

Mating Dances and the Evolution of Language: What's the Next Step?

****Accepted at *Biology & Philosophy*****
Please cite final published version only

Cameron Buckner
Assistant Professor
Department of Philosophy
University of Houston
cjbuckner@uh.edu

Keyao Yang
Adjunct Instructor
Houston Community College

Abstract: The Darwinian protolanguage hypothesis is one of the most popular theories of the evolution of human language. According to this hypothesis, language evolved through a three stage process involving general increases in intelligence, the emergence of grammatical structure as a result of sexual selection on protomusical songs, and finally the attachment of meaning to the components of those songs. The strongest evidence for the second stage of this process has been considered to be birdsong, and as a result researchers have investigated the existence of various forms of grammar in the production and comprehension of songs by birds. Here, we argue that mating dances are another relevant source of sexually-selected complexity that has until now been largely overlooked by proponents of Darwinian protolanguage, focusing especially on the dances of long-tailed manakins. We end by sketching several lines of research that should be pursued to determine the relevance of mating dances to the evolution of language.

1. Introduction

Comparisons between humans and animals have long been guided by the idea that only humans possess a genuine or full-blown language faculty—though if we want to understand how human language evolved, we must study its precursors in non-human animals. Spoken language leaves no direct archaeological evidence, so biologists since Darwin have tried to puzzle out the trajectory of language's evolution in hominids by looking for its components and prerequisites in other species. Reflection on homologies and analogies can ground inferences as to which components of our language faculty presuppose others, which selection pressures favor the evolution of each component, and which components are uniquely human.

In a landmark article, Hauser, Chomsky, and Fitch (2002) marshal empirical evidence in favor of a particular set of answers to these questions. They hypothesize that the faculty of language can be understood in a broad and narrow sense. The Faculty of Language in the Broad sense (FLB) includes a sensory-motor system, a conceptual-intentional system, and recursive grammatical competence. They further propose that

only the recursive grammatical competence is uniquely human, dubbing it the Faculty of Language in the Narrow sense (FLN). Their favored theory is thus that our hominid ancestors excelled at the other components of FLB, and later the recursive computational core of FLN evolved uniquely in hominids, allowing our ancestors to modify and integrate these earlier components to produce and understand the infinite variety of human linguistic productions. Their hypothesis thus presents a somewhat linear account of the evolution of language, beginning with the earlier components of FLB and continuing on through FLN (though further adaptations for the “externalization” of FLN could have come still later—Berwick & Chomsky, 2017).

An obvious but still unresolved question about this evolutionary story concerns the centrality of vocalization to the evolution of FLB and FLN. The classic Darwinian account—developed more recently by Fitch (2010, 2013)—grants pride of place to vocalization, offering birdsong as the most promising place to look for selection pressures favoring grammatical competence in nonhuman animals. This “musical protolanguage” hypothesis deploys comparative analyses of birdsong to justify a three-stage story about the evolution of human language: first, selective pressures spurred a general increase in intelligence and associative memory; second, sexual selection favored the production and comprehension of complex patterns in mating calls (with only the most complex or beautiful calls earning copulation privileges); and third—this, perhaps being a uniquely human step in the trajectory—some “unusually wise ape-like animal” would notice that symbolic meanings could be attached to the components of structured songs, so that when recombined in novel but rule-governed ways, our ancestors could consider and discuss never-before-experienced situations and experiences.

Despite the vivacity and popularity of this account, it is not yet established whether the components of FLB or FLN evolved to serve specifically vocal-auditory modalities in hominids. Hauser, Chomsky, & Fitch emphasize a number of specifically vocal/auditory adaptations involved in human language, especially our excellent capacities for phonemic discrimination, vocal control (with a descended larynx), and vocal imitation. However, the most promising work on protolinguistic communication in chimpanzees involves instead gestures, suggesting that many of the components of FLB might have evolved to guide non-vocal

behaviors in hominids and only later have been co-opted for use in spoken language (Scott-Phillips et al., 2015).¹ Hauser, Chomsky, and Fitch also review work in this area, and further consider the possibility that components of FLB originally performed non-linguistic functions, and that the recursive core of FLN might have evolved to serve non-vocal purposes like numerical understanding, tool-use, or social cognition. Thus, perhaps the strongest line of evidence that vocalization is “special” in the evolution of language (and specifically grammatical complexity) remains the comparative analysis of birdsong.

Our goal in the present paper is to argue that a broader survey of evidence from non-mammals could also support non-vocal hypotheses about the evolution of language. Specifically, the same kinds of sexual selection pressures thought to favor grammatical complexity in songs apply to a variety of other non-vocal/auditory modalities, from plumage displays (Bortolotti, Blas, Negro, & Tella, 2006), to the highly articulated bower decoration strategies of bowerbirds (Borgia, 1995; Day, Westcott, & Olster, 2005), to the layered olfactory signals produced by orchid bees (Eltz, Sager, & Lunau, 2005). Here, we review and investigate the prospects for a particularly alluring source of sexually-selected but non-vocal complexity in birds: the elaborate mating “dances” of *Pipridae*, or manakin birds.² The Pipridae are comprised of fifty-two species of bird distributed through the American tropics. As a result of advances in high-speed photography and new field observations (Fusani, Giordano, Day, & Schlinger, 2007), researchers have learned that the males of many manakin species perform a variety of extended, elaborately organized, and fast-paced courtship dances, ranging from traditional leks, to long solitary displays, to cooperative displays involving multiple males (Lukianchuk and Doucet, 2014). As a comparison case for the evolution of language, of particular interest are the complexity and flexibility of the dances, especially whether their structures follow hierarchical patterns or rules. We will here primarily focus on the acquisition and organization of dance-patterns in long-tailed manakins (*Chiroxiphia linearis*), which exhibit some of the most complex dances.

¹ Notably for the present thesis, chimpanzee gestural communication has been analyzed as a form of dance (Donald, 1991, 1998; King, 2009).

² Following practice in the empirical literature, we will often refer to the manakin mating displays as “dances”, keeping in mind that the degree of similarity between these displays and the most elaborate forms of human dance is one of the issues to be empirically and philosophically evaluated in the present paper. There is no ideal terminology with which to begin, as some of the other options (such as “courtship” or “mating” display) may also be deemed unduly anthropomorphic. So long as we do not begin by interpreting “dance” in an “anthropofabulous” manner (Buckner, 2013), it is as good a starting point as any.

As we are just now beginning to characterize the “moves” and “choreography” of these dances, it is currently unknown whether manakin courtship displays are organized according to grammatical rules. Our goal in this paper is two-fold: 1) to make the case that manakin courtship displays provide a fruitful comparison in the study of the evolution of language, and 2) to provide specific suggestions for empirical researchers who wish to investigate the relevance of manakin mating dances to the evolution of language. To be clear, our goal is not to argue that uniquely human language traits (i.e., FLN or more specifically Merge—Berwick & Chomsky, 2015) appear in mating displays, nor do we here offer new arguments for a gradualist/bootstrapping account of the evolution of grammar (as opposed to a more minimalist/saltational account now favored by some Chomskyans). Rather, we here evaluate the evidential parity of song and dance for the precursors of full-blown human language in the second, sexual-selection stage of the Darwinian protolanguage hypothesis.

In short, our case is directed at readers who already find grammatical analysis of birdsong to be relevant to the evolution of hominid language, and we urge those readers to also consider the support that similar analyses of mating dances may provide for non-vocal origins hypotheses. Of particular interest will be the ways that the dances differ in their development, performance, and supporting neural mechanisms compared to vocal displays in songbirds, and whether these differences add up to the irrelevance of courtship dances to language. Our verdict on this question will ultimately be mixed: the dances do differ in some critical ways, but are also similar in other subtle but significant ways. As a result, our view is that comparisons between dances and protolinguistic vocal behaviors mutually enrich our understanding of the evolutionary forces governing language-like complexity in both vocal and non-vocal motor productions, and this question merits much more research.

The structure of this paper is as follows. Section 2 reviews the musical protolanguage hypothesis and the role that birdsong has played in our biological understanding of language, focusing especially on the different classes of grammar that birdsong can exhibit. Section 3 describes in more detail the relevant features of courtship dances of long-tailed manakins. Section 4 explores the similarities and differences between full-body dances and vocal performances, arguing that dances could be grammatically structured in

relevant ways—with some important caveats. Section 5 concludes by describing several variants of three broad classes of empirical investigation—roughly, on performer effects, audience effects, and neural correlates—that could be conducted as the next step to determine whether manakin dances are grammatically structured in the relevant ways.

2. Musical Protolanguage and the Evolution of Grammar

The evolution of language was a central focus of Darwin's attention in *The Descent of Man* (1871), because he knew the staunchest opposition to his theory of evolution by natural selection would come from students of the “science of language” like Max Müller. Müller (1861) held that language was the “Rubicon which divides man from beast,” the feature of human cognition that would never be amenable to evolutionary analysis. The reasoning behind this skepticism is traceable at least back to Descartes, and probably even further back to the Stoics, who emphasized the infinite variety of meaningful linguistic constructions that humans can easily produce and understand. Behind this purportedly systematic ability to produce an infinite variety of structured thoughts and sentences was an understanding of grammar, supposed by these thinkers to be the core of the uniquely rational human soul. Just as Descartes assumed that our infinite productive capacity had to outstrip the capacities of any physical mechanism, Müllerians supposed it must outstrip the means of any cumulative modification and selection, creating a qualitative gap that no gradualist theory of evolution could bridge.

Darwin took the skeptical challenge posed by this line of reasoning seriously, and rebutting it was one of the main goals of the *Descent*. The first step of his rebuttal recast language as a faculty involving the cooperation of several different mechanisms, rather than the operation of monolithic rational soul. These mechanisms included general neural resources for intelligence and associative learning, anatomical structures for vocal production and control, and an instinct to learn structured patterns in vocal productions during a critical period in development. This multi-component view of language was inspired by Darwin's extensive observations of animal communication in nature—especially of birdsong.

