and Cop Within, which closely resemble each other, regardless of male or female identity. Vertical lines on rightmost four bars are ± 1 sd. None of the values in the Observed Mal-Within resampled distribution overlapped with any in the Observed Cop-Within resampled distribution. Likewise none of the values in the Observed Mal-Between resampled distribution overlapped with any of those in the Observed Cop-Between resampled distribution (see Methods and text for more details), meaning that the small sample size of Cop Jaro distances did not affect the conclusions.

## **DISCUSSION**

## Our major result is that context dependence and simplicity, not individual differences ("personalities") or complexity characterize variation in the courtship displays of male *Masius*. The pattern holds, across our three contexts (Mal, Fem, Cop), whether assessed in terms of weighted networks (Fig. 1), entropy (Fig. 2), compression ratio (Figs. 3 and 4) or Jaro string distance (Fig. 5). The emergent feature is a progression, from diverse (more random) display sequences that were dissimilar (from one bout to the next) when no female was present (Mal bouts), to redundant (predictable, organized) and similar to one another (within or between males) in displays leading to copulations (Cop bouts). In contrast, variation between males was relatively minor, especially in displays that included a copulation. Each of our analyses (networks, entropy, compression ratio, Jaro distances) contributes to a fuller understanding of how simplicity and context-dependence characterize the organization of *Masius* courtship displays.

## We assessed the comparative simplicity of the displays by analyzing them as networks (Fig. 1). Network density (proportion of all possible edges actually occurring) decreased from Mal to Fem to Cop, meaning that Mal bouts (density = 0.8) had many different kinds of transitions between display elements, while Cop bouts, especially (density = 0.27), had relatively few types of transitions. Low effective degree corresponds to high variance in edge weight, with one or a few edges dominating the transitions. With no variance in edge weights, effective degree equals simple binary degree (i.e., a simple count of the edges, disregarding weights). Cop bouts had low effective degree (2.8), meaning that most of the edge weight was concentrated on a few edges, particularly the repetition (heavy self-loops in Fig. 1) of Bows (before audible log-approach displays, ALADs) and Neck twists (after ALADs). In contrast, the Mal network had much higher effective degree (24.8), meaning a more even distribution of edge weights, and lower rates of repetition for any particular display element. All-simple-paths analysis showed a dramatic difference in the number of possible paths among nodes when comparing the Cop network (1,530 simple paths) to the Mal network (mean of 649,610 paths). That difference certainly underestimates the true difference by a wide margin, because we used a set of reduced-edge Mal networks to make the analysis computationally feasible.

## Commonly, increased song complexity, often assessed by entropy (Da Silva, Piqueria, & Vielliard, 2000; Palmero, Espelosín, Laiolo, & Illera, 2014) has been considered a positive attribute, associated with increased mating success. Here, we find that, to the contrary, Cop (successful courtship) bouts had significantly lower entropy (Fig. 2, mean = 1.57) than did Fem bouts (mean = 2.21) or Mal bouts (mean = 2.61). Thus, display bouts that include a copulation tended to be more predictable, with greater repetition of display elements, particularly Bows and Neck twists than do either Fem or Mal bouts. Concordant with our results, Vanderbilt, Kelley and DuVal (2015) found that in the dual-male courtship displays of Lance-tailed Manakins, *Chiroxiphia lanceolata*, displays for females had lower entropy than did displays in the absence of females.

## Compression ratio (Fig. 3) increased dramatically from Mal bouts to Fem bouts and thence to Cop bouts. Compression ratio was highest in Cop bouts, which featured long strings of repeated behaviours—particularly Bows and Neck twists—and the repeated motifs of elements detected by the LZ77 algorithm.

## Overall, Jaro string analysis (Fig. 5) indicated that Mal display bouts were variable within and between males, and that Cop bouts were similar, whether within or between males. Cop bouts (mean Jaro distance = 0.24) were more similar to each other than Fem bouts were to each other (mean Jaro distance = 0.38) or Mal bouts were to each other (mean Jaro distance = 0.41; the bars in Fig. S2). Thus, structural differences in the organization of courtship displays were due largely to the context—female present or not, and progressing toward copulation or not— and not to consistent differences between males in how they organized their displays. Although they measured similarity differently, Botero et al. (2009) stressed the importance of consistent song repertoires in Tropical Mockingbirds (*Mimus gilvus*), both as an ontogenetic trajectory and to explain reproductive success. Similarly, Janisch, Perinot & Fusani (2020) found that male *Manacus vitellinus* developed consistent courtship display sequences, in terms of movements around their courts.

