Element Transitions in the Phonemic Communication System of a Non-Human Animal

Submitted by Eleanor Ann Tew to the University of Exeter as a thesis for the degree of Masters by Research in Biological Sciences, July 2021.

(Signature) ……………………………………

ACKNOWLEDGMENTS

My first and biggest thank you has to go to my supervisors, Dr Camille Coye and Professor Andy Russell, who have offered endless guidance and knowledge throughout this project. Without their teaching this research would been simply impossible. Next my thanks goes to Louis O'Neill, Chiara L'Herpinere and Karen Keegan who along with Camille Coye made up the 2018 babbler field team. Not only did they contribute to my data collection and introduced me to new field skills, they made the whole experience a joy. This thanks extends to everyone at the Fowlers Gap Arid Zone Research Station in 2018. Next, I must thank Helen Mylne who worked tirelessly to code and extract data from our nest videos with me. Her cheery enthusiasm was much appreciated during those long hours! Thank you to Lucy Browning who contributed the video data from 2007 and 2008. Finally, I acknowledge my friends and family who have offered endless support and encouragement during this time. Without this unwavering support, I know this experience would have been immeasurably harder. Funding was provided, via Professor Andy Russel and the University of Exeter, by NERC, The Royal Society, and the Australian Research Council.

ABSTRACT

Traditionally, language is a trait which has been considered as “uniquely human”, without too much consideration as to what parallels may exist in animal communication. However, in recent years this has begun to change, and great leaps are being made to study animal vocalisations on a deeper linguistic level. In doing so it has become apparent that they do use sound units as building blocks to form longer more complex calls, a parallel to human morphosyntax. However, in the case of phonemes, where these sound units are meaningless when produced in isolation, only one species has so far been demonstrated to use a comparable system: the chestnut-crowned babbler. Here we build on previous research, supporting the parallel to rudimentary phoneme use by this cooperative passerine. We demonstrate these two meaningless sound units, the A and B notes, are produced by all individuals and across different environmental contexts. We propose that their production as single elements in the babblers’ repertoire to be mistakes as the birds search for the correct call, or switch between different calls, and thus still hold no meaning or behavioural cue to the birds. We continue to demonstrate further parallels to human language through the presence of the linguistic laws of compression and coarticulation in the A-B call complexes. These two notes of the babbler’s repertoire act as an example as to the depth of complexity that exists in animal communication. Further research in other species is crucial to identify these linguistic parallels with human language in order to give greater insights into the evolution of language and communication as a whole.

CONTENTS

.

CONTENTS OF TABLES AND FIGURES

INTRODUCTION

Communication is a vital component of life and comes in several forms: acoustic, visual, chemical, electrical and tactile (Bradbury & Vehrencamp 1998). Acoustic signals are a particularly efficient form of communication over longer distances due to their ability to overcome many physical obstacles (Bradbury and Vehrencamp, 1998). Humans rely heavily on language for communication, and it has therefore evolved into an incredibly complex and extensive system. Language consists of several crucial components: phonemes, meaningless elements which can be combined in different combinations to form morphemes (words), and syntax, which is the rearrangement of these morphemes (words) into different complex structures (sentences) (Collier *et al.*, 2014). The combination of morphemes, meaningful sounds, into longer structures can be called morphosyntax.

Traditionally, language has been widely considered to be something that is uniquely human. However, the origins of human language remain unclear and the theories as to its evolution are highly debated (Scott-Phillips, 2015). The discontinuous theory of language evolution supports the idea that certain elements of language are limited to just humans and are not present in any form, however rudimentary, in animal communication (Hauser *et al*., 2002). On the contrary, the continuous view of language evolution suggests that language must evolve following the rules of natural selection, suggesting that perhaps the evolution of human language should not be considered qualitatively different to that of non-human animals (Fedurek & Slocombe, 2011; Pinker & Bloom, 1990). In this framework, adopted by most biologists, any differences between the taxa are purely quantitative in nature.

According to the continuous view of language evolution, language evolved gradually from pre-existing trait which could be co-opted for communication. These traits can be seen in their rudimentary forms in our closest non-human relatives. This is illustrated in the case of proto-conversation, such as turn-taking, where individuals pause to allow alternation between being the speaker and receiver. In human language, the timing in turn taking is universal, repeatably approximately 200 milliseconds (Stivers *et al*., 2009). It also has very early onset, with 3-month-old infants able to take turns, and then becoming more refined with age (Hilbrink *et al.*, 2015). Parallels with animal communication have been studied in primates (Levison *et al*., 2016). Common marmoset (*Callithrix jacchus*) infants are able to demonstrate turn taking and continue to make fewer mistakes as they get older (Chow *et al*., 2015). Similar conversation rules are observed in Campbell’s monkeys (*Cercopithecus campbelli*) (Lemasson *et al*., 2005). The evolution of conversation rules, and specifically turn-taking, in multiple primate species, and on separate evolutionary branches (e.g., dolphins (Janik, 2000), elephants (Solik *et al.*, 2005), lesser skylarks (Gochfeld, 1978) and nightingales (Naguib, 1990), see also Henry *et al*., 2015 for a review) to humans, suggests it is a homologous trait and likely to be related to social bonds and necessary for communication in closely bonded groups (Takahashi *et al*., 2013).

Studies into the parallels between human language and animal communication have progressed significantly over the last decade. There are several steps to drawing these parallels. Firstly, the specific parallels must be identified. Once this has occurred, these parallels can be quantified. Next it is important to evaluate when these parallels may have emerged. Lastly the specific evolutionary pressures can be considered. An example of these four steps can be seen in the research into vocal learning. Vocal learning is a process in which individuals modify their vocalisations based on their experience and interactions of those of other individuals (Janik & Slater, 1997). It has been identified to be a parallel between human language and animal communication. Now this parallel has been identified, it can be quantified. It is likely that both humans and animals rely on similar neural structures and developmental paths and that the FoxP2 gene has been co-opted in several bird and mammal species as well as humans (Fee & Scharff, 2010; Scharff & Petri 2011). The FoxP2 gene codes for transcription factors which have been found to be required for the acquisition of language (Enard *et al*., 2002). Vocal learning was then found to be distributed across a wide range of species, including hummingbirds, parrots, elephants and pinnipeds (Baptista & Schuchmann, 1990; Pepperberg, 2010; Poole *et al*., 2005; Schusterman 2008). This presence across such a wide range of taxa suggests that this linguistic aspect is most likely an analogous trait and evolved as convergent evolution (Varella & Ghazanfar 2021). This completes the third step, allowing an assessment of when the parallel may have emerged. Finally, evolutionary pressures can be considered. It is believed to have more likely evolved in highly social species, due to the possible benefits it would have to group adhesion, affiliation, and individual identification (see Tyack, 2008 for a review).

As outlined above, combination of sounds is an integral component of human language and it occurs on two levels: phonology and syntax. To date no animal species has shown parallels on both levels. However, comparisons have already between drawn to morphosyntax. Certain animal species have been shown to combine meaningful calls into meaningful call combinations. Such morphosyntax-like systems have been found in southern pied babblers (*Turdoides bicolor*), Japanese great tits (*Parus minor*) and primates including Campbell's and Diana (*C. diana*) monkeys. Southern pied babblers, a cooperatively breeding bird endemic to South Africa, include two calls in their call repertoire: a broadband alert call and tonal recruitment call. However, they are seen to combine the two in a bi-component call complex to produce a mobbing call (Engesser *et al*., 2016). Playback experiments using natural and artificially combined mobbing calls confirm that this combination does utilise the separate alert and recruitment, ensuring that this is an example of rudimentary morphosyntax. Similarly, Diana monkeys have four meaningful calls used within a social context, H, L, R and A. The H, L and R calls convey contextual information while the A call is given in a wide variety of contexts and strongly signals caller identity (Candiotti *et al*., 2012). However, the A can be used as a suffix and combined with the L and R calls to create a larger call structure which also conveys caller identity as well as social context (Coye *et al*., 2016)

When it comes to phoneme-like systems, there is only one non-human animal to have yet been shown to use any rudimentary system; the chestnut-crowned babbler (*Pomatostomus ruficeps*). These cooperatively breeding birds, endemic to the Australian outback, combine and reuse meaningless elements in several meaningful calls (Engesser *et al*., 2015, 2019). Several of the calls in their repertoire are candidates for a phoneme-like structure but it is the A/B call complex which has so far received attention. When arranged in an A-B order, it constitutes a flight call, when in a B-A-B arrangement it is a prompt call, each with very different meanings and behavioural responses (Engesser *et al*., 2015). Previous playback experiments have demonstrated that it is the same elements reused in both calls (Engesser *et al*., 2015).

The question of meaning is at the core of the above-mentioned definitions of phonemic and morphosyntactic systems. While in humans the meaning of utterances is often straightforward and accessible directly through questions to the speaker, this is not the case in animals. Here, we propose to access the meaning of animal signals via indirect observations of context and receiver's reaction. Is the signal associated with a clear context of emission? Does the sound in question produce a predictable and consistent response in receivers? Thanks to detailed observations, one can delineate the most likely set of circumstances feature that appear at a rate greater than chance across the signal utterances, a definition for meaning proposed by Berthet *et al*., (in prep) and adapted from Dezecache & Berthet (2018). When it comes to the flight and prompt calls of chestnut-crowned babblers, previous research has already established their meaning. 61% of flight movement is accompanied by a flight call (Crane *et al*., 2016) and prompt calls are given almost exclusively in the nest (90%, Crane *et al*., 2016). Furthermore, in aviary playback experiments, recordings of flight calls caused birds to spend more time looking outside the aviary (49% increase, compared to natural prompt calls), as if searching for the receiver, and more time hopping or flying about (36% increase, compared to natural prompt calls) (Engesser *et al*., 2015). Prompt calls caused birds to spend more time looking towards the nest (81% increase, compared to natural flight calls) (Engesser *et al*., 2015). In opposition to flight and prompt calls, single elements have not been associated with a consistent behavioural response or context of emission, and playback experiments only prompt general arousal to unknown stimuli (Engesser *et al*., 2015).

In this study I have investigated the A and B notes produced by chestnut-crowned babblers, focusing both on the flight and prompt call structures and particularly on the production of these two notes as single elements, an aspect which has so far received no attention. In chapter one, the production of these single elements is outlined, detailing when single elements are most commonly produced and who by. Furthermore, Markov models have been used to confirm the occurrence on non-random transitions between elements, a crucial consideration if hoping to draw parallels between this communication system and language. Chapter two will be used to address the question of compression laws and coarticulation on A and B elements when produced as single elements and as part of call structures. When combined, the results from both chapters will give clearer insights into the A-B call complex and allow an evaluation to the extent to which parallels can be drawn between this communication system of the chestnut-crowned babbler and human language.

CHAPTER ONE

COMBINATORIAL CALLS ARE DIVISIBLE INTO CONSTITUENT SOUND ELEMENTS IN CHESTNUT-CROWNED BABBLERS

ABSTRACT

A distinct component of human language is the ability to confer meaning through the combination of meaningless elements, the phonemes. This combining of sounds allows for a more complex level of communication, which until recently was not seen to be replicated in non-human animals. However recent research focused on *Pomatostomus ruficeps*, chestnut-crowned babblers, a cooperatively breeding passerine endemic to Australia, has highlighted it as a candidate for utilising a phonemic-like communication system. Two notes, A and B, are reused and rearranged in both flight and prompt calls, hinting at a phoneme-like system. Firstly, A and B notes are shown to be arranged non-randomly and most note sequences follow a strict order. Secondly, single A and B elements are shown to be commonly produced outside of the flight and prompt calls. Furthermore, single A and B elements are produced by all individuals, they are not idiosyncrasies of certain individuals. These analyses continue to suggest that it is the A and B notes, rather than the flight and prompt calls they form, which are unitary. We suggest two hypotheses to explain the prevalence of single element production, despite the absence of meaning of these A and B notes in isolation. Our early termination hypothesis proposes these elements are the result of flight or prompt calls initiated but not completed. Alternatively, our protracted time hypothesis suggests that it is perhaps due to exceedingly long inter-note intervals within calls which causes the A and B notes to appear as single elements. Crucially both hypotheses would continue to support the theory that these single elements are meaningless, acting as building blocks to form higher meaningful structures, and thus should continue to be considered one of the first identifiable phoneme-like systems in non-human animals.

INTRODUCTION

The ability to confer meaning through the generation of words is an integral component of language (Hurford, 2002). Words are generated using a finite number of meaningless sounds (phonemes) in specific arrangements. An association is then learnt between this phoneme arrangement and its contextual meaning. For example, the word *pup* contains two phonemes in the order /p/ /u/ /p/, while the word *up* uses the same two phonemes, but with each represented just once (/u/ /p/). Through this process of combination and recombination, hundreds of thousands of words can be generated from just ~40 phonemes in the English language. Importantly, there is no connection between the phonemes /u/ or /p/ and the meaning of the words in which they are used and nor, despite lexical similarity, is one word generated from the other (i.e. by adding or dropping the /p/ phoneme). Put another way, both words are generated independently from each other by sourcing the phonemic building blocks from the potential pool, which in this case, just so happen to contain the same two blocks in contrasting arrangements. Vocal communication systems in animals can also comprise combinations of sound elements, but whether or not such sound elements constitute individual phonemes in the linguistic sense is still unclear. Nevertheless, comparing and contrasting the basic process of word generation with the make-up of animal vocalisations offers one of the few means of generating insights into early candidate forms of language pre-cursors (Marler, 1998).