Darwin's theory of evolution made it possible to articulate two different ways that language-like complexity in birdsong (and, as we shall see, in dance) could be relevant to human language: 1) the capacities

producing the complexity could be homologies, deriving from a common ancestor, or 2) they could be analogies, having evolved convergently in different lineages. Since humans and birds diverged from their last common ancestor over 300 million years ago, homology in this case is unlikely for any traits but very general precursors of language such as motor control or domain-general learning mechanisms. Darwin himself spoke of birdsong as the “nearest analogy to language,” but even analogies can tell us much about developmental trajectories and general selection pressures that could have favored grammatical complexity in our own ancestors. Here, Darwin was particularly drawn to the way that juvenile songbirds pass through a critical developmental period involving “babbling”—an “instinctive tendency to speak” (55)—and later experiment with novel combinations of vocal components (and their effects on listeners) before mature adult patterns emerge. This construal deftly provided a role for both nature and nurture in the emergence of language, articulating a middle-ground in the stark divide between nativist and “blank slate” positions that stymied earlier debates. Furthermore, by casting language as a gradually developing multi-component faculty rather than focusing on its mature adult manifestations, he rendered gradualism about the phylogeny of language more plausible. Each component in the language faculty could have its own evolutionary scenario with its own selection pressures (which Fitch pithily dubs “evolutionarios”—Fitch, 2010).

With these components in place, Darwin was then in a position to sketch a multi-stage sequence for the evolution of human language that explained not only the presence of analogous traits in birds, but also the absence of them in our nearest primate ancestors. Darwin first posited a stage of general increase in intelligence, perhaps arising from social and technological factors. Next, Darwin posited that our ancestors went through a period of intense selective pressure on vocal calls. As with birdsong, these pressures would have favored more complex, music-like calls—perhaps for the purposes of territorial defense, but even more likely for purposes of mating. This selection pressure would have fueled the rapid evolution of ever-more-structured—and perhaps, ever more beautiful (Prum, 2012)—songs. In the third and final stage, our ancestors would have learned to attach meaning to the components of these complex song-structures—originally through onomatopoeia and acoustic imitation, but eventually through arbitrary and culturally-acquired symbolic associations—such that we could use these songs to talk about novel states of affairs. This

finally marks the transition to meaningful, fully human speech—as von Humboldt put it, the stage at which man became “a singing creature, only associating thoughts with tones” (von Humboldt, 1836, 76).

Each stage of this story could be elaborated into further, more gradual steps, each grounding further increases in the faculties presupposed in the previous steps; but the key stage for present purposes is the sexual selection for more complex song structure in Stage 2 (and so we will not discuss Stage 1 or 3 further here). One form that more gradual iteration might take here would concern grammatical transitions and relations between song components. If the female birds could somehow distinguish and prefer mate candidates whose displays exhibited greater grammatical complexity, this would induce a pressure for not just longer songs with more varied components, but also for those components to be organized into more sophisticated grammatical structures. Fitch and other researchers have here mapped birdsong into different classes of grammar studied in linguistics. A primary question of the present investigation is whether it makes sense to look for such grammars in mating dances as well, and whether we in fact see selection pressures for more complex dance-grammars in manakins.

For readers without a background in linguistics, it will help to review a key distinction amongst different classes of grammar (readers already familiar with the literature on grammar in animal vocalizations can probably skip to the next section). The weakest class in this hierarchy is finite state grammars (FSGs), which can be fully specified by transition probabilities between a fixed number of elements. On the next level of this hierarchy are phrase structure grammars (PSGs—also called “context-free” grammars). A PSG can embed sequences within other sequences, and thereby create complex hierarchical structures that respect long-distance dependencies between phrases in a sequence (Fitch & Hauser, 2004). Both kinds of grammar are commonly denoted by symbolic shorthand, where capital letters like “A” and “B” indicate syntactic component types (such as particular phonemes or repeatable song components) and exponents indicate the potential for a pattern to be repeated. Because each element with an exponent can be repeated different numbers of times, a simple finite specification is capable of fitting an infinite number of distinct sequences.

A simple example of an FSG is $(AB)^n$, which can produce sequences of symbols such as AB, ABABAB, ABABABAB...and so on. The basic pattern “AB” can be repeated an indefinite number of times,

but there is no structural dependency between any two instances of the pattern. In other words, the structure of such sequences possesses only one layer, as shown in Figure 1. The corresponding example of PSG is A^nB^n , where the shared exponent indicates that the phrase must have the same number of As that it does of Bs. This specification can produce sequences of symbols such as AB, AABB, AAABBB...and so on. In this case, there is a long-distance dependency between the sub-patterns—to follow this rule, the overall sequence must contain the same number of As and Bs. Thought of in generative terms, the grammar can be given an inductive specification:

1. AB is grammatical (base case)
2. If XY is grammatical, then X+AB+Y is grammatical (inductive step).
3. No other sequences are grammatical (closure step).

Because the product of step 2 can be re-applied as input to its next iteration, the structure of such sequences can thus involve multiple layers of recursion, with each layer “center-embedded” in the phrase of the previous layer, as demonstrated by Figure 2.



Figure 1. A parse tree for the FSG for the $(AB)^n$ pattern. (Reproduced from Fitch & Hauser 2004, 378)



Figure 2. A PSG parse tree for the A^nB^n grammar. (Reproduced from Fitch & Hauser 2004, 378)

FSGs are not complex enough to generate all structures of human language, some of which involve at least PSG-level complexity. For example, in English we have higher-order phrase such as “if...then”. When

we use phrases like this, we can embed an indefinite number of grammatical phrases within the ellipses; but so long as we use one antecedent “if” phrase in the sentence, we must complete the higher-order pattern by using a consequent “then” phrase later in the sentence.³ The number of antecedents and consequents in the sentence must match. This kind of long-distance dependence and hierarchical structure between “if” and “then” cannot be captured by any FSG. Therefore, when Hauser, Chomsky, & Fitch say that recursion is the key component of human language, what they mean is that the recursion generated by at least a grammar on the PSG level is the key component of FLN.

This hypothesis would thus be challenged by the discovery that some nonhuman species are capable of producing or perceiving recursive sequences with a grammar higher than the FSG level. However, even if they are right that humans alone possess PSG and higher grammars, convergently evolved precursors of these grammars might be of interest in understanding how our ancestors evolved more elaborate grammars—whether through periods of iterative bootstrapping from less to more complex grammars, or by providing a selection pressure to drive the spread of FLN/Merge after a single, saltational mutation. Furthermore, their hypothesis implies a further question: did the human capacity for producing and perceiving recursion evolve specifically for spoken or even gestural language, or to support entirely different forms of structured behavior—navigation, tool use, social cognition, or dance choreography—and was only later exapted for use in spoken language?

Both lines of question have been investigated empirically by looking for grammatical understanding in animals such as tamarins, keas, pigeons, and starlings. Such investigations can focus on two different aspects of grammatical understanding: production and comprehension. The first line studies behavior producers to determine if any animal can structure behavior according to a grammar higher than FSG level. However, merely finding such patterns in behavior would not prove that they were produced by recursive rules or generative forms of representation, for the patterns could merely be the product of reflex or some other inflexible mechanism. A second line of investigation intended to resolve this question would focus on comprehension, by building artificial stimuli that follow grammars higher than the FSG level, and testing

³ Of course in actual spoken language, either the “if” or the “then” can be implicit (i.e. “I’ll go to the store if we are out of milk.”)

animals' sensitivity to whether a variety of novel patterns follow those grammatical rules. These experiments typically take the form of playback of grammatical artificial stimuli until animals should have habituated to the new grammar, and then playing similar but novel sequences that violate the grammar to see if the animals notice the mistakes.

The typical artificial FSG learned and tested in such experiments is $(AB)^n$, and the typical artificial PSG learned and tested in the experiment is $(A)^n(B)^n$. In order to be able to compare the test results between FSG and PSG, and to eliminate the influence of any factors other than the grammar, the testers must control and match the other aspects of the signals, such as their number and length (both grammars possess only two elements, A and B) and the overall frequency with which each component appears in the sequence. So far, many animals have passed the test with an artificial FSG. Results for PSGs, however, have been more mixed. Cotton-top tamarins apparently failed to discriminate the PSG violations in matched artificial stimuli, whereas European starlings and other birds appeared to succeed at this task (Abe & Watanabe, 2011; Gentner, Fenn, Margoliash, & Nusbaum, 2006).

However, critics have recently responded to the positive PSG results in animals by noting that the $(A)^n(B)^n$ tests are still ambiguous, because the comprehension results could also be generated by alternative, first-order parse-trees that do not involve recursive center-embedding (Corballis, 2007; Fitch, 2010--and see Figure 3). The debate over the application of the $(A)^n(B)^n$ test to birdsong raises important epistemological issues and calls for more research, but we need not resolve these questions here. At any rate, center-embedding has probably been overemphasized in this literature; it is relatively rare in human languages, and it reflects only one amongst many different kinds of recursive grammatical construction (Corballis, 2007). Moreover, forms of grammatical complexity short of context-free recursion should still be of biological interest if our goal is to understand the full evolutionary trajectory of language. As such, we here argue that we should explore a wider variety of grammatical patterns and open this search to non-vocal animal behaviors as well.