## Why are Mal bouts (no female present) more random and disorganized than either Fem or Cop bouts? Displays with no female present (Mal bouts) could represent practicing, identifying the best trajectory for log-approach displays (ALADs and, particularly, SLADs), or maintaining control of a display log, to name but a few possibilities. None of these behaviours requires strict adherence to a fixed routine, possibly increasing the complexity and diversity of male-only displays (Mal) when compared to Fem or Cop bouts. The simple presence of a female appeared to suffice in galvanizing males to perform simpler and more organized display sequences.

## A feature that still puzzles us is how females choose which display logs to visit. In many lek-mating species, females can use fairly long-distance visual or acoustic cues to locate the courts of displaying males. In *Masius*, the "nurrt" advertising call is of surprisingly low-amplitude and is performed at a low rate overall, even at the height of display activity in Nov and Dec (2.2 nurrts/min at the two most active display logs) and within-bout vocalizations averaged only 4.2/min, reasonably similar to the calling rates documented by Prum & Johnson (1989). The call rate for *Masius* contrasts sharply with the far-carrying dual-male *toledos* of male Long-tailed Manakins, *Chiroxiphia linearis,* which occurred at overall rates > 5.4 toledos/min (McDonald, 1989) over the course of entire days (McDonald 2010, Fig. 1), with within-bout rates averaging 15.6/min (Trainer and McDonald, 1993). We conclude, provisionally, that males rely mostly on regular traffic of females to promote female visitation, as might be predicted under a hotspot model of lek placement (Bradbury, 1981; Beehler & Foster 1988). Further, our proportion of banded females and revisitation rates by banded females were far higher than in Long-tailed Manakins (pers. obs.), suggesting that female *Masius* may have much smaller home ranges that largely overlap the display logs of the males they sample.

## Our results do provide some insights into the bases and patterns of female mate choice, once they have chosen to visit a display log. The similarity of Cop displays across three distinct males, despite their being performed for at least 14 different females, suggests that a simple, rather invariant display worked well when females were ready to copulate, as indicated by females' steady presence on the display log during Cop bouts as compared to Fem bouts. A simple sequence of many Bows, followed by an ALAD, immediately followed by a copulation and then a series of Neck twists, was the consistent pattern for Cop bouts. One might suppose, that once having succeeded in copulating, males would modify their Mal and Fem displays to more closely resemble the simple rhythm of Cop displays. Our data do not support that hypothesis. Fig. 4 shows that M296's Mal displays did not become more organized and less complex as the breeding season progressed. Even immediately after Cop bouts with high compression ratios, M296's Fem and Mal bouts had low compression ratios was similar to those of bouts long before the first copulation.

## Once females chose to visit, an obvious feature of Fem and Cop displays was that male behaviour depended on the relative position of the female. When she was below him on the log, he would perform Bows until she moved or he left the log to perform an audible log-approach display (ALAD). When she was above him, or off the display log, he would perform Neck twists. If a female was not in the immediate vicinity of the display log, the male might still perform those same behaviours, but would tend not to repeat them for as long, and would perform them in broken series, in a less predictable order. Further, the male would intersperse other display elements that he did not perform while the female was on the log, such as Metronome displays. The result was long strings of disjointed behaviours and frequent male flights to and from the display log. Although it was not the primary focus of this paper, we found that female as well as male behaviour varied across contexts. Females moved on and off the display log more during Fem bouts, perhaps because it was their first visit to a particular display log that season, or because they lacked experience in evaluating males. The results suggest that when females were prepared to copulate, they moved much less on the log, providing some feedback to the male on the female's interest level, and causing males to settle into the simple and predictable rhythm of the peri-copulatory displays.