To qualify as a phoneme, at least four features that must be met in animal vocalisations. First, sound units or elements must be contextually meaningless, in that the meaning of the whole cannot be gleaned directly from the inclusion of specific constituent sound units. Second, the same sound units must be used in different arrangements within a species' vocal repertoire. Third, changing the arrangement of sound elements within vocal sequences must change the contextual meaning of the whole. Finally, each sound unit must be demonstrably independent of the whole for it to operate as a building block (i.e. words pup and up are generated by the differential use of independent phonemes rather than existing as two unitary words). Bird and mammal song, usually defined as vocalisations that function in mate attraction and territoriality (Bradbury and Vehrencamp, 1998), often comprise sequences of apparently meaningless sound elements in different combinations. In addition, although songs can be simple and with a fixed sequence of sounds (e.g. great tits, *Parus major*; Lambrechts & Dhondt, 1990), in others, song sequences can be so variable so as to be rarely repeated (e.g. nightingales, *Luscinia megarhynchos*; Weiss *et al.*, 2014). The most obvious explanation for such complex singers is that songs are generated by the variable use of a finite set of individual sounds, otherwise they would be relying on learning hundreds or thousands of songs in their lifetimes. Nevertheless, studies have yet to demonstrate that changing song structure changes the function or contextual meaning of information. By contrast, animal calls, defined as vocalisations typically directed towards partners and group members, are demonstrably context specific (such as for contact, aggression or predation; Manser *et al*., 2014; Searcy & Marler 1981). However, in this case, the same sound elements are rarely shown to be meaningless or to exist in different arrangements in contrasting contexts; so again, call sounds appear rarely to be comparable to phonemes in structure.

There are at least three intriguing exceptions that suggest a possibility that animal calls can comprise phoneme-like entities. First, black-capped and Mexican chickadee (*Parus atricapillus* and *P. sclateri*) calls typically comprise variable organisations of four acoustically distinct elements (A-D) (Hailman, 1989; Moscicki *et al.,* 2010). Although element organisation in calls varies across contexts (e.g. foraging *vs*. alarm), it has yet to be demonstrated that elements are shared and contextually meaningless, that information is underpinned by precise sequence structure of calls and elements exist as independent sound units. Secondly, acoustic elements also appear to be shared across calls used in distinct contexts in white-handed gibbons (*Hylobates lar*). For example, in this case, some acoustically similar elements are more common and occur in different positions within calls in the context of predators versus inter-group interactions (Andrieu *et al.,* 2020; Clarke *et al.*, 2006). However, here again, experiments confirming that meaningless and independent elements are shared across context, and that a precise relationship underpins element organisation and information, are yet to be conducted. Finally, chestnut-crowned babblers (*Pomatostomus ruficeps*) have been shown to use similar acoustic elements across calls in the context of flight and offspring provisioning (Crane *et al*., 2016). Playback experiments confirm that the two calls comprise the same two sound elements in different organisations and that context is determined by element organisation within the two calls (AB – flight calls; BAB – prompt calls) (Engesser *et al*., 2015). Further, separate experiments confirm that the two elements are contextually meaningless (Engesser *et al*., 2019). However, what is not clear in this case is whether or not the two calls exist as unitary entities or whether A and B elements are independent of the calls in which they occur, as is the case with phonemes.

Here I investigate whether or not A and B elements are likely to be independent of the calls in which they are found in the chestnut-crowned babbler system, the final requirement to meet the definition of phoneme-like entities. This highly social cooperatively breeding bird from inland south-eastern Australia has a rich vocal repertoire of at least 18 distinct calls. Although several calls offer candidates of phoneme-like elements (Crane *et al*. 2017), I concentrate here on the A and B elements that comprise flight and prompt calls since these are the only elements for which all other requirements for phoneme-like entities have been met through previous experimentation (Engesser *et al*. 2015, 2019). Flight calls are uttered during >60% of flights, including to and from nests, whereas prompt calls are almost exclusively given in the nest when provisioning nestlings (Crane *et al*., 2016); particularly when they are not begging (Young *et al*., 2013). Demonstrating that just two elements across two calls constitute phoneme-like entities, however, is challenging since in human language phonemes are identified by their meaning-bearing impacts in multiple words. This leads to a paradox: in animal communication with limited element sharing, demonstrating independent elements needs to both confirm that elements are associated non-randomly within calls but also can occur alone or in aberrant arrangements. There will be two steps to this chapter. Firstly, I will investigate whether elements are combined non-randomly in calls, an important step in the comparison to human language where phenome ordering is crucially non-random. Secondly, I will investigate whether A and B elements are likely to be independent building blocks of the larger non-random structures that they generate, the flight and prompt calls.

If the two elements exist as independent entities, I hypothesise that they should sometimes be produced outside of precise flight or prompt call arrangement. If A and B elements are independent units then the two calls, flight and prompt calls, are not rigid structures that occur in in these configurations only due to some production constraints and the birds should be able to produce these elements outside of the traditional flight and prompt arrangements. However, if the two calls are “unitary” structures that are produced in a very rigid way because the elements in it are not really independent notes, then we should not observe single A and B elements nor partial calls, (i.e. the emission of the call would be a strictly stereotyped behavioural pattern). For example, if individuals are conflicted about whether and when to transition between flight calls (AB) and prompt calls (BAB) when travelling to and from the nest to feed young, and elements occur as independent units, we might expect some B elements to occur in isolation or truncated prompt calls to occur (i.e. BA). Similarly, if elements are independent entities, we might also expect uncertainty whether or not to fly or especially short flights to lead to truncated flight calls (i.e isolated A elements). The elements would be considered independent entities as they act as building blocks to form the more complex meaningful calls. To test these possibilities, I first quantify the production of single A and B elements outside of structured calls. Next, I investigate the influence of bird status and external factors such as chick age and feeding activity on single element production. This will allow us to see whether certain individuals or situations encourage higher rates of these vocalisations. Comparisons will also be made between rates of vocalisation in nest environments and in aviary environments to confirm single element production is not limited to one situation and is the widespread phenomenon we would expect if these elements are to be considered independent entities and building blocks of the widely used flight call and prompt call.

METHODS

Data was collected from two sources on wild chestnut-crowned babblers at Fowlers Gap Arid Zone Research Station in New South Wales, Australia (141°42′E, 31°06′S) (Russell, 2016). The study population was established in 2004 and the identity of the vast majority of individuals is known. For the purposes of this study, this includes approximate age immigration status, with individuals classified as yearlings if they were born in the previous breeding season (where known) and immigrants if they were not natal to the social group (Russell, 2016). Sex was determined molecularly from blood samples (Rollins *et al*., 2012). Additionally, the vast majority of individuals carry 2x12mm Trovan PIT tag (TROVAN^{TM} Ltd, NL) transponder chips inserted subcutaneously in their flanks either just prior to fledgling or as free-flying juveniles or adults (Nomano et al. 2014). Free-flying individuals are captured relatively easily by setting mistnest across the creek-beds along which they generally travel (Portelli *et al*. 2009) and nests are found primarily by following females that utter a distinct ‘piping’ call during breeding (Russell et al. 2010; Crane et al. 2017). In some years, 10 mm diameter MO-S408 pen cameras with audio recording capacity (Misumi Electronics Corporation, New Taipei City, Taiwan) were installed ~5cm from the nest cup to record nest visits by group members (see Browning *et al*. 2012 for further details). Whilst in others, birds were temporarily removed to on-site aviaries during group size manipulation experiments (max = 48 h) and vocalisations recorded using tie-clip microphones. No more than a third of group members were removed and breeding females were never removed. Specific details of removals, housing conditions and successful releases is provided elsewhere (Engesser *et al*., 2015; Nomano *et al*., 2015). Here I make use of recordings obtained in each context to address the aims of this study.

(i) Data Collection

From the nest recordings collected in 2007 and 2008, I extracted 4732 and 7409 A and B elements, respectively from high quality recordings (i.e. with reduced wind interference) from 50 individuals in 10 groups over a total of 20 days (see Appendices 1 for further details). All recordings were obtained within 2 seconds of nest entrance and departure to ensure sound elements could be reliably determined. Individual identity was known in the majority of visits owing to the unique transponder code being registered along with date and time as the bird passed through the copper coil antenna positioned in the entrance of the dome-shaped nests and linked to a concealed LID650 decoder at the base of the tree (TROVANTM Ltd) (Browning *et al*., 2012; Nomano *et al*., 2014). There is no evidence that the equipment in the nest had any effect on the babblers' behaviour and birds routinely use nests with equipment already in position (Russell 2016).

From aviary recordings collected in 2018, I extracted 69 single A elements and 18 single B elements. When also including the elements produced as part of flight calls, prompt calls and truncated prompt calls, I extracted 307 and 269 A and B elements respectively, for 8 individuals from 7 groups (see Table 1).

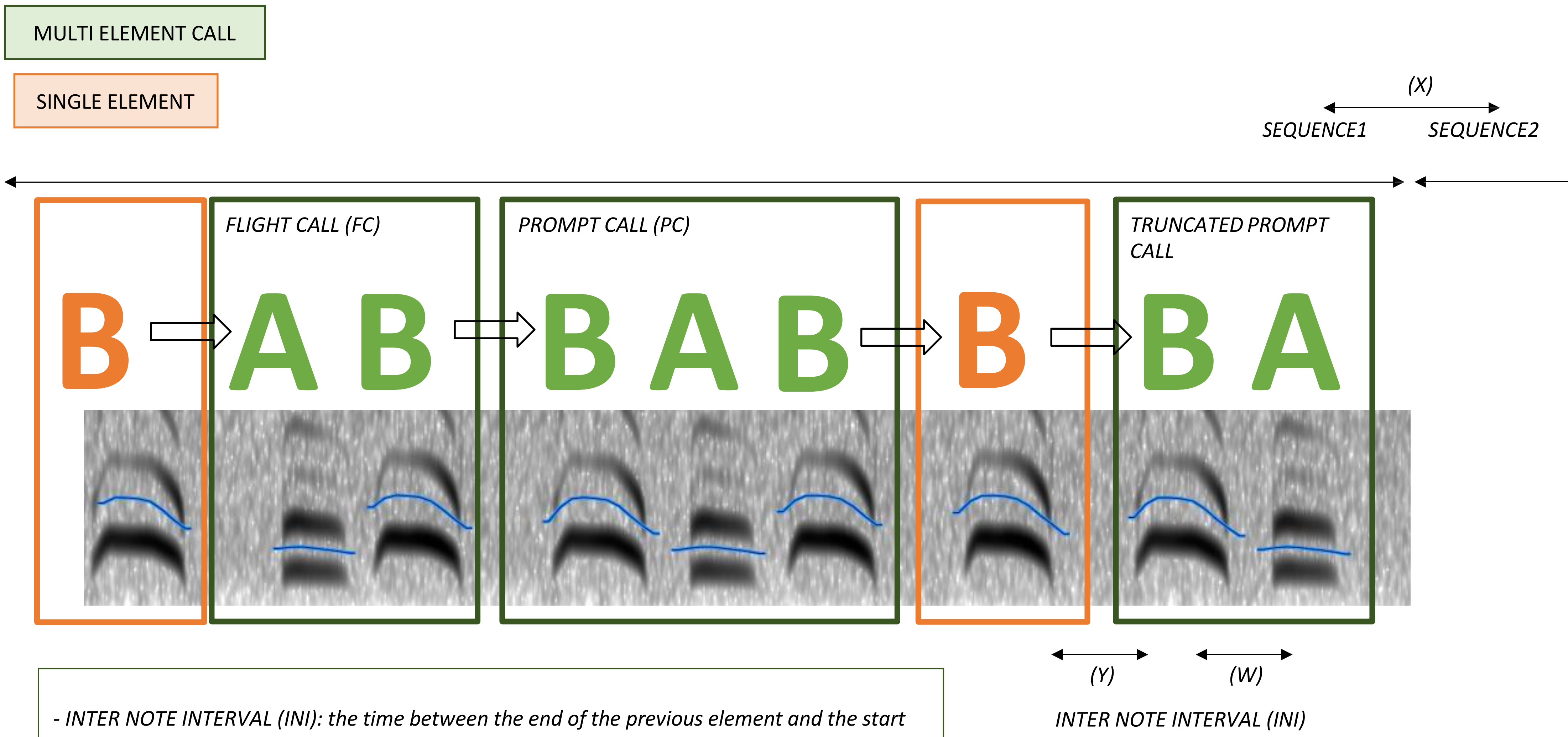
Table 1.1: Summary of aviary single notes and calls. *This table summaries our aviary data, giving an overview of the number of individuals per group and the number of each call we recorded for each of them. FC stands for flight call and PC for prompt call.*

GROUP	No. Individuals	A	B	FC	PC	Truncated PC	TOTAL
BFD	**2**	**1**		**1**			**2**
62356				1			1
79281/62203		1					1
EFG	**1**	**20**	**4**	**22**	**10**	**9**	**65**
62650		20	4	22	1	9	56
EFH	**1**						**0**
79278							0
NFD	**1**	**16**	**1**	**1**			**18**
79341		16	1	1			18
SAG	**2**	**11**	**4**	**149**	**3**		**167**
79173		2	3	84	2		91
79194		9	1	65	1		76
TGF	**1**	**21**	**9**	**34**	**2**	**4**	**70**
79811		21	9	34	2	4	70
TOTALS	**8**	**69**	**18**	**207**	**15**	**13**	**322**

(ii) Video Coding and Note Extraction

In total 15 videos from 10 different nests were used in my analyses. The data was extracted by Helen Mylne and myself and totalled 24 hours and 28 minutes of video footage, including calls from 50 individual birds and a further 42 unidentifiable entries. The videos from nest pen cameras were recorded in MP3 format. Element of relevance were identified by converting videos were converted to WAV format, at 16-bit amplitude resolution, using VLC Media Player (Version 3.0.8 Vetinari, Intel 64bit). VLC Media Player was also used to watch the MP4 files and ensure they matched with the PIT tag data and crosscheck with behavioural and prey data from Lucy Browning. Although recording directly in WAV format would have been preferred due to the higher quality, findings by Matt Doyle (Doyle, Masters Thesis, 2020) demonstrate that the conversion of MP3 to WAV does not significantly alter measured frequency parameters. These WAV files were then entered into the Praat software (Version 5.3.55 DSP Package (Boersma and Weenink, 2009) for analysis. Calls and notes were identified from audio-visual examination of spectrograms (FFT method, window length = 0.4 s, time steps = 1000, frequency steps = 250, Guassian window shape, dynamic range = 60dB). Calls or notes with very high levels of background noise, usually when birds were flying to or away from the nest from a considerable distance from the microphone, were not used. If multiple birds gave overlapping calls where it was not possible to accurately distinguish individual notes, or chicks begged loudly over the visiting adults, these notes were also ignored. In Praat, I used the annotation tool to mark each vocalisation. Boundaries were placed at the beginning and end of each note (tier 1) and call (tier 2). All notes which I considered to be A and B notes were selected. A and B notes were considered single elements when their inter note interval (INI), the duration from to the beginning of the next note, was double the mean of the INI of known calls (see figure 1 for more details). Both notes and calls were extracted automatically using a praat script (produced by Crosswhite and edited by Antoniou (2013) and Coye (2019)).