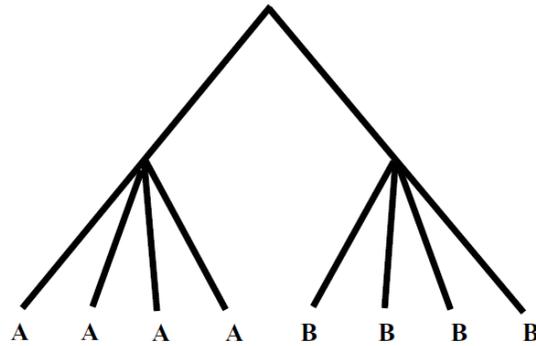


Figure 3. Alternative parse-tree for $A^n B^n$ sequence that does not involve long-distance dependencies, (reproduced from Corballis, 2007, 700).

To begin the transition from vocal productions to dance, an argument for the relevance of grammatical recursion in intentional motor movements is offered by Vicari & Adenzato (2014). They list the four features exhibited by recursion in human language—self-embedding, long-distance dependence, identity preservation, and discrete infinity—and argue that all of these features can be found in non-vocal forms of intentional action. For present purposes, the most important aspect of Vicari and Adenzato’s argument is the idea that motor intentions can possess hierarchically-structured, recursive goals. The plans for achieving these goals can be structured like grammatical vocal productions, with components that can only be specified in terms of layered sub-intentions that may have to be repeatedly iterated until a sub-goal is achieved.

To provide another relevant comparison, this is one of the reasons that tool-use has also been studied as an alternative domain in which recursive faculties might evolve; archaeologists have sought such layered sub-intentions in the archeological record of tool manufacture, using advances in flint-knapping techniques as indirect evidence for the evolution of recursive thinking in our Paleolithic forebears. Notably, around the same time period that modern human language is thought to have evolved, the archaeological record shows the spread of a new “Levallois flake” technique for flint-knapping, which archaeologists argue requires forms of recursive thinking. Levallois tools appear to show that our ancestors suddenly developed the ability to implement a complex knapping strategy that requires repeating steps in sequence until a desired subgoal is achieved, as nested in a hierarchically structured pattern (Figure 4). Like grammatical patterns, such motor plans have repeatable, reidentifiable subcomponents, hierarchical organization, and can be used to generate a

potentially infinite number of variations by changing the number of times that patterns (and subpatterns) are iterated. We should thus ask many of the same empirical questions—what kind of grammatical organization do they possess, is the organization flexible, and how is the structure represented by the organisms producing and comprehending those patterns—about complex motor movements as well.

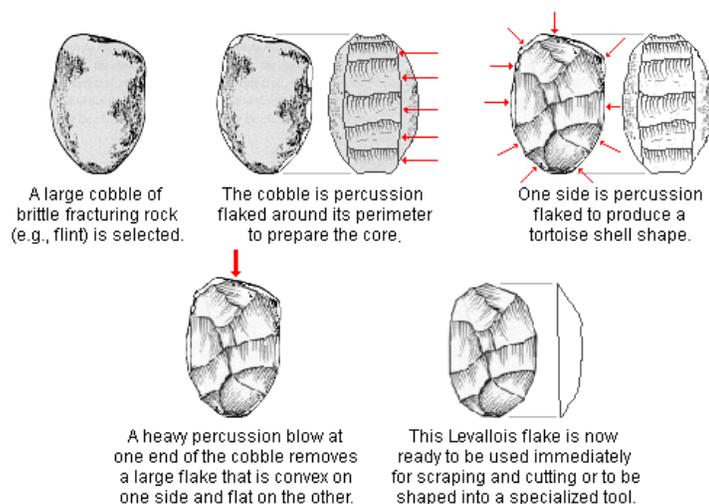


Figure 4. The steps of a Levallois flake production; each step may involve repetition in context to achieve a desired outcome before moving on to the next step, and many variations in end product are available to produce different kinds of tools or points. Illustration reprinted from (O’Neil, 2012) with permission from the author, who retains all rights.

3. The Mating Dances of the Long-Tailed Manakin

There are at least two places we can look for grammatical structure in non-vocal behavior in animals. The first place, as noted above, is animal tool construction, and this line of thought has already been explored in the case of New Caledonian crows (Taylor, Hunt, Holzhaider, & Gray, 2007). The second place is the production of long sequences of complex actions, as is found in the mating displays of many birds. Notably, mating dances slot into a Darwinian protolanguage story even more neatly than tool use, for the forms of sexual selection acting on the dances may be extremely similar to those acting on songs. Research on humans has shown that females observing male dances can make fine discriminations in the motor performance and variation with which moves can be performed, lending support to the idea that dance structure might have played an important role in sexual selection in hominids as well (Barske, Schlinger, Wikelski, & Fusani, 2011; Neave et al., 2011). In evaluating the hypothesis that grammatical competence found in language may have

evolved in mating dances, we will focus especially on the long-tailed manakin, which produces some of the most complex mating dances yet found in nature.

As with many lekking species, male long-tailed manakins gather in a display area, where they exhibit their courtship dances to onlooking females. The center of the courtship zone is a primary display perch, where the dances usually begin. While long-tailed manakins sometimes perform solo routines, a particularly interesting feature of their dances is that they are often and most successfully performed in pairs. Each pair involves a dominant alpha male and a subordinate beta male. The full display consists of three or more “phases”: first, if partners are being recruited for the dance, the alpha male will sing a “*teeamoo*” call from the canopy above a perch, which attracts a beta to a perch; second, the alpha (and beta, if present) will engage in “*toledo*” calls which summon females to watch the dance; third, performing males descend to the display perch and begin a “hopping” display, which can consist of many highly-skilled, coordinated movements; fourth, the dance typically proceeds to a butterfly display, which involves a distinct category of skilled physical maneuvers and typically ends with the alpha performing an “angel flight” movement followed by a “bow” motion; and fifth, if the female has remained through the entire display, copulation will occur. Females also make some movements during the male’s performance, which may provide feedback on the female’s likelihood of copulation (Ward, 2012). Notably, only the alpha male is allowed to breed at the end of a successful display, which raises questions about the adaptive benefits of the display for betas (DuVal, 2013).

The different phases of the dance are of particular interest, for they may indicate that the display possesses a hierarchical grammatical organization. While previous ethologists lacking current video recording technology only decomposed the dance into hopping and butterfly components, Lukianchuk and Doucet (2014) analyzed 2500 hours of video footage to extract 16 re-identifiable “moves” in the displays. Their criteria for typing a display element required that the behavior i) always be executed in a similar manner across males and display areas, ii) never be executed partially unless disrupted, iii) be performable in different combinations with other discrete behaviors immediately before or afterwards. These criteria are important to the current investigation, for they provide us with the analogy of “words” for a dance grammar. Moreover, the behaviors can be grouped into different grammatical supercategories of hopping and butterfly elements.

Elements performed during the hopping phase include the “stationary display”, “popcorn hop”, “leapfrog”, “buzz-weent”, and “tucked wing-flick”, whereas elements performed during the butterfly phase include “butterfly flight”, “dart”, “bounce”, “bill wipe”, “upright posture”, “about face”, “lightning bolt”, “back-and-forth”, “angel flight”, and “bows” (see Figure 5 for diagrams). Though we do not emphasize this aspect here, the displays are also multi-modal; long-tailed manakins can produce 14 different vocalizations, and vocalizations during the display appear to be integrated with dance moves, rather than serving as separate or parallel components—though they may play some additional role in coordinating partners or enticing females. The speed with which various components are performed also appears to be relevant to copulation success, though we will not discuss its influence here.

Analyses of 92 different recorded displays allowed Lukianchuk and Doucet to build a probability matrix of transition likelihoods between different components of the dance moves, akin to the kind of transition matrices one might find in analyses of birdsong (see Figure 6). They found that some components—leapfrogs, back-and-forths, and tucked wing flicks—exhibited a high probability of being repeated, whereas others—lightning bolts and stationary displays—exhibited high probabilities of transitioning to specific other components. All but one of the elements showed a high probability of directly preceding at least one other element, and many elements were not easily predicted by any preceding elements. Some components were more likely to mark a transition between the hopping phase to the butterfly phase (i.e. buzz-weent or tucked wing flick), or between the butterfly phase to the copulation phase (which was 83% of the time preceded by a bow). Though some aspects of the dance are highly stereotyped, other elements were not highly associated with any others, suggesting a considerable amount of flexibility in the displays.⁴ Moreover, there was a high degree of individual variation in the patterns presented in different displays, suggesting that females may have significant diversity of patterns to choose from when selecting mates. Other research has shown that just a few males monopolize most matings (Stein & Uy, 2006), so if females are choosing on the basis of dance characteristics, then there is extreme pressure to produce the best dance.

⁴ For an overview of the cognitive and evolutionary significance of such behavioral flexibility, see (Buckner, 2015; Mikhalevich, Powell, & Logan, 2017).

