

The role of acoustic templates, contextual information and visual representation in the combinatorial communication system of chestnut-crowned babblers

Submitted by Joseph Giacomo Mine to the University of Exeter
as a thesis for the degree of
MSc by Research in Biological Sciences
In October 2019

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: *Joseph Giacomo Mine*

Abstract

Many non-human animals produce vocalisations via the combination of multiple acoustic elements into larger sequences, a property referred to as combinatoriality. Chestnut crowned babblers (*Pomatostomus ruficeps*) are the only known species other than humans to use phonemic contrasts, i.e. the rearrangement of meaningless acoustic elements into meaningful calls, in the generation of vocal signals. Identifying the similarities and differences in phoneme use between humans and babblers, therefore, can potentially grant valuable insight into the evolution of this linguistic ability in humans. The primary goal of this thesis is to understand the cognitive mechanisms that underlie the reception of vocalisations generated via phonemic contrasts in chestnut crowned babblers. Specifically, I address the following research questions: 1) whether babblers are able to respond appropriately to vocal signals even when the constituent acoustic sub-units have been rearranged; 2) whether individuals are capable of integrating previously acquired contextual information in order to produce appropriate behavioural responses to ambiguous signals; and 3) whether babblers generate a visual mental image of signal referents upon reception of context-specific acoustic signals. I find that subjects respond appropriately to vocal stimuli even when the acoustic sub-units have been rearranged to differ from any existing call template, but only when the modified signal exhibits a high degree of acoustic similarity to the natural call. This suggests that babblers, as opposed to humans, possess flexible acoustic templates for their phoneme-based vocalisations. I also observe that subjects respond appropriately to ambiguous signals if they have been previously exposed to information which may disambiguate the meaning of future, uncertain stimuli. This extends the evidence for the integration of contextual information to an avian species exhibiting a combinatorial repertoire. Finally, I do not find evidence in support of the hypothesis that babblers generate visual mental images of signal referents upon reception of context-specific calls. Thus it remains as of yet unclear whether babblers respond to calls via a process of affect-conditioning, associative learning or conceptual semanticity. The findings of this thesis contribute to the field of comparative and combinatorial communication by describing some of the cognitive mechanisms which underlie the reception of phoneme-based vocalisations in an avian species.

List of Contents

- Title page (1)
- Abstract (2)
- List of Contents (3)
- Author's Declaration (4)
- Text (5-120)
 1. Chapter 1 – Introduction (5-15)
 2. Chapter 2 – Overcoming uncertainty in a non-human phonemic communication system (16-38)
 3. Chapter 3 – An empirical test of the generation of visual mental images upon signal reception in an avian species (39-58)
 4. Chapter 4 – Conclusion (59-66)
- Acknowledgements (67)
- Bibliography (68-81)

Author's Declaration

The experiments outlined in this thesis were conducted between July-September 2016. I did not become involved in this project until September 2018, thus I declare that I did not participate in any field work specifically relevant to this project. The extent of my contribution to this work is the following:

- Coding of 100% of the behavioural data from video recordings of the experiments
- Re-evaluation of the hypotheses under consideration in the experiments
- Statistical analyses of data
- Write-up of project

Chapter 1

Introduction

Study species

The studies described in this thesis were conducted on an avian species, the chestnut crowned babbler (*Pomatostomus ruficeps*). Chestnut crowned babblers are 50g, sexually monomorphic birds. They are a member of the ancient family Pomatostomidae, i.e. Australo-Papuan babblers, of which *P. ruficeps* are one of five extant members (Russell 2015). This species, which is endemic to south-east inland Australia, emerged ~11Ma, at a time when the Australian continent was covered by forests. Thus although the current habitat of this species mainly consists of arid and semi-arid inland scrubs and open woodland, it likely originated as a forest-dwelling species.

The first and only long-term study on chestnut crowned babblers is based at Fowlers Gap Arid Zone Research Station, in far western New South Wales (31° 05' S, 141° 43' E). The study area is 64km², dominated by open chenopod shrubland which also exhibits areas of denser vegetation along numerous dry creek beds. In this region, the annual temperature variation is extreme: from 0°C nocturnal minimums during winter to 45°C daytime maximums during summer. The habitat is classified as arid, as annual rainfall averages 215mm/y, with high annual variation and no seasonal pattern.

Chestnut crowned babblers are highly dependent on group living (Portelli *et al.* 2009). Outside the breeding season these groups, referred to as “social units”, are large, comprising 7-16 (mean 11) individuals, of which ~50% are juveniles born during the previous breeding season. The sex ratio of social groups is on average 64% male. Male babblers remain philopatric for life, while females typically disperse after a maximum of one year. The members of a social group typically forage together, and roost within the same nest. Babblers are weakly territorial (Sorato *et al.* 2015): groups commonly forage over an area of ~1km², but may venture into the territory of neighbouring social units. Between-group encounters result in vocalisation or physical interaction only 55% of the time, and often these interactions are amicable rather than aggressive. Greatest foraging activity occurs in the early morning or late afternoon, and groups spend on average 64% of time foraging. The preferred habitat for foraging is along dry creek beds, which account for 21% of the group’s territory

on average (Portelli *et al.* 2009). This habitat represents the ideal foraging location as it combines maximum prey availability with minimum predation risk. Babblers are not strong flyers, and are most commonly observed exhibiting short bouts of swooping flight between low perches. By contrast, these birds are surprisingly agile when moving on the ground, where they prefer to forage via manipulation of the substrate. Although the members of a social group roost together in a single nest, the territory of the group typically exhibits multiple nests. During nonbreeding, groups regularly alternate between roosting nests and may roost in a different nest each night. Nests used for roosting are actively maintained, as shown by the presence of fresh twigs along the nest opening (Louis O'Neill, pers. obs).

Like all of the other Pomatostomids, chestnut crowned babblers are cooperative breeders, i.e. they rely on the provisioning efforts of helpers outside the breeding pair for successful reproduction (Russell 2015). Prior to the onset of breeding (mid-late July), babbler social units fission into smaller groups, referred to as "breeding units". These groups comprise on average 6 individuals (range 2-13), and typically include 1-8 nonbreeding helpers in addition to the breeding pair. The home ranges of breeding units are $\sim 0.5\text{km}^2$, but may vary as a function of unit size. Babblers make up to 3 breeding attempts per season, the first attempt generally occurring in early August. 50% of breeding units that make one attempt also make a second attempt with a mean interval between attempts of 31 days; third attempts are rare. The mean clutch size is 4 in first and 3.5 in subsequent attempts, with a maximum clutch size of 6. Nestling mortality is high, and on average, 3.5 offspring are reared to fledging per breeding unit throughout the entire breeding season.

Breeding success in chestnut crowned babblers is strongly influenced by the size of the breeding unit (Russell 2010, Liebl *et al.* 2016). Firstly, larger breeding units breed earlier and re-nest sooner than smaller ones, which is likely to be mediated by: a) a reduction in the amount of time that females need to spend off-roost with increasing number of helpers, b) greater direct provisioning of the incubating female, and c) increased provisioning of the offspring, thereby freeing the female from caring for the previous brood. Second, nestling starvation decreases with increasing size of the breeding unit, as a result of fully additive care. On average, each additional helper results in a roughly 35% increase in

chick per capita food intake, despite females showing significant load-lightening. Finally, larger breeding units are less susceptible to nest depredation.

Within a typical babbler breeding group, 70% of non-breeders are first-order relatives to the breeding male, female or both (Russell 2015). The remaining 30% are second-order relatives or unrelated to the breeding pair. Almost all first-order relatives contribute to offspring provisioning, and do so at an average rate 90% higher than second order relatives, who only contribute to offspring provisioning in one-third of cases. This highlights a strong kin-selected component to cooperative breeding in chestnut crowned babblers (Browning *et al.* 2012a,b). Indeed, inclusive fitness benefits play a key role in shaping both the occurrence and patterns of cooperative breeding in this species.

In chestnut crowned babblers, males exhibit lifetime philopatry, while females disperse out of their natal territory to join other babbler groups. The most commonly invoked explanation for delayed dispersal involves ecological constraints, i.e. a lack of available territories or suitable mates for independent breeding (Emlen 1982). In babblers, dispersal patterns reflect neither the availability of suitable territories nor potential mates (Russell 2015), however, babbler philopatry can be explained by understanding the importance of indirect fitness benefits. Only 21% of adult males gain direct fitness in their lifetime, so for the majority of individuals, indirect benefits are the only avenue available for promoting the spread of their genes in the population. In order to reap these indirect benefits, delaying dispersal is of fundamental importance, as it grants access to close relatives who are in a position to breed independently. Therefore, in this system, delayed dispersal represents a consequence, rather than a cause, of cooperative breeding.

The importance of inclusive fitness benefits in babblers also provides an explanation for the presence of fully additive care among nonbreeding helpers (Liebl *et al* 2016). In this species, nestling starvation is the primary form of offspring mortality. However, increases in the number of male helpers are linearly associated with the rates at which whole broods and individual nestlings are provisioned. Together, these results imply that additional care can substantially increase the productivity of breeding attempts. As most group members derive their only fitness benefits from the successful rearing of the offspring of related breeders, it is unsurprising that they should provision maximally in order to guarantee the highest brood productivity, rather than

reduce their contributions with an increasing number of additional helpers. The dynamics of cooperative breeding in arid zones have received comparatively little attention, thus chestnut crowned babblers provide a valuable opportunity to elucidate the selective pressures that operate on cooperative breeders inhabiting harsh environments, such as the Australian arid zone.

Acoustic communication in chestnut crowned babblers

Chestnut crowned babblers possess a rich repertoire of vocalisations (Crane *et al.* 2016). The first thorough analysis of the diversity of the babbler repertoire revealed that this species produces at least 18 distinct calls, 13 of which are specific to a certain context, while the remaining 5 are not reliably associated with a unique function. The call types of babblers reflect their species-specific ecological and behavioural features. The 13 context-specific calls can be assigned to four broad functional categories:

1. General arousal: includes only the chatter call. This short, monosyllabic call exhibits two variants, one harsh and fully broadband, the other harmonically rich and squeaky. The call, which is typically produced in bouts, denotes a combination of mild anxiety, threat or excitement: for example, it is produced reliably upon encountering people, as well as when the group exits the roost nest in the morning.
2. Threat: includes the aerial alarm, alert and distress calls. These are all monosyllabic, often produced in bouts, and among the calls in the repertoire which exhibit the highest amplitude (Joseph Mine, unpublished data). Babblers have several nest and aerial predators, so a multitude of alarm calls associated with various levels of urgency are highly beneficial for this species. Furthermore, the distress call elicits mobbing behaviour in group members, an indication of the highly cooperative tendencies of babblers.
3. Contact: includes the flight, maternal contact, long-distance contact, short-distance contact and *peow pee* calls. These vocalisations are used to mediate a number of within and between-group interactions. The flight, short-distance and long-distance contact calls are produced frequently, and function to maintain group cohesion and limit foraging competition during both static foraging as well as during individual or group movements. The long distance contact and flight calls are multisyllabic,

the first element of each exhibiting a lower frequency than the following acoustic elements. Interestingly, the long-distance contact call exhibits individual-specific variation in acoustic properties, and is used for individual recognition by group members (Crane *et al.* 2015). The *peow pee* is also multisyllabic, and utilised to maintain group cohesion, but produced predominantly by juveniles. The maternal contact call is one of the most complex and variable vocalisations in the repertoire. It is multisyllabic with a variable first element, followed by the repetition of high-amplitude and high-frequency elements among which there is also considerable variation, resulting in a sequence referred to as “piping”. This call is produced exclusively by the breeding female, to communicate personal or nestling hunger and to elicit provisioning efforts by helpers. Females produce this call before breeding, as well as throughout the duration of the breeding attempt.

4. Social: Includes the adult begging, provisioning, chase and conflict calls. Begging and provisioning calls are common among avian cooperative breeders (Otter *et al.* 2007; Ellis *et al.* 2009), and are used by babblers to elicit and coordinate cooperative efforts to provision offspring. The provisioning call is multisyllabic, comprising a total of three acoustic elements with a harmonic structure: the first and last elements are identical and exhibit a higher fundamental frequency than the middle element. The chase and conflict calls, on the other hand, are important for the formation and maintenance of social hierarchies, and are also produced during encounters with other groups. These calls mediate agonistic interactions such as dominance displays during breeding or roosting, or escalated inter-group conflicts.

Chestnut crowned babblers combine large group sizes with a strict dependence on cooperation. Consistent with the hypothesis that vocal complexity reflects social complexity (Freeberg 2006), as well as the hypothesis that cooperative breeding influences the size of avian vocal repertoires (Leighton 2018), the large size of the babbler repertoire reflects its high level of sociality as a species, along with its dependence on cooperation for successful reproduction. Indeed, in the case of babblers, a high proportion of the entire repertoire is devoted to cooperative contexts, such as breeding and alarm calling, highlighting the

fundamental importance of vocal communication in mediating cooperative behaviour in this species. This high level of cooperation has been suggested to play a significant role in the colonisation of harsh environments, such as the arid and unpredictable habitats commonly occupied by chestnut crowned babblers (Cornwallis *et al.* 2017). Considering this interdependence between cooperation and acoustic communication, in addition to the suitability of open habitats to the propagation of acoustic signals (Morton 1975), it is consequently unsurprising that the babbler repertoire has evolved to exhibit such a diversity in form and function of vocalisations.