Figure 1.1 (pg. 24): Typical chestnut-crowned babbler A-B sequence: An artificial sequence created for this figure shows a sequence typical of what may be seen at the nest. Single elements and calls are both present. Inter note intervals are described in the figure, including the mean and standard deviation values which have been used for defining single notes for analysis.



- INTER NOTE INTERVAL (INI): the time between the end of the previous element and the start of the next element (Y / W).

- MULTI ELEMENT CALL: A and B notes produced within either a flight call, prompt call or a truncated prompt call

- SINGLE ELEMENT: single A or B note produced in isolation. Classified as single if INI is more than double the average multi element INI.

- SEQUENCE: a string of single notes and calls separated by at least a 2 seconds at the beginning and end. A sequence can exist as just one single element if produced in total isolation, with no other call or element either two seconds before or two seconds after

INTER NOTE INTERVALS

X = new sequences define when X is more than 2 seconds. Mean = 37.39 seconds, SD = 85.06

W = INI between element within calls. Mean = 0.064 seconds, SD = 0.035

Y = INI between single elements. Mean = 0.49 seconds, SD = 0.47

Statistical Analyses

(i) Element Transitions

I examined the probability of note transitions using zero order markov processes (ZOMP) and first order markov processes (FOMP) (Kershenbaum *et al*., 2016). These analyses will indicate whether or not the transition between A and B notes in our dataset occur at a rate more or less than you would expect if the transitions were random. Any sequences which included a note which could not be accurately labelled (which in the previous analyses just the note would have been ignored), the entire sequence was removed. The actual number of A and B notes in the new dataset was recorded and from that the proportions of A and B notes calculated. This is the ZOMP. Next the FOMP can be calculated. There were four possible transitions in our dataset: A-A, A-B, B-A, B-B. The FOMP will result in two values for each transition, an observed value, and a theoretical value. The observed value is simply extracted from the dataset using excel. These observed counts are then easily converted to proportions and percentages. The theoretical value is calculated by multiplying the ZOMP values together for both note types involved in the transition, ie. A-A = A ZOMP x A ZOMP. This means the FOMP for the transitions A-B and B-A are theoretically the same. ZOMP and FOMP values were then calculated for on a five-level call factor, looking at the transitions between different call types. The five different calls were: single A notes, single B notes, flight calls (FC), prompt calls (PC) and the truncated prompt call. Chi2 contingency tables were then carried out to test for statistical significance between the actual and theoretical values on both a ZOMP and FOMP level.

(ii) ***Descriptive Statistics and Proportion of Calls***

For all future statistics a reduced, error checked data set was used, to remove poor quality notes. Predominantly this will have removed calls made as birds flew away from the nest, the sound faded fast, and interference was very high, leaving us with very poor recordings. This reduced our data set to 262 single A notes, 1393 single B notes, 2320 flight calls, 1543 prompt calls and 512 truncated prompt calls. Importantly it reduced our sample size of known individuals to 48. All statistical analyses were carried out in R (Version 3.5.3) and R Studio (Version 1.1.463) (R Core Team, 2020). All figures were produced in R Studio. To test the probability of production of single notes vs multi note calls, I used a generalised linear mixed model (GLMM) (using the glmer function of the lme4 package (version 1.1.23) (Bates et al., 2015)). The model included bird status (as either dominant female, dominant male or helper), food (i.e. yes or no) and chick age (a continuous variable, measured per day) as fixed effects and group ID and individual ID were included as random factors. Multi-note calls are expected to be more common than single note calls. To account for these pre-existing biases, if a single note was given during a visit it was coded as 1, regardless of how many multi-note calls are also present. Only if a visit contained only multi-note calls, was it coded as 0.

(iii) Repeatability Analysis

Repeatability analysis was carried out to examine the pattern of within- and between-individual variations in vocal production. If there is low within individual variability but high between there are likely to be individual characteristics influencing vocal production. If it is the opposite, higher individual variability and low between individual variability, then it could be influence by a context which regularly effects all the birds in the sample size. If both within- and between- show low variability, then something is fixed in the species and population. If both show high variability it would cause us to question whether we may have missed something important or that the production of single notes is entirely random. This analysis used the rpt function from the rptR package (version 0.9.22) (Stoffel, Nakagawa and Schielzeth, 2017). Individual ID and group ID were controlled for as random effects.

(iv) Comparison Between Contexts

To assess whether the patterns of vocal production described above were maintained in other contexts, I compared the proportion and rate of different call and element types in a nest and in an aviary setting. Mann-whitney tests using the wilcox.test were used for these comparisons. The analysis used the wilcox.test function from the dplyr package (version 3.6.2). Both proportion and rate were calculated. Proportion allows insight into how the different calls make up the repertoire. Rate allows insight into how this proportion may be skewed by a high rate of one specific call. Proportions were calculated on several levels. Looking at overall calls (including flight calls, prompt calls and reverse flight calls) and elements (single A and B notes). At a call level, looking at just flight and prompt calls and excluding the small number of reverse flight calls. And finally, on an element level, looking at both A and B single notes.

Similarly, I computed calling rates for each individual call and note type by dividing the number of elements of each type over the total observation time (in seconds). This was done separately for each bird. If the same bird was recorded over multiple days, a separate rate was calculated for each day. The daily rates were the averaged to generate a mean call rate for the individual. Initially a call rate was calculated for each call and element type. I then computed a rate for all calls (including flight calls, prompt calls and reverse flight calls) and all elements (including single A and B elements).

RESULTS

(i) Element organisation in vocal sequences

Whilst phonemes are meaningless, independent entities their ability to generate meaning requires predictable associations with other phonemes within words. Four results from nest recordings are consistent with A and B elements being distributed non-randomly within vocal sequences. First, on average, only 0.8% (± 3.4% SD) of A elements and 1.0% (± 1.8% SD) of B elements were produced out of sequence (defined as > 2 seconds between elements, N = 18063 A and B elements, 48 individuals). Second, in vocal sequences (as defined) the 50 individuals produced B elements 56% more often than A elements, which is statistically equivalent to the 40% random expectation given the ratio of flight: prompt calls recorded at the nest (3:2 in favour of flight calls) and both calls consisting of a single A element, but prompt calls containing two B elements (contingency table; $\chi^2 = 1.00$, $P = 0.32$, $n = 161$ average transitions; Fig 2A). Third, an average of 80% of element transitions by the 48 individuals were A-B or B-A (against combined expected of 48% based on random transitions), whereas 19% of transition were B-B (expected 37%) and just 1% were A-A (expected 15%) ($\chi^2 = 29.68$, d.f = 3, $P < 0.0001$; $n = 126$ average transition; Fig 2B). These transition percentages make sense given: (a) the forms of (AB) flight calls and (BAB) prompt calls; (b) both flight call-prompt call and prompt call-prompt call sequences result in B-B transitions; and (c) A-A transitions are unexpected from both call structure and sequences.

Finally, individuals produced an average of 85% (±16% SD) of A elements and 83% (±15% SD) of B elements within flight and prompt calls (see Fig 1 for definition of call).

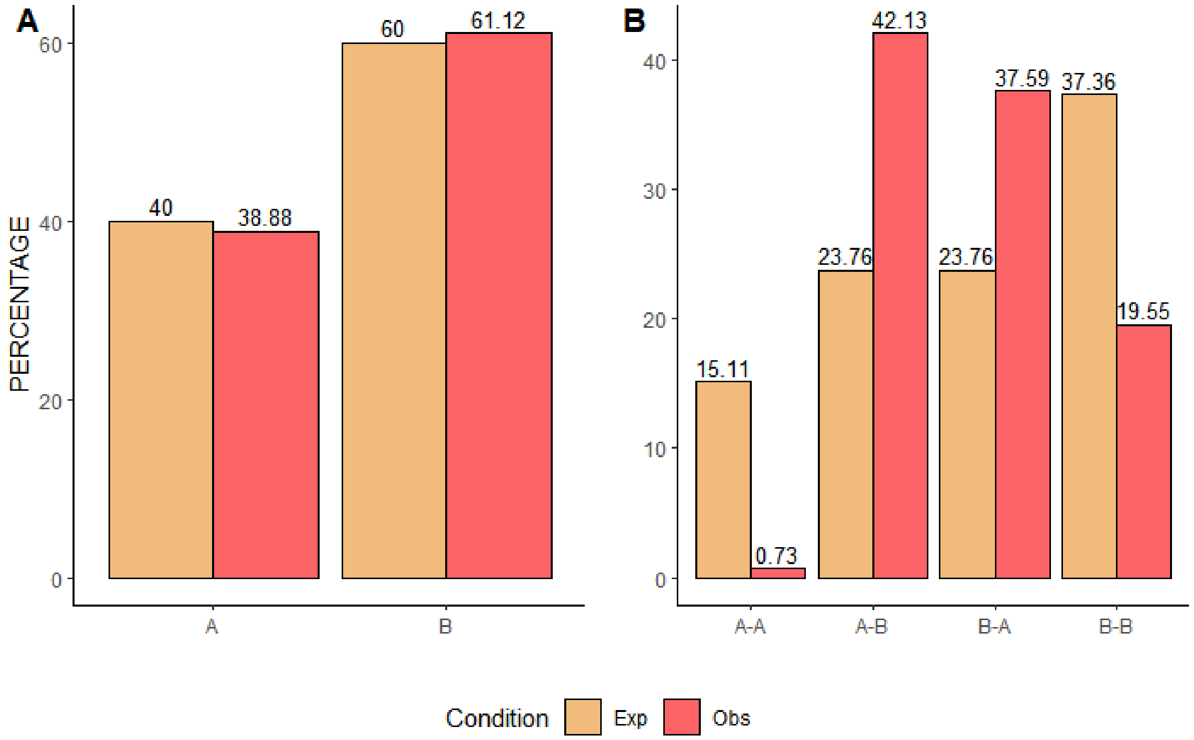


Figure 1.2: Comparison of theoretical and actual element transitions. *Bar plots displaying the element organisation and transition probabilities. Plot 1.2A displays the proportion of A and B notes across the entire dataset, regardless of whether these notes occur as single elements or within calls. As displayed, they occur at an approximately 40:60 split in the observed data "Obs", matching the expected 40:60 split "Exp" ($\chi^2 = 1.00$, $P = 0.32$, $n = 161$ average transitions). Plot 1.2B displays the transition probabilities. A-B and B-A transitions are by far the most common, together making up approximately 80% of all transitions and all differing from the expected percentages ($\chi^2 = 29.68$, d.f = 3, $P < 0.0001$; $n = 126$ average transition).*

Whilst the above results from nest recordings clearly demonstrate that A and B elements are primarily produced in non-random sequences matching the forms of flight and prompt calls, exceptions suggest that A and B elements are independent units. In addition to being occasionally produced in isolation (see above), A and B elements were found to occur in sequences with protracted time intervals, as well as in aberrant sequences. For example, on average individuals produced A elements in sequences with protracted time intervals or in BA sequences in 3% (±5% SD) and 9% (±9% SD) of sequences, respectively. In addition, individuals produced an average of 11% (±14% SD) of B elements in sequences with protracted time intervals, and a further average of 8% (±11% SD) in aberrant BA sequences. Interestingly, although the frequency of A and B elements in BA sequences was obviously the same, B elements were six times more often produced alone than A elements (Z = 3.08, N=48 pairs, P = <0.001) and fivefold more often produced in sequences with protracted time intervals than A elements (Z = 5.17, N= 48 pairs, P = <0.001). Together, these results suggest that not only are A and B elements independent units, but in the contexts of nests, their production outside of calls is non-random, with the production of B elements more prevalent.

(ii) Factors affecting lone element production in nests

The relatively large standard deviations surrounding average productions of A and B elements outside of calls hints at significant among-female variation but might also be driven by variation in observation time and/or call rates. Repeatability analyses, controlling for these confounding possibilities, suggested that some individuals were significantly more likely to produce elements out of call sequence than others at the nest (R = 0.13, p-value <

0.001), but most (~86%) variation occurred within individuals. Further, linear mixed effects models showed that the probability of producing elements out of call sequence during a nest visit was significantly affected by carer status, whether or not carers arrived at the nest with food and brood age (Table 2). Specifically, breeding females were more likely to produce single elements than breeding males (103.7%) or helpers (60.7%) (Fig 3A). Single elements increased significantly during nest visits with food (Est = 0.506, n = 629, df = 622, p-value = 0.0397) and decreased as chick age increased (Est = -0.389, n = 629, df = 622, p-value = 0.0056) (Fig 3B). These results show that the production of single elements varies consistently with external factors pertaining to nestlings (see Discussion for explanations).