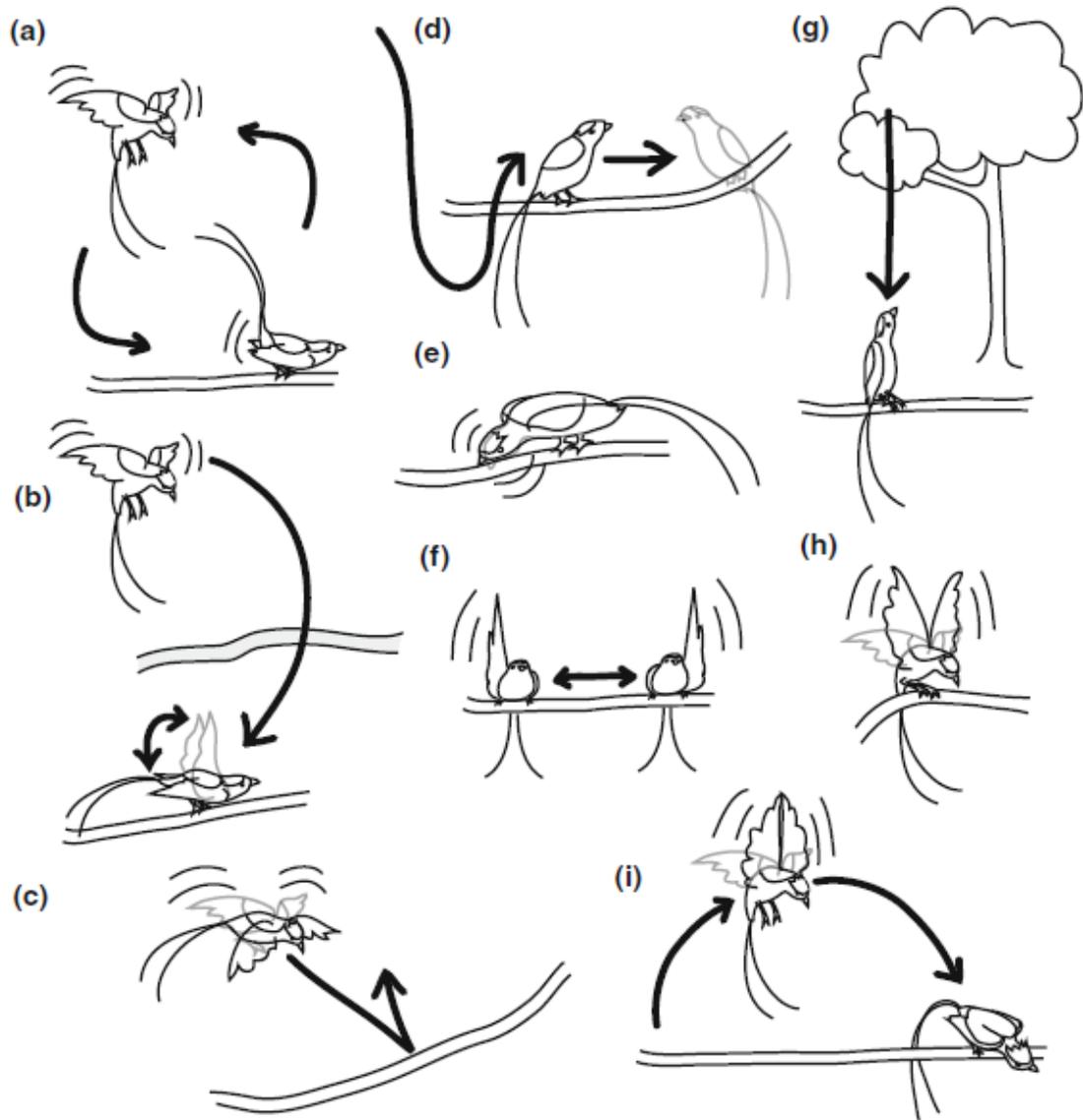


Figure 5. Elements of the long-tailed manakin *Chiroxiphia linearis* courtship display: a) two males performing leapfrogs, with lower male performing stationary display; b) male exhibiting buzz-weent display on primary display perch (shaded), followed by a set of tucked wing flicks on a perch below; c) male performing butterfly flight towards perch, followed by a bounce upon landing; d) arrow depicting flight pattern of s-shaped dart, followed by an about face upon landing; e) male exhibiting bill wipes; f) male exhibiting back and-forth displays; g) arrow depicting flight pattern of lightning bolt, followed by an upright posture upon landing; h) male performing frenzied flutter; and i) male performing an angel flight and landing in a bow position (Illustrations by K.C. Lukianchuk, reproduced from Lukianchuk & Doucet, 2014, p733).

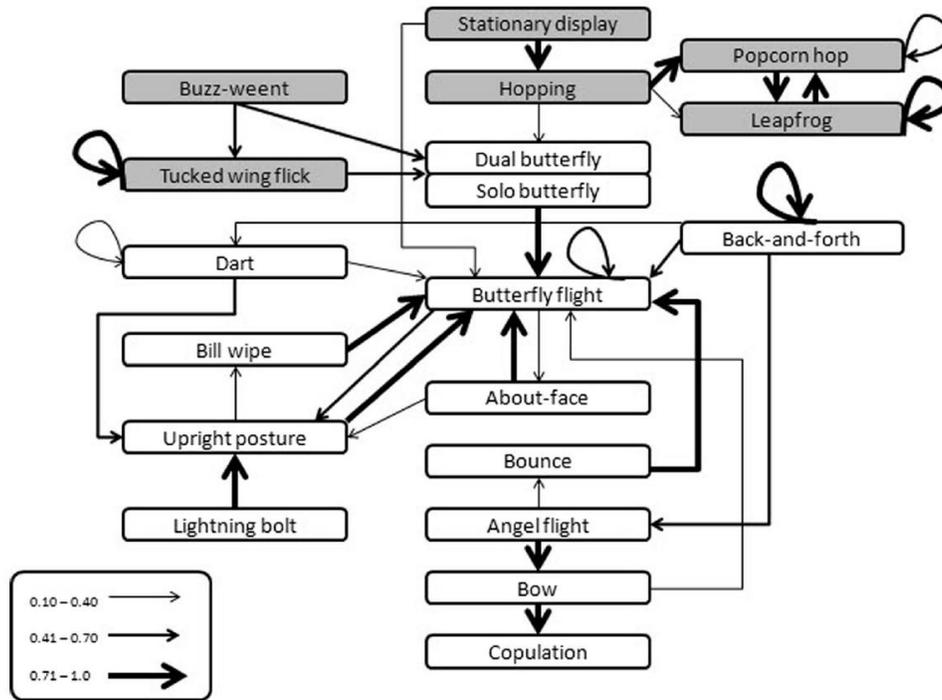


Figure 6. First-order transition probabilities amongst dance elements during successful dances, reproduced from Lukianchuk & Doucet (2014, 738). Shaded boxes are elements in the “hopping” portion of the display, and unshaded boxes are in the “butterfly” portion of the display.

4. Grammar in Dance? Promise and Problems

Are such dances organized according to grammatical rules? At present, we simply do not yet know, as researchers have only begun to analyze the structure of the dances. However, it is tempting based on the first-order transition probabilities already documented to suggest that the structure of such dances might be captured by a generative grammar. For example, consider the following set of rules (specified in Backus-Naur Form):

1. <Dance> ::= <teemo call> <toledo call> <hopping display> <butterfly display> <copulation>
2. <hopping display> ::= <stationary display> | <popcorn hop> | <leapfrog> | <buzz-weent>
| <tucked wing-flick> | <hopping display>
3. <butterfly display> ::= <dart> | <bounce> | <bill wipe> | <upright posture > | <about face> |
<lightning bolt> | <back-and-forth> | <angel flight> | <bow> | <butterfly display>

These specifications already display FSG-level recursive complexity, allowing indefinite iterations of moves within a display phase. The transition probabilities displayed in Figure 6, however, suggest further

complexity. For example, is it a “rule” that the hopping display always ends with a butterfly flight? Should the butterfly display always end with a bow? Is there a distinction between components that are to be multiply iterated and those that are not? Is there any relationship between the number of iterations of a repeatable component in one section of the dance and another repeatable component that occurs in a later section? If such long-distance dependency exists, how much must be planned by the performer, how much is determined by responses of the females during the dance, and how much is simply spontaneous? Moreover, the fact that the high-speed dances are coordinated between a dominant and subordinate male makes it especially likely that any complexity must be represented and communicated by the cooperators to one another before the performance. So to these questions we should add more about the means of coordination between the alpha and beta male, and whether their dominance relationship plays any role in planning, directing, and practicing the choreography of the dance.

Does it even make sense, however, to look for linguistic grammars in a sequenced, full-body motor activity like dance; or is this extension of linguistic terminology to a different domain likely to lead only to confusion (Bowling & Fitch, 2015)? We have already described the way in which the dances may be broken down into repeatable syntactic components that can be recombined in flexible ways. In fact, such analyses mirror grammatical analyses of explicitly choreographed dances in humans, such as classical ballet, where dance components are similarly broken down into a lexicon of moves so that phrases and patterns can be explicitly formulated (Graham, 1973; Starr & Cunningham, 1968; Stevens & McKechnie, 2005).

Anthropologists have also speculated that human dance can serve as a form of symbolic communication, with physical motions especially suited to kinematic onomatopoeia in its beginnings, but later moving into a more symbolic and ritualized form of communication, as is found for example in the hula story-forms of the Polynesian diaspora, or the war haka of the Maori (Armstrong, 1964; Pukui, 1942). When working out evolutionaries for the components of modern human language, mating dances thus possess at least as much *prima facie* anthropological relevance as birdsong.

We must, however, acknowledge the differences in the anatomical and neural demands of dance as compared to vocalizations—though these differences can be a “bug” or a “feature”, depending upon how

one stands on wider debates concerning the embodied and embedded nature of cognition (Barrett, 2016; Clark, 1999). In most cases, vocalizations are a comparatively low-energy activity involving only a small portion of the body: the facial muscles, vocal tract, and breathing system. The vocal production system is also very flexible; once trained up, there is little energetic cost to experimenting with novel combinations, which can be produced on the fly with little interference between components. When there is such interference—for example, if one finds it necessary to quickly assert that “Sally sells sea shells by the sea shore”—it is a notable exception that proves the rule. Dance, by comparison, is an energetically expensive activity that involves the entire body. The moves themselves are difficult to perform, and it may require years of practice to perfect a particular form or combination. Moreover, there is kinematic interference between components, involving factors like balance, positioning, and muscle fatigue. While a skilled speaker can often utter novel sentences for hours without physical fatigue, even an expert dancer may be exhausted after nine or ten double *Fouettés en Tournant* or *Grand Jeté*. For these reasons, concerns about embodiment and environmental embedding of dance activities should receive special attention in the comparative study of mating dances as a source of grammatical understanding. It may be, for example, that certain elements go unreported not because the birds lack the grammatical competence to do so, but rather because it would be physiologically or spatially impractical to repeat the move. Correspondingly, grammatical understanding involving such components may take different forms than it does with speech. Though we may occasionally think about the context in which we speak—a loud, crowded train station may make a poor site for an important conversation—in most cases, the background conditions for speech are ignored.