Apart from its diversity, the babbler repertoire exhibits another characteristic that is indicative of vocal complexity: combinatoriality, i.e. the combination of multiple acoustic elements to create larger acoustic structures. But what determines the uniqueness of babbler sound combinations with respect to any other non-human animal is the use of phonemic contrasts (Engesser *et al.* 2015). This term refers to the rearrangement of specific acoustic elements to generate functionally distinct signals, which is the same principle that underlies the generative power of most human languages. For example, within the babbler repertoire, the flight and provisioning calls utilise the same acoustic elements, but in different arrangements: the flight call is a bisyllabic signal comprising the elements labelled A + B, whereas the provisioning call is a trisyllabic signal made up of B + A + B. Indeed, empirical evidence has been obtained in support of the following statements (Engesser *et al.* 2015, Engesser *et al.* 2019): a) the A and B elements are perceptibly distinct, and neither of these conveys functionally-relevant information when used in isolation; b) the element labelled A in the flight call is statistically and perceptibly equivalent to the A in the provisioning call, as is the element labelled B; c) the flight and provisioning calls are context-specific signals, therefore possessing the potential for reliable information transfer; d) the addition or omission of a single B element at a specific position results in the generation of two functionally distinct signals. The importance of these results is twofold: on one hand, they represent the first demonstration that non-human animals have the capacity to produce phoneme-like contrasts; on the other, they highlight similarities with the mechanism that underlies word formation in human language. For example, the basic principles described above equally apply to the English words “on” and “non”, or the Italian words “la” and “ala”.

Despite a superficial resemblance, significant differences exist between babbling and human phoneme use. The first is that the elements used in human words, such as the sound /n/, are re-used productively across a number of other words, while the elements A and B in babblers are present only in the flight and provisioning calls. Moreover, the use of phonemic contrasts in babblers involves the addition or omission of a single element, and not the complete rearrangement of the units that compose the sequence, such as in the example “tap” and “apt”. This suggests that the phoneme-like elements present in the babbling repertoire represent a simple precursor for a more elaborate use of phoneme-based structures. It is likely that the phonological layer of acoustic combination present in many human languages is the result of gradual increases in complexity (Zuidema & de Boer 2009, 2018), and yet little evidence exists regarding the early stages of language evolution. Understanding the precise form and function of babbling vocalisations, therefore, may grant valuable insight into the emergence and evolution of phoneme structuring. Although the babbling vocal repertoire has been carefully described, there is still considerable uncertainty regarding the contextual correlates, combinatorial rules and computational mechanisms that underpin the production and reception of many of the babbling calls (Crane *et al.* 2016).

Aims and methods

The aim of the studies outlined in this thesis is to investigate the reception of combinatorial acoustic signals in chestnut crowned babblers. Specifically, I wish to obtain further information regarding the combinatorial rules underlying babbling vocalisations, and to understand how these complex signals are interpreted by receivers. I tackle these questions by conducting playback experiments on wild-caught babblers, utilising both natural and artificial acoustic stimuli, and measuring behavioural responses to these signals.

Chestnut crowned babblers currently represent the sole opportunity outside of human language to investigate the use of phonemic contrasts in vocal communication (Engesser *et al.* 2015). In this regard, the first question that is addressed in this study of babbling vocalisations concerns the strictness of the relationship between vocal signals and specific acoustic templates. Many babbling calls are multisyllabic, i.e. composed of multiple consecutive acoustic elements. The arrangement of the individual elements within specific babbling

calls is relatively consistent, however there are also certain calls in the repertoire, such as the maternal contact and flight calls, that exhibit more within and between-individual variation than would be expected if these calls adhered to very strict acoustic templates. For example, the maternal contact call is highly variable in both the temporal and spectral features of the first element of the call, as well as the subsequent repeated “piping” elements. The bisyllabic flight call, which commonly consists of element A followed by element B, is occasionally produced in reverse order, i.e. BA. And yet one of the key features which determines the reliability of information transfer in vocal communication is the strength of association between a specific vocalisation and a single context: when this association is loose, i.e. multiple signals are associated with one context, or one signal is associated with multiple contexts, the potential for information transfer is reduced (Seyfarth & Cheney 2017). So understanding the precise relationship between vocalisation, template and context in babbler calls will provide information regarding the potential for information transfer of these calls. In order to assess the strictness of the relationship between signals and their respective acoustic templates, I conduct a playback experiment involving both natural babbler vocalisations and artificially modified versions of existing babbler calls. I specifically focus on two calls of the babbler repertoire, i.e. the flight and provisioning calls, which are both composed of the acoustic elements labelled A and B. As a result of sharing acoustic elements, these two calls naturally exhibit a certain measure of acoustic similarity. However, babblers have been shown to exhibit functionally different responses to playbacks of flight vs prompt calls (Engesser *et al.* 2015), indicating that these call types, despite their similarity, adhere to two distinct acoustic templates. In this experiment, I expose wild babblers to playback stimuli with increasing levels of acoustic similarity to the provisioning call, in order to determine the threshold similarity that results in a provisioning-related response, and thereby assess the strictness of the relationship between signal and acoustic template.

The second research question addressed in this study regards a cognitive mechanism underlying signal reception known as pragmatic inference (Seyfarth & Cheney 2014a, 2017), which consists in integrating additional contextual information during the reception of an acoustic stimulus. Indeed, among the cognitive mechanisms that increase the efficiency of information transfer in vocal communication, pragmatic inference has been shown to play a significant role in

several mammalian species, including humans (Miller *et al.* 1951). For example, the alarm call repertoires of putty-nosed monkeys (*Cercopithecus nictitans*) contain individual vocalisations that are produced in multiple different contexts, and thereby exhibit a reduced potential for information transfer (Arnold & Zuberbuhler 2013, Fischer *et al.* 2001, Hammerschmidt & Fischer 1998). However, these and other primate species have empirically demonstrated the ability to acquire information from sources other than the signal itself, during a communicative event between sender and receiver (Price & Fischer 2014, Scott-Phillips 2015, Crockford *et al.* 2007). Indeed, subjects are able to extract additional information, in the form of both visual and acoustic stimuli, from the context in which the signal is emitted, and utilise this information to refine behavioural responses to signals exhibiting low information content. Therefore, pragmatic inference represents a powerful tool to overcome the challenges posed by signals associated with low information content. However, the investigation of this cognitive ability has mostly been limited to mammalian species. In the past, pragmatic inference has commonly been tested via playback experiments in which the availability of contextual information is manipulated, and behavioural responses to playbacks with and without contextual information are measured (Arnold & Zuberbuhler 2013, Price & Fischer 2014). I apply this same experimental design to chestnut crowned babblers, utilising contextual information in the form of additional acoustic stimuli. Specifically, I expose subjects to artificial playback stimuli resembling erroneous provisioning calls, both with and without prior exposure to natural provisioning call stimuli. By comparing responses to artificial playbacks experienced after exposure to the provisioning call with responses to the same playbacks experienced without a priming stimulus, I evaluate the ability of babblers to integrate information obtained during past communicative events when responding to a novel acoustic stimulus, thereby extending the search for pragmatic inference to avian species.

The final research question addressed in this thesis regards another cognitive mechanism that potentially underlies acoustic communication in chestnut crowned babblers: the visual mental representation of signal referents (Suzuki 2018). When processing an acoustic stimulus, subjects may either attend only to the physical features of the call or, alternatively, they may also attend to the meaning of the signal as well as its acoustic properties

(Zuberbuhler *et al.* 1999). In the first scenario, referred to as perceptual semanticity, behavioural responses result from the direct effect of the acoustic properties of the stimulus on the internal state of the receiver (Owren & Rendall 2001), while in the second, known as conceptual semanticity, responses are mediated by a mental representation of the referent of the acoustic signal (Seyfarth *et al.* 2010). My goal is to determine whether babbler communication represents a case of perceptual or conceptual semanticity. Conceptual semanticity provides a way to increase the robustness of information transfer between individuals, as it involves the detailed internal representation of features of the external environment (Premack 1984). However, it is a complex cognitive mechanism which in humans and other primates requires interactions between specialised areas of the brain (Pan & Sakagami 2012, Logothetis *et al.* 1995). To date, only two other studies have attempted to find evidence for the visual or conceptual mental representation of signal referents in avian species, and obtained positive results (Evans & Evans 2007, Suzuki 2018), suggesting that this ability may not be limited to vocal communication in primates. Chestnut crowned babblers are highly dependent on acoustic communication to mediate the crucial cooperative behaviours that enable their survival and reproduction in a harsh environment. Furthermore, their vocal repertoire is complex in terms of both diversity and combinatorial features. Therefore, I consider babblers to be a prime candidate for the investigation of mental representation of signal referents, as the selective pressure to increase the robustness of acoustically-mediated information transfer is likely to be high in this species.

Tests of conceptual semanticity in the scientific literature are scarce, due to the inherent difficulty of investigating the mental processes of non-human animals. However, a novel methodology for the empirical investigation of visual mental images evoked by acoustic signals has been recently developed by Suzuki (2018). This design involves playback of specific acoustic stimuli combined with physical model referents of target acoustic signals. By observing the subject's attention to specific visual stimuli in response to playback, Suzuki (2018) proposes that the presence of visual mental images of signal referents can be inferred. In this experiment, implementing additional experimental features including those suggested by Bond (2019), I utilise the following playback approach: I expose subjects to playbacks of provisioning calls, which are reliably associated with babbler nests, and give them the choice to interact

with nests that differ in their component visual features. I predict that if stimuli do not evoke visual mental images of nests, receivers should not preferentially interact with nests that differ in their component visual features. However, if reception of the provisioning call evokes the visual mental image of a nest, I predict that babblers should be primed to detect a nest that matches the component visual features established by the mental image.

The questions addressed in this thesis may generate important discoveries in the field of comparative vocal communication. First, it is not yet known whether birds possess strict or loose internal templates for their context-specific calls, due to a lack of species that exhibit phonemic contrasts and a paucity of studies that examine responses to imperfect calls. Second, the ability to integrate contextual information upon reception of an acoustic stimulus is a powerful cognitive mechanism that underlies human and primate vocal communication, yet this ability has been comparatively under-investigated in avian species to date. Finally, the mental representation of signal referents is a hotly debated topic in the field of animal communication, with exponents of both perceptual and conceptual semanticity offering valid arguments in defence of their claims (Seyfarth *et al.* 2010, Rendall *et al.* 2009); yet a consistent base of empirical evidence for either view is still lacking. Thus, the studies described in this thesis represent a valuable opportunity to provide novel experimental evidence concerning all three of these important questions in the field of acoustic communication.

Chapter 2

Overcoming uncertainty in a non-human phonemic communication system

Summary

The production of combinatorial vocalisations, wherein multiple acoustic elements are combined to create larger structures, is an example of vocal complexity exhibited by a number of mammals and birds. However, in only one non-human species to date has empirical evidence for phonemic contrasts, i.e. the recombination of meaningless acoustic elements to produce meaningful signals, been reported: the chestnut-crowned babbler (*Pomatostomus ruficeps*). Therefore, babblers currently represent the only opportunity to compare and contrast the cognitive abilities associated with a phoneme-like communication system, and thus possibly gain insight into the evolution of this particular trait in humans. For example, such a combinatorial design may require specialised cognitive mechanisms for overcoming the ambiguity inherent in complex combinatorial signals. In this study I examine two cognitive mechanisms that may aid this species in processing vocalisations generated via phonemic contrasts. Specifically, I test for a) the ability to interpret erroneous phoneme-based sequences via the assessment of acoustic similarity between signal and template, and b) the ability to integrate additional, contextual information upon reception of such erroneous sequences. Using a playback experiment involving both natural and artificial acoustic stimuli, I find empirical support for the presence of both of these cognitive abilities in this species.

Introduction

Communication, defined as the production of a costly signal by a sender to influence the behaviour of a receiver (Seyfarth & Cheney 2003), is probably a near-universal feature of animals. Throughout the diversity of life, communication occurs in a variety of sensory modalities: among these, chemical signalling is believed to be the most ancient, as it occurs even in prokaryotic bacteria (Mashburn & Whiteley 2005, Waters & Bassler 2005). More derived forms include visual displays, such as the waggle dances of honey bees (*Apis mellifera*) or the aerial courtship acrobatics of Anna's hummingbirds (*Calypte anna*) (Gruter & Farina 2009, Clark 2009), as well as acoustic signals, including the complex songs of passerine birds, or the low-frequency rumbles of elephants

(*Loxodonta africana*), used to communicate with conspecifics up to 10km away (Garstang 2004). These signals, whether acoustic, visual, chemical or other, share one fundamental outcome: they reduce the receiver's uncertainty regarding features of the environment (Kight *et al.* 2013). They accomplish this by shifting the probability of occurrence of certain events in predictable directions. Indeed, when a signal associated with a particular context or event occurs, the statistical probability of the event it is associated with increases, while the probability of contrasting events diminishes. In this sense, signals transmit probability-related information from sender to receiver (Seyfarth *et al.* 2010, Wiley 2013, Fischer 2013).