## The most dramatic behavioural element in the courtship display of *Masius* is the audible log-approach display (ALAD); it occurred just prior to all 24 observed copulations, and is perceived as a blur by the human eye (see slow-motion video at http://www.uwyo.edu/dbmcd/masiusvideo.html). Why should females respond to such an odd and rapid maneuver? One intriguing possibility is that it requires extreme agility that could be useful in a foraging context, for females as well as their offspring. Howe (1979) argued that small frugivores might be particularly vulnerable to predation at fruit trees, which are reliable sites for prey detection. Consistent with Howe's hypothesis, Uriarte et al. (2011) found that four species of manakins spent a mean of only 13 ± 4.3 secs at their fruit trees. Many of the visits had zero perch time, because birds often gulped (Levey, 1987) the fruits and departed, without ever perching. For larger frugivores, perch times can be considerably longer (median visit length 225 secs, Wheelwright, 1991). Ultra-rapid maneuvers while taking fruit might significantly reduce predation risk at these predictable locations. High-speed video of *Masius* at fruit trees, particularly in the event of a predation attempt, would be an interesting test of this hypothesis.

## Although we found that greater simplicity and uniformity characterized successful courtship displays, we are not arguing that simplicity or uniformity, per se, increases reproductive success in *Masius*. Instead, we conclude that context (presence or absence of a female), and behavioural feedback from female visitors, as demonstrated in the elegant experiment by Patricelli, Coleman & Borgia (2006; see also Rodríguez et al., 2012), together create a progression from diverse (unpatterned, with low repetition of elements), to orderly (highly patterned with repeated elements and motifs) in displays leading to copulations. Displays leading to successful copulations are simple and uniform from one iteration to the next, whether within the Cop displays of an individual male or between the Cop displays of different males. What remains an intriguing question is how and why females transition from fidgety (many landings and departures from the log in Fem bouts) to focused (steadily remaining on the log during Cop bouts). Gibson (1996) found that the inter-pop interval was positively correlated with the number of females visiting a sage grouse (*Centrocercus urophasianus*) male, but that eventual copulation was related to display rate. Similarly, McDonald (1989) found that female visitation correlated with the output of dual-male unison *toledo* song in Long-tailed Manakins. Copulatory success, however, correlated with the duration of the dual-male butterfly flights, which occur between bouts of dual-male backwards leapfrog displays. Those studies suggest that female preferences are based, not on a single male trait, but instead vary with a female's sampling context, experience and, very likely, her reproductive state. Thus, understanding prior female experience (both within and across years) may be critical to understanding how female response to courtship drives context-dependent variation in displays. The ball is back in the court of the difficult problem of assessing female sampling histories, often stretching back several years, and of ascertaining female motivational states.

*Conclusions*

## We conclude that *Masius* courtship displays progress from disorderly jumbles to tightly organized sequences characterized by repetition of major display elements (particularly Bows) as the context moves from male-only displays to those leading to copulations. We have provided a worked example of the utility of network, lossless compression and Jaro string-distance metrics that could be applied to many other sorts of sequences of behaviours. For example, the methods used here could be applied to the frame of reference, in terms of motor activity (sensu Miller, 1988), variation in song repertoires, other sorts of courtship display, animal movement trajectories and even sequences of environmental variables such as time series of temperature data. These metrics were powerful enough to differentiate between Fem bouts (female present) and Cop bouts (including one or more copulations) and to demonstrate that those differences arose due to differences in context, and not to individual differences between males. Despite the lack of individual variation in Cop displays, we do not conclude that behavioural syndromes are absent in *Masius*. They are just not a driving force behind the organization of and variation in courtship displays. Likewise, the simplicity of successful courtship displays (Cop bouts) does not mean that complexity and non-linearity (Bradbury & Vehrencamp, 2014; Patricelli & Hebets, 2016) are not important to many other aspects of the behaviour of *Masius* or other manakins. Nevertheless, for *Masius* displays, as found by Cardoso and Hu (2011) for wood warbler song, we conclude that "simple does it."

## **LITERATURE CITED**

Barske, J., Schlinger, B. A., Wikelski, M., & Fusani, L. (2011). Female choice for male motor skills. *Proceedings of the Royal Society of London Series B: Biological Sciences*, *278*, 3523–3528.

Beehler, B. M., & Foster, M. S. (1988). Hotshots, hotspots, and female preference in the organization of lek mating systems. *American Naturalist, 131*, 203–219.

Benedict, L., & Najar, N. A. (2019). Are commonly used metrics of bird song complexity concordant? *Auk, 136*, 1–11.