Conversely, though, apparent grammatical understanding may be a byproduct of physiological demands, spatial demands, or occurrent signaling not performed intentionally as part of a dance structure, or even represented anywhere in the manakin’s nervous system. One of the central motivations for the cognitive revolution was the purported “stimulus-independence” of language, so if apparent grammatical complexity is found to be controlled by external stimuli or spatial/kinesthetic constraints, this might be seen to obviate the need for any cognitive representation of grammar. For example, it may be that moves involving wingflaps are repeated simply because flapping wings is such a common and automatic activity in

the bird's life, or hops between branches may be repeated in even numbers because it is a spatial constraint of the mating display that the bird must return to the original perch to proceed. Even if such constraints do drive the apparent structure of the dances, this should not automatically rule out the possibility that the computational level of description is appropriate or that the behavior exhibits a grammar. However, proponents of this argument should explicitly consider these possibilities by noting the required concessions to embedding and embodiment, and offer rebuttals that preserve the relevance of externalized dance structure to the evolution of vocal or gestural human speech.

One useful comparison here might be between an expert orator and an expert gymnast: what are the differing demands that their expertise places on their bodies, and how do the constraints differ in the way that they can demonstrate their expertise? An expert orator may be able to perform for hours, extemporize on a variety of different topics with ease, and show little physiological specialization to her task. An expert gymnast, on the other hand, may have great difficulty improvising a routine on the fly, and expertise in one gymnastic category may not generalize well to another. For example, an Olympic gymnast spends months specializing a routine for the parallel bars, and this degree of specialization will typically not generalize easily to another routine or another activity like the uneven bars (indeed, we already know that the quality of dances in golden-collared manakins is affected by the removal of single perch from their display area—Coccon, Schlinger, & Fusani, 2012). Their musculature, flexibility, reflexes, “muscle memory”, and physiological responses will be highly attuned to the specific demands of their practiced activity. A gymnast's routine may thus indeed follow a grammar, but the influence of that grammar would be expressed more in the choreography and practice of that routine, rather than the relatively stereotyped performance. It can thus appear that the demands of speech and dance are quite different indeed.

A rebuttal can be offered on the part of a dance-based evolutionario, however, that similar tensions are present on the original Darwinian musical protolanguage hypothesis—for the fair comparison here is not an expert orator, but rather an expert crooner or bard. By hypothesis, the “songs” in the second stage of this evolutionario that earned mates would have been much more difficult to perform than pedestrian speech, and only the songs of the most skilled performers would end in copulation. The highest levels of human artistic

vocal performance—the fair comparisons with classical ballet—still require years of practice to attain, are often practiced extensively before performance, and cannot be performed without the right anatomy (with a few Dylanesque exceptions aside). Perhaps only later in the third stage (and beyond) of the Darwinian story was speech exapted into a more relaxed form of performance where only the clear transmission of phonemes (together with perhaps other features like intonation and mood) mattered for successful communication.

Finally, we should consider differences in the neural correlates of skilled speech and dance. These demands are already known to vary in both humans and birds, and these differences may be relevant to the arguments of homology or analogy.⁵ Some of the strongest evidence for the uniqueness of human grammatical processing, for example, comes from evidence that Broca's Area in the left interior frontal gyrus is specialized for syntactic processing (Grodzinsky, 2000). Supposing that the human form of Broca's area evolved uniquely in our near ancestors raises more questions about its evolutionario, however, and other animals possess homologues and analogues in "Area 44" that have similar functions. A wide examination of the imaging literature on Area 44 in humans and animals reveals a range of possible functions that are more broadly consistent with the idea that syntax can emerge in both language and motor planning. For example, in humans and monkeys, activation in Area 44 has also been correlated with non-vocal activities like motor imagery, object manipulation, motor preparation, digit sequence learning, and complex motor planning in "Tower-of-Hanoi" type problems (Müller & Basho, 2004).

Recent "rewiring" and "reuse" hypotheses in the study of brain evolution lend further credence to the compatibility of the dance hypothesis with a functional role for syntactic processing of Area 44 in adult humans. These views hold that neural plasticity routinely results in brain areas with older, more general "workings" being pressed into service for a variety of different domain-specific functional roles through plasticity in evolution and development (Anderson, 2010). Work on the motor theory of vocal learning in

⁵ Despite the vast evolutionary distance between humans and birds, we should entirely foreclose the possibility that neural structures controlling components of FLB or even important precursors to FLN could be homologous. Even though the architecture of the avian brain is radically different from that of mammalian brains—birds lack a neocortex and white matter—some brain areas like the hippocampus (implicated in higher cognitive functions like declarative memory, transitive inference, sequence learning, and spatial cognition) do appear to be homologous, and there may be more subtle functional or architectural similarities. For reviews of such comparisons, see (Güntürkün & Bugnyar, 2016; Reiner et al., 2004).

songbirds—which finds that the song nuclei in three distantly related songbird groups evolved as a specialization of a pre-existing motor pathway (Feenders et al., 2008)—provides evidence that such rewiring of motor learning areas did in fact occur in the evolution of three distantly related species, strongly supporting comparative evidential parity between song and dance. Current neuroscientific evidence on language processing in humans is thus more compatible with the dance hypothesis than might initially be supposed, and raises further questions as to why the same brain region would be involved in both vocal and non-vocal tasks, questions that could be answered by the dance hypothesis.

Another line of relevant evidence can be obtained from comparative studies of the respective masses of brain areas across songbirds and manakins. These studies assess correlations between the size of the brain or brain regions and increased complexity in song or dance (Airey, Castillo-Juarez, Casella, Pollak, & DeVogd, 2000; Lindsay, Houck, Giuliano, & Day, 2015; Nowicki, Searcy, & Peters, 2002). In the case of songbirds, increased song complexity is correlated with increased brain size in the “song system”, which in zebra finches includes “area X” of the avian striatum (X), the high vocal center (HVC), and the robust nucleus of the archistriatum (RA) (Mooney, 1999). Manakins, by contrast, show specializations in the male motor control circuitry and female visual processing centers of the brain (Day, Fusani, Kim, & Schlinger, 2011), such as the cerebellum (CB) and arcopallium (AP), respectively. Overall brain size is correlated with increased complexity of dances across a variety of manakin species (Lindsay et al., 2015), though it is not currently known whether the size-complexity relationship is confined to particular motor regions of the brain or reflects a more coordinated brain-wide expansion. “Complexity” in these studies was also operationalized in terms of an aggregated combination of number of moves in a dance, presence of cooperation, complexity of display arena, and number of concomitant sounds in repertoire. To our knowledge, no studies have yet looked specifically for the neural correlates of grammatical complexity in dances, and it may be that the specific neural correlate of sequence pattern complexity is more constrained.

5. Experiments: The Next Steps

Though the evidence so far is suggestive, the current verdict must be agnosticism: we do not yet know how close a comparison can be made between mating dances and protolinguistic vocal productions in a Darwinian

approach to language. As such, we end the paper by sketching future lines of investigation that should be pursued—in parallel to the existing lines on vocal production described above—to determine the relevance of mating dances to the evolution of language. As mentioned above, these experiments are broadly broken down into producer effects, observer effects, and neuroscientific studies.

5.1 Producer effects

The first line of investigation into producer effects would be to perform more thorough observation and analysis of existing dance structure, looking especially for higher-order and long-distance dependencies between dance elements. There is already good evidence that some forms of FSG grammar exist in the manakin dances, such as a distinction between repeatable and non-repeatable elements. Given that “lexicons” of dance moves have already been proposed by various researchers for particular species, we could now construct a centralized database of dance patterns encoded in this lexicon from collected video recordings. These pattern databases could then be statistically analyzed to look for more complex forms of grammatical structure in their transition probabilities.

Crucially, however, a different kind of analysis is needed than that already performed by Lukianchuk & Doucet. Their analysis aggregated first-order transition probabilities across the dances of many different individuals; but Chomskyans have long stressed the irrelevance of this kind of probabilistic analysis to their program, as it would not detect hierarchical organization, especially if such organization varied from individual to individual. More powerful kinds of analysis are needed, especially forms that could test hypotheses about different ways to generate sequences hierarchically. Two promising avenues would be data-oriented parsing methods like Bayesian program induction and deep learning, both of which excel at discovering hierarchical organization in complex input sequences (Bod, 2009; Lake, Salakhutdinov, & Tenenbaum, 2015). A disadvantage of such methods is that they can require large input corpora, which we do not yet possess.

Exploring the details of these methods here would consume an entire paper, however; more useful for present purposes would be elaboration on the kind of structures these analyses might expect to find, such as:

A. Iterable Multi-Move Sequences

Is there a dance variation that repeats a hop and a wing flick a given number of times—a substructure like (<hop><wing flick>)ⁿ—in a context before moving on to another move sequence? How common are such patterns? Do all individuals display the same iterable patterns and simply combine them in different orders, or is there a large degree of intra-individual variation in pattern choice?

B. Sub-Sequence Openers and Closers

Are there “rules” regarding moves that must begin or end particular phases or sequences of the dance? If iterable components occur in a given context, but must be preceded or succeeded by particular moves, this suggests hierarchical sequence organization. Existing observations in long-tailed manakins strongly suggest that the *teemoo* and *toledo* calls must precede the hopping phase, and that a bow usually precedes copulation. However, it is unknown whether these patterns reflect arbitrary grammatical preferences, or rather universal features of dances for a given species induced by organizational and spatial demands of the cooperative display or copulation.