In social species, increased pressure on communication drives the evolution of a broader range of signals, and in the case of acoustic communication, of larger vocal repertoires (Leighton 2017). However, as repertoire size increases, the likelihood of exhibiting calls that are acoustically similar is also greater, especially when combinatorial structures are used, i.e. composite vocalisations generated via the combination of smaller acoustic elements (Nowak & Krakauer 1999, Zuidema & de Boer 2009, 2018). Therefore, all species that exhibit large or combinatorial repertoires must overcome ambiguity associated with acoustic stimuli. To do so, receivers may depend on specialised cognitive mechanisms to aid in signal reception. For example, receivers might: a) possess a very strict acoustic template for each call, responding appropriately only if the call is a correct match to the template; or b) exhibit a graded response, which is increasingly appropriate the closer the call is to a given template. Following recent studies on human speech perception, it emerges that fine-grain similarity between a perceived stimulus and a target word is required to identify the picture of the target word (Toscano *et al.* 2013, Gregg *et al.* 2019). This suggests that in humans, where the productive combination of phonemes results in boundless generative power (Fitch 2018), acoustic templates are necessarily strict. Conversely, whether animals rely on strict or loose templates is not yet known, due to the scarcity of playback experiments that assess the response to imperfect stimuli. Thus the first question addressed in this study regards the importance of strict vs loose acoustic templates in the reception of combinatorial vocalisations in an avian species.

In order to evaluate the ability of animals to process stimuli that are a sub-optimal match to the corresponding acoustic template, it is necessary to

artificially modify the organisation of the acoustic sub-elements that make up the playback stimulus. This is a difficult task when a) vocalisations belonging to a specific call type consist of a single vocal emission, or b) the sub-elements of a single call type are not clearly defined. Therefore, to tackle these questions, I have focused on the only known non-human species whose calls are generated utilising phoneme-like elements (Engesser *et al.* 2015), i.e. acoustic items that are not used in isolation but combined with others to produce context-specific calls (Coye *et al.* 2017): the chestnut-crowned babbler (*Pomatostomus ruficeps*). This cooperative breeder of the Australian outback exhibits a unique vocal repertoire wherein distinct calls are seemingly produced via the re-arrangement of specific acoustic sub-units. Two such calls are the flight and nest provisioning calls: the flight call is typically produced during aerial movement, while the provisioning call is produced when entering or leaving the nest for the provisioning of nestlings; moreover, playback experiments have shown that subjects respond to the former by scanning the surroundings and to the latter by orienting towards an available babbler nest (Engesser *et al.* 2015). Both the flight and provisioning call comprise the acoustic elements labelled A and B: where the flight call consists only of element A followed by element B, the nest provisioning call contains B, A and B again. Indeed, Engesser *et al.* (2015) showed that the presence/absence of a single B element generates functionally distinct behavioural responses to playback, thus constituting unique empirical evidence for phonemic contrasts in a non-human animal. Here, by rearranging the A and B elements into new sequences and exposing subjects to both natural and aberrant versions of the provisioning call, I test whether note sequence alteration results in a complete loss of functional response, or if instead the modified provisioning calls still elicit nest-related behavioural responses. The latter outcome would suggest the ability to respond to coarse-grain similarity to a call template (Gregg *et al.* 2019), which may be advantageous in the context of complex and error-prone phonemic constructs.

Another well-established mechanism used by both humans and non-human animals to disambiguate the meaning of uncertain calls is the integration of contextual information (Norris *et al.* 2003, Eisner & McQueen 2005, Arnold & Zuberbuhler 2013, Fischer *et al.* 2001). This ability, known as pragmatic inference (Cheney & Seyfarth 2014a, 2016), is based on the acquisition of information from sources other than the signal itself, during a communicative

event between sender and receiver (Price & Fischer 2014, Scott-Phillips 2015, Crockford et al. 2007). Any event that is prior to or concurrent with the signal, whose association with the signal is recognised, may contribute to refining the receiver's behavioural response – the most direct proxy for 'meaning' in studies on animal communication (Marler 1961, Smith 1980, Wiley 2013) – particularly when the stimulus itself is ambiguous (Fischer *et al.* 2001). In articulation tests on humans, where receivers are tasked with identifying and writing down the acoustic test stimuli to which they are exposed, words are more often transcribed correctly if they are heard in a sentence than if they are heard in isolation (Miller *et al.* 1951). Furthermore, the receiver's previously acquired knowledge of the vocabulary from which the test stimuli are chosen influences his/her ability to identify words correctly. For example, if the receiver is told in advance that he/she will be hearing a digit from 1 to 10, then the test stimulus "four" is more frequently recognised than if it were presented without previous notification.

Non-human animals are also capable of integrating information originating from sources other than the acoustic stimuli themselves, to guide behavioural responses in the appropriate direction (Seyfarth & Cheney 2017). Many primate species exhibit calls that are used in multiple different contexts, and thus have low context specificity and are relatively ambiguous (Meise *et al.* 2010, Price *et al.* 2015). It has been shown that when exposed to a priming stimulus, such as an eagle shriek or the sound of a falling branch, followed by a multi-purpose alarm call, subjects exhibit more appropriate behavioural responses than if the call is presented alone (Arnold & Zuberbuhler 2013, Price & Fischer 2014). In addition, previously acquired knowledge concerning the social bonds between group members has been shown to influence subjects' responses to playbacks simulating social interactions in baboons (*Papio ursinus*) (Crockford *et al.* 2007). Although the investigation of pragmatic inference has involved a number of mammalian species (e.g. Miller *et al.* 1951, Norris *et al.* 2003, Cheney & Seyfarth 1988, Price & Fischer 2014, Townsend *et al.* 2011), it has received comparatively little attention with regards to other taxa (Wheeler & Fischer 2012). Furthermore, the experimental designs implemented to research this cognitive ability range from direct to indirect, and differ between mammals and birds. To my knowledge, only two avian studies have obtained results suggesting the ability to combine the information from signal and context. In one case, male song sparrows (*Melospiza melodia*) were exposed to intrusive vs non-intrusive

playback stimuli from neighbouring males, and later exhibited variation in aggressiveness to further playbacks from those same males, consistent with the conditional retaliation hypothesis (Akçay *et al.* 2009, 2010). On the other hand, a study on fowl (*Gallus gallus domesticus*) by Evans & Evans (2006) was designed to test whether acoustic signals in birds are associated with mental representations of signal referents (Arnold & Zuberbühler 1999, Townsend & Manser 2012). It emerged that chickens which had been recently familiarised with the presence of food later exhibited a weaker response to playback of chicken food calls; the authors hence concluded that specific external referents such as food are represented mentally upon signal reception. Thus although the investigation of pragmatic inference was not stated among the questions addressed in the paper, the results of Evans & Evans (2006) also reveal that signalling in birds is influenced by previously acquired information. As a consequence, previous studies of pragmatic inference in birds are limited by both: a) the absence of replication and b) variation in explicit research aims, and thus a deeper understanding of this issue in comparative cognition is lacking (Fischer 2013). Furthermore, the relevance of this cognitive mechanism to forms of combinatorial acoustic communication has not yet been fully appreciated, as most studies have focused on species in which acoustic signalling is constrained rather than flexible (e.g. Zuberbühler *et al.* 2013).

Thus the second aim of this study is to extend to an avian species, which exhibits a complex combinatorial repertoire, the important research questions addressed mainly by linguists and primatologists regarding the ability to integrate contextual information upon reception of acoustic stimuli (e.g. Norris *et al.* 2003, Arnold & Zuberbühler 2013, Price & Fischer 2014). To do so, I conform to the experimental procedure outlined by Arnold & Zuberbühler (2013) and Price & Fischer (2014), in which subjects are exposed to a target playback stimulus (e.g. general-purpose alarm call) either with or without previous exposure to a priming stimulus (e.g. predator model or vocalisation). As the priming stimulus provides additional contextual information which is related to the target stimulus, the use of pragmatic inference predicts that behavioural responses to the target stimulus should be more appropriate in the primed vs non-primed condition. In this study, I utilise the correct provisioning call sequence (i.e. BAB) as the priming stimulus, and measure responses to erroneous sequences, which constitute the target stimuli (e.g. AAB), both with and without prior exposure to the priming stimulus.

Chestnut-crowned babblers have been observed to produce signals in which the arrangement of acoustic elements differs from that of the most common templates. For example, babblers occasionally combine the A and B elements into aberrant sequences, such as reverse flight calls “BA” (Crane *et al.* 2016) or unusual 3-element sequences, such as “AAB” and “ABA” (Joseph Mine, unpublished data). Furthermore, babblers frequently produce calls in bouts, such that multiple flight or provisioning calls may follow one another in rapid succession, making it possible for the sequence “unambiguous call – ambiguous call” to occur in the wild. Thus I ask the following question: if subjects are first exposed to the correct combination of provisioning call elements, and subsequently played the aberrant version, do they respond to the ambiguous signal differently than if it were played alone, i.e. without the priming stimulus? I predict that subjects that are previously exposed to the correct provisioning call should be more likely to interact with an available babbler nest upon reception of a subsequent ambiguous provisioning call stimulus than those which are not primed with the correct call. If so, this would suggest that the challenge of processing ambiguous acoustic stimuli may be overcome by referring to unambiguous stimuli heard previously (Davis *et al.* 2005). I propose that this capacity would be particularly important for a species with such a complex combinatorial repertoire, where errors in call production are likely to occur.

Methods

Study site and species

The study was conducted during August to November 2016 at the Fowlers Gap Arid Zone Research Station in far western New South Wales, Australia (141°42'E, 31°06'S). Chestnut-crowned babblers are 50g, sexually monomorphic passerine birds endemic to south-eastern Australia (Russell 2015). This cooperatively breeding species lives in groups of up to 23 individuals (mean 10), and is highly vocal, with each adult possessing a repertoire of at least 18 functionally distinct calls (Crane *et al.* 2016). The repertoire of this species is one of only two known cases in the animal kingdom (including humans) in which phonemic contrasts are involved in the production of complex vocalisations (Engesser *et al.* 2015). The calls of the babbler repertoire are used in various contexts including social, threat, contact and arousal (Crane *et al.* 2016). The acoustic signals under investigation in this study are the flight call and the

provisioning call, both of which are composed of the two acoustically distinct elements labelled A and B, in the order AB and BAB, respectively (Engesser *et al.* 2015).

Test-subject Selection and Housing

Birds were captured using mist-nets in creek beds and transported in bird bags no more than 5km away by vehicle to onsite aviaries. No more than 4 group members were captured at a time, and no more than 50% of the group was removed. Juveniles and breeding females were not taken for experiments. All birds were released back to their groups within 48 hours of capture. The aviaries consisted of four single compartments each of 2 m long, 2.5 m deep and 2 m high with standardised artificial perches on natural substrate with *ad lib* water. Each aviary compartment also contained two previously used babbler nests in which the birds could roost overnight, as is typical of this species. The nests were extracted from territories within the field site known to contain babbler groups, and were selected for their qualities by experienced team members. For the purposes of a further experiment described elsewhere (see Chapter 3), two nests exhibiting different visual features were chosen: one nest was completely intact, of the highest standard of quality (large size, curved dome, protruding neck, small opening) and recently maintained, while the other was older and exhibited signs of natural degradation such as a flattened dome, smaller neck and larger entrance. Babbler nests are highly durable structures, and may retain functionality in the wild for several years (Louis O'Neill, *pers. obs.*), thus it was not possible to determine the precise age of the nests. 100% of birds roosted within one of the available nests. The two sides of the aviary were opaque metal, the front consisted of one-way Perspex allowing observers to record behaviour, while the back side of the aviary exhibited a metal mesh of 1cm² allowing test subjects a view to the outside. During the day, each bird was provided 20 mealworms every 2-3 hours. When more than one bird was removed from a group, these birds were housed in the same compartment overnight.

Experimental protocol: playback stimuli and video recordings

The calls used in the playback experiments were obtained from recordings of 6 groups of wild chestnut-crowned babblers. Vocalising individuals were adult male and female babblers of at least 1 year of age. Flight calls were produced during

flights to and from the nest, while provisioning calls were produced while provisioning chicks within the nest. In each case, calls were recorded using a Sennheiser directional microphone (ME66/K6) connected to a Marantz solid-state recorder (PMD660, sampling frequency 48 KHz, 24 bits), positioned within 1 m of a nest. Playbacks, including the construction of artificial calls, were created with Adobe Audition CS6 (Version 6 Build 732, Adobe Systems). From each of the six groups recorded, utilising only high-quality recordings (high signal-to-noise ratio, no obscuring vocalisations, low background noise), multiple sets of seven playback stimuli were created. Each set included a natural flight call (AB), a natural prompt call (BAB), a reverse flight call (BA), three triple-element combinations of A and B notes (AAB, ABA, BAA), and a control stimulus consisting in the call of a heterospecific species (spiny-cheeked honeyeater). When acoustic elements for the generation of artificial calls were added and/or replaced, it was ensured that inter-element distance and amplitude matched the original call. A new call-set was played in randomized order for each subject or pair of subjects, and birds never received a call-set from their own group. In 13 cases, two individuals of the same group were placed in separate, non-adjacent aviary compartments and tested simultaneously with the same playback-set. Subjects tested in pairs did not have visual access to each other, which reduced the confounding effect of the presence of another individual during testing. This procedure was implemented in order to increase the sample size of the experiment within the available time frame.

Playbacks were broadcast at an amplitude of 50 dB at 2 meters, matched to the natural amplitude of flight and provisioning calls as measured with a Castle GA206 sound level meter. The speaker was placed immediately outside the aviary compartment(s) containing the test subject(s). During each playback, a stimulus was repeated six times over a total of 10 s; a break of at least 10 min was given for focal individuals to resume pre-stimuli behaviour before the initiation of another playback sequence. Playback experiments were conducted on the day following capture.