Berglund, A., Bisazza, A., & Pilastro, A. (1996). Armaments and ornaments: an evolutionary explanation of traits of dual utility. *Biological Journal of the Linnean Society, 58*, 385–399. https://doi.org/10.1006/bijl.1996.0043

Bostwick, K. S., & Prum, R. O. (2003). High-speed video analysis of wing-snapping in two manakin clades (Pipridae: Aves). *Journal of Experimental Biology, 206*, 3693–3706.

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

Bradbury, J. W. (1981). The evolution of leks. In R. D. Alexander & D. W. Tinkle (Eds.), *Natural selection and social behaviour: recent research and new theory* (pp. 138–169). Chiron Press.

Bradbury, J. W., & Vehrencamp, S. L. (2014). Complexity and behavioural ecology. *Behavioural Ecology, 25*, 435–442.

Byers, B. E., & Kroodsma, D. E. (2009). Female mate choice and songbird song repertoires. *Animal Behaviour, 77*, 13–22.

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

Costa, M., Goldberger, A.L., & Peng, C.-K. (2005). Multiscale entropy analysis of biological signals. *Physical Review E, 71*, 13–22.

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature, 433*, 513–516.

Csardi G., & Nepusz T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems 1695*.

Da Silva, M. L., Piqueria, J. R. C., & Vielliard, J. M. E. (2000). Using Shannon entropy on measuring the individual variability in the rufous-bellied thrush (*Turdus rufiventris*) vocal communication. *Journal of Theoretical Biology, 207*, 57–64.

Deneubourg, J. L., Aron, S., Goss, S., & Pasteels, J. M. (1990). The self-organizing exploratory pattern of the Argentine ant. *Journal of Insect Behaviour, 3*, 159–168.

Emlen, S. T., & Oring, L. W. (1977). Ecology, sexual selection, and the evolution of mating systems. *Science, 197*, 215–223.

Fellegi, I. P., & Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association, 64*, 1183-1210.

Freeberg, T. M., & Lucas. J. R. (2012). Information theoretical approaches to chick-a-dee calls of Carolina Chickadees (*Poecile carolinensis*). *Journal of Comparative Psychology 126*, 68–81.

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology & Evolution, 7*, 1325–1330.

Fusani, L., Giordano, M., Day, L. B., & Schlinger, B. A. (2007). High-speed video analysis reveals individual variability in the courtship displays of male Golden-Collared Manakins. *Ethology, 113*, 964–972.

Gibson., R.M. (1996). Female choice in sage grouse: The roles of attracting and active comparison. *Behavioural Ecology and Sociobiology, 39*, 55–59.

Gibson, R. M., & Bradbury, J. W. (1985). Sexual selection in lekking sage grouse: phenotypic correlates of male mating success. *Behavioural Ecology & Sociobiology, 18*, 117–123.

Hewitt, S. E., Macdonald, D. W., & Dugdale, H.L. (2009). Context-dependent linear dominance hierarchies in social groups of European badgers, *Meles meles*. *Animal Behaviour, 77*, 161–169.

Hobson, E. A., John, D. J., Mcintosh, T. L., Avery, M.L. & Wright, T. F. (2015). The effect of social context and social scale on the perception of relationships in monk parakeets. *Current Zoology, 61*, 55–69.

Howe, H. F. (1979). Fear and frugivory. *American Naturalist, 114*, 925–31.

Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the I.R.E, 40*, 1098-1101.

Huxley, J. S. (1914). The courtship habits of the great crested grebe (*Podiceps cristatus*); with an addition to the theory of sexual selection. *Proceedings of the Zoological Society of London, 84*, 491–562.

Janisch, J., Perinot, E., & Fusani, L. Behavioural flexibility in the courtship dance of Golden-Collared Manakins, *Manacus vitellinus*. *Animal Behaviour 166*, 61–71.

Jaro, M. A. (1989). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association, 84*, 414–420.

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

Kodric-Brown, A., & Nicoletto, P. F. (2001). Female choice in the guppy (*Poecilia reticulata*): the interaction between male colour and display. *Behavioural Ecology & Sociobiology, 50*, 346–351.

Levey, D. J. (1987). Seed size and fruit-handling techniques of avian frugivores. *American Naturalist, 129,* 471–485.

Lill, A. (1974). Sexual behaviour of the lek-forming White-bearded Manakin (*Manacus manacus trinitatis* Hartert). *Ethology, 36,* 1–36.