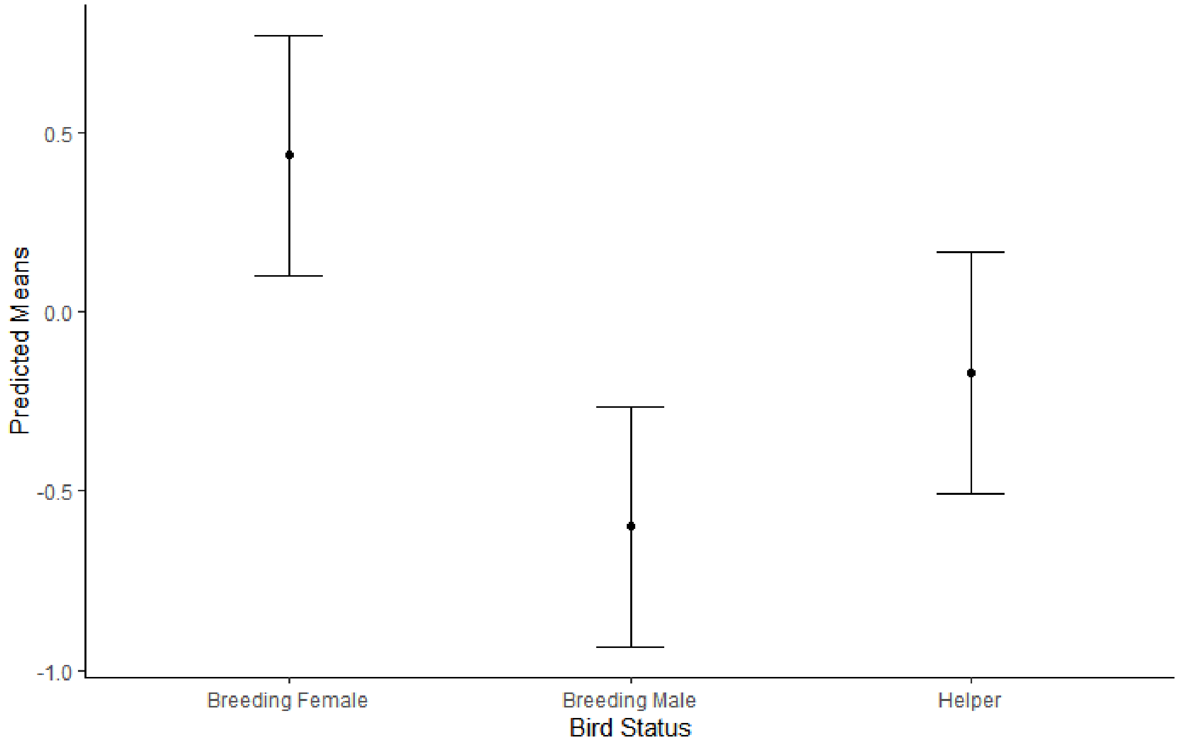


Figure 1.3A: The influence of status on single note production. *Plotted predicted means of the generalised linear model. The points plot the means and the error bars the standard error. As displayed breeding females are more likely to produce single elements than both breeding males (Est = -1.037, n = 629, df = 622, p-value = 0.002) or helpers (Est = -0.607, n = 629, df = 622, p-value = 0.051).*

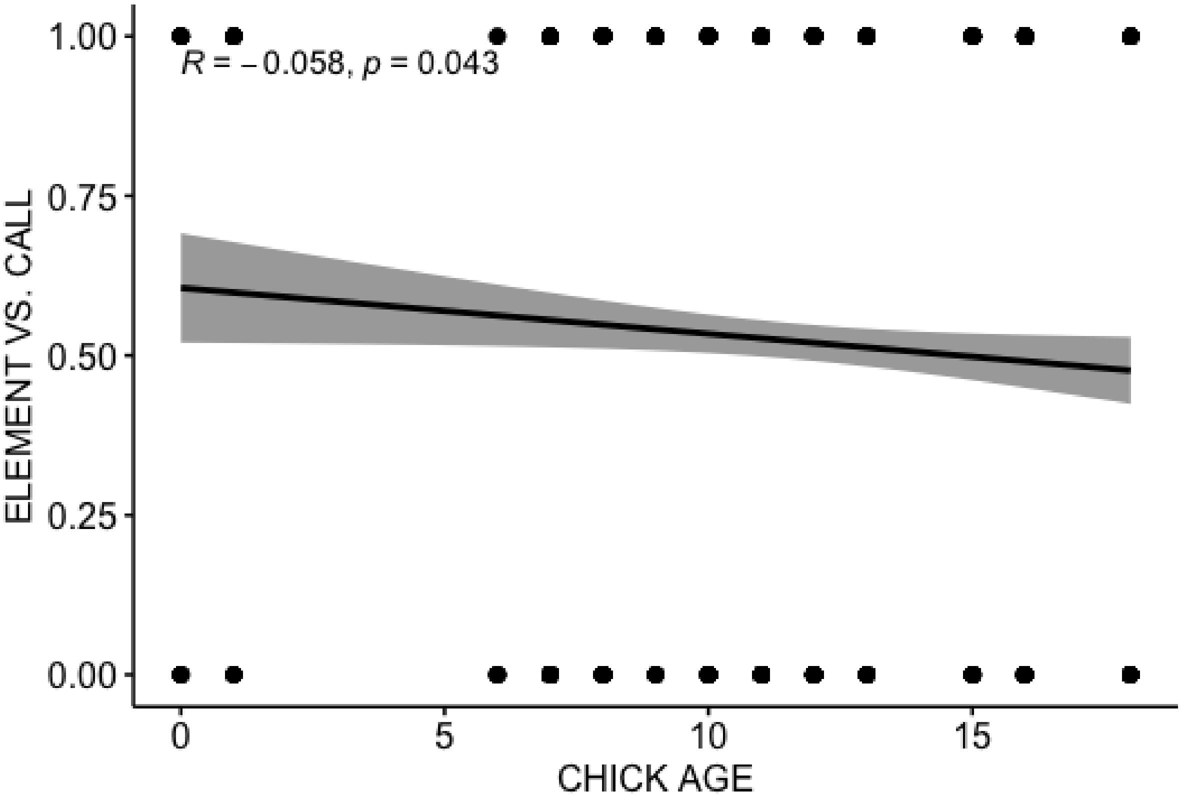
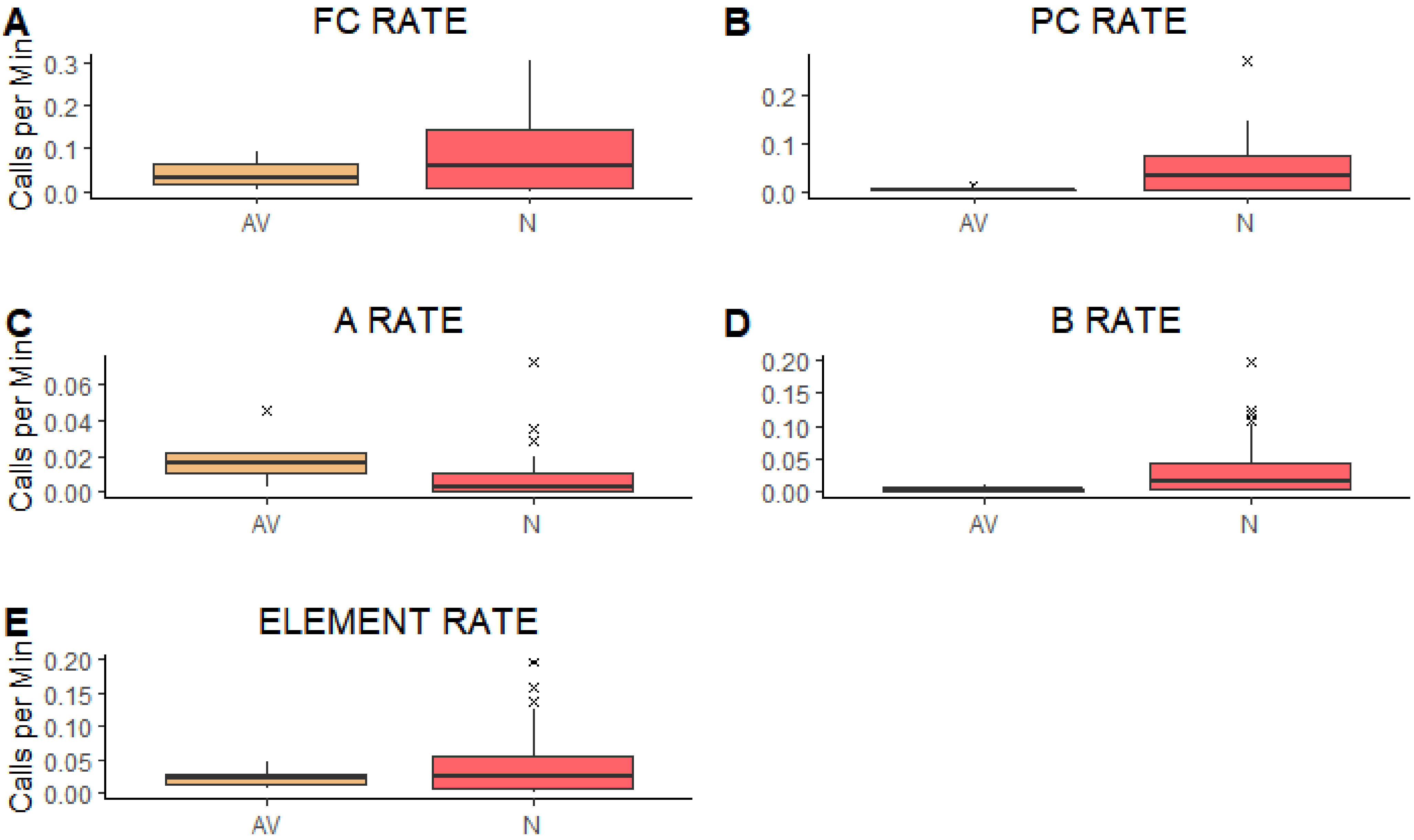


Figure 1.3B: The relationship between chick age and single note production. *Birds are more likely to produce a call vs. an element as chick age increases. Single elements and elements in calls have been coded binomially: elements as 0, elements in calls as 1. As displayed, there is a significant negative correlation between element production and chick age (R = -1.058, p = 0.043). Chick age was also incorporated into the generalised linear model and had a significant effect, single element production decreasing with chick age (Est = -0.389, n = 629, p-value = 0.0056).*

(iii) Lone element production in nests versus aviaries

In recordings at the nest, flight and prompt calls were split with a ratio of 50:50, meaning that nest visitors commonly made decisions about whether and when to transition between the two calls in the nest environment. By contrast, the flight: prompt call ratio in aviaries was over 9:1, meaning that transitioning between the two calls was rare. This difference in ratio was not driven by changes in the rate of flight calling across the two environments (Wilcoxon test: W = 182, n = 57, P = 0.40), but a significant reduction in the rate of prompt calling outside of the nest context (W = 246, n = 57, P = 0.012). Again, in this environment, lone elements were produced, and indeed the rate at which they were so was comparable to the rate observed at the nest (W = 160, n = 57, P = 0.80). However, this time, the production of A rather than B elements predominated. For example, while lone B elements were 7-fold more frequent than lone A elements in nest recordings (89% vs. 11% of lone elements), the reverse was true in the aviary, with A elements making up 87% of the lone A or B elements produced (W = 5, n = 57, P = 0.0002). Similarly, B elements were produced at a two-fold elevated rate in nests (W = 233, n = 57, P = 0.029), but A elements were produced at a similarly elevated rate in aviaries (W= 56, n = 57, P = 0.012). The results from aviaries not only further confirm that call elements can be produced independently, but that the dominant element produced alone varies across contexts allowing insights into why lone elements are produced.

Figure 1.4 (pg. 35): Nest and aviary rate comparisons*. Boxplots comparing rates of element and production in the aviary environment (“AV”) and the nest environment (“N”). As displayed, flight calls did not differ significantly between the two environments (1.4A), while prompt calls were drastically reduced in the aviary (1.4B). Furthermore, lone element production (1.4E) was at similar rates in both the nest and aviary but single A notes were produced at higher rates in the aviary (1.4C), while single B notes were produced at higher rates in the nest (1.4D)*


A
FC RATE
Calls per Min
0.3
0.2
0.1
0.0
AV
N
B
PC RATE
0.2
0.1
0.0
AV
N
C
A RATE
Calls per Min
0.06
0.04
0.02
0.00
AV
N
D
B RATE
0.20
0.15
0.10
0.05
0.00
AV
N
E
ELEMENT RATE
Calls per Min
0.20
0.15
0.10
0.05
0.00
AV
N

DISCUSSION

Chestnut-crowned babblers produce two multi-element calls comprising the same sound elements (A & B); wherein the precise arrangement of the sound elements within each call generates its specific meaning (Engesser *et al*., 2015). One explanation is that the two calls are unitary (i.e. each call represents a minimal unit), and just so happen to comprise the same sounds in different arrangements. Another is that, as with phonemes in word generation, sound elements represent the minimal units and that the two calls are 'built' from the same independent sound units. We reasoned that if the latter were true, the elements should sometimes be uttered alone or with aberrant sequence transitions. In support, although A and B elements were primarily produced in non-random sequences matching flight (AB) and prompt (BAB) calls, in both nest and aviary contexts A and B elements were commonly recorded with protracted inter-element time intervals and in aberrant sequences. Although such tendencies were more prevalent in some birds than others (e.g. breeding females in nests), the overwhelming tendency was for all birds to produce aberrant element sequences and intervals. In nests, lone elements were more common when birds arrived without food and declined with increasing brood age. In addition, the production of lone A and B elements was non-random with respect to recording environment, with lone B elements dominating in nests and lone A elements dominating in aviaries. These results demonstrate that it is the elements not the calls that are unitary, providing further evidence to suggest that chestnut-crowned babblers communicate using a phoneme-like system with implications for early forms of word generation.

Fundamental to a phonemic communication system is the use of independent and meaningless sounds in different arrangements to generate new meaning. That A and B elements were commonly produced alone in both nest and aviary contexts strongly suggests that the elements are independent units, but raises the question (again) of whether or not the elements are meaningless. Although lone A and particularly B elements were commonly produced during nest visits, in only 15 visits (1.3%) were A elements recorded without also recording a flight or prompt call whilst in only 69 visits (5.9%) was this true of B elements. In addition, such rare instances invariably arose at the beginning and end of recordings at the nest, suggesting full flight calls were given but only A elements were recorded as individuals left the nest and B elements as they arrived. Thus, the production of lone A and B elements is rarely (if ever) independent of the production of flight or prompt calls. Furthermore, we have previously demonstrated that behavioural responses to playbacks of lone A and B elements are comparable in aviary environments, which indicates that neither element carries propositional meaning (Engesser *et al*. 2019). However, these results lead to an obvious question: if neither A nor B elements carries meaning, then why are they commonly produced alone or in aberrant sequences, albeit invariably in conjunction with flight and prompt calls?