During testing, individuals were recorded using digital Sony handycams (HDR-CX220 and HDR-CX160). Visual recordings of 4 s from playback onset were analysed frame by frame using Adobe Audition CC (Version 6 Build 732, Adobe Systems). The behavioural parameters recorded were time (s) spent in camera view (mean = 3.12 s, range = 0.99-4s), looking at the nest, looking

outside, flying or hopping, looking other. Following Engesser et al. (2015), gaze direction was evaluated as orientation of the beak following a head movement. Behavioural observations began with the first head movement following the first acoustic element of the playback, which occurred within a mean of 0.5s from the beginning of the playback. Marker lists created in Adobe Audition were extracted into txt-files by using CueListTool (Version 1.7). Scoring was conducted blind to experimental condition.

The experiment was conducted on 70 adult chestnut-crowned babblers. Some of the individuals were removed from statistical analyses due to confounding variables during the experiment: for example, in 2 cases, an aerial predator was perched within view of the aviary, which resulted in the test subject remaining entirely motionless throughout the test.

Experimental protocol: Rationale

The primary goal of this experiment was to rearrange the A and B elements found uniquely in the flight and provisioning call of the babbler repertoire into new combinations, in order to examine the cognitive mechanisms associated with signal reception. Specifically, I tested for: a) the ability to respond to imperfect stimuli based on acoustic similarity to a natural template, and b) the ability to integrate contextual information upon reception of ambiguous calls.

Each bird was exposed to a series of 7 unique playback treatments with a break of 10 min between treatments. The order of the treatment stimuli was randomised. 6 out of 7 treatments consisted in playing back two- and three-element combinations of the A and B elements typical of the flight and provisioning calls. Of these, 2 treatments exhibited naturally-occurring combinations of elements, i.e. AB (flight call) and BAB (provisioning call). In the other four treatments, attempting to span the variation of potential similarity to the provisioning call, the elements were rearranged to produce the following aberrant combinations: BA, ABA, AAB, BAA. The remaining treatment was a control, where the subject was played the call of a sympatric species (spiny-cheeked honeyeater, *Acanthagenys rufogularis*). Thus the 6 experimental stimuli (AB, BA, ABA, AAB, BAA, BAB) exhibited variation in two key parameters: a) the total number of A and B elements (i.e. 2 or 3), and b) the position of these elements relative to the correct BAB sequence. I predicted that 3-element combinations, and combinations where the position of the A and B elements

resembled that of the correct sequence more closely, would be perceived as more similar to the correct sequence than 2-element combinations and those where the order of A and B elements differed more substantially from the correct sequence (see below for a quantification of acoustic similarity between playback stimuli).

The provisioning call of the babbler repertoire is produced exclusively during interactions with the nest, and in addition, it has been shown to elicit nest-looking behaviour following playback (Crane *et al.* 2016, Engesser *et al.* 2015). Therefore the key data that was collected regarded the probability that babblers performed nest-looking behaviour and the duration of this behaviour, in response to different playback stimuli. As mentioned above, the experimental aviaries contained two previously used babbler nests exhibiting variation in their component visual features. Despite this variation, babblers interacted visually with both nests during testing, thus nest-looking response was measured irrespective of nest quality for this experiment.

Based on previous studies on human speech perception (Davis *et al.* 2005), I assumed that exposure to the correct provisioning call combination (i.e. BAB) prior to reception of an incorrect combination (treatments BA, ABA, AAB and BAA) would provide contextual information that could be used to disambiguate the meaning of the uncertain signal heard subsequently. However, I predicted that this effect should not be present when the correct BAB combination was followed by an unambiguous combination, i.e. the flight call (AB). Thus I first divided the non-BAB experimental stimuli into two categories: 1) ambiguous stimuli (BA, ABA, AAB and BAA), i.e. where the elements had been reorganised to differ from any existing call template; and 2) unambiguous stimulus (AB), where modifications to natural calls had not been made. Via the randomisation of treatment order, I ensured that some birds were exposed to the correct provisioning call treatment (BAB) early within the playback set, such that all subsequent treatments were categorised as post-priming, and some late, such that all previous treatments were effectively not primed with the correct call combination. This allowed me to compare the response to ambiguous and unambiguous signals heard before and after the focal signal, i.e. BAB, and thus to evaluate whether exposure to the correct provisioning call combination could be used as contextual information to disambiguate the meaning of future, uncertain signals.

During playback experiments, a common risk is that birds may habituate to playback, decreasing their behavioural responses as more playbacks are presented during the experiment (Engesser *et al.* 2015). I tested for habituation by assessing the relationship between behavioural response and playback number. There was no significant effect of treatment number on the probability or duration of time spent looking at the nest (GLMM, $\chi^2_1 = 2.49$, $p = 0.1$; GLMM, $\chi^2_7 = 6.87$, $p = 0.4$).

Statistical Analyses

Statistical analyses were conducted using the software programme Rstudio (Engesser *et al.* 2017, Ihaka & Gentleman 1996). The packages used were: boot, car, tidyverse, ggplot2, tidyr, dplyr, broom, lme4. Analyses of behavioural responses from the playback experiments were conducted using Generalized Linear Mixed Models (GLMM), some of which were carried out within a two-step hurdle model, and chi-square tests for given probabilities. The choice of a hurdle model, in which the behavioural response is first assessed qualitatively and then quantitatively, was made necessary by the characteristics of the dataset, wherein 68% of entries for the behaviour of interest (looking at the nest) were 0.

The first set of analyses regarded the ability to respond to imperfect stimuli based on acoustic similarity to a natural template, i.e. the provisioning call. To model the differences in behavioural responses among the 7 playback treatments I used GLMMs executed within two-step hurdle models. The first step of the hurdle model consisted in determining the probability of looking at the nest as a function of treatment. The occurrence of nest-looking behaviour, coded as 0 (did not look at nest) and 1 (looked at nest) represented the response term, fitted to a binomial error structure with logit link function. Time spent in camera view was fitted as the binomial denominator, treatment type (CTR, AB, BA, ABA, BAA, AAB, BAB) was fitted as a seven-level factor, treatment number was fitted as an additional explanatory term, and individual identity nested within group identity were fitted as random terms. The second step of the hurdle model assessed the differences in nest-gaze duration (s) among playback treatments and restricted its scope to all the non-zero entries in the dataset. The duration of nest-looking behaviour represented the response term, fitted to a binomial error structure with logit link function. Time spent in camera view was fitted as the binomial denominator, treatment type (CTR, BA, AB, ABA, BAA, AAB, BAB) was fitted as

a seven-level factor, treatment number was fitted as an additional explanatory term, and individual identity nested within group identity were fitted as random terms.

Once the differences among treatments in the probability and duration of nest-looking behaviour had been modelled, I assessed how these differences were affected by variation in acoustic similarity to the BAB call template. Classic methods for evaluating acoustic similarity, such as spectrographic cross-correlation (SPCC), were deemed unsuited to the experimental design given that all playback stimuli (excluding the control treatment CTR) consisted of rearranged versions of two acoustic elements (i.e. A and B), the spectral and temporal features of which were maintained constant across all experimental stimuli. The playback stimuli nonetheless differed acoustically in two key parameters: 1) the total number of A and B elements in the sequence, and 2) the number of elements in the correct position relative to the BAB sequence. As both of these measures could be directly compared to the correct provisioning call sequence, I computed a similarity index for each non-BAB playback treatment which accounted for the variation in these two parameters, relative to the BAB sequence. Thus each non-BAB treatment was first assigned a score for both a) total number of A and B elements relative to the number of A and B elements in the BAB sequence (range = 0-1, mean = 0.72); and b) number of elements in correct position relative to BAB (range = 0-0.66, mean = 0.44). Finally, a composite index was generated for each treatment by calculating the mean of the two individual scores (range = 0-0.83, mean = 0.58). Using these composite similarity scores, I then examined the effect of acoustic similarity to the BAB sequence on the qualitative nest response (i.e. probability of looking at the nest) via a further GLMM. The occurrence of nest-looking behaviour, coded as 0 (did not look at nest) and 1 (looked at nest) represented the response term, fitted to a binomial error structure with logit link function. Time spent in camera view was fitted as the binomial denominator, acoustic similarity to the BAB sequence was fitted as the primary explanatory term, treatment number was fitted as an additional explanatory term, and individual identity nested within group identity were fitted as random terms.

The second and final set of analyses regarded the integration of contextual information upon reception of acoustic stimuli. To model the variation in nest response as a function of variation in exposure to contextual information I

used GLMMs. Separate GLMMs were carried out for responses to ambiguous stimuli, unambiguous stimuli and control stimuli. In all cases, the occurrence of nest-looking behaviour, coded as 0 (did not look at nest) and 1 (looked at nest) represented the response term, fitted to a binomial error structure with logit link function. Position of the BAB playback stimulus within the playback sequence was fitted as a binomial denominator, exposure to the BAB sequence (Before vs After) was fitted as a two-level factor, and individual identity nested within group identity were fitted as random terms.

Results

Effect of call structure

Overall, 32% of the 258 playbacks resulted in birds looking at the nest. For those individuals that looked at the nest, the average duration of nest looks was 0.41s (SD = 0.22, range = 0.078s - 1.22s).

The probability of looking at a nest was significantly affected by treatment type (GLMM, $\chi^2_6 = 16.54$, $p = 0.01$) (Fig. 1A). The control stimulus, as predicted, elicited the lowest nest response, with only 13% of subjects looking at the nest during such playbacks. The AB and BA stimuli showed a mean 11.5% increase in nest response compared to the control, while the incorrect 3-element combinations (i.e. ABA, AAB and BAA) showed a mean 22% increase in nest attentiveness. Finally, the stimulus where the elements of the provisioning call were played in the correct order (i.e. BAB) evoked the highest qualitative nest response, with a 37% increase compared to the control. The BAB treatment was fitted as the reference level for all subsequent analyses: overall, the probability of looking at a nest was significantly greater in the BAB treatment compared to Control, AB and BA (GLMM, effect size (ES) = -2.08, SE = 0.61, z value₂₅₇ = -3.28, $p = 0.0006$ [CTR]; ES = -1.19, SE = 0.53, z value₂₅₇ = -2.4, $p = 0.02$ [AB]; ES = -1.15, SE = 0.53, z value₂₅₇ = -2.13, $p = 0.03$ [BA]) but not compared to ABA, AAB and BAA stimuli (GLMM, ES = -0.84, SE = 0.51, z value₂₅₇ = -1.65, $p = 0.09$ [ABA]; ES = -0.54, SE = 0.5, z value₂₅₇ = 1.08, $p = 0.2$ [AAB]; ES = -0.63, SE = 0.49, z value₂₅₇ = -1.27, $p = 0.2$ [BAA]). These results suggest a graded response, such that playbacks are more likely to induce nest looking behaviour when they are acoustically more similar to the correct provisioning call sequence. In support of this suggestion, acoustic similarity to the provisioning call, quantified via a composite similarity index as described in Methods section, had

a significant positive effect on the probability of looking at the nest (GLMM, $\chi^2_1 = 6.87$, $p = 0.008$): a greater probability of looking at the nest was associated with higher acoustic similarity to the BAB sequence (Fig 1B).

In contrast to the probability of looking at the nest, there was no effect of call type on the duration of time spent looking at the nest (GLMM, $\chi^2_6 = 3.54$, $p = 0.7$) (Fig. 1C). This result may have been influenced by the small sample sizes of certain treatments (e.g. Control), during which very few birds looked at the nest (total = 4). However, the effect of treatment type on duration of time spent looking was non-significant even when the treatments with fewest nest looks were removed from the analysis (GLMM, $\chi^2_3 = 0.69$, $p = 0.8$). These results suggest that a gradation in similarity to the correct call induces the nest looking response, but does not affect the duration of looking behaviour.

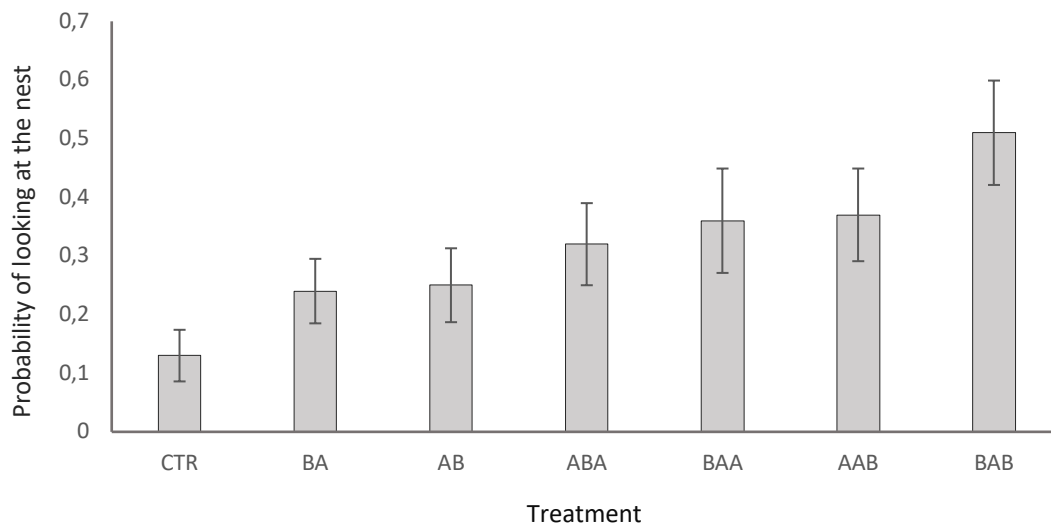


Fig 1A. The probability of exhibiting the behaviour of functional relevance (i.e. looking at the nest) differed significantly among treatment types. Figure 1A shows the probability for each treatment type (\pm Standard Error [SE]), generated from a Generalized Linear Mixed Model (GLMM), in which the occurrence of nest-looking behaviour was fitted as the independent response term. The response term was fitted to a binomial error structure with logit link function, time spent in camera view was fitted as the binomial denominator, treatment type (CTR, AB, BA, ABA, BAA, AAB, BAB) was fitted as a seven-level factor, treatment number was fitted as an additional explanatory term, and individual identity nested within group identity were fitted as random terms. CTR represents the control treatment, wherein the call of a different bird species was played. The

other treatments are labelled according to the combination of acoustic elements specific to each treatment, i.e. AB treatment consisted of elements A + B, BA consisted of elements B + A, etc. As predicted, the control treatment exhibited the lowest probability of looking at the nest, while the correct provisioning call combination (i.e. BAB treatment) elicited the highest probability of looking at the nest.