It may prove difficult to determine a unique analytical category for openers and closers—i.e. whether they should be seen as serving a grammatical role as a syntactic type in a dance’s structure, or serve merely as a pragmatic requirement in recruiting other birds to view or partner in the dance. However, such confusion over an element’s linguistic category would not be unique to dance. For example, linguists have similarly debated whether prosody should be seen as playing primarily grammatical, semantic, or pragmatic roles. Speakers of many languages raise tone when beginning a statement and drop tone at its end; should these prosodic elements be seen as syntactic components of the spoken sentence’s grammatical structure, or pragmatic moves in auditory segmentation to grab listener’s attention and then signal the termination of a message, to aid in conversational turn-taking (Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991; Selting, 1996)? In fact, theorists should consider the possibility that dance openers and closers (and prosody) play roles in multiple distinct analytical categories simultaneously.

C. Long Distance Dependencies

For particular dances, are there higher-order correlations between the number of times that an iterable move or sequence is repeated in one section of the dance, and the number of times a different iterable move or sequence is repeated in a later section of the dance? For example, is there any structure of the A^nB^n form that would begin to suggest PSG-level grammatical complexity? Given the fact that such patterns should be expected to vary from performer to performer, the kind of analysis needed is one that could use free parameters to fit higher-order patterns, to determine whether a higher-order hypothesis fits recorded dance structures better than first-order alternatives.

The discovery of any of these patterns in the dance databases would support the position that mating displays deserve as much attention in the evolution of grammatical structure as birdsong. However, not even the discovery of all of these patterns would definitively demonstrate that the manakin dances exhibit PSG-level grammatical complexity, for these analyses would run into the same kind of epistemological problems as the A^nB^n tests in birdsong. The problem is that the PSG-level grammar hypothesis is ultimately about the functional or computational organization of a system—how it represents the patterns it can produce—and general considerations about the infinite number of ways that one might produce the same pattern show that no amount of analysis of output patterns alone can definitively establish the function that the manakins compute. These doubts raise more general considerations of parsimony, idealization, functional abstraction, and inference to the best explanation that would take us well beyond the scope of this paper (Franks, 1995; Zednik & Jäkel, 2016). We can, however, discuss other forms of experimentation that might provide stronger evidence concerning the representational format of the manakins dance-plans, should they exist.

D. Flexible Recovery from Interruption

One class of experiment would be those that interrupt the male manakin's display while already underway to see how they recover. Do they need to start back from the beginning of a display, or can they pick up where they left off? Is there evidence that the remainder of the display continues to fit patterns expected from long-distance dependencies (should any exist) to patterns in the earlier part of the display? There is a further ecological concern about the experiment that the male manakins may opt not to pick up where they left off

not because they do not represent the rest of the intended pattern, but rather because they know that only a flawless, completed dance will impress a female.

One kind of preliminary experiment in this vein has already been performed on golden-collared manakins. Coccon, Schlinger, and Fusani (2012) disrupted the dances of male golden-collared manakins (a solo-performing species) by placing a branch on one of their display perches as a form of obstruction. The idea behind the experiment was to test whether the manakins could adapt their display to the presence of the obstruction. The males in this case performed longer displays that omitted a key element inviting females to copulate, suggesting that the dances cannot be adapted on the fly. Though other species of manakin may be more flexible, this experiment shows that we should not expect to find grammatical flexibility at the time of performance; if present, it would rather have to be at the time that the moves are planned or practiced. An interesting precedent for experimental design is provided by this experiment, however—they used the same males' behavior before the court was disrupted as a control for their post-disruption behavior. This is one important way that experimental methodologies may overcome the challenge that each individual may favor a differently choreographed pattern to its dance. The experiment further suggests that in at least this species, once practiced the dance patterns are relatively stable for that individual at the time of performance.

E. Flexibility at Learning/Practice

The final kind of producer effect discussed here pertains to structural flexibility exhibited at the time that the dances are learned or practiced. Unfortunately, little is currently known about the development of dance structure. Male long-tailed manakins have a relatively long lifespan of 12-18 years, but do not develop full display plumage or reach sexual maturity until 3-5 years of age (McDonald, 1993)—though they begin practicing dance elements alone and with partners much earlier. Prior observations suggest that male manakins begin to choreograph and practice their dances at the beginning of each mating season, first at low speed, gradually increasing speed through practice until they achieve extreme speed and precision (Coccon et al., 2012). Upon reaching sexual maturity, males engage in an extensive process of partner selection and bonding. Dominance at dance sites is largely determined by age and experience; it typically takes males 6-10

years establish a stable relationship as an “apprentice” to an alpha who routinely performs dances for females (Trainer, McDonald, & Learn, 2002).

In this and other cooperatively displaying species, there is speculation that the subordinate males improve their own ability to perform and sequence dance moves during their apprenticeships (DuVal, 2013; Foster, 1981). Though to our knowledge no long-term longitudinal studies have been performed on coordination or learning of specific dance structures by betas from alphas, one study was conducted on call frequency matching in long-tailed manakin displays (Trainer et al., 2002). This study found no evidence of short-term accommodation in partners’ song frequencies, though it did find that with increased age non-alpha song competence increased, variability in song structure amongst partners decreased, and frequency matching among partners increased. This supports a hypothesis not of rapid learning and accommodation, but more long-term adjustments akin to the acquisition of expertise. It is not known, however, whether the congruence amongst partners develops as a result of imitation, sheer practice, or mutual attunement; and this study did not focus on grammatical structure in either song or dance.

Another important form of flexibility to investigate at the time of learning/practice is thus whether and how younger males acquire dance-patterns from watching or participating in dances involving more dominant males. Are patterns acquired and added to prior routines one move at a time, all-at-once, or in sequenced chunks? Perhaps the latter would be the most suggestive in overcoming skepticism about mere patterns found in behavioral output, because it would show that the birds themselves, in the process of learning the patterns, chunk those elements together, rather than learning to sequence individual moves together by rote. For example, if an onlooker observed a male performing a dance with an A^nB^n pattern in the hopping phase of the dance, and modified its own pattern by adding this whole chunk at once rather than by adding one A and B element at a time after repeated observations until the dances matched, then this would suggest that the onlooker understood that pattern as a chunk, and could plan to perform it as a chunk before practicing it as a series of individual moves. Conversely, observations that onlookers added such patterns one move at a time would provide evidence against grammatical understanding of the dance sequences, as it would be like a child

learning to play a piano melody by rote practice gradually adding one note at a time, with no knowledge of the overall sequence or musical structure of the phrase.

F. “Negative Evidence”

Some of the strongest evidence for hierarchical grammatical structure in mating displays might be derived from what we fail to find in exhaustive search. Some of the strongest evidence for hierarchical grammar in human language comes from the fact that certain patterns or parsings that would cohere with first-order probabilistic expectations do not occur (so-called “negative evidence”—Penke & Rosenbach, 2004). For example, in English certain strings like “John thinks Sally is taller than himself” do not occur despite cohering with local probabilistic expectations. Moreover, certain parsings of sentences are grammatically impossible; for example, we can parse the sentence “John saw the man from Texas” in exactly two ways—either John saw the man from a Texan vantage point, or John saw a Texan man, but there is no reading of the sentence that would assert that John is Texan.⁶ In short, some of the strongest evidence for hierarchical dependence in dance might come from the absence of certain patterns that might otherwise be expected on the basis of first-order transition probabilities.

5.2 Observer Effects

An even more promising source of evidence for grammatical understanding might come from the way that females process the dances in making copulation decisions. Since the females do not perform the highly-skilled dances themselves, we may face fewer confounds between grammatical understanding and motor specializations (though again this would only be a deal-breaker on the assumption, contra embodied-embedded ideas, that such motor specializations could not constitute grammatical understanding). Moreover, though females have the opportunity to observe many dances by males, they evaluate finished products from many different onlookers, and they may select a pattern for copulation the first time they see it. Thus, even if it is unlikely for physiological reasons that the males would improvise a highly novel pattern in a display at the time of performance (as opposed to the time of acquisition/practice), females may prefer a highly novel pattern the first time they have seen it if it demonstrates more grammatical complexity. We thus move to

⁶ We are grateful to an anonymous reviewer for suggesting these examples.

discuss different ways to look for such flexible understanding in onlooking females, as well as forms of data that would speak against such grammatical understanding.

A. Dependencies Between Grammatical Patterns and Copulation Success

The most obvious source of evidence for grammatical understanding in females would be the finding that the forms of grammatical pattern and sequencing discussed in section 5.1 are highly determinative of copulation decisions by females. For example, are females more likely to offer copulation in response to some particular pattern, even on the first time they see it? Are they more likely to fly away from the observation position or otherwise demonstrate a lack of interest when grammatical patterns are violated? Existing research already suggests that females prefer more complex sequences of moves (Lindsay et al., 2015), but other studies have found that courtship preferences in some manakin species can be predicted largely by milliseconds-level speed differences with which components of the display are performed (Barske et al., 2011). In other words, the dances can be seen to display information to potential mates about both neuromuscular coordination/power and cognitive flexibility, and part of what remains to be determined is the degree to which each factor determines copulation success in each species.

Though it is debated whether the female manakin's courtship decisions are based on the ability of the dances to broadcast fitness or simply because they provide the females with aesthetic pleasure (Prum, 2012), an initial adaptationist hypothesis about this tradeoff between neuromuscular fitness and cognitive complexity might be that both features will be relevant in most species, though the balance will be determined by the distinctive ecological pressures each species confronts. For example, in species facing high predation risk, mere motor strength and speed might be most important. For species that need to hunt or escape predators by darting quickly through dense canopy, coordination between motor and visual processing might be most important. For more social species or those that face difficult foraging challenges, cognitive complexity might dominate, on the assumption that cognitive demands of mastering dominance hierarchies or foraging maps are more important to reproductive fitness (and the social and spatial lives of long-tailed manakins may be quite complex indeed—Edelman & McDonald, 2014; McDonald, 2010). These are merely speculative “just so” stories, but they are offered to illustrate that we should not expect a simple answer to this question,

and the most suggestive evidence might come from comparing predictors of copulation success to ecological pressures across species.