Indeed, in both nests and aviaries, all transition possibilities among elements and calls were recorded. There are three non-mutually exclusive possibilities that appear consistent with the majority of the variation. First, and most commonly, lone elements and aberrant sequences appear to occur when calls are initiated but not completed owing to a change of context. This ‘early termination’ hypothesis is the primary reason behind the relatively high frequency of lone A elements uttered in the aviary: birds starting a flight call during movement but landing before the call could be completed due to the restricted distances over which they typically flew. In addition, it explains the rarer instances of isolated A elements at the nest, with birds landing at the nest before flight-call completion. Finally, this hypothesis is also consistent with 1.4% of isolated B elements and 5% of BA sequences, with birds possibly beginning prompt calls but stopping early due to the onset of nestling begging or departure from the nest (N=10076 pair-wise sequences & 168 isolated elements, i.e. those outside of sequences). Second, because element production within sequences but outside of calls was defined by their inter-element time intervals (Figure 1), elements could be produced in call-sequence but with protracted time intervals. However, in contrast with this ‘protracted time-interval’ hypothesis, in aviaries just 1% of occasions lone A elements were followed by a lone B (as expected under element transitions in flight calls) or B elements followed by AB sequences, while BA sequences were never followed by lone B elements (the latter two expected under prompt calls) (N=370 pair-wise element transitions). Similarly, in nests, the corresponding values are 0.6%, 0.9% and 1%, respectively (N=10076 pair-wise transitions). Finally, in both aviaries and particularly nests with the considerably larger numbers of element transitions recorded, almost all conceivable combinations of element transitions were recorded outside of calls at low levels, suggesting that there was often confusion over exactly what call to produce. Although further work is required here, the key point is that these results reinforce the broad hypothesis that sound elements, not calls, represent the basic units in babbler communication, at least in the context of this study.

Whilst this conclusion is perhaps unsurprising, it has a profound bearing on our understanding of when and how phonemic use evolved. Evidently single elements are regularly used as part of chestnut-crowned babblers natural vocalisations. Our analyses demonstrate that these elements are produced not only by all individuals, not being influenced by breeding status, but across different environmental contexts. Despite their common production we suggest and support two theories as to why they are seen at such frequency. Both the “protracted time-interval” hypothesis and “early termination” hypothesis, offer an explanation for their production and continue to support the theory that they’re meaningless entities, as proposed by Engesser *et al*.’s (2015) playback experiments. This level of phonemic-like element reuse in an animal communication system is extremely rare. Steps to draw parallels between human linguistics and animal communication have so far been largely limited to morphosyntax, whereby the units of reuse are larger and hold they own meaning in isolation. Drawing parallels at this finer scale is a step which is only beginning to occur. Although element reuse has begun to be demonstrated in certain species, (in white handed gibbons, (Clarke *et al.*, 2006) chickadees (Moscicki *et al.,* 2010)), so far it is yet to be proven whether these elements are meaningless in their single entities. The babbler communication system is the first to demonstrate that animals may use complex call organisation akin to human phonemes. They should not be considered anomalies in this, and further and deeper study should be considered in other species to grasp the parallels that may exist between animal communication and human language.

CHAPTER TWO

ADHERENCE TO LINGUISTIC LAWS IN THE RUDIMENTARY PHONEMIC COMMUNICATION SYSTEM OF A NON-HUMAN ANIMAL

ABSTRACT

Human language is underpinned by several linguistic rules and for non-human animal communication to be compared to language, these rules must continue to be met. The vocal system of chestnut-crowned babblers, *Pomatostomus ruficeps*, is the first non-human animal system to display rudimentary forms of phonology, where meaningless elements are combined and rearranged to form meaningful sounds. These elements, named the A and B notes, have been shown to be arranged in a non-random order. Further linguistic rules relate to the effects of coarticulation and compression. First, coarticulation refers to the effect neighbouring elements have on a single elements' acoustic parameters, an area of research which has so far received almost no attention in the field of animal communication. Second compression laws characterise the structure of utterances. Specifically, that known as Menzerath's law; a rule stretching across biology stating that information elements decrease in duration as sequence length increases. Our analyses of the chestnut-crowned babblers A and B notes, display support for both coarticulation and compression. A and B notes produced as single elements have lower maximum frequency than notes within a call. Additionally, B notes specifically follow Menzerath's law, with single notes being shorter in duration than those within structured calls. The results of this study demonstrate the importance of a deeper linguistic analysis when it comes to drawing parallels between human language and non-human animal communication. Examining the communication systems of other animals is likely to reveal similar parallels as those suggested in chestnut-crowned babblers, expanding the field of bioacoustics and providing scope for a more holistic conversation surrounding language evolution.

INTRODUCTION

Several parallels have been drawn between animal communication and human language. This can include vocal learning, where individuals are able to obtain new vocalisations (Snowdon & Hausberger, 1997), as observed from passerines (Thorpe, 1958; Marler 1970) to killer whales (Deecke *et al*., 2000). Furthermore, parallels extend to conversational rules, such as turn-taking and overlap avoidance, as observed in common marmosets and Campbell’s monkeys, where conversers alternate in listener and speaker roles to avoid speaking over each other (Chow *et al*., 2015; Lemasson *et al*., 2011). Promising evidence also exists, in the case of animal vocal sequences, where in several cases the structure would appear to parallel that of human language (Arnold & Zuberbuhler, 2006; Coye *et al*., 2015; Engesser *et al*., 2016). However, it is often difficult to assess the depth of these parallels and more specific, targeted research is needed.

Several key rules and principles underpin language. Firstly, language relies on sequences of elements reused and rearranged throughout the sequence. This reuse occurs on multiple levels: phonemes are recombined to produced different words (e.g.bat is b/æ/t, cat is c/æ/t), a system known as phonology. Further, words are recombined to form sentences (e.g. “ship” and “wreck” can be combined to from “shipwreck” which holds a different meaning) according to morphology. Importantly, at each level combination is not random and both the type of elements and their order is key for meaning. Secondly, elements’ acoustic parameters can be influenced by their surrounding elements and calls via coarticulation (Daniloff and Hammarberg, 1973). The interaction of different neighbouring elements results in a “smoothing out” of acoustic features, thus resulting in a continuum of sound, rather than a series of segregated elements (Daniloff and Hammarberg, 1973). Coarticulation relies on the interaction of neighbouring sound elements to influence an element’s own acoustic properties. This interaction is bidirectional (Daniloff and Hammarberg, 1973): an element can be influenced by either its preceding or following element. Coarticulation can occur on a phonological (Fowler & Saltzman, 1993; Ohala, 1993) or morphological level (Cho, 2001; Mousikou *et al*., 2021). Despite the different processing of morphologically simple and morphologically complex words, with simple words being retrieved quicker in speech production, no difference in coarticulation has yet been found (Mousikou *et al*., 2021). Finally, human language is influenced by the compression principle. This principle derived from the information theoretic principle of minimising the length of a code ensures more efficient information translation, compressing information. Menzerath’s Law explains that larger linguistic constructs are made up of smaller components (Menzerath, 1954). For instance, in human language, longer words will be composed of shorter syllables (Altmann, 1980), and in longer sentences the clauses will be shorter (Teupenhayn and Altmann, 1984). This law clearly has a strong role to play in shaping human language; evidence for the same significance in animal communication is much more varied (Watson *et al*., 2020; James *et al*., 2021).

In order to examine animal communication, and especially when drawing parallels to human language, communication must be analysed on increasing levels of complexity. Firstly, are distinct elements present, and are they reused and rearranged in a language-like capacity? Element reuse across calls has been demonstrated in several species, such as female Diana monkeys (*Cercopithecus diana diana*) (Coye *et al.,* 2016), southern pied babblers

((*Turdoides bicolor)* (Engesser *et al.,* 2016) and Campbell's monkeys whose Krak call is reused to create Krak-oo calls by attaching an "oo" suffix to existing calls to make predator alert calls less specific (Ouattara *et al*., 2009; Coye *et al*., 2015). Once element reuse is established, the transitions between these elements must be investigated. Element transitions must be non-random. Previous research of this in animal communication is limited, with non-random element transitions demonstrated in pied babblers *(Turdoides bicolor)* (Engesser *et al.,* 2016) and Mexican chickadees *(Poecile sclateri)* (Moscicki *et al.,* 2010). Moscicki *et al*. examined the transition probabilities in the four-note *chick-a-dee* call. They analysed this firstly by calculating the probability of a certain note type occurring in a specific position in a call. Secondly the probability of a certain note position containing a specific note type was calculated. They demonstrated that the birds consistently produced the four notes of the *chick-a-dee* call in the same order; A - C - D_h - D. Continuing this, to determine whether the system parallels phonology or morphosyntax, the meaning of elements and their relation to the meaning of the sequence must be assessed. If meaningful elements are combined into sequences whose meaning is derived from that of the elements involved, a parallel with morphosyntax will be preferred. This is for instance the case in southern pied babblers (*Turdoides bicolor*) and Campbell's monkeys (*Cercopithecus campbelli*) (Ouattara *et al.,* 2009). On the contrary, if meaningless elements are combined into a meaningful structure, a parallel with phonology will be considered. The only evidence for such a parallel has been found in chestnut-crowned babblers (*Pomatostomus ruficeps)* (Engesser *et al.,* 2015), through the use of a playback experiment, which we will develop below.

The final step is to examine the coarticulation and compression principles. Evidence for coarticulation in animal communication is currently very limited and the field lacking targeted study. However, one study using Japanese quail (*Coturnix coturnix japonica*) trained these birds to peck at a series of human syllable stimuli (Lotto *et al.,* 1997). They showed an ability to distinguish between gradating human syllables and respond to novel syllables, which Lotto *et al*., conclude demonstrates that perceptual compensation for coarticulation is species general. This supports the notion that there is huge scope for further research on this topic especially since shedding light on the presence (or absence) of coarticulation in non-human species may help us understand the depths of the parallels that can be drawn with human language. Empirical evidence of compression law, specifically Menzerath's Law, in animal communication comes primarily from primate species. The first case of Menzerath's Law in a nonhuman species was outlined in geladas (*Theropithecus gelada*), with longer sequences, sequences containing a higher number of calls, being composed of calls which were shorter in duration (Gustison *et al.,* 2016). This negative relationship between sequence length and call duration has since been further demonstrated in other primates. The male long-distance contact calls of two gibbon species, the cao vit gibbon (*Nomascus nasutus*) and western black crested gibbon (*Nomascus concolor*), display this phenomenon (Huang *et al.,* 2020). Additionally, in longer sequences specific note types were seen to shorten, and shorter calls increased in proportion (Huang *et al.,* 2020). Chimpanzees (*Pan troglodytes schweinfurthii*) use distinct pant hoot vocalisations as part of their communication system. Again, as sequence length increases, the duration of calls decreases (Fedurek *et al.,* 2017); following the principles of compression law. Support for Menzerath's Law also comes from other forms of communication. Whole body gestural communication made by chimpanzees displays the same pattern; longer sequences of communication are made up of shorter gestures (Heesen *et al.,* 2019). Limited research into linguistic laws outside of primate species exists. African penguins (*Spheniscus demersus*) have been shown to display a negative correlation between length of call sequence and duration of call constituents, conforming to Menzerath's Law (Favaro *et al.,* 2020). Further support for Menzerath's Law is seen in at least 15 different songbird species (James *et al*., 2021).

Although animal research into linguistic laws has previously been largely limited to primates, chestnut-crowned babblers are providing a new model for these studies. Chestnut-crowned babblers (*Pomatostomus ruficeps*) are small cooperatively breeding passerines endemic to the Australian outback (Russell, 2016). Although their call repertoire is made up of at least 18 distinct calls (Crane *et al.,* 2016), they have two calls which have been a particular focus of interest for the field of linguistics. These calls, which have received the most interest and investigation, are the flight and prompt calls. Importantly, these calls have been shown to reuse the same two notes in different rearrangements, a process likened to phoneme use in human language. The two notes composing them have been labelled the “A” and “B” notes. The flight call is bi-element call, arranging its two notes in an A – B structure. As the name suggest, the flight call is given during or immediately preceding flight. The prompt call is a tri-element call, with the note arrangement B – A – B. The role of this call is less clear, but 90% of prompt calls are given inside the nest (Crane *et al.,* 2016) and it is believed it plays a significant role in the provisioning of chicks. Crucially, previous research suggests both A and B notes hold no meaning and prompt the same behavioural response as a non-natural stimulus (e.g. non-existing C-A-B call). In addition, my work on transition between elements has highlighted that transitions are non-random in this call system, thus further supporting a potential parallel with human phoneme structuring. However as outlined in chapter one, new research demonstrates that single A and B elements are in fact commonly observed as part of the birds' natural vocalisations, despite having no apparent meaning or purpose. While this questions the depth of the parallel to be found with human phonology, it provides an interesting basis to study the effects of both compression law and coarticulation, allowing for a comparison between the multi-element calls and the single element vocalisations given in the same sequences. Examining both these linguistic laws in chestnut-crowned babblers allows us to look further at the possible parallels between human language and animal communication.

I will thus continue to examine the communication system of chestnut-crowned babblers in further depth and specifically investigate sequences of A and B notes, produced both as single elements and as parts of the flight and prompt calls. Here I will evaluate the relevance of both coarticulation and compression law in this communication system. If the vocalisations are influenced by coarticulation I expect a significant difference in the acoustic parameters of notes within calls compared to single elements produced in isolation. Notes within a call unit will be more open to the influence of coarticulation due to the proximity of neighbouring notes. Not only will these provide further insights into the comparison to human language, but it will also broaden a field in animal communication which is currently vastly lacking research. Secondly, compression law, and specifically Menzerath's Law, dictates that the longer the sequence the shorter the comprising components. Therefore, I expect that notes within a call will be shorter in duration than notes produced as single elements. Together these analyses will allow us to further clarify the extent to which parallels between the vocal system of chestnut-crowned babblers and human language can be made.