McCowan, B., Doyle, L. R., & Hanser, S. F. (2002). Using information theory to assess the diversity, complexity, and development of communicative repertoires. *Journal of Comparative Psychology, 116*, 166–172.

McDonald, D. B. (1989). Correlates of male mating success in a lekking bird with male-male cooperation. *Animal Behaviour, 37*, 1007–1022.

McDonald, D. B. (2007). Predicting fate from early connectivity in a social network. *Proceedings of the National Academy of Sciences, 104*, 10910–10914.

McDonald, D. B. (2010). A spatial dance to the music of time in the leks of Long-Tailed Manakins. *Advances in the Study of Behaviour 42*, 55–81.

McDonald, D. B., & Hobson, E. A. (2018). Edge weight variance: population genetic metrics for social network analysis. *Animal Behaviour, 136*, 239–250.

McDonald, D. B., & Potts, W. K. (1994). Cooperative display and relatedness among males in a lek-mating bird. *Science, 266*, 1030–1032.

McDonald, D. B., & Shizuka, D. (2013). Comparative transitive and temporal orderliness in dominance networks. *Behavioural Ecology 24*, 511-520.

Miles, M. C., & Fuxjager, M. J. (2018). Synergistic selection regimens drive the evolution of display complexity in birds of paradise. *Journal of Animal Ecology, 87*, 1149–1159.

Miller, E.H. (1988). Description of bird behaviour for comparative purposes. *Current Ornithology, 5*, 347–394.

Morales-Betancourt, J.A., & Castaño-Villa, G.J. (2018). Males in seemingly female-like plumage do not mimic females. *Journal of Avian Biology, e01467*.

Muller, Z., Cantor, M., I.C. Cuthill, I. C., & Harris, S. (2018). Giraffe social preferences are context dependent. *Animal Behaviour 146*, 37–49.

Nelson, X. J., Wilson, D. R., & Evans, C. S. (2008). Behavioural syndromes in stable social groups: an artifact of external constraints? *Ethology*, 114, 1154–1165.

Nooker, J. K., & Sandercock, B. K. (2008). Phenotypic correlates and survival consequences of male mating success in lek-mating greater prairie-chickens (*Tympanuchus cupido*). *Behavioural Ecology & Sociobiology, 62*, 1377–1388.

Ooms, J., & Google, Inc. (2018). brotli: A compression format optimized for the web. R package version 1.2. https://CRAN.R-project.org/package=brotli

Palmero, A. M., Espelosín, J., Laiolo, P., & Illera, J. C. (2014). Information theory reveals that individual birds do not alter song complexity when varying song length. *Animal Behaviour, 87*, 153–163.

Patricelli, G. L., & Hebets, E. A. (2016). New dimensions in animal communication: the case for complexity. *Current Opinion in Behavioural Sciences, 12*, 80–89.

Patricelli, G. L., Coleman, S. W. & Borgia, G. (2006). Male Satin Bowerbirds, *Ptilonorhynchus violaceus*, adjust their display intensity in response to female startling: an experiment with robotic females. *Animal Behaviour* 71, 49–59.

Pinter-Wollman, N., Hobson, E. A., Smith, J. E., Edelman, A. J., Shizuka, D., de Silva, S., Waters, J. S., Prager, S. D., Sasaki, T., Wittemyer, G., Fewell, J., & McDonald, D. B. (2014). The dynamics of animal social networks: analytical, conceptual, and theoretical advances. *Behavioural Ecology, 25,* 242–255.

Potts, W. K. (1984). The chorus-line hypothesis of maneuver coordination in avian flocks. *Nature, 309*, 344.

Prum, R., & Johnson, A. (1987). Display behaviour, foraging ecology, and systematics of the Golden-Winged Manakin (*Masius chrysopterus*). *Wilson Bulletin, 99*, 521–539.

R Core Team, (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reding, L., & Cummings, M. E. (2017). Context-dependent preferences vary by multicomponent signals in a swordtail. *Animal Behaviour, 129*, 237–247.

Reynolds, J. D., & Gross, M. R. (1990). Costs and benefits of female mate choice: is there a lek paradox? *American Naturalist, 136*, 230–243.

Rodríguez, R. L., Haen, C., Cocroft, R. B. & Fowler-Finn, K. D. 2012. Males adjust signaling effort based on female mate-preference cues. *Behavioural Ecology* 23, 1218–1225.