It is worth explicitly connecting this point to those about negative evidence in the previous section. Specifically, one obstacle to extending methods from human linguistics to the study of mating dances is that mating dances almost surely have no semantic significance, so some of the ways that linguists analyze human speech for semantic coherence may be inapplicable. However, there may be a useful analogy between grammaticality and copulation success that could be exploited in future research. Namely, copulation success could provide a proxy as to whether a female judged a male's dance to be "grammatically correct". Are there some patterns that fit first-order probabilistic expectations, but never result in copulation success? This linkage is at present speculative, but researchers might keep it in mind when studying dance-patterns.

B. Guidance/Control of Female Responses Over Male Dance Structure

Another important factor to investigate would be the ongoing influence of female responses on the structure of the dance, should any exist. Females in many manakin species actively participate in the courtship display, either by performing their own moves, shifting position, or merely by continuing to watch instead of flying away. The degree to which males can modulate their performance in response to female cues is not currently known. For example, it might be that males move on to the next phase of a dance once receiving sufficiently positive feedback from females, or it may be that they increase the number or speed of moves when females begin to appear disinterested. Evidence that dance structure is largely determined spontaneously by female responses should be seen as evidence against the centralized representation of grammatical structure by males or females, as such structure would only emerge from the embodied cooperation between the birds. Again, however, it bears repeating that fans of embodied/embedded influences on cognition, and "second-person" interactions in human social cognition might see this emergent quality as a feature rather than a bug (Barrett, 2016; De Jaegher, Di Paolo, & Gallagher, 2010).

C. Playback/Artificial Stimuli

One of the most alluring possibilities and challenges in this line of research would be playing videos of dances to females in controlled conditions to gauge their responses. This kind of experiment would mirror those

already conducted on birdsong—for example, by habituating females to dance-grammars from playbacks of real or artificially constructed stimuli, to determine whether they treat dances that match or violate the patterns from various grammars differently. However, there are serious practical challenges facing the execution of analogous studies with dance. For one, it is already known that females are very good at discriminating slight differences in male dances. It is unknown whether animals are sensitive to an “uncanny valley” for artificial stimuli or specifically where it might exist (Mori, MacDorman, & Kageki, 2012), and it could be a challenging practical problem to replay real stimuli or construct artificial stimuli that the females did not reject for grammatically-irrelevant reasons.

Such experiments would involve showing females recordings or animations on video displays. There is general reason to suppose that birds can process visual information presented in this manner, even in a mate-choice context (Lea & Dittrich, 2000; Ophir & Galef, 2003; Patterson-Kane, Nicol, Foster, & Temple, 1997). If the stimuli consist of recorded male dances, however, then experimenters would have to simply wait until males produced the relevant patterns and replay those dances over and over again to habituate the females. It would also be difficult to control for the non-grammatical features of the stimuli such as length and speed of movement performances, since there is so much variation in natural dances and females are able to make such fine discriminations along these variables. This raises the question of artificial stimuli. It might be possible to create artificial dances with the desired qualities by stitching together frames from real dances or producing CGI animations of dancing males. The generation of such artificial dance videos would be more challenging than artificial sound stimuli, for which considerable expertise already exists in biology and ethology, and it is unknown whether female manakins would reject these videos. However, a number of different labs in animal psychology and ethology have recently recruited individuals with expertise in video editing and computer animation to generate animations that have been successfully deployed in experiments on animal cognition, including research on mating preferences with intelligent birds like parrots (Bian, Chandler, Laird, Pinilla, & Peters, 2017; Moravec, Striedter, & Burley, 2010; Scherer, Godin, & Schuett, 2017; Woo & Rieucou, 2011). Thus, while it would be difficult to exactly replicate the structure of experiments

performed on artificial grammar in birdsong, animations could provide one of the most powerful methods to test hypotheses about grammar in dance.

5.3 Neural Comparisons

Because it would be difficult to exactly replicate the designs of experiments on grammatical structure in birdsong, some of the strongest support for the dance protolanguage hypothesis might come from study of birds' brains. Specifically, we should compare neuroanatomical metrics of various brain areas (whether size or number of neurons--Olkowicz et al., 2016) between manakins exhibiting more grammatically complex dances and those with less grammatically complex dances. The dance grammar hypothesis would hold that we should expect to find higher scores on neuroanatomical metrics in areas correlated with cognitive flexibility and grammatical understanding in birds with more grammatically complex dances, and higher scores on neuroanatomical metrics for cerebellar and motor areas in birds with less grammatically complex dances. Such candidate areas associated with cognitive and grammatical flexibility include Area X of the LPO and the NCL (Güntürkün, 2005; Watanabe, 2006).

6. Conclusion

We have here made the case that the elaborate mating dances of manakins and other lekking birds may be—like birdsong—an important comparison case in the study of the evolution of language. These dances appear to demonstrate many of the same selection pressures favoring increased grammatical complexity as birdsong, and exhibit as much *prima facie* anthropological relevance. Though current findings are suggestive, we do not yet see anything to challenge the dominant view that PSG-level complexity is uniquely human, probably supported by a kind of push-down stack architecture in Broca's Area (Fitch, 2014). However, particular tests for this kind of computational structure in animals—such as tests for A^nB^n center-embedding design in auditory stimuli—have been overemphasized, and other relevant sources of sexually-selected complexity like dance have been neglected. The more interesting comparative questions now concern the precursors for the evolution of such a computational structure, the selection pressures that favor them, and the specific kinds of grammatical structure that might emerge as a result (even if they fall short of center-embedding). In this

essay, we hope to have brought together the considerations and concerns that would support fruitful investigation of these questions in a promising, non-vocal domain.

Acknowledgments

The authors would like to thank Susan Balenger, Simon Brown, Lainy Day, Carrie Figdor, Jo Anne Fleischhauer, Nathan Gabriel, Bruce Rushing, and two anonymous reviewers for discussions and comments on earlier versions of this paper.

References

- Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, *14*(8), 1067–1074. <https://doi.org/10.1038/nn.2869>
- Airey, D. C., Castillo-Juarez, H., Casella, G., Pollak, E. J., & DeVoogd, T. J. (2000). Variation in the volume of zebra finch song control nuclei is heritable: developmental and evolutionary implications. *Proceedings of the Royal Society of London B: Biological Sciences*, *267*(1457), 2099–2104.
- Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *The Behavioral and Brain Sciences*, *33*(4), 245–66; discussion 266–313. <https://doi.org/10.1017/S0140525X10000853>
- Armstrong, A. (1964). *Maori Games and Hakas: Instructions, Words, and Actions*. Wellington: AH & AW Reed.
- Barrett, L. (2016). Why Brains Are Not Computers, Why Behaviorism Is Not Satanism, and Why Dolphins Are Not Aquatic Apes. *The Behavior Analyst*, *39*(1), 9–23. <https://doi.org/10.1007/s40614-015-0047-0>
- Barske, J., Schlinger, B. A., Wikelski, M., & Fusani, L. (2011). Female choice for male motor skills. *Proceedings of the Royal Society of London B: Biological Sciences*, *rsob20110382*. <https://doi.org/10.1098/rspb.2011.0382>
- Berwick, R. C., & Chomsky, N. (2015). *Why only us: Language and evolution*. MIT press.
- Berwick, R. C., & Chomsky, N. (2017). Why only us: Recent questions and answers. *Journal of Neurolinguistics*, *43*, 166–177. <https://doi.org/10.1016/j.jneuroling.2016.12.002>
- Bian, X., Chandler, T., Laird, W., Pinilla, A., & Peters, R. (2017). Integrating evolutionary biology with digital arts to quantify ecological constraints on vision-based behaviour. *Methods in Ecology and Evolution*, *Early View*. <https://doi.org/10.1111/2041-210X.12912>
- Bod, R. (2009). From Exemplar to Grammar: A Probabilistic Analogy-Based Model of Language Learning. *Cognitive Science*, *33*(5), 752–793. <https://doi.org/10.1111/j.1551-6709.2009.01031.x>
- Borgia, G. (1995). Complex male display and female choice in the spotted bowerbird: specialized functions for different bower decorations. *Animal Behaviour*, *49*(5), 1291–1301.
- Bortolotti, G. R., Blas, J., Negro, J. J., & Tella, J. L. (2006). A complex plumage pattern as an honest social signal. *Animal Behaviour*, *72*(2), 423–430.
- Bowling, D. L., & Fitch, W. T. (2015). Do Animal Communication Systems Have Phonemes? *Trends in Cognitive Sciences*, *19*(10), 555–557.
- Buckner, C. (2013). Morgan’s Canon, meet Hume’s Dictum: avoiding anthropofabulation in cross-species comparisons. *Biology & Philosophy*, *28*(5), 853–871.
- Buckner, C. (2015). A property cluster theory of cognition. *Philosophical Psychology*, *28*(3), 307–336.
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, *3*(9), 345–351.
- Coccon, F., Schlinger, B. A., & Fusani, L. (2012). Male Golden-collared Manakins *Manacus vitellinus* do not adapt their courtship display to spatial alteration of their court. *Ibis*, *154*(1), 173–176.
- Corballis, M. C. (2007). Recursion, Language, and Starlings. *Cognitive Science*, *31*(4), 697–704. <https://doi.org/10.1080/15326900701399947>
- Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex* (First Ed.). London: John Murray.
- Day, L. B., Fusani, L., Kim, C., & Schlinger, B. A. (2011). Sexually dimorphic neural phenotypes in golden-collared manakins (*Manacus vitellinus*). *Brain, Behavior and Evolution*, *77*(3), 206–218.