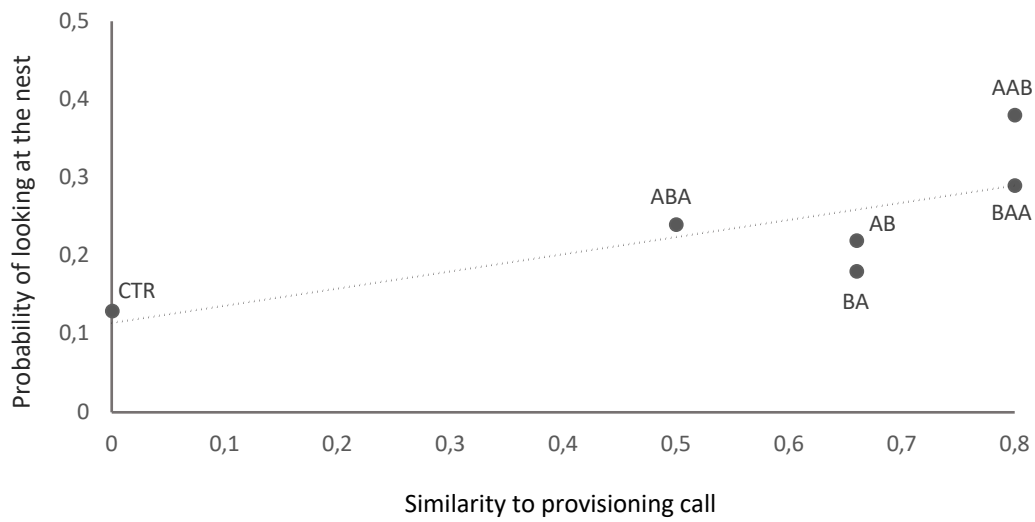


Fig 1B. The probability of looking at the nest for each playback treatment increases with greater acoustic similarity to the provisioning call, i.e. BAB. The similarity to the provisioning call for each treatment type was calculated utilising a composite similarity index generated according to two specific criteria: a) total number of A and B elements in the call relative to the number of A and B elements in the provisioning call, b) number of A and B elements in correct position relative to the provisioning call.

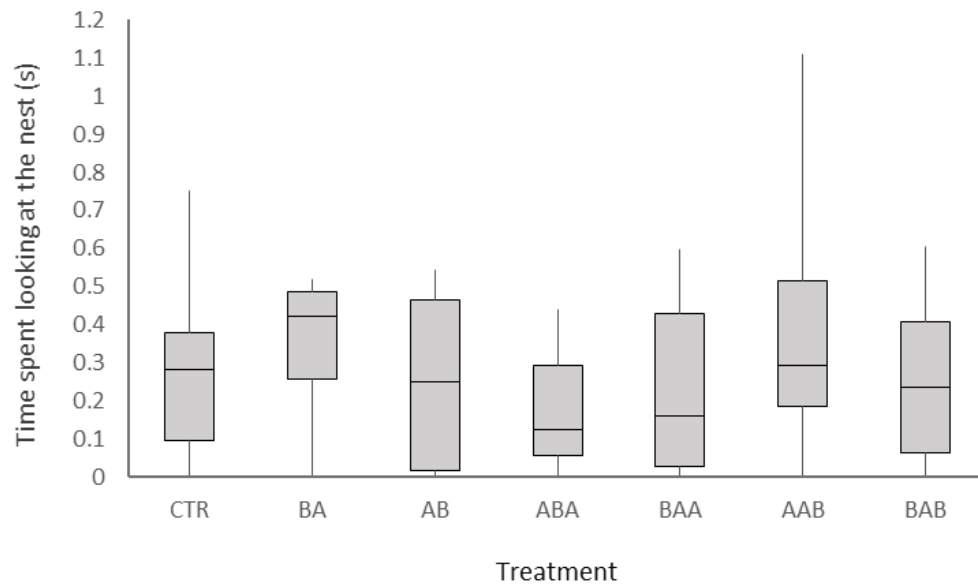


Fig 1C. The duration of nest looks did not differ significantly among treatments. Figure shows means and quartiles for each treatment type, generated from a Generalized Linear Mixed Model (GLMM), in which the duration of nest-looking behaviour was fitted as the independent response term. The response term was fitted to a binomial error structure with logit link function, time spent in camera view was fitted as the binomial denominator, treatment type (CTR, BA, AB, ABA, BAA, AAB, BAB) was fitted as a seven-level factor, treatment number was fitted as an additional explanatory variable, and individual identity nested within group identity were fitted as random terms.

Effect of context

Birds might differentially perceive the reorganised playbacks as provisioning calls in the artificial arena depending on whether or not they have already heard the provisioning call. The hypothesis that prior playbacks of BAB have a priming effect would be upheld if birds responded differentially to ambiguous treatments but not unambiguous treatments after previously hearing the correct call. In accordance with these predictions, prior exposure to the BAB treatment was associated with a significant increase in the likelihood of looking at the nest during subsequent ambiguous stimuli (GLMM, $\chi^2 = 6.024$, $p = 0.01$). Furthermore, exposure to the BAB treatment did not significantly influence the probability of looking at the nest during subsequent unambiguous stimuli (GLMM, $\chi^2 = 6e-04$, $p = 0.9$) and during subsequent control stimuli (GLMM, $\chi^2 = 3e-04$, $p = 0.9$) (Fig. 2).

At this point, it must be noted that these results have potential confounding implications on the previous analyses regarding the ability to respond to imperfect signals based on perceived acoustic similarity. Indeed, the finding that birds demonstrated significantly greater nest-looking behaviour in response to certain treatments which exhibited increased acoustic similarity to the provisioning call (e.g. AAB, BAA), could have potentially been confounded by the fact that these treatments were experienced after exposure to the correct sequence of provisioning call elements, as described above. Therefore, to control for the confounding effect of previous exposure to the BAB sequence, I re-ran all statistical analyses assessing the variation in nest-looking response as a function of playback treatment and acoustic similarity according to the following procedure: all unambiguous playback treatments (i.e. AB) and the control treatment (i.e. CTR), which had not shown to be significantly influenced by previous exposure to the BAB sequence, were maintained without modification; while all ambiguous playback treatments (i.e. BA, ABA, AAB, BAA) experienced after exposure to the BAB sequence were removed from the new analyses, leaving only the cases in which ambiguous stimuli were experienced prior to exposure to the BAB sequence. These further tests confirmed the previous findings regarding the response to imperfect stimuli: a) the probability of looking at a nest was still significantly affected by treatment type (GLMM, $\chi^2_6 = 14.7$, $p = 0.02$); and b) treatments with greater acoustic similarity, as measured by the composite similarity index described above, were still associated with a significantly greater probability of looking at the nest (GLMM, $\chi^2_1 = 3.87$, $p = 0.04$). Thus the effect of playback treatment on the probability of looking at the nest, via the variation in acoustic similarity to the provisioning call, was significant regardless of the integration of contextual information from previous acoustic stimuli.

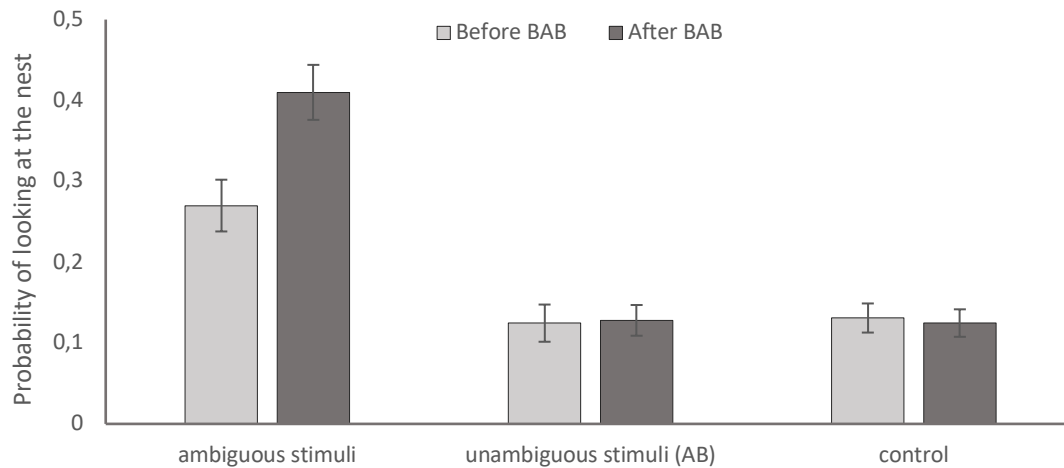


Fig 2. Previous exposure to the BAB treatment induced greater nest looking behaviour in subsequent ambiguous treatments (i.e. BA, ABA, AAB, BAA), but not in subsequent unambiguous ones (AB), or in the control treatment (CTR). Figure shows probability of looking at the nest for each treatment category (\pm Standard Error [SE]). Ambiguous treatments were defined as those where the element combination had been rearranged so as not to match any existing call template, while unambiguous treatments represented unmodified templates of existing bird calls.

Discussion

I have provided evidence that chestnut crowned babblers respond to acoustic signals constructed via phonemic contrasts according to the similarity of the perceived call to its call type template. Furthermore, the results suggest that babblers are capable of integrating contextual information upon reception of an acoustic signal. I propose that the ability to assess acoustic similarity between signal and template, and to integrate contextual information during vocal communication, may have played an important role in the evolution of a phoneme-like communication system in chestnut crowned babblers. These cognitive abilities are not necessarily specialised adaptations to a phoneme-based communication system: indeed, the facultative adjustment of behaviour based on acoustic similarities and the integration of contextual information have been documented in species exhibiting non-phoneme based repertoires (e.g. McDonald & Wright 2011, Crockford et al. 2007). However, whether these cognitive abilities in chestnut crowned babblers represent specialised adaptations for phoneme-based signalling, or pre-existing mechanisms which

were then co-opted for use in this specific form of combinatorial communication, this study highlights the importance of such abilities in a species exhibiting a rudimentary form of phonemic contrasts. As chestnut crowned babblers are the only non-human animal known to utilise such a combinatorial design, understanding which cognitive mechanisms are shared or unique in the comparison of human and babbler phoneme use provides valuable insight into the evolution of combinatorial communication.

In communication systems such as the one investigated in this study, as well as in species that exhibit graded repertoires, vocalisations belonging to a single call type may nonetheless exhibit within-individual variation in acoustic properties, resulting in ambiguity (Hammerschmidt & Fischer 1998, Keenan *et al.* 2013). For example, a phoneme-like combinatorial design is associated with a high chance of error in element sequence generation, and thus of producing ambiguous signals, in speech development in human children (Jakobson 2012, Smit 1993). Similarly, the chestnut crowned babbler provisioning call normally follows the B + A + B element combination template, yet several instances of incorrect A and B element combinations have been recorded in both natural and experimental circumstances (H. Mylne, E. Tew, J. Mine, unpublished data). These aberrant signals could potentially impede the acquisition of information, and thus have detrimental effects on both sender and receiver. To circumvent this problem, information acquisition may depend not only on the recognition of associations between signals and events, but also between signals and natural templates, via the assessment of acoustic similarity. Indeed, the ability to adjust behaviour based on the assessment of acoustic similarity between different vocalisations is not without precedent, as this mechanism is used for kin discrimination in multiple avian species (Sharp *et al.* 2005, McDonald & Wright 2011). The results of this study indicate that babblers are sensitive to the similarity between a call type template, i.e. the provisioning call, and the actual stimulus that is heard. Not only do they perceive this similarity, but the provisioning-related behavioural response associated with each stimulus is greater when the similarity of the call to the template is also high. As highlighted by this study in the case of variation in sequences of A and B elements, the ability to assess acoustic similarity to the BAB call template allows the babblers to produce appropriate behavioural responses to signals even when these signals match their corresponding template imprecisely.

Interestingly, recent studies show that human phoneme combination appears to rely on fine-grain, as opposed to coarse-grain similarity between a perceived signal and the target template for the identification of signal referents (Gregg *et al.* 2019). Therefore, where phoneme-based templates used in human speech may be referred to as “strict”, babbler templates appear to exhibit greater flexibility. It is possible that this discrepancy may be explained by differences in the productivity of the phoneme-based system: humans recombine single phonemes, such as the sound /n/, into a vast array of different words; in babblers, on the other hand, the use of the A and B elements is limited to only two functionally distinct calls. As a result, the ambiguity experienced upon reception of erroneous combinations of A and B elements is likely to be lower than in the case of phonemes that are used productively. Thus the limited number of potential signals comprising the A and B elements, as a result of non-productivity in phoneme-like combination, may facilitate adaptive decision-making based on coarse-grain acoustic similarity. This study has limited its scope to the A and B elements of the babbler repertoire, and it is therefore unknown, although plausible, whether the ability to respond to ambiguous phoneme-based sequences via coarse-grain acoustic similarity to a template also applies to other calls produced by this species. Replicating this experimental procedure across the entire repertoire would help to shed light on the importance of this cognitive ability for the interpretation of phoneme-based combinatorial signals.