Ryder, T. B., & Durães, R. (2005). It’s not easy being green: using molt and morphological criteria to age and sex green-plumage manakins (Aves: Pipridae). *Ornitologia Neotropical, 16*, 481–491.

Ryder, T. B, McDonald, D. B., Blake, J. G., Parker, P. G., & Loiselle, B. A. (2008). Social networks in the lek-mating Wire-tailed Manakin (*Pipra filicauda*). *Proceedings of the Royal Society of London Series B: Biological Sciences, 275*, 1367–1374.

Sibly, R. M., Nott, H. M. R., & Fletcher, D. J. (1990). Splitting behaviour into bouts. *Animal Behaviour 39*, 63-69.

Sih, A., Bell, A, & and Johnson, J. C. (2004). Behavioural syndromes: an ecological and evolutionary overview. *Trends in Ecology & Evolution 19*, 372–378.

Snow, D. W., & Snow, B. K. (1992). Display of the Golden-winged Manakin *Masius* *chrysopterus*. *Bulletin of the British Ornithologists’ Club, 112*, 264–270.

Sonderegger, D. (2020). SiZer: Significant zero crossings. R package version 1.1-7. https://CRAN.R-project.org/package=SiZer

Tinbergen, N. (1963). On aims and methods of Ethology. *Ethology, 20*, 410–433.

Trainer, J. M., & McDonald, D. B. (1993). Vocal repertoire of the Long-Tailed Manakin and its relation to male-male cooperation. *Condor, 95,* 769–81.

Uriarte, M., Anciães, M. da Silva, M. T. B., Rubim, P., E. Johnson, E., & Bruna, E. M. (2011). Disentangling the drivers of reduced long-distance seed dispersal by birds in an experimentally fragmented landscape. *Ecology 92*, 924–937.

van der Loo, M. P. J. (2014). The stringdist package for approximate sting matching. *The R Journal 6*, 111–122.

Vanderbilt, C. C., J.P. Kelley, J. P., & DuVal, E. H. (2015). Variation in the performance of cross-contextual displays suggests selection on dual-male phenotypes in a lekking bird. *Animal Behaviour, 107*, 213–19.

Wheelwright, N. T. (1991). How long do fruit-eating birds stay in the plants where they feed? *Biotropica 23*, 29–40.

Ziv, J., & Lempel, A. (1977). A universal algorithm for sequential data compression. *IEEE Transactions of Information Theory, 12*, 337–343.

Zuk, M., Thornhill, R., Ligon, J. D., Johnson, K., Austad, S., Ligon, S. H., Thornhill, N. W., & Costin, C. (1990). The role of male ornaments and courtship behaviour in female mate choice of red jungle fowl. *American Naturalist, 136*, 459–473.

**ACKNOWLEDGMENTS**

The Mindo Cloudforest Foundation (MCF) allowed us to use their Milpe Sanctuary for our study. Brian Krohnke and Maria José Guerrero of MCF were extremely helpful in providing logistical support. Support from the Fulbright Foundation, and the University of Wyoming's Flittie sabbatical fund helped enhance the field efforts. We are grateful to Grady Harris, Will McDonald, and many field assistants for hard work in the field that provided the scaffold on which we built our research. The University of Wyoming Education Abroad Program provided some of the travel funds for the authors. Gracias tambien a la vida silvestre, al pueblo y al gobierno del Ecuador.

**Compliance with Ethical Standards**

*Conflict of Interest*

The authors declare that they have no conflict of interest.

*Ethical Approval*

All applicable international, national, and institutional guidelines for the care and use of animals were followed. Capture by mist nets, handling and banding were approved by the University of Wyoming's Institutional Animal Use Committee (IACUC protocol #20160602DM00242-01) for each of the years of fieldwork.