METHODS

(i) Study Species

Chestnut-crowned babblers (*Pomatostomus ruficeps*) are small (average = 50g), cooperatively breeding passerine endemic to Australia. All the data for this study was gathered from our field site; Fowlers Gap Arid Zone Research Station in New South Wales, Australia, 141°42′E, 31°06′S (Russell, 2016) which covers a 53km^2 area. Fowlers Gap consists of open saltbush and bluebush chenopod shrubland habitat, with annual rainfall <260 mm and breeding season temperature ranging between 0 - 45 ^{o}C (Russell, 2016). During the non-breeding season (January – June) the babblers form large social groups (2-23 individuals, mean = 11, Russell, 2016) which split into smaller breeding groups come the breeding season (July – December). These breeding groups are made up of 2 – 17 individuals (mean=7.5), including one breeding female and 1 – 4 (mean = 1.3) breeding males. The remaining individuals are made up of related and unrelated helpers (Russell 2016). The call repertoire of chestnut-crowned babblers is comprised of at least 18 distinct calls which can be categorised into three groups; threats calls, which alert the groups, social calls, used for both bonding and aggression, and contact calls, used for group cohesion (Crane *et al*., 2016).

(ii) Nest Data Collection

All data for this study was acquired from nest recordings taken during the 2007 and 2008 field seasons, between August and September (Browning *et al*., 2012). During these seasons the study groups comprised of 53 social groups and 88 breeding units (Young *et al*., 2012). Using mist netting procedures, individuals were caught and processed, including fitting: a three wrap-around colour leg-bands, a numbered leg band (Australian Bird and Band Banding Scheme issued), tagged with a 2 x 12 mm passive integrated transporter (PIT tag; TROVAN^{TM} Ltd, UK), blood tests to determine sex (Rollins *et al*., 2012) and immigration status determined from the last three years of data collections. 10 x 15 mm pen cameras were positioned in active breeding nests, which collected both acoustic and visual data. Videos were either continuous and recorded for 100 minutes, regardless of whether adult individuals were present in the nest (n=3) or discontinuous, where the camera was triggered by tagged individuals entering the nest (n=12). The camera then recorded for 1 minute. Entry and exit of each individual bird was recorded using PIT-tag technology (Gibbons and Andrews, 2004). The entrances to the breeding nests were fitted with copper coils. On entry to the nest the birds pass this coil which detects a unique alphanumeric code. The code, with date and time, is recorded on a LID650 decoder (TROVANTM Ltd), connected to the copper coil. The same information is gathered when the bird exits the nest. There is no evidence that the equipment in the nest had any effect on the babblers' behaviour (A. Russell, personal observation).

(iii) Video Coding and Note Extraction

In total 15 videos from 10 different nests were used in my analyses. Calls came from 50 identifiable birds and numerous untagged birds were also present in our data set. These birds were included in our analyses. The pen cameras in the nest recorded the videos in MP3 format, so these were converted to 16-bit amplitude resolution WAV format using VLC

Media Player (Version 3.0.8 Vetinari, Intel 64bit). The videos were also watched on VLC and, in addition to Lucy Browning's data, crosschecked with the PIT tag data. Initially recording the videos on WAV format would have been ideal, rather than subsequently converting their format, but equipment constraints at the time prevented this. However, analysis by Matt Doyle (Doyle, Masters Thesis, 2020), demonstrates that the conversion of MP3 to WAV does not significantly alter the measured frequency parameters in the vocalisations of chestnut-crowned babblers. Praat software (Version 5.3.55 DSP Package) was then used for note extraction and analysis (Boersma and Weenink, 2009). Calls and notes were identified from audio-visual examination of spectrograms (FFT method, window length = 0.4 s, time steps = 1000, frequency steps = 250, Guassian window shape, dynamic range = 60dB). Calls or notes with very high levels of background noise, usually when birds were flying to or away from the nest from a considerable distance from the microphone, were not used. If multiple birds gave overlapping calls where it was not possible to accurately distinguish individual notes, or chicks begged loudly over the visiting adults, these notes were also ignored. In these cases, if one note of a call was considered of too poor quality to be recorded, the entire call would be removed from the dataset. The annotation and tier tool in Praat was used to label vocalisations. Boundaries were placed at the beginning and end of every note and call. All notes which were considered to be either an A or a B note were selected and labelled as either single notes, or as part of a flight call, prompt call or truncated prompt call. The notes were extracted from Praat automatically using a script (produced by Crosswhite and edited by Antoniou (2013) and Coye (2019)). The data set was then put together in Excel (Version 16.39). The extracted note data was matched to PIT tag data using temporal data so individual identity could be assigned to each note and call. The strings of calls and notes were able to be separated into sequences. A new sequence was considered when the inter note interval exceeded 2 seconds.

(iv) Statistical Analyses

In total 8573 notes were extracted, as part of 1565 sequences, from 50 identifiable individuals and unidentifiable (those without PIT tags) individuals were present in every group. Overall, the dataset contained 183 single A notes, 987 single B notes, 3111 flight calls and 3574 prompt calls (Table 1).

GROUP	NO. VIDEOS	NO. INDIVIDUALS	NO. SEQUENCES	A	B	FC	PC	Truncated PC
CEB	1	3	192	9	187	121	217	55
CEJ	1	4	133	45	39	312	133	179
CEK	4	13	479	64	279	1283	1141	221
EFG	2	6	198	29	157	317	332	47
EFP	1	10	243	10	172	685	1140	28
HSB	1	3	139	9	36	164	359	56
HSF	1	4	31	1	13	14	20	37
OFQ	2	3	50	3	15	130	106	8
SFG	1	2	64	8	53	64	88	86
SFN	1	2	36	5	36	21	38	1

***Table 2.1: Summary of notes, calls and sequences.** This table summaries our sample size, giving an overview of the number of individuals per group, how many vocal sequences we analysed per group and the number of each call that was recorded for each of them. All groups all contained unidentifiable individuals.*

All statistical analyses were carried out in R (Version 3.5.3) and R Studio (Version 1.1.463) (R Core Team, 2020). All figures were produced in R Studio. To test for the presence of coarticulation between single A and B notes and A and B note within calls, I used linear mixed models (LMMs) (all used the lmer function of the lme4 package (version 1.1.23) (Bates et al., 2015)). The models included the note position (as either single note or note within call) as the fixed effects and individual ID was included as the random factor. Three parameters were tested in six separate models, one testing for the effect of the previous note and one for the effect from the following note: start frequency, end frequency and maximum frequency. All parameters were scaled in the model using the scale() function. To test for compression, Mann-Whitney tests using the wilcox.test function from the dplyr package in R were used (version 3.6.2). Separate Mann-Whitney tests were carried out for A elements and B elements.

RESULTS

(i) Coarticulation

The results confirm that coarticulation does occur in both A and B notes, consistent with my hypothesis that single elements will be under less coarticulation constraints than elements within calls (Figure 1). Maximum frequency was increased in single B elements in comparison to B notes in calls which were preceded by an A note (t-value = 3.227, df = 1731.757, p-value = 0.00128). This A-B arrangement could include flight calls or the second half of a prompt calls. Similarly, maximum frequency was increased in single A notes compared to A notes which were preceded by a B (t-value = 2.039, df = 370.30021, p-value = 0.0422). This B-A arrangement would be made up predominantly by the first half of prompt calls, and less commonly by reverse flight calls. No significant coarticulation effect was seen on start or end frequencies.

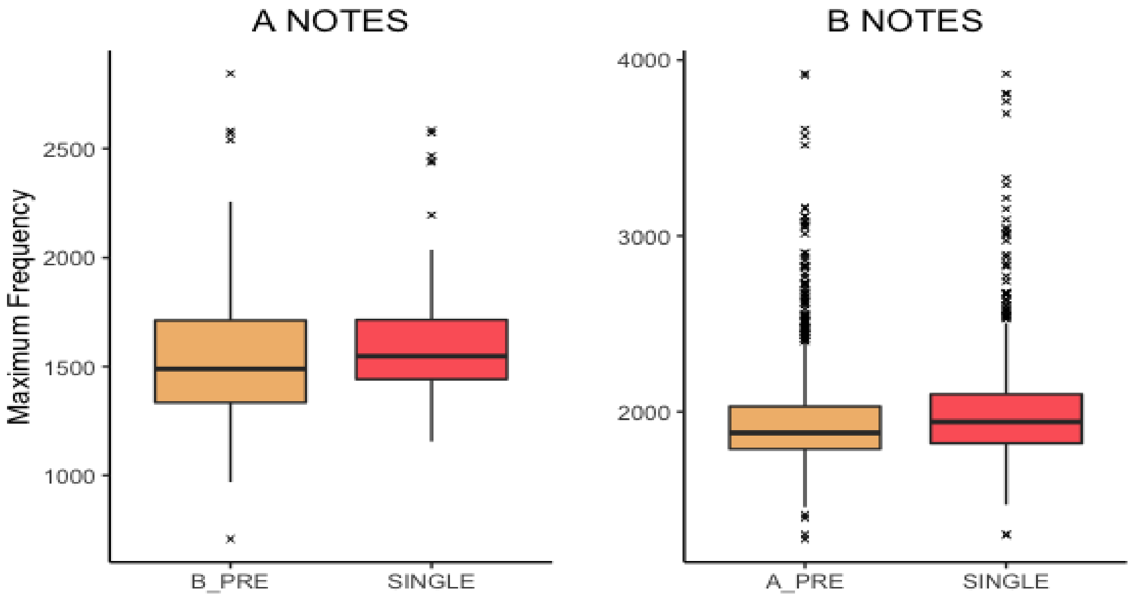


Figure 2.1: Coarticulation effects of A and B notes, a comparison of single notes and notes within calls. *Box plots show the maximum frequency of single elements ("SINGLE") and elements within calls (either "B_PRE" or "A_PRE". "B_PRE" indicates that the A note being studied is preceded by a B note, "A_PRE" indicates that the B note being studied is preceded by an A note). The mean maximum frequency is denoted by the middle horizontal line in the box and the upper and lower inter-quartile ranges (IQR) are shown by the top and bottom of the box. The top and bottom the line illustrates the IQR x 1.5. Any additional data point, considered as outliers, are plotted as crosses.*

(ii) Compression

Compression law would appear to be more complicated (Figure 2). It is supported in B notes, with single B elements being longer in duration than B notes that are produced within a call (W = 1820700, n = 4290 p-value = 0.00000000000696). On the contrary, although A notes show no significance, they trend in the opposite direction, with single A elements being shorter in duration that A notes produced within a call (W = 282514, n = 3114, p-value = 0.8526).

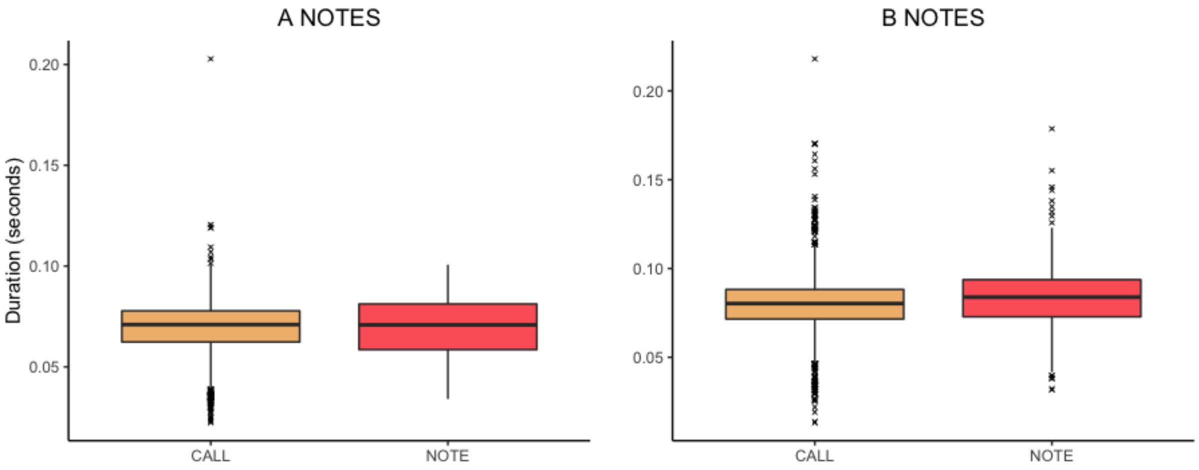


Figure 2.2: Compression effects on A and B notes, a comparison of single notes and notes within calls. *Box plots show the duration of single elements ("NOTE") and elements within calls ("CALL"). The mean duration is denoted by the middle horizontal line in the box and the upper and lower inter-quartile ranges (IQR) are shown by the top and bottom of the box. The top and bottom the line illustrates the IQR x 1.5. Any additional data point, considered as outliers, are plotted as crosses.*

DISCUSSION

Chestnut-crowned babblers have become an excellent study species for researching animal communication and its parallels to human language. Coarticulation and the compression principle are key components of human language and must be considered when looking for these parallels. Here I demonstrate that both these phenomena exist to varying extents within the A-B note sequences.

I found evidence for coarticulation in the acoustic structure of both A and B notes. Specifically, the maximum frequency is increased in notes preceded by another element in comparison to single elements. The same effect is seen in both A and B notes. However maximum frequency is not significantly affected by following elements, and the start and end frequency parameters are not affected by surrounding elements at all. Both note types, A and B, give the same coarticulation response suggesting some form of articulatory constraints may be acting on the calls, which are independent of the specific notes. The findings of these coarticulation analyses must be carefully considered in future chestnut-crowned babbler communication studies. Previous playback experiments used A and B notes from both flight and prompt calls, rearranging them as needed to create artificial sequences. The origin of the notes and their position within a call was not taken into consideration when arranging and using these artificial sequences as this new coarticulation response was previously unknown. However as maximum frequency appears to vary depending on the position of the note within a call, this is a crucial finding which can be used to guide the interpretation of future playback studies. Coarticulation is present on notes from inside calls, compared to single element notes, prompting questions as to whether or not the birds can perceive these slight acoustic differences. This opens up a very interesting avenue for further research; habituation-dishabituation experiments could be used to further investigate the perception of single elements versus call elements and consequently provide deeper insights into previous playback experiment results.