With this study, I have also presented novel empirical evidence for the ability to integrate contextual information upon reception of acoustic stimuli in chestnut crowned babblers, a cooperatively breeding bird. Specifically, exposure to incorrect combinations of the acoustic elements that make up the provisioning call resulted in a heightened nest-looking response when subjects had previously been exposed the correct sequence of provisioning call elements. In essence, subjects were more likely to respond to incorrect arrangements if they had previously heard the correct arrangement, than if they had not. These findings suggest that chestnut crowned babblers are capable of integrating previous stimuli in order to produce appropriate behavioural responses to future, ambiguous signals. This ability, known as pragmatic inference, appears to be beneficial in the context of vocal communication in this species by allowing receivers to reduce their uncertainty regarding error-prone, combinatorial

acoustic stimuli. I thus propose that the role of contextual information in acoustic communication is as important for vocal birds as it is for mammals.

In primates, where pragmatic inference has received substantial attention, the most commonly invoked pressure driving its evolution is the inflexibility of signal production due to vocal constraints (e.g. Arnold & Zuberbuhler 2013). Indeed, the vocal repertoires of non-human primates are constrained by both the neurobiological circuits involved in the vocal pathway, as well as the morphology of the vocal apparatus (Hammerschmidt & Fischer 2008). Due to such constraints, a limited number of calls may be produced in association with a wide range of contexts. Putty-nosed monkeys (*Cercopithecus nictitans*), for instance, possess two distinct alarm calls, which are used in contexts ranging from aerial to terrestrial predators, falling trees or branches, encounters with non-predatory heterospecifics, as well as within and between-group interactions (Arnold & Zuberbuhler 2013). But as in the case of putty-nosed monkeys and other primate species, despite constraints on signal production, behavioural responses to vocal signals are often flexible and complex (Wheeler & Fischer 2012). This behavioural flexibility has been shown to depend largely on the capacity to integrate contextual information upon reception of acoustic stimuli. For instance, when such general-purpose alarm calls are combined with contextual stimuli such as the behaviour of other group members or environmental cues, the associated behavioural responses reflect the available contextual information (Arnold & Zuberbuhler 2013, Price & Fischer 2014). Thus in primates, receivers acquire additional information before or during the communicative event, in order to overcome ambiguity associated with vocal constraints.

In many avian species, the vocal apparatus is not constrained, but highly specialised and flexible (Suthers *et al.* 2016). Many passerines exhibit large and complex song and call repertoires (Catchpole & Slater 2003), and in some cases, new signals can be continuously learned throughout the lifetime (Tyack 2016). Thus where pragmatic inference occurs in birds, the selective pressures on its development are likely to be different from those operating in primates. As birds exhibit a much greater scope for acoustic flexibility, ambiguities in vocal communication are not likely to arise from vocal constraints. Yet a high degree of signal complexity, whether in terms of repertoire size or combinatorial properties, may also represent a source of ambiguity (Engesser & Townsend 2019). When repertoires are large, or when multiple calls share a proportion of their

combinatorial sub-units, different signals are more likely to resemble each other (Coye et al. 2017), as demonstrated by flight/prompt calls and long distance/maternal contact calls in chestnut crowned babblers (Crane et al. 2016). Furthermore, when sequences of acoustic elements are underpinned by combinatorial rules, such as in the stochastic structures of black fronted titi monkeys (*Callicebus nigrifrons*) or the temporal structures of southern pied babblers (*Turdoides bicolor*) (Engesser & Townsend 2019), errors in production and reception may occur (Bradbury & Vehrencamp 1998). Thus in birds, specialised cognitive mechanisms may be required to overcome ambiguity associated with complex, rather than constrained, vocal signals. Indeed, as shown by this study, pragmatic inference in chestnut-crowned babblers is useful to circumvent the loss of information arising from ambiguities in the production and reception of signals generated via phonemic contrasts.

This study constitutes one of the first explicit tests of pragmatic inference in an avian species, and hence the experimental procedure warrants replication. Further research on pragmatic inference in birds may confirm that although distinct taxa have converged upon this cognitive mechanism to aid in vocal communication, the specific selective pressures driving its emergence are distinct. In particular, this study suggests a role of pragmatic inference in mediating the reception of highly specialised and flexible forms of vocal communication, as opposed to constrained vocal output. Furthermore, this constitutes an important step forward for the field of comparative communication, as it highlights the link between pragmatic inference and combinatorial signalling in a non-human species. It is widely accepted that the acquisition of contextual information is an important component of vocal communication in humans (Noveck et al. 1991, Goodman & Frank 2016, Seyfarth & Cheney 2017). However, human language is often regarded as an occurrence whose distinct features are entirely unique within the animal kingdom (Hauser, Fitch & Chomsky 2002). Yet bit by bit, combinatorial abilities once considered unique to language such as affixation, phonemic contrasts and compositional processing are increasingly finding parallels in animal vocal communication, especially in avian species (Outtara et al. 2009, Engesser et al. 2015, Suzuki et al. 2016, Engesser et al. 2016). Furthermore, as detailed by this study, the use of combinatorial acoustic signals in vocal communication is accompanied by a cognitive mechanism known as pragmatic inference, which enables signallers and

receivers to overcome the ambiguity inherent in vocal communication by acquiring information from sources other than the signals themselves. This suggests that the ability to acquire contextual information during a communicative event may be a prerequisite for the evolution of complex forms of combinatorial signalling, with specific reference to phonemic contrasts.

In conclusion, there is certainly much to be learned from renewed and heightened interest in animal combinatorial communication. This field has the potential to shed light on the evolutionary origins of human language, which has played such a key role in developing the highly adaptive social structures and behaviours typical of our species, and yet remains shrouded in mystery preceding the adoption of written language. For example, the link between pragmatic inference and phonemic contrasts highlighted here may have important implications for understanding the cognitive mechanisms which promoted the emergence of phoneme-based signals in human language. In addition, species with large and complex vocal repertoires, such as birds and marine mammals, provide a valuable opportunity to explore the diversity of combinatorial structures and cognitive mechanisms related to vocal communication (Engesser & Townsend 2019). This endeavour is sure to enlighten us as to the myriad ways in which vocal communication has evolved, and rescale our anthropocentric views while broadening our comprehension of what communication really is.

Chapter 3

An empirical test of the generation visual mental images upon signal reception in an avian species

Summary

In non-human vocal communication, behavioural responses to acoustic signals may either be mediated by reflexive, motivation-induced reactions to the physical features of the signal, by associative learning between signals and contexts, or by processing the signal's meaning through an internal representation of the signal's referent. Here, I investigate the processing of a context-specific signal, i.e. the nest provisioning call, in cooperatively breeding chestnut-crowned babblers (*Pomatostomus ruficeps*), by exposing subjects to playbacks of natural nest provisioning calls and to artificial stimuli designed to exhibit varying degrees of acoustic similarity to the provisioning call. I hypothesise that upon reception of provisioning calls, babblers generate a visual mental image of the signal's presumed referent, i.e. the nest. I predict that if babblers generate a mental image of the nest, they should be primed to visually detect a nest that matches their internal representation rather than other available nests. Furthermore, I predict that this preferential detection should be dependent upon the similarity of the perceived signal to the provisioning call template. The results however show no significant tendency to interact preferentially with any individual nest during playback of provisioning calls or similar acoustic stimuli. This suggests that the reception of provisioning calls does not evoke the visual mental image of a nest in chestnut crowned babblers. Alternative explanations for the processing of acoustic signals are discussed.

Introduction

The ability to process information extracted from the environment is a fundamental characterisation of all animals (Thornton *et al.* 2012). The capacity for information processing exhibits spectacular variation in complexity, from high-speed and fine-scale perception of sensory information (Wagner *et al.* 1987; Schnitzler & Denzinger 2011), to detailed spatial memorisation of environmental features such as food or water resources (Thornton & Boogert 2019; Polansky *et al.* 2015). Furthermore, information extracted from the environment may be stored in the animal's cognitive machinery and re-used during future behavioural

events. For example, when brown-eared bulbuls (*Hypsipetes amaurotis pryeri*) learn to associate the butterfly species (*Pachliopta aristolochiae*) and its Batesian mimic (*Papilio polytes*) with an uncomfortable taste, they subsequently refuse to eat these species in later encounters (Uèstugi 1996). However, the precise form in which this information is perceived, stored and retrieved within animal minds is unclear, and represents a topic of hot debate (Stegmann 2013). In humans, visual mental images, defined as the mental representation of visual and accompanying sensory information, are a pervasive aspect of cognition (Kreiman et al. 2000, Pearson et al. 2015). However, it is currently unknown whether non-human animals also generate mental images depicting specific behaviourally-relevant features of the environment, and particularly, whether these mental images can be evoked by specific sensory stimuli such as vocal signals (Seyfarth et al. 2010, Stegmann 2013, Rendall & Owren 2009).

In human language, it is natural for senders and receivers to associate a vocal stimulus, such as the noun 'leopard', with an internal depiction of the referent (Hurford 2007, Fitch 2010). Crucially, we do not only interact with signals on the basis of their acoustic properties, we also attend to their semantic content via an internal representation of the signal's meaning (Zuberbuhler et al. 1999). But what happens in the mind of a non-human animal upon reception of a context-specific signal? On one hand, some suggest that animal signals could represent a case of perceptual semanticity: receivers only attend to the physical features of a call and this sensory percept alone, via its influence on the receiver's internal motivational state, drives subsequent behaviour (Owren & Rendall 2001; Rendall et al. 2009). This approach excludes any implication of information, meaning or visual mental images. Others argue in favour of conceptual semanticity, i.e. where the acoustic properties of vocal signals are only relevant insofar as they relate to associated cognitive structures such as visual mental images or concepts (Seyfarth *et al.* 2010; Seyfarth & Cheney 2017). Prompted by the seminal work of Seyfarth, Cheney and Marler (1980a,b) on the alarm call system of vervet monkeys, over the past four decades, a flurry of research in the field of animal communication has produced evidence across multiple taxa for so-called functionally referential signals, i.e. those that contain information referring to specific internal or external events, insofar as they: a) are produced predictably in certain contexts, and b) elicit specific behavioural responses in the receiver (reviewed in Townsend & Manser 2013, Wheeler &

Fischer 2012). Yet despite representing promising candidates for conceptual semanticity, even signals that meet these two key criteria for functional reference do not imply visual mental images, as the resulting behavioural responses could be explained more parsimoniously via associative learning (Wheeler & Fischer 2012). Therefore, direct testing of conceptual semanticity, as opposed to perceptual semanticity, requires a careful experimental approach.

Few empirical tests have been conducted which specifically address the presence and quality of visual or conceptual representations of signal referents in non-human animals, and much of what we know, we owe to the tradition of Peter Marler. One such study was conducted by Zuberbuhler, Cheney & Seyfarth (1999), on the alarm call repertoire of diana monkeys (*Cercopithecus diana diana*). The authors examined whether receivers simply attend to the perceptual, acoustic features of alarm calls, or whether, on the other hand, they form a concept regarding the meaning of calls upon signal reception. To do so, they used a prime-probe playback experiment, subjecting individuals to a two-step playback sequence involving vocalisations related to specific predators and recording responses in terms of subsequent alarm calling behaviour. In the baseline treatment, a crowned eagle (*Stephanoaetus coronatus*) or leopard (*Panthera pardus*) alarm call was followed by another, identical alarm call, thus the two stimuli were similar in both their acoustic and purported semantic properties. The authors predicted a weaker response to the probe due to a habituation effect. In the test treatment, the priming stimulus in the form of an eagle or leopard alarm call was followed by an actual eagle or leopard vocalisation as the probe stimulus. In this case, the two stimuli differed in their acoustic properties but were similar in their semantic properties, i.e. in indicating the presence of a specific predator. The authors predicted that if receivers attended only to acoustic properties, they would not transfer habituation between the prime and probe stimulus, thus they would exhibit a strong response to both stimuli. However, if they also attended to the semantic properties of the signal, via an internal conceptualisation of the signal's meaning, then they would exhibit a weaker response to the probe stimulus due to its similarity in semantic content to the priming stimulus. Indeed, monkeys exposed to eagle alarm calls followed by eagle vocalisations produced significantly fewer alarm calls to the probe than to the prime stimulus. This result suggests that diana monkeys generate an

internal representation of signal referents, either in the form of a visual mental image or a concept, upon reception of predator-specific alarm calls.

Recently, a study by Suzuki (2018) on an avian species, the Japanese tit (*parus minor*), has provided a novel empirical approach and promising results for the study of visual mental images evoked by vocal signals. Japanese tits, like the diana monkeys described above, produce distinct alarm calls for different predators, including as a specific “jar” call produced only when encountering predatory snakes. The author hypothesised that this predator-specific alarm call evokes a visual mental image of a snake in receivers, and consequently, predicted that reception of this call would enhance detection of the target object, i.e. a snake. In Suzuki’s experimental design, tits were exposed to a model snake, in the form of a short stick pulled by a string, and to three kinds of acoustic stimuli: 1) the snake-specific “jar” call, 2) a general alarm call used for a wider range of predators, and 3) a recruitment call, used to attract conspecifics in non-predatory contexts (Suzuki 2011, 2012, 2014). Conforming to his predictions, Suzuki found that tits were more likely to approach the stick when hearing the snake-specific alarm call than when hearing any of the other playback stimuli, and moreover, that tits would only approach the stick when it appeared to move in a snake-like fashion (i.e. being pulled along the ground or up a tree), rather than when the movement of the stick did not resemble the movement of a real snake (i.e. swinging from a low branch). Thus from the observation that receivers become more visually perceptive to objects resembling snakes only when hearing a snake-specific alarm call, the author concluded that this increased visual detection of snake-like stimuli is mediated by a visual mental image of a snake.