**Supplementary Materials**

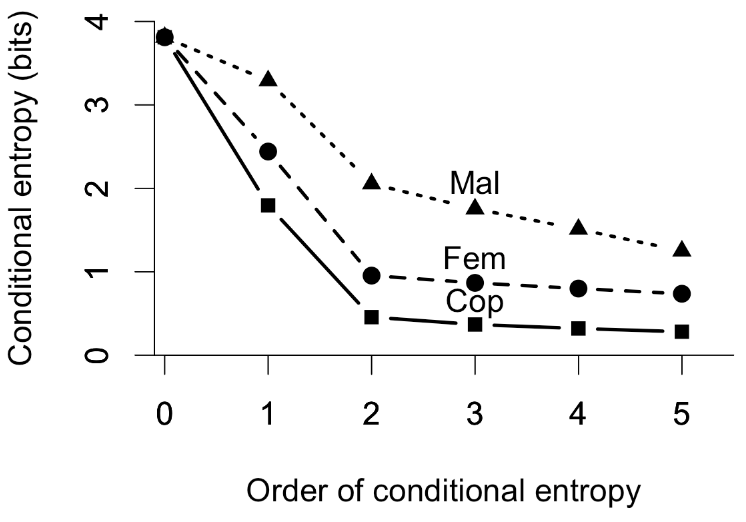


Fig. S1. Higher order (conditional) entropies for Mal, Fem, and Cop bouts. Order 1 is the first-order entropy of Fig. 2. Note that Mal continues to drop for higher-order entropies beyond Order 2. That difference suggests that in Fem and Cop bouts, the current element depends largely on the preceding element, whereas in Mal bouts, elements depend, in complex ways, on earlier elements. Nevertheless, the sharpest drop for all three is from Order 1 to Order 2, suggesting that the sequences are a reasonable fit to a first-order Markov chain process.

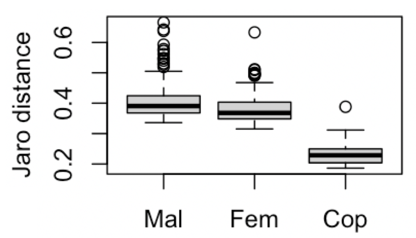


Fig. S2. Boxplots for Jaro distances in 198 Mal bouts, 100 Fem bouts and 14 Cop bouts. Cop bouts are considerably more similar to each other (low Jaro distances), than are either Mal or Fem bouts.

Table S1. List of 41 unmodified BORIS behavioural elements that were abbreviated, omitted or combined (indented in italics) to create the 21 numbered variables (ethogram elements) listed in Table 1.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Behaviour value in BORIS | Frequency | Abbreviation for analyses |
| 1 | Attempted Copulation | 11 | 1. AttCop |
| 2 | Bird2 ALAD | 55 | 2. B2AL |
| 3 | Bird2 BowLeft | 695 | 3. B2Bw |
| 4 | Bird2 BowRight | 677 | B2Bw |
| 5 | Bird2 Half-bow Left | 26 |  |
| 6 | Bird2 Half-bow Right | 21 |  |
| 7 | Bird2 HeadDownBowing | 56 |  |
| 8 | Bird2 Mixed Element | 13 |  |
| 9 | Bird2 NeckTwist | 1693 | 4. B2Nk |
| 10 | Bird2 On-log NO display | 79 |  |
| 11 | Bird2 SLAD | 1 |  |
| 12 | Bird2 TafLF\_Off | 28 |  |
| 13 | Bird2 TafLF\_On | 26 |  |
| 14 | Copulation | 24 | 5. Cop |
| 15 | End | 486 | 6. End |
| 16 | Female Looking Away | 1736 |  |
| 17 | Female Movement | 978 |  |
| 18 | Female Off Log | 320 | 7. Fff |
| 19 | Female On Log | 320 | 8. FOn |
| 20 | Female ResponseToALAD | 188 |  |
| 21 | Female Tracking Male | 1924 |  |
| 22 | FemaleSwitch | 406 |  |
| 23 | Male1 ALAD | 1104 | 9. ALAD |
| 24 | Male1 BowLeft | 8088 | 10. Bow |
| 25 | Male1 BowRight | 7950 | Bow |
| 26 | Male1 Half-bow Left | 296 | 11. HafB |
| 27 | Male1 Half-bow Right | 258 | HafB |
| 28 | Male1 HeadDownBowing | 2361 | 12. HdBw |
| 29 | Male1 Metronome\_Left | 11 | 13. Metr |
| 30 | Male1 Metronome\_Right | 25 | Metr |
| 31 | Male1 Mixed Element | 361 | 14. Mix |
| 32 | Male1 NeckTwist | 5618 | 15. Neck |
| 33 | Male1 Off Log | 41 |  |
| 34 | Male1 On Log | 39 |  |
| 35 | Male1 On Log No Display | 2225 | 16. Zero |
| 36 | Male1 Other Behaviour | 2142 | 17. OthX |
| 37 | Male1 SLAD | 655 | 18. SLAD |
| 38 | Male1 Switch | 1145 | 19. Swtc |
| 39 | Male1 TafLF\_Off | 1193 | 20. Taf |
| 40 | Male1 TafLF\_On | 1223 | Taf |
| 41 | Start | 486 | 21. Start |

Notes for 2022-2023 revision.