Examining A and B notes and their surrounding elements for the implications of the compression principle presents conflicting results. From my analyses it is apparent that in B notes, single elements are longer than elements within calls. This aligns with the compression principle, and specifically Menzerath's Law which states that as sequences, sentences or words get longer, the elements or words that make up these sequences get shorter (Altmann, 1980; Teupenhayn & Altmann, 1984). Conversely in A elements, a slight trend in the opposite direction was observed, single elements are marginally shorter than elements within calls. This result does not align with the compression principle, but it needs to be considered carefully as the sample size for A elements was far smaller than for B elements and as the trend observed was not significant. Interestingly, the results obtained here align with a recent finding in the chestnut-crowned babbler; other calls in their repertoire have been examined in relation to the compression principle (Coye, Townsend & Russell, In Prep). Through this study calls were broadly grouped into 'social calls' and 'alarm calls' and found to behave differently. Firstly, alarm calls are not compressed as strongly as social calls. Furthermore, within social calls there is variation in compression response, with some notes following Menzerath's Law but others not. It is suggested that in combined calls, distinct elements serve distinct functions, perhaps to reach receivers at differing distances or convey identity or emotional state, and thus these elements may be under different selection pressures (Coye, Townsend & Russell, In Prep).

Previous research into the compression principle in animal communication has found mixed evidence. Not all calls and signals appear to behave in the same way. In duetting primates, specifically tarsiers, titi monkeys and gibbons, it was uncovered that Menzerath's Law should be considered the exception not the rule (Clink & Lau, 2020), with many call types not exhibiting compression principle at all. Clink & Lau suggests that the ability of calls to be able to travel over long distances outweighs the benefits of compression in these species and thus the principle is abandoned. Interestingly further study of gibbon communication by Clink and Lau (2020) indicates that although their communication system largely follows Menzerath's Law, whereby longer sequences are made up of shorter component, it fails to follow the second compression principle Zipf's Law, where the shorter the signal the more frequently it occurs. This highlights the interplaying selection pressures which may be acting, and the variation that can occur not only between different species but within the same species as well.

As the field of animal communication continues to grow and expand, it becomes more important that we follow rigorous steps in drawing parallels with human language. Human linguistic research does not suffer from the same constraints often encountered in the field of animal communication. Feedback from human participants can provide much greater clarity on syntaxical meaning, a method which is clearly impossible in animal research. Further, any potential propagation relations problems can be overcome by ensuring participants are at set distances from microphones and speakers, a task made much more difficult by fast moving animal species. It is important that this comparison is completed on several levels. The most basic level begins with comparing frequency parameters and confirming that non-random transitions between elements occurs. I have already confirmed these non-random transitions exist in the A-B note sequences of chestnut-crowned babblers, in my first chapter. Furthermore, previous research demonstrates that A and B elements within the chestnut-crowned babblers' repertoire are reused and rearranged in multiple calls (Engesser 2015). The single elements have been confirmed to be meaningless when produced in isolation and prompt no behavioural response, a very rudimentary parallel to human phoneme use. In this chapter I outline the parallels to human languages on a further level, both through coarticulation and the compression principle. Perhaps as the field develops, animal communication and human language should not be assumed to be completely separate entities and there could be more to learn by drawing and investigating the parallels between the two. As demonstrated by my own results, the compression principle in animal communication appears more ambiguous than in humans and there is huge scope for further research into why this may be. The results of coarticulation are, at least in the case of chestnut-crowned babblers, a little clearer. The presence of coarticulation suggests that notes are dependent on each other within a call, they do not exist as completely separate entities. Similarly, to human language this coarticulation may allow the sequence of notes to blend into one and other, rather than existing just as sequential signals.

This study acts as an excellent prompt for similar research in other species as chestnut-crowned babblers will unlikely be the only species to follow these communication principles. Other social species may be the best candidates for further research, particularly if they rely on vocal communication with limited visual stimuli to disambiguate communication signals. Alternatively, another interesting avenue could be looking for similarities in the communication systems of the other members of the Pomatostomidae family, all of which are poorly studied. These four species (chestnut-crowned babblers, white crowned babblers *P. superciliosus*, Hall's babblers, *P. halli*, and grey crowned babblers, *P. temporalis*) are the closest relatives of the chestnut-crowned babbler and may be a good basis for further research. For instance, exploring the extent to which their communication systems differ and the likely selective pressures at play (eg. environment, social structure) could shed light onto the evolutionary path of combinatorial communication in the chestnut-crowned babbler. As the field of animal communication continues to grow it is becoming apparent that human language should not be excluded from the conversation. The same principles that govern our own language would appear to hold a place in many other species as well.

CONCLUSION

The continuous theory of language supports the school of thought that language evolution must follow the rules of natural selection and thus has the capacity to be seen across a range of taxa as similar pre-existing traits are co-opted for the same purpose (Fedurek and Slocombe, 2011). Support for this comes from the presence of vocal learning and turn-taking in both humans and certain animal species (Chow *et al*., 2015; Lemasson *et al*., 2005). As we examine the parallels between human language and animal communication on a deeper linguistic level, the similarities become sparser. Rudimentary morphosyntax has been demonstrated in both birds (Engesser *et al*., 2016) and primates (Coye *et al*., 2015), but the only parallels to human phonology have been found in chestnut-crowned babblers. I have built on the previous research identifying the use of meaningless elements (Engesser *et al*., 2015, 2019), going further to describe their occurrence and the linguistic laws that are at play in their communication system.

Evidently, chestnut-crowned babblers regularly emit single A and B notes as part of their normal vocalisations. This is component of their repertoire which has previously received little attention. Previous research identified A and B notes as part of call complexes and demonstrated their importance as meaningless elements (Engesser et al., 2015). My analysis demonstrates that this single note production is not limited to a nest setting, nor is it skewed by vocalisations from a few specific individuals; it is a common and widespread phenomenon throughout our study groups. The sheer quantity of single A and B production had caused us to question the meaninglessness of these two single notes. However, we suggest two hypotheses to explain the production of single elements, both of which continue to support the idea of rudimentary phoneme use proposed by Engesser et al. (2015). Firstly, the ‘early termination hypothesis’ suggests that birds begin a call and cut it short, thus single B notes are produced when a prompt call is initiated and then abandoned, and the same for A notes and flight calls. This hypothesis also explains the presence of truncated prompt calls, where the birds have produced two out of three notes in the three-note prompt call before termination. Secondly, we propose the ‘protracted time-interval’ hypothesis. This explains that the inter note interval between the A and B notes has become so long to no longer be considered as part of the same call complex. This was a constraint we placed on the defining of calls and single elements and further research needs to be done to establish whether or not the vocalisations can still prompt the same behavioural response with such disjointed sound units. Crucially these two hypotheses do not take away from the fact that the A and B elements remain meaningless when produced in isolation and so can continued to be consideration as parallels to human phonemes, as the birds use them as building blocks for their larger, more complex call structures.

We continue to clarify these parallels to human language by demonstrating that both A and B notes are subject to linguistic laws: coarticulation and compression. Currently the research into coarticulation in animal communication is incredibly limited, this study highlights the relevance it may have, and it should be an important consideration in future studies. The compression analyses present mixed findings, with B notes abiding to Menzerath’s law and A notes ignoring it. Although several studies have found support for Menzerath’s law (primarily in primates, (Huang *et al*., 2020; Gustison *et al*., 2016), but also in penguins (Favaro *et al*., 2020)) there is also evidence that it may not be a universal law amongst animal species (Clink & Lau, 2020). If compression is acting differently on distinct notes, it prompts the question as to why? The distinct notes within a call may be subject to distinct selection pressures if they have distinct functions. For instance, the A and B notes may function to reach audiences at different distances or may respectively convey information about caller's identity and emotional state. Further research investigating the propagation of distinct notes in the babbler repertoire, their potential to convey caller's identity or emotional state would be useful to clarify this question. Further, we can wonder why the two A and B notes are used in both flight and prompt call when each note could have been used as a single call instead (e.g. A during flight and B to prompt begging). Further research would have to investigate but it is quite possible that the frequency gap could help with call discrimination. Nevertheless, the presence of these linguistic law, in any form, continues to highlight the similarities to language and emphasises the importance of further study going forward.

Our concentration of Menzerath's law in our compression analysis paves the way for further analysis focusing on Zipf's law. Zipf's law dictates that more frequent calls are shorter in duration and would be an interesting avenue to investigate in the babbler repertoire. It has been shown to be present in several animal species including gibbons (Huang *et al*., 2020), penguins, (Favaro *et al*., 2020), Formosan macaques (Semple *et al*., 2010) and bats (Luo *et al*., 2013). Further research in the chestnut-crowned babblers could also extend to several of their other calls, such as the conflict call or the maternal contact call. Both of these are multi element calls and the same playback experiments and analyses could be carried out as in the A-B calls to establish whether we have similar rudimentary phonemic structure occurring. Our findings highlight the need for both deeper and wider study into the parallels of human language and animal communication. Following the continuous theory of language evolution there will be more candidates for the comparison to human phonology. Currently white handed gibbons and gorillas have shown promising evidence (Clarke *et al*. 2006, Hedwig *et al*., 2015) but more in-depth research is needed to clarify the extent of their parallels.

It is clear that human language and animal communication may not be the quantitatively different entities that were once traditionally assumed, and our research highlights the comparisons that can be made. We have outlined the clear steps that can be followed to allow for these parallels to be drawn. Not only are there further avenues and unanswered questions to explore in the babbler's extensive acoustic repertoire, we have highlighted the great scope for further research in other species who may well possess the same level of complexity we have demonstrated here. The field of animal communication is only expanding and the parallels to human language displayed in the chestnut-crowned babbler's communication lays an exciting corner stone for broadening our knowledge of language evolution.

APPENDICES

Appendices 1: Call counts extracted from nest videos.

This table summaries our dataset, giving an overview of the number of individuals per group and the number of each call we recorded for each of them. Note, the "UNTAGGED" count for each group may account for more than one individual. per group. There is also the small possibility that it could be the same individual visiting different groups, we cannot rule it out due to the lack of PIT tag. FC stands for flight call and PC for prompt call

GROUP	No. Individuals	A	B	FC	PC	Truncated PC	TOTAL
CEB	**3**	**14**	**285**	**114**	**98**	**40**	**551**
0006730B5E		2	25	22	17	5	71
00067377AE		4	36	28	15	11	94
0006739D71		0	103	27	32	12	174
UNTAGGED		8	121	37	34	12	212
CEJ	**4**	**74**	**74**	**250**	**86**	**148**	**632**
0006736851		1	6	40	22	0	69
000672F58F		10	35	48	46	78	217
000682D8D0		36	12	59	5	31	143
0006B8A11E		3	0	18	1	1	23
UNTAGGED		24	21	85	12	38	180
CEK	**13**	**91**	**393**	**1033**	**533**	**167**	**2217**
0006731536		2	13	19	3	18	55
0006736851		1	9	64	15	3	92
000682E450		13	74	168	51	19	325
000672F725		6	47	66	56	14	189
000672F9C2		3	35	27	75	5	145
0006730E6C		3	41	62	46	3	155

00067322F1		12	9	99	10	14	144
00067373FF		0	9	30	28	2	69
000682D8D0		13	10	14	2	19	58
0006B88336		3	11	72	11	4	101
0006B8986F		3	2	65	11	0	81
0006B89D85		1	12	19	24	9	65
0006B8A69C		4	4	35	18	6	67
UNTAGGED		27	117	293	183	51	671
EFG	**6**	**38**	**240**	**229**	**150**	**29**	**686**
000699FF16		4	16	18	8	5	51
00069A0A46		7	60	43	65	12	187
00069A1228		0	38	39	7	1	85
00069A17DB		0	4	28	14	3	49
00069A1FC5		10	35	15	10	3	73
00069A27F9		3	2	3	0	0	8
UNTAGGED		14	85	83	46	5	233
EFP	**10**	**10**	**175**	**340**	**390**	**14**	**929**
00067323B6		1	4	19	13	0	37
000699FA4F		1	23	20	24	0	68
000699FE7D		0	3	35	20	1	59
000699FF62		0	3	10	16	0	29
00069A00F9		0	8	27	39	1	75
00069A2211		1	33	58	43	0	135
00069A2FC4		0	1	1	4	0	6
0006B88540		1	2	28	31	2	64
0006B8A5B7		1	13	27	40	3	84
0006B8A8A4		2	2	26	4	3	37
UNTAGGED		3	83	89	156	4	335
HSB	**3**	**12**	**57**	**112**	**154**	**36**	**371**
000682C76D		5	33	25	44	8	115
000682E1BA		3	7	31	39	19	99

00069A2816		0	2	23	51	3	79
UNTAGGED		4	15	33	20	6	78
HSF	**4**	**2**	**20**	**19**	**9**	**21**	**71**
000699F771		0	7	8	3	8	26
00069A0B8C		1	6	3	1	4	15
00069A273D		0	1	3	2	0	6
00069A2F76		0	4	5	2	6	17
UNTAGGED		1	2	0	1	3	7
OFQ	**3**	**6**	**52**	**179**	**94**	**9**	**340**
000699F81C		0	5	43	17	6	71
00069A0D34		0	11	24	23	1	59
00069A15E2		1	10	31	10	1	53
UNTAGGED		5	26	81	44	1	157
SFG	**2**	**10**	**57**	**52**	**34**	**47**	**200**
00069A18D0		1	1	2	1	3	8
00069A3110		2	20	7	9	12	50
UNTAGGED		7	36	43	24	32	142
SFN	**2**	**7**	**39**	**19**	**15**	**2**	**82**
00069A1901		4	10	3	1	0	18
00069A2A16		3	24	3	10	1	41
UNTAGGED		0	5	13	4	1	23
TOTALS	**50**	**264**	**1392**	**2346**	**1562**	**513**	**6077**

REFERENCES

Altmann, G. 1980. “Prolegomena to Menzerath's law”. In *Glottometrika 2*, Edited by: Grotjahn, R. 1–10. Bochum: Studienverlag Dr. N. Brockmeyer.

Andrieu, J., Penny, S.G., Bouchet, H., Malaivijitnond, S., Reichard, U.H. and Zuberbühler, K. 2020. White-handed gibbons discriminate context-specific song compositions. *PeerJ*, **8**, e9477.

Arnold, K. and Zuberbühler, K. 2006. Semantic combinations in primate calls. *Nature*, **441**(7091), pp.303-303.

Arnold, K., & Zuberbühler, K. 2006. Language evolution: Semantic combinations in primate calls. *Nature*, **441**(7091), e.303. doi:10.1038/441303a

Bates, D., Maechler, M., Bolker, B. and Walker, S. 2015. Fitting Linear Mixed Effects Models Using lme4. *Journal of Statistical Software*, **67**(1), 1-48

Baptista, L.F. and Schuchmann, K.L. 1990. Song learning in the Anna hummingbird (Calypte anna). *Ethology*, **84**(1), pp.15-26.