Although encouraging, the results of Suzuki (2018) spark consideration as to whether the experiment can be unequivocally interpreted as evidence for signal-evoked visual mental images in receivers, as detailed in Bond’s (2019) subsequent criticism of the study. Crucially, Bond notes that the findings cannot rule out the alternative explanation that the behavioural responses observed could merely represent a chain of innate responses, driven by associative learning: when the bird hears a snake-specific call, it automatically looks down; by doing so, it sees the snake, which induces a learned mobbing response. However, Bond does not propose to abandon the novel approach outlined by Suzuki altogether, but to refine it so as to permit a closer examination of the link

between a context-specific call and particular attention to specific visual features. To successfully validate the hypothesis of a visual mental image, Bond suggests integrating an analysis of the component features of the presumed referent, by utilising a fully realistic model, as well as additional models that differ from the true model in certain visual stimulus features e.g. shape/color. Thus in future playback experiments, by combining a highly referential call with models that differ in their component visual features, and observing the resulting variation in behavioural responses, more detailed information can be acquired regarding the presence, as well as the quality of signal-evoked visual mental images.

With this study, I continue the investigation of visual mental images associated with vocal communication in non-human animals. To this end, I conduct a playback experiment on an avian species, the chestnut-crowned babbler (*Pomatostomus ruficeps*). This highly social and cooperatively-breeding bird exhibits a complex vocal repertoire comprising at least 13 distinct signals used in specific contexts (Crane *et al.* 2016), and relies on such a complex system of vocal communication to cooperate effectively within a harsh environment, the Australian outback (Russell 2016). I limit my investigation to the process underlying the reception of one babbler call, the provisioning call. This signal, which consists of the combination of the acoustic elements labelled A and B in the arrangement B+A+B, is produced by babblers carrying food items when approaching and leaving the breeding nest (Engeser *et al.* 2015). Babblers are obligate cooperative breeders, and thus rely heavily on the provisioning efforts of helpers outside the breeding pair for successful offspring development (Russell *et al.* 2010). As the provisioning call plays a significant role in mediating this cooperative behaviour, I consider this signal to be sufficiently important to be an appropriate target stimulus for the investigation of visual mental images. I propose that the evolution of visual mental images of signal referents could have been driven by selection for increased robustness of information transfer during vocal communication.

My experimental design draws from that of Suzuki (2018): I conduct a playback experiment in which subjects are exposed to multiple distinct stimuli, and allowed to interact with the presumed referents of the acoustic stimuli. However, my design implements several novel features compared to previous work. First, I use multiple signal referent models with different component visual features, as suggested by Bond (2019). The presumed referent of the chestnut

crowned babbler provisioning call is a nest, as this call is produced solely during interactions with the breeding nest. I thus allow subjects the opportunity to interact with nests differing in their component visual features. I utilise both a completely intact, high-quality babbler nest, as well as a nest which has undergone natural degradation: the two nests differ in overall shape and size, as well as other visual features such as the size of the nest opening and the length of the protruding neck. Finally, adding a completely novel perspective to the approach outlined by Suzuki (2018) and Bond (2019), I deconstruct the vocal signal in question, i.e. the provisioning call, into its core acoustic components, and rearrange these structural units into novel combinations to use as experimental playback stimuli. I propose that by artificially generating acoustic variation in the target signal, and observing how this variation relates to behavioural responses, it is possible to gain an understanding of the underlying mental process that results in the formation of a visual mental image upon signal reception. I envision two possible outcomes: 1) increasing acoustic similarity to the provisioning call results in enhanced detection of target referents with specific component visual features, suggesting a graded mechanism; or 2) only the correct, unmodified provisioning call results in enhanced detection of target referents, suggesting a threshold mechanism. I propose that the empirical investigation of visual mental images generated upon signal reception can greatly benefit from procedures that involve the deconstruction of both the acoustic signal and its presumed referent into their component acoustic and visual features.

Methods

Study site and species

The study was conducted during August to November 2016 at the Fowlers Gap Arid Zone Research Station in far western New South Wales, Australia (141°42'E, 31°06'S). Chestnut-crowned babblers are 50g, sexually monomorphic passerine birds endemic to south-eastern Australia (Russell 2015). This cooperatively breeding species lives in groups of up to 23 individuals (mean 10), and is highly vocal, with each adult possessing a repertoire of at least 18 functionally distinct calls (Crane *et al.* 2016). The repertoire of this species is one of only two known cases in the animal kingdom (including humans) in which

phonemic contrasts are involved in the production of complex vocalisations (Engesser *et al.* 2015). The calls of the babbler repertoire are used in various contexts including social, threat, contact and arousal (Crane *et al.* 2016). The acoustic signals under investigation in this study were the flight call and the provisioning call, both of which are composed of the two acoustically distinct elements labelled A and B, in the order AB and BAB, respectively (Engesser *et al.* 2015).

Test-subject Selection and Housing

Birds were captured using mist-nets in creek beds and transported in bird bags no more than 5km away by vehicle to onsite aviaries. No more than 4 group members were captured at a time, and no more than 50% of the group was removed. Juveniles and breeding females were not taken for experiments. All birds were released back to their groups within 48 hours of capture. The aviaries consisted of four single compartments each of 2 m long, 2.5 m deep and 2 m high with standardised artificial perches on natural substrate with ad lib water. The two sides of the aviary were opaque metal, the front consisted of one-way Perspex allowing observers to record behaviour, while the back side of the aviary exhibited a metal mesh of 1cm² allowing test subjects a view to the outside. During the day, each bird was provided 20 mealworms every 2-3 hours. When more than one bird was removed from a group, these birds were housed in the same compartment overnight. Crucially, each aviary compartment also contained two previously used babbler nests in which the birds could roost overnight, as is typical of this species. The nests were extracted from territories within the field site known to contain babbler groups, and were selected for their qualities by experienced team members. For the purposes of this experiment, two nests of different quality were chosen: one nest, referred to as the high quality nest, was completely intact and recently maintained, while the other, referred to as the low quality nest, was older and exhibited signs of natural degradation (Fig 1). Specifically, the nests differed in 4 key visual features: 1) size (large vs small), 2) shape (domed vs flattened), 3) presence of protruding neck (present vs absent), and 4) size of nest opening (large vs small). The fully intact, high quality nest was large (34x45x15cm), domed, with a protruding neck and small nest opening (3x4cm). All of these features contribute to increasing the functionality of the nest: larger and domed nests are more spacious, while the protruding neck and

small opening hinder the access of larger avian predators such as the nankeen kestrel. Thus the high quality nest represents a fully realistic model of the presumed referent of the provisioning call, via the biological relevance of its naturalistic features. The degraded, low quality nest was small (23x27x10cm), flattened, with no protruding neck and a larger nest opening (5x7cm). Babbler nests are highly durable structures (Louis O'Neill, pers. obs.), thus it was not possible to determine the precise age of the nests. 100% of birds roosted within one of the available nests.

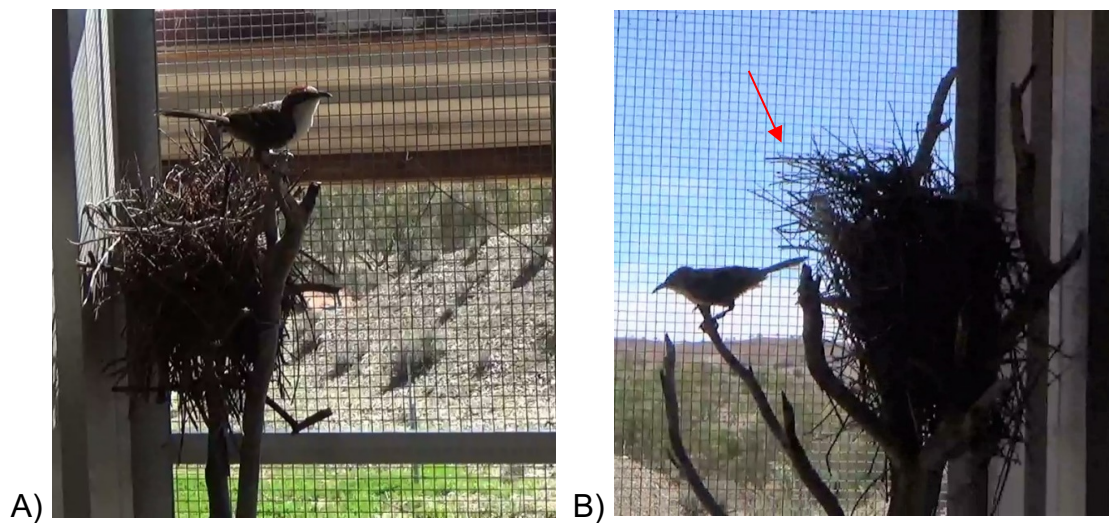


Fig 1. The two natural babbler nests used as presumed referents for the provisioning call in the experiment. Nest A) is the low quality nest, measuring 23x27x10cm while nest B) is the high quality nest, measuring 34x45x18cm. Apart from differences in size and overall shape (flattened vs domed), the nests differ in the presence of a protruding neck (highlighted by the red arrow) and the size of the opening, which measured 5x7cm in the low quality nest and 3x4cm in the high quality nest.

Experimental protocol: playback stimuli and video recordings

The calls used in the playback experiments were obtained from recordings of 6 groups of wild chestnut-crowned babblers. Vocalising individuals were adult male and female babblers of at least 1 year of age. Flight calls were produced during flights to and from the nest, while provisioning calls were produced while provisioning chicks within the nest. In each case, calls were recorded using a Sennheiser directional microphone (ME66/K6) connected to a Marantz solid-state recorder (PMD660, sampling frequency 48 KHz, 24 bits), positioned within

1 m of a nest. Playbacks, including the construction of artificial calls, were created with Adobe Audition CS6 (Version 6 Build 732, Adobe Systems). From each of the six groups recorded, utilising only high-quality recordings (high signal-to-noise ratio, no obscuring vocalisations, low background noise), multiple sets of seven playback stimuli were created. Each set included a natural flight call (AB), a natural prompt call (BAB), a reverse flight call (BA), three triple-element combinations of A and B notes (AAB, ABA, BAA), and a control stimulus consisting in the call of a heterospecific species (spiny-cheeked honeyeater, *Acanthagenys rufogularis*). When acoustic elements for the generation of artificial calls were added and/or replaced, it was ensured that inter-element distance and amplitude matched the original call. A new call-set was played in randomized order for each subject or pair of subjects, and birds never received a call-set from their own group. In 13 cases, two individuals of the same group were placed in separate, non-adjacent aviary compartments and tested simultaneously with the same playback-set. Subjects tested in pairs did not have visual access to each other, which reduced the confounding effect of the presence of another individual during testing. This procedure was implemented in order to increase the sample size of the experiment within the available time frame.

Playbacks were broadcast at an amplitude of 50 dB at 2 meters, matched to the natural amplitude of flight and provisioning calls as measured with a Castle GA206 sound level meter. The speaker was placed immediately outside the aviary compartment(s) containing the test subject(s). During each playback, a stimulus was repeated six times over a total of 10 s; a break of at least 10 min was given for focal individuals to resume pre-stimuli behaviour before the initiation of another playback sequence. Playback experiments were conducted on the day following capture.

During testing, individuals were recorded using digital Sony handycams (HDR-CX220 and HDR-CX160). Visual recordings of 4 s from playback onset were analysed frame by frame using Adobe Audition CC (Version 6 Build 732, Adobe Systems). The behavioural parameters recorded were time (s) spent in camera view (mean = 3.12 s, range = 0.99-4s), looking at the nest, looking outside, flying or hopping, looking other. Following Engesser et al. (2015), gaze direction was evaluated as orientation of the beak following a head movement. Behavioural observations began with the first head movement following the first

acoustic element of the playback, which occurred within a mean of 0.5s from the beginning of the playback. Marker lists created in Adobe Audition were extracted into txt-files by using CueListTool (Version 1.7). Scoring was conducted blind to experimental condition.

The experiment was conducted on 70 adult chestnut-crowned babblers. Some of the individuals were removed from statistical analyses due to confounding variables during the experiment: for example, in 2 cases, an aerial predator was perched within view of the aviary, which resulted in the test subject remaining entirely motionless throughout the test.

Rationale

The primary goal of this experiment is to determine whether reception of the provisioning call, a signal produced during interactions with a working babbler nest, is associated with a visual mental image of the nest. To this end, I offered subjects the choice to interact with two nests that differed in their component visual features upon reception of various natural and artificial playback stimuli, and observed their preferences of visual interaction in relation to different playbacks. The experimental aviary compartments contained both a recently used, fully intact babbler nest, referred to as the high quality nest, as well as a previously used but partially degraded babbler nest, referred to as the low quality nest. To a trained eye, the two nests are visually distinguishable: the high quality nest has a more rounded, domed structure with a slightly protruding neck, while the old nest is flatter at the top, without a visible neck and with a much more prominent opening. Furthermore, years of field observations have led to the understanding that high quality nests are fully functional, i.e. used for both roosting and breeding, while low quality nests, once degraded to the condition described above, are no longer used by babblers (L. O'Neill, pers. obs.).