Fig. 1. Four measures of the gradient of complexity from complex male-only courtship display bouts (Mal) to courtship displays with one or more females present but no copulation (Fem) to least complex in courtship displays ending in a copulation (Cop)  
a. Elem count/BoutLength ratio (output of 10-Oct-22). Ratio of (distinct element count)/(string length) here.   
  
 Chart, box and whisker chart

Description automatically generated

b. Entropy is as much a measure of surprise as of complexity. Entropy goes from low (few surprises) in display bouts ending in a copulation (Cop) to high in display bouts involving only males (Mal), with intermediate entropy for display bouts attended by one or more females that did not end in a copulation (Fem).   
  
 Chart, box and whisker chart

Description automatically generated

c. Compression string length (raw or ratio?). Ratio here. A courtship bout whose string of display elements can be greatly compressed (high compression ratio on the ordinate), such as the Cop display bouts, can be viewed as less complex than ones that are not very compressible, such as the Mal display bouts.   
  
 Chart, box and whisker chart

Description automatically generated  
  
 d. Algorithmic complexity, as an approximation of Kolmogorov complexity. Again, the   
 complexity gradient goes from low complexity for display bouts ending in a copulation   
 (Cop), to high complexity in display bouts involving only males (Mal).

Chart, box and whisker chart

Description automatically generated

From Liam's email of 6-Oct-22

"Focusing on the compressibility score as the central value and explicitly comparing it to simple measures like "repertoire size" might make for a very simple+impactful presentation of the data.

A central figure might be as simple as a four subplot bar plot:

(A) Repertoire sizes. Mal vs. Fem (exclude cop) vs. Cop.

(B) Entropy. Mal vs. Fem (exclude cop) vs. Cop

(C) Length of compression string. Mal vs. Fem (exclude cop) vs. Cop

(D) If any different than (C), Algorithmic complexity from acss. M vs. F vs. C

Then another figure (or worked into the above) parsing within vs. between individuals.

Plot (A) gives us the background value that most investigators might have looked for.

Plot (B) tells us something about predictability -- from the view of the female, how "surprised" she might possibly be, given the available male display elements.

Plot (C) tells us something about an approximation of complexity -- from the view of the male, how simple is the display to produce (aesthetic structure, repetition, etc.).

Plot (D) bolsters plot (C) with a slightly different estimator of complexity."

Questions/comments for Liam (22-Oct-22)

Should we call the 11 elements "events" rather than "elements"? Some (Cop, AttC, FOn, Fff, Mix, Othr, Zro) are not display elements per se (especially in the Schaedler comparative scheme).

Are the (almost impercepible, except in slomo) neck bobs and tail wags after an ALAD worth attomizing?

Vault (male vaults over female when both are on the log): does it occur in context other than ALAD?

Gradient from Neck twist (*Neck*) to Head-down Bow (*HdBw*)

Nick's lower canopy flybys?

"Kichup" sound in ALAD is a vocalization. See "1519\_Taf&kichup ALAD voc.mp4" & "1519\_CreekLog\_11\_18\_17\_02,mp4"

I need to check on "Gaping", Nick's fly-bys and wing-flicks. Check email of 22-Oct-22 and new voucher videos in Dropbox.

Hasenjager, Matthew J., Victoria R. Franks, and Ellouise Leadbeater. “From Dyads to Collectives: A Review of Honeybee Signalling.” Behavioral Ecology and Sociobiology 76, no. 9 (September 2022): 124. <https://doi.org/10.1007/s00265-022-03218-1> Zotero.

Bosshard, Alexandra B., Maël Leroux, Nicholas A. Lester, Balthasar Bickel, Sabine Stoll, and Simon W. Townsend. “From Collocations to Call-Ocations: Using Linguistic Methods to Quantify Animal Call Combinations.” Behavioral Ecology and Sociobiology 76, no. 9 (September 2022): 122. <https://doi.org/10.1007/s00265-022-03224-3>. Zotero