- Day, L. B., Westcott, D. A., & Olster, D. H. (2005). Evolution of bower complexity and cerebellum size in bowerbirds. *Brain, Behavior and Evolution*, 66(1), 62–72.
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14(10), 441–447.
- Donald, M. (1991). *Origins of the modern mind: Three stages in the evolution of culture and cognition*. Harvard University Press.
- Donald, M. (1998). Mimesis and the executive suite: Missing links in language evolution. *Approaches to the Evolution of Language: Social and Cognitive Bases*, 44–67.
- DuVal, E. H. (2013). Does cooperation increase helpers' later success as breeders? A test of the skills hypothesis in the cooperatively displaying lance-tailed manakin. *Journal of Animal Ecology*, 82(4), 884–893.
- Edelman, A. J., & McDonald, D. B. (2014). Structure of male cooperation networks at long-tailed manakin leks. *Animal Behaviour*, 97, 125–133. <https://doi.org/10.1016/j.anbehav.2014.09.004>
- Eltz, T., Sager, A., & Lunau, K. (2005). Juggling with volatiles: exposure of perfumes by displaying male orchid bees. *Journal of Comparative Physiology A*, 191(7), 575–581.
- Feenders, G., Liedvogel, M., Rivas, M., Zapka, M., Horita, H., Hara, E., ... Jarvis, E. D. (2008). Molecular Mapping of Movement-Associated Areas in the Avian Brain: A Motor Theory for Vocal Learning Origin. *PLOS ONE*, 3(3), e1768. <https://doi.org/10.1371/journal.pone.0001768>
- Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.
- Fitch, W. T. (2013). Musical protolanguage: Darwin's theory of language evolution revisited. *Birdsong, Speech, and Language: Exploring the Evolution of Mind and Brain*, 489–504.
- Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, 11(3), 329–364. <https://doi.org/10.1016/j.plrev.2014.04.005>
- Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656), 377–380.
- Foster, M. S. (1981). Cooperative behavior and social organization of the Swallow-tailed Manakin (*Chiroxiphia caudata*). *Behavioral Ecology and Sociobiology*, 9(3), 167–177.
- Franks, B. (1995). On explanation in the cognitive sciences: Competence, idealization, and the failure of the classical cascade. *The British Journal for the Philosophy of Science*, 46(4), 475–502.
- 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(10), 964–972.
- Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440(7088), 1204–1207. <https://doi.org/10.1038/nature04675>
- Graham, M. (1973). *The Notebooks of Martha Graham*. Harcourt.
- Grodzinsky, Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23(01), 1–21.
- Güntürkün, O. (2005). The avian “prefrontal cortex” and cognition. *Current Opinion in Neurobiology*, 15(6), 686–693.
- Güntürkün, O., & Bugnyar, T. (2016). Cognition without cortex. *Trends in Cognitive Sciences*, 20(4), 291–303.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598), 1569–1579.
- King, B. J. (2009). *The dynamic dance: nonvocal communication in African great apes*. Harvard University Press.
- Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, 350(6266), 1332–1338.
- Lea, S. E., & Dittrich, W. H. (2000). What do birds see in moving video images. *Picture Perception in Animals*, 143–180.
- Lindsay, W. R., Houck, J. T., Giuliano, C. E., & Day, L. B. (2015). Acrobatic courtship display coevolves with brain size in manakins (Pipridae). *Brain, Behavior and Evolution*, 85(1), 29–36.

- Lukianchuk, K. C., & Doucet, S. M. (2014). Cooperative courtship display in Long-tailed Manakins *Chiroxiphia linearis*: predictors of courtship success revealed through full characterization of display. *Journal of Ornithology*, 155(3), 729–743.
- McDonald, D. B. (1993). Delayed Plumage Maturation and Orderly Queues for Status: A Manakin Mannequin Experiment. *Ethology*, 94(1), 31–45. <https://doi.org/10.1111/j.1439-0310.1993.tb00545.x>
- McDonald, D. B. (2010). A Spatial Dance to the Music of Time in the Leks of Long-Tailed Manakins. In R. Macedo (Ed.), *Advances in the Study of Behavior* (Vol. 42, pp. 55–81). Academic Press. [https://doi.org/10.1016/S0065-3454\(10\)42002-1](https://doi.org/10.1016/S0065-3454(10)42002-1)
- Mikhalevich, I., Powell, R., & Logan, C. (2017). Is behavioural flexibility evidence of cognitive complexity? How evolution can inform comparative cognition. *Interface Focus*, 7(3), 20160121.
- Mooney, R. (1999). Sensitive periods and circuits for learned birdsong. *Current Opinion in Neurobiology*, 9(1), 121–127.
- Moravec, M. L., Striedter, G. F., & Burley, N. T. (2010). “Virtual Parrots” Confirm Mating Preferences of Female Budgerigars. *Ethology*, 116(10), 961–971. <https://doi.org/10.1111/j.1439-0310.2010.01809.x>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100.
- Müller, F. M. (1861). The theoretical stage, and the origin of language. *Lectures on the Science of Language*, 7–8.
- Müller, R.-A., & Basho, S. (2004). Are nonlinguistic functions in “Broca’s area” prerequisites for language acquisition? fMRI findings from an ontogenetic viewpoint. *Brain and Language*, 89(2), 329–336.
- Neave, N., McCarty, K., Freynik, J., Caplan, N., Hönekopp, J., & Fink, B. (2011). Male dance moves that catch a woman’s eye. *Biology Letters*, 7(2), 221–224. <https://doi.org/10.1098/rsbl.2010.0619>
- Nowicki, S., Searcy, W. A., & Peters, S. (2002). Brain development, song learning and mate choice in birds: a review and experimental test of the “nutritional stress hypothesis.” *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, 188(11), 1003–1014.
- Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences*, 201517131.
- O’Neil, D. (2012). Evolution of Modern Humans: Archaic Human Culture. Retrieved November 14, 2017, from https://www2.palomar.edu/anthro/homo2/mod_homo_3.htm
- Ophir, A. G., & Galef, B. G. (2003). Female Japanese quail affiliate with live males that they have seen mate on video. *Animal Behaviour*, 66(2), 369–375.
- Patterson-Kane, E., Nicol, C. J., Foster, T. M., & Temple, W. (1997). Limited perception of video images by domestic hens. *Animal Behaviour*, 53(5), 951–963.
- Penke, M., & Rosenbach, A. (2004). What counts as evidence in linguistics?: An introduction. *Studies in Language. International Journal Sponsored by the Foundation “Foundations of Language,”* 28(3), 480–526. <https://doi.org/10.1075/sl.28.3.03pen>
- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, 90(6), 2956–2970.
- Prum, R. O. (2012). Aesthetic evolution by mate choice: Darwin’s really dangerous idea. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 367(1600), 2253–2265.
- Pukui, M. K. (1942). The Hula, Hawaii’s Own Dance. *Hula: Historical Perspectives*, 30, 70–73.
- Reiner, A., Perkel, D. J., Bruce, L. L., Butler, A. B., Csillag, A., Kuenzel, W., ... Jarvis, E. D. (2004). Revised nomenclature for avian telencephalon and some related brainstem nuclei. *Journal of Comparative Neurology*, 473(3), 377–414.
- Scherer, U., Godin, J.-G. J., & Schuett, W. (2017). Validation of 2D-animated pictures as an investigative tool in the behavioural sciences: A case study with a West African cichlid fish, *Pelvicachromis pulcher*. *Ethology*, 123(8), 560–570. <https://doi.org/10.1111/eth.12630>
- Scott-Phillips, T. C., Cartmill, E. A., Crockford, C., Gärdenfors, P., Gómez, J. C., Luef, E. M., ... others. (2015). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, 56(1), 000–000.

- Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructive units and turns in conversation. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 6(3), 371–388.
- Starr, F., & Cunningham, M. (1968). *Changes: notes on choreography*. Something Else Press.
- Stein, A. C., & Uy, J. A. C. (2006). Plumage brightness predicts male mating success in the lekking golden-collared manakin, *Manacus vitellinus*. *Behavioral Ecology*, 17(1), 41–47.
- Stevens, C., & McKechnie, S. (2005). Thinking in action: thought made visible in contemporary dance. *Cognitive Processing*, 6(4), 243–252.
- Taylor, A. H., Hunt, G. R., Holzhaider, J. C., & Gray, R. D. (2007). Spontaneous metatool use by New Caledonian crows. *Current Biology*, 17(17), 1504–1507.
- Trainer, J. M., McDonald, D. B., & Learn, W. A. (2002). The development of coordinated singing in cooperatively displaying long-tailed manakins. *Behavioral Ecology*, 13(1), 65–69.
<https://doi.org/10.1093/beheco/13.1.65>
- Vicari, G., & Adenzato, M. (2014). Is recursion language-specific? Evidence of recursive mechanisms in the structure of intentional action. *Consciousness and Cognition*, 26, 169–188.
- von Humboldt, W. (1836). *Über die Kawi-Sprache auf der Insel Java*. Berlin: Druckerei der Königlichen Akademie der Wissenschaften.
- Ward, K.-A. (2012). *Female mate-searching strategies and behavioural correlates of copulation success in lekking Long-tailed Manakins (Chiroxiphia linearis)* (MSc). University of Windsor. Retrieved from <http://scholar.uwindsor.ca/etd/301/>
- Watanabe, S. (2006). The neural basis of cognitive flexibility in birds. In E. A. Wasserman & T. R. Zentall (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 619–639). Oxford: Oxford University Press.
- Woo, K. L., & Rieucou, G. (2011). From dummies to animations: a review of computer-animated stimuli used in animal behavior studies. *Behavioral Ecology and Sociobiology*, 65(9), 1671.
- Zednik, C., & Jäkel, F. (2016). Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, 193(12), 3951–3985.