The null hypothesis is that reception of the provisioning call does not generate in the receiver any visual mental image of the nest. This hypothesis predicts that the receiver should not discriminate between nests that differ in their component visual features when exposed to the provisioning call. Conversely, the two complementary alternative hypotheses under investigation in this experiment are outlined as follows. Hypothesis 1: The provisioning call generates in the receiver a visual mental image of the nest. Prediction 1: Reception of the provisioning call primes the receiver to detect a nest that

matches the image established by the mental representation, and thus to visually interact with the high quality nest more than the low quality nest. Hypothesis 2: The occurrence of a visual mental image depends on the reception of a signal adhering to a specific acoustic template, i.e. the provisioning call. Prediction 2: Upon hearing the provisioning call, or artificially generated variants which exhibit high acoustic similarity to the provisioning call, receivers should preferentially associate with a nest which exhibits specific component visual features, i.e. the high quality nest; conversely, such a preference for the high quality nest should not be apparent in response to playback stimuli which exhibit a reduced acoustic similarity to the provisioning call.

Each bird was exposed to a series of 7 unique playback treatments with a break of 10 min between treatments. 6 out of 7 treatments consisted in playing back two- and three-element combinations of the A and B elements typical of the flight and provisioning calls. Of these, 2 treatments exhibited naturally-occurring combinations of elements, i.e. AB (flight call) and BAB (provisioning call). In the other four treatments, attempting to span the variation of potential similarity to the provisioning call, the elements were rearranged to produce the following aberrant combinations: BA, ABA, AAB, BAA. The remaining treatment was a control, where the subject was played the call of a different species (spiny-cheeked honey-eater). Thus the 6 experimental stimuli (AB, BA, ABA, AAB, BAA, BAB) exhibited variation in two key parameters: a) the total number of A and B elements (i.e. 2 or 3), and b) the position of these elements relative to the correct BAB sequence. I predicted that 3-element combinations, and combinations where the position of the A and B elements resembled that of the correct sequence more closely, would be perceived as more similar to the correct sequence than 2-element combinations and those where the order of A and B elements differed more substantially from the correct sequence (see below).

The provisioning call of the babbler repertoire is produced exclusively during interactions with the nest, and in addition, it has been shown to elicit nest-looking behaviour following playback (Engesser *et al.* 2015). Therefore the key data that was collected regarded the probability that babblers performed nest-looking behaviour towards a specific nest, and the duration of this behaviour, in response to different playback stimuli. The order of the treatment stimuli was randomised.

During playback experiments, a common risk is that birds may habituate to playback, decreasing their behavioural responses as more playbacks are presented during the experiment (Engesser *et al.* 2015). I tested for habituation by assessing the relationship between behavioural response and playback number. There was no significant effect of treatment number on the probability or duration of time spent looking at the nest (GLMM, $\chi^2_1 = 2.49$, $p = 0.1$, GLMM, $\chi^2_7 = 6.87$, $p = 0.4$).

Statistical Analyses

Statistical analyses were conducted using the software programme Rstudio (Engesser *et al.* 2017, Ihaka & Gentleman 1996). The packages used were: boot, car, tidyverse, ggplot2, tidyr, dplyr, broom, lme4. Analyses of behavioural responses from the playback experiments were conducted using Generalized Linear Mixed Models (GLMM).

Hypothesis 1 was tested on the basis of behavioural responses to playback of the provisioning call treatment (i.e. BAB), while Hypothesis 2 was tested via the comparison of behavioural responses across all treatments, considering their various degrees of similarity to the provisioning call template. Classic methods for quantifying acoustic similarity, such as spectrographic cross-correlation (SPCC), were deemed unsuited to the experimental design given that all playback stimuli (excluding the control treatment CTR) consisted of rearranged versions of two acoustic elements (i.e. A and B), the spectral and temporal features of which were maintained constant across all experimental stimuli. The playback stimuli nonetheless differed acoustically in two key parameters: 1) the total number of A and B elements in the sequence, and 2) the number of elements in the correct position relative to the BAB sequence. As both of these measures could be directly compared to the correct provisioning call sequence, I computed a similarity index for each non-BAB playback treatment which accounted for the variation in these two parameters, relative to the BAB sequence. Thus each non-BAB treatment was first assigned a score for both a) total number of A and B elements relative to the number of A and B elements in the BAB sequence (range = 0-1, mean = 0.72); and b) number of elements in correct position relative to BAB (range = 0-0.66, mean = 0.44). Finally, a composite index was generated for each treatment by calculating the mean of the two individual scores (range = 0-0.83, mean = 0.58).

To analyse nest-looking responses to the BAB treatment (Hypothesis 1), and all other treatment types (Hypothesis 2) I used a Generalised Linear Mixed Model (GLMM). First, I examined responses in terms of the probability of looking at the high quality nest vs the low quality nest. The orientation of nest-looking behaviour, coded as 0 (looked at low quality nest) and 1 (looked at high quality nest) represented the response term, fitted to a binomial error structure with logit link function. Time spent in camera view was fitted as the binomial denominator, treatment type (CTR, AB, BA, ABA, BAA, AAB, BAB) was fitted as a seven-level factor, treatment number was fitted as an additional explanatory term (to account for repeated measures of individuals), and individual identity nested within group identity were fitted as random terms. Next, I examined responses in terms of differences in nest-gaze duration towards the high vs low quality nest. The duration of nest-looking behaviour (s) towards high vs low quality nest represented the response term, fitted to a binomial error structure with logit link function. Time spent in camera view was fitted as the binomial denominator, treatment type (CTR, BA, AB, ABA, BAA, AAB, BAB) was fitted as a seven-level factor, treatment number was fitted as an additional explanatory term, and individual identity nested within group identity were fitted as random terms.

Results

Overall, 32% of the 258 playbacks resulted in birds looking at either nest. Of these, looks to the high quality nest occurred in 77% of cases, and looks to the low quality nest in 23% of cases. For those individuals that looked at the nest, the average duration of looks to the good quality nest was 0.41s (SD= 0.22, range= 0.11-1.22s), while the average duration of looks to the bad quality nest was 0.38s (SD= 0.18, range= 0.07-0.67s). The control stimulus elicited the lowest overall nest response, with only 4 subjects (10%) looking at either nest during such playbacks. The 2-element combinations (i.e. AB and BA stimuli) showed a mean 11.5% increase in total nest response compared to the control, while the incorrect 3-element combinations (i.e. ABA, AAB and BAA) showed a mean 22% increase in nest attentiveness. Finally, the stimulus where the provisioning call (i.e. BAB) was played without modification evoked the highest qualitative nest response, with a 37% increase compared to the control. The nest-looking response in the control treatment was deemed insufficient to provide a reliable assessment of preference for the high or low quality nest, and

was thus removed from further statistical analyses. In contrast with Predictions 1 and 2, birds were not significantly more likely to look at the good quality nest during the BAB treatment nor any of the other playback treatments (GLMM, $\chi^2_6 = 3.19$, all p values > 0.1) (Fig. 1). Furthermore, I found no significant effect of treatment on the duration of nest looks toward the good vs bad quality nests (GLMM, $\chi^2_6 = 4.24$, $p = 0.64$). These results suggest that acoustic stimuli that either match or closely resemble the provisioning call do not generate in receivers a preference for nests exhibiting specific visual component features.

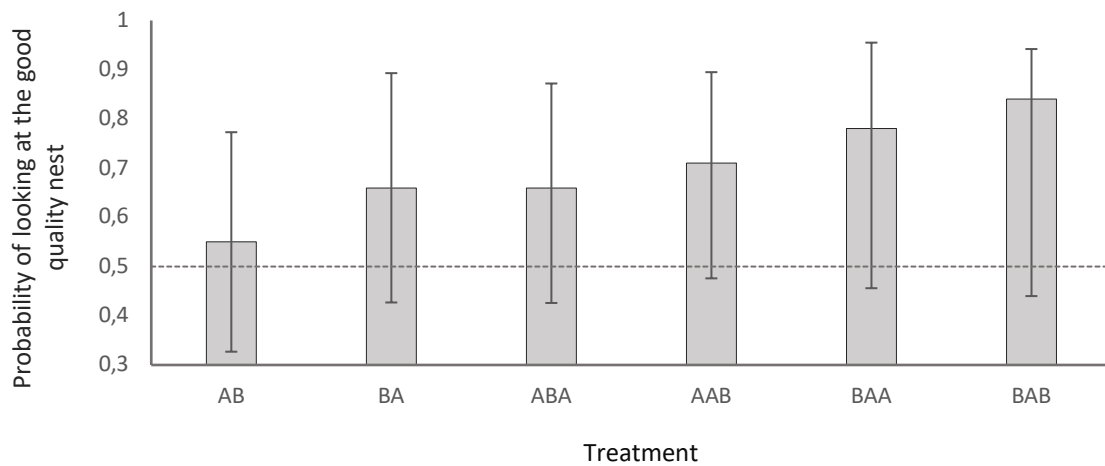


Fig 1. The probability of looking at the good quality nest was not significantly greater than chance in any of the playback treatments. Figure 1A shows back-transformed predicted means (\pm Standard Error [SE]), generated from a Generalized Linear Mixed Model (GLMM), in which the orientation of nest- looking behaviour (looking at the good quality nest, looking at the bad quality nest) was fitted as the independent response term. Time spent in camera view was fitted as the binomial denominator, treatment type (CTR, AB, BA, ABA, BAA, AAB, BAB) was fitted as a seven-level factor, treatment number was fitted as an additional explanatory term (to account for repeated measures of individuals), and individual identity nested within group identity were fitted as random terms. The CTR treatment was removed from statistical analyses of nest preference due to insufficient nest response. The other treatments are labelled according to the combination of acoustic elements specific to each treatment, i.e. AB treatment consisted of elements A + B, BA consisted of elements B + A, etc.

Discussion

The results of this study do not conform to the predictions formulated on the basis of the hypothesis that upon reception of a context-specific acoustic signal, i.e. the provisioning call, chestnut crowned babblers generate a visual mental image of the signal's presumed referent, i.e. the nest. The birds in this study were exposed to playbacks of natural nest provisioning calls and to artificial stimuli designed to exhibit varying degrees of acoustic similarity to the provisioning call. Upon stimulus reception, babblers were given the choice to interact with two previously used babbler nests which differed in their component visual features: one was fully intact, while the other was partially degraded. I hypothesised that the visual component features of the mental image generated in receivers would resemble those of a fully intact and functional babbler nest. I predicted that if the reception of a provisioning call was accompanied by a visual mental image of the nest, then babblers should discriminate between nests that differ in their visual features, preferentially interacting with a nest that matches the component visual features established by the mental image. The results however showed no significant tendency to interact preferentially with the fully intact nest during any of the playback treatments. This suggests that the reception of provisioning calls does not evoke the visual mental image of a nest in chestnut crowned babblers.

If indeed the visual mental image hypothesis is to be excluded, then understanding the precise cognitive mechanism underlying signal reception in chestnut crowned babblers requires the formulation of alternative operant definitions of the "meaning" of babbler vocal signals. On one account, this call could merely represent the direct outcome of a specific internal state in the sender, via motivationally-driven changes in vocal fold tension and respiration; in turn, the physical properties of this acoustic signal could affect the emotional or motivational state of the receiver, triggering a specific response (Owren & Rendall 1997; Owren *et al.* 2003). This idea, championed by Rendall *et al.* (2009), abandons the implication of a conceptual elaboration of the meaning of the signal altogether, in favour of an affect-driven response. In support of this view, the direct link between vocal signals and emotional states has long been established in the non-human vocal communication literature. Decades ago, Morton (1977) and Tembrock (1975) identified a series of acoustic correlates of affective-motivational states that are consistently observed across different

species and even distinct taxa. For example, relaxation and contentment are reliably associated with short sounds and low frequencies, while fear and defensive behaviour correlate with short, high-amplitude, tonelike calls with high frequency onset and upward frequency modulation (Scherer 1992).

Not only has it been shown that certain emotional and motivational states consistently result in the production of specific acoustic features, it has also been empirically demonstrated that receivers attend to these emotionally-driven acoustic features and respond with specific behaviours. For instance, the rate at which fallow bucks (*Dama dama*) groan varies in relation to levels of arousal and fatigue. Moreover, bucks on the receiving end produce different behavioural responses to playbacks of high-rate vs low-rate groans, indicating that groans are perceived as honest sources of information regarding the motivational state of the sender (Pitcher *et al.* 2014). Therefore, the view that some animal vocal signals are produced on the basis of variation in the affective-motivational state of the sender, and interpreted via influence on the receiver's own motivational state, without an intervening semantic representation of the referent, is firmly grounded in the scientific literature. However, such a purely affective model fails to adequately explain the context-specificity of vocal signals and sender-receiver dynamics in chestnut crowned babblers: indeed, provisioning calls are specific to an external, rather than internal context, i.e. chick provisioning events at babbler nests (Crane *et al.* 2016). In addition, these calls have shown to reliably elicit nest-looking behaviour in receivers, further indicating an external referent (Engesser *et al.* 2015; see also Chapter 2).

As in chestnut crowned babblers, many vocal signals used by mammals and other birds are reliably associated with specific stimuli or events that are external to both sender and receiver (reviews in Townsend & Manser 2013, Smith 2017). For example, the three types of alarm calls of vervet monkeys (*Chlorocebus pygerythrus*) are highly