Boersma, P. and Weenink, D., 2009. Praat: doing phonetics by computer (Version 5.1. 05) [Computer program]. Retrieved May 1, 2009.

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

Browning, L.E., Young, C.M., Savage, J.L., Russell, D.J.F., Barclay, H., Griffith, S.C. and Russell, A.F. 2012. Carer provisioning rules in an obligate cooperative breeder: prey type, size and delivery rate. *Behavioral Ecology and Sociobiology*, **66**(12), pp.1639-1649.

Candiotti, A., Zuberbühler, K. and Lemasson, A., 2012. Convergence and divergence in Diana monkey vocalizations. *Biology Letters*, **8**(3), pp.382-385.

Cho, T., 2001. Effects of morpheme boundaries on intergestural timing: Evidence from Korean. *Phonetica*, **58**(3), pp.129-162.

Chow, C.P., Mitchell, J.F. and Miller, C.T., 2015. Vocal turn-taking in a non-human primate is learned during ontogeny. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1807), e.20150069.

Clarke, E., Reichard, U.H. and Zuberbühler, K., 2006. The syntax and meaning of wild gibbon songs. *PloS one*, **1**(1), e.73.

Clink, D.J. and Lau, A.R., 2020. Adherence to Menzerath's Law is the exception (not the rule) in three duetting primate species. *Royal Society Open Science*, **7**(11), e.201557.

Collier, K., Bickel, B., van Schaik, C.P., Manser, M.B. and Townsend, S.W. 2014. Language evolution: syntax before phonology?. *Proceedings of the Royal Society B: Biological Sciences*, **281**(1788), e.20140263.

Coye, C., Ouattara, K., Zuberbühler, K. and Lemasson, A. 2015. Suffixation influences receivers' behaviour in non-human primates. *Proceedings of the Royal Society B: Biological Sciences*, **282**(1807), e.20150265.

Coye, C., Zuberbühler, K. and Lemasson, A., 2016. Morphologically structured vocalizations in female Diana monkeys. Animal Behaviour, 115, pp.97-105.

Crane, J.M., Savage, J.L. and Russell, A.F. 2016. Diversity and function of vocalisations in the cooperatively breeding Chestnut-crowned Babbler. *Emu*, **116**(3), pp.241-253.

Crosswhite, K. [Year unknown]. Modified by Antoniou, M. (2013) and Coye, C. (2019). Original script available online from www.ling.upenn.edu/courses/Fall_2013/ling520/DurationLogger.praat.

Daniloff, R.G. and Hammarberg, R.E. 1973. On defining coarticulation. *Journal of Phonetics*, **1**(3), pp.239-248.

Deecke, V.B., Ford, J.K. and Spong, P., 2000. Dialect change in resident killer whales: implications for vocal learning and cultural transmission. *Animal Behaviour*, **60**(5), pp.629-638.

Dezecache, G. and Berthet, M., 2018. Working hypotheses on the meaning of general alarm calls. *Animal Behaviour*, **142**, pp.113-118.

Enard, W., Przworski, M., Fisher, S.E., Lai, C.S., Wiebe, V., Kitano, T., Monaco, A.P. and Pääbo, S. 2002 Molecular evolution of FOXP2, a gene involved in speech a language. *Nature,* **418**(69000, pp.869-872.

Engesser, S., Crane, JMS., Savage, JL., Russell, AF. and Townsend, SW., 2015. Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PloS Biology*, **13**(6): e1002171.

Engesser, S., Holub, J.L., O'Neill, L.G., Russell, A.F. and Townsend, S.W., 2019. Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proceedings of the National Academy of Sciences,* **116**(39), pp.19579-19584.

Engesser, S., Ridley, A.R. and Townsend, S.W., 2016. Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences,* **113**(21), pp.5976-5981.

Favaro, L., Gamba, M., Cresta, E., Fumagalli, E., Bandoli, F., Pilenga, C., Isaja, V., Mathevon, N. and Reby, D., 2020. Do penguins' vocal sequences conform to linguistic laws?. *Biology Letters*, **16**(2), e.20190589.

Fedurek, P. and Slocombe, K.E., 2011. Primate vocal communication: a useful tool for understanding human speech and language evolution?. *Human Biology*, **83**(2), pp.153-173.

Fedurek, P., Zuberbühler, K. and Semple, S., 2017. Trade-offs in the production of animal vocal sequences: insights from the structure of wild chimpanzee pant hoots. *Frontiers in Zoology*, **14**(1), pp.1-10.

Fee, M.S. and Scharff, C., 2010. The songbird as a model for the generation and learning of complex sequential behaviors*. ILAR journal*, **51**(4), pp.362-377.

Fowler, C.A. and Saltzman, E., 1993. Coordination and coarticulation in speech production. *Language and speech*, **36**(2-3), pp.171-195.

Gibbons, W.J. and Andrews, K.M., 2004. PIT tagging: simple technology at its best. *Bioscience*, **54**(5), pp.447-454.

Gochfeld, M. (1978). Intraspecific social stimulation and temporal displacement of songs in the lesser skylark. *Zeitschrift für Tierpsychologie*. **48**(4), 337–344.

Gustison, M.L., Semple, S., Ferrer-i-Cancho, R. and Bergman, T.J., 2016. Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, **113**(19), .e2750-e2758.

Hailman, J.P. (1989) The Organization of Major Vocalizations in the Paridae. T*he Wilson Bulletin* **101**(2): pp.305–343.

Hauser, M.D., Chomsky, N. and Fitch, W.T., 2002. The faculty of language: what is it, who has it, and how did it evolve?. *Science*, **298**(5598), pp.1569-1579.

Heesen, R., Hobaiter, C., Ferrer-i-Cancho, R. and Semple, S., 2019. Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B*, **286**(1896), e.20182900.

Henry, L., Craig, A.J.F.K., Lemasson, A. and Hausberger, M., 2016. Social coordination in animal vocal interactions. Is there any evidence of turn-taking? The starling as an animal model. In: Casillas, M., Levinson, SC, (Eds), Turn-Taking in Human Communicative Interaction. Lausanne: Frontiers Media. **4**, pp.39-59.

Hilbrink, E.E., Gattis, M. and Levinson, S.C., 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in psychology*, **6**(1), p.1492.

Huang, M., Ma, H., Ma, C., Garber, P.A. and Fan, P., 2020. Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: insights into the origin of human language. *Animal Behaviour*, **160**(1), pp.145-155.

Hurford J. 2008 The evolution of human communication and language. In: D’Ettore. P. and Hughes, D.P. (Eds), Sociobiology of communication: an interdisciplinary perspective. Oxford, UK: Oxford University Press. pp. 249 – 264.

James, L.S., Mori, C., Wada, K. and Sakata, J.T. 2021. Phylogeny and mechanisms of shared hierarchical patterns in birdsong. *Current Biology*. **31**(13), 2796-2808

Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins (Tursiops truncatus). *Science*, **289**(1), pp.1355–1357.

Janik, V.M. and Slater, P.J. 1997. Vocal learning in mammals. *Advances in the Study of Behaviour*, **26**(1), pp.59-100.

Kershenbaum, A., Blumstein, D.T., Roch, M.A., Akçay, Ç., Backus, G., Bee, M.A., Bohn, K., Cao, Y., Carter, G., Cäsar, C. and Coen, M. 2016. Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biological Reviews*, **91**(1), pp.13-52.

Lambrechts, M.M. and Dhondt, A.A., 1990. A relationship between the composition and size of great tit song repertoires. *Animal Behaviour*, **39**(2), pp.213-218.

Lemasson, A., Glas, L., Barbu, S., Larcroiz, A., Guilloux, M., Remeuf, K. and Koda, H. 2011. Youngsters do not pay attention to conversational rules: is this so for nonhuman primates? *Scientific Reports*, **1**(1), pp.1-4.

Levinson, S. C. (2016). Turn-taking in human communication-origins and implications for language processing. *Trends in Cognitive Sciences,* **20**, pp.6–14

Lotto, A.J., Kluender, K.R. and Holt, L.L. 1997. Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *The Journal of the Acoustical Society of America,* **102**(2), pp.1134-1140.

Manser, M.B., Jansen, D.A., Graw, B., Hollén, L.I., Bousquet, C.A., Furrer, R.D. and Le Roux, A. 2014. Vocal complexity in meerkats and other mongoose species. *Advances in the Study of Behavior*, **46**(1), pp.281-310.

Marler P. 1998 Animal communication and human language. In: Jablonski, N.G. and Aiello, L. (Eds), The origin and diversification of language. San Francisco, CA: California Academy of Sciences. pp. 1 – 20.

Marler, P. 1955. Characteristics of some animal calls. *Nature*. **176**(1) (4470): 6–8.

Marler, P., 1970. A comparative approach to vocal learning: song development in white-crowned sparrows. *Journal of Comparative and Physiological Psychology,* **71**(2, pt. 2), pp.1-25.

Moscicki, M.K., Hoeschele, M. and Sturdy, C.B., 2010. Note types and coding in parid vocalizations: the chick-a-dee call of the Mexican chickadee Poecile sclateri. *Acta Ornithologica,* **45**(2), pp.147-160.

Mousikou, P., Strycharczuk, P., Turk, A. and Scobbie, J.M., 2021. Coarticulation across morpheme boundaries: An ultrasound study of past-tense inflection in Scottish English. *Journal of Phonetics*, **88**, pp.101101

Naguib, M. 1990. Effects of song overlapping and alternating in nocturnally singing nightingales. *Animal Behaviour*. **88**(1), 1061–1067.

Weiss, M., Hultsch, H., Adam, I., Scharff, C. and Kipper, S., 2014. The use of network analysis to study complex animal communication systems: a study on nightingale song. *Proceedings of the Royal Society B: Biological Sciences*, **281**(1785), e.20140460.

Nomano, F.Y., Browning, L.E., Savage, J.L., Rollins, L.A., Griffith, S.C. and Russell, A.F. 2015. Unrelated helpers neither signal contributions nor suffer retribution in chestnut-crowed babblers. *Behavioral Ecology*, **26**(4), pp.986-995.

Ohala, J.J., 1993. Coarticulation and phonology. *Language and speech*, **36**(2-3), pp.155-170.

Ouattara, K., Lemasson, A. and Zuberbühler, K. 2009. Campbell's monkeys use affixation to alter call meaning. PloS *One*, **4**(11), p.e7808.

Pepperberg, I.M., 2010. Vocal learning in Grey parrots: A brief review of perception, production, and cross-species comparisons. *Brain and Language*, **115**(1), pp.81-91.

Pinker, S. and Bloom, P., 1990. Natural language and natural selection. *Behavioral and Brain Sciences*, **13**(4), pp.707-727.

Poole, J.H., Tyack, P.L., Stoeger-Horwath, A.S. and Watwood, S. 2005. Elephants are capable of vocal learning. *Nature*, **434**(7032), pp.455-456.

R Core Team, 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.Rproject.org/.

Rollins, L.A., Browning, L.E., Holleley, C.E., Savage, J.L., Russell, A.F. and Griffith, S.C. 2012. Building genetic networks using relatedness information: a novel approach for the estimation of dispersal and characterization of group structure in social animals. *Molecular Ecology*, **21**(7), pp.1727-1740.

Russell, AF., 2016. Chestnut-crowned babblers: dealing with climate adversity and uncertainty in the Australian arid zone. In: W. Koenig, J. Dickinson (Eds.), Cooperative breeding in vertebrates: Studies in ecology, evolution and behaviour. Cambridge University Press, Cambridge, MA.

Scharff, C. and Petri, J. 2011. Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366**(1574), pp.2124-2140.

Schusterman, R.J. 2008. Vocal learning in mammals with special emphasis on pinnipeds. In: Oller, D.K. and Gribel, U. (Eds). The evolution of communicative flexibility: Complexity, creativity, and adaptability in human and animal communication. MIT Press: Cambridge, MA. pp.41-70.

Scott-Phillips, T.C. 2015. Nonhuman primate communication, pragmatics, and the origin of language. *Current Anthropology,* **56**(1), pp.56-80.

Searcy, W.A. and Marler, P. 1981. A test for responsiveness to song structure and programming in female sparrows. *Science*, **213**(4510), pp.926-928.

Snowdon, C.T. and Hausberger, M. (Eds), 1997. Social influences on vocal development. Cambridge University Press.

Soltis, J., Leong, K., and Savage, A. (2005). African elephant vocal communication I: antiphonal calling behaviour among affiliated females. *Animal Behaviour,* **70**(1), 579–587.

Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J.P., Yoon., K.E. and Levison, S.E. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, **106**(26), pp.10587–10592.

Suzuki, T.N., Wheatcroft, D. and Griesser, M., 2016. Experimental evidence for compositional syntax in bird calls. *Nature communications*, **7**(1), pp.1-7.

Takahashi, D.Y., Fenley, A.R. and Ghazanfar, A.A. 2016. Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **371**(1693), e.20150370.

Teupenhayn, R. and Altmann, G. 1984. Clause length and Menzerath’s law. In: Boy, J. and Köhler, R. (Eds). Glottmetricka 6, Brockmeyer, Bochum. pp.127-138.

Thorpe, W.H., 1958. The learning of song patterns by birds, with especial reference to the song of the chaffinch Fringilla coelebs. *Ibis*, **100**(4), pp.535-570.

Tyack, P.L. 2008. Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *Journal of Comparative Psychology*, **122**(3), p.319.

Varella, T.T. and Ghazanfar, A.A. 2021. Cooperative care and the evolution of the prelinguistic vocal learning. *Developmental Psychobiology*.

Watson, S.K., Heesen, R., Hedwig, D., Robbins, M.M. and Townsend, S.W. 2020. An exploration of Menzerath's law in wild mountain gorilla vocal sequences. *Biology Letters*, **16**(10), e.20200380.

Young, C.M., Browning, L.E., Savage, J.L., Griffith, S.C. and Russell, A.F. 2013. No evidence for deception over allocation to brood care in a cooperative bird, *Behavioural Ecology*, **24**, pp.70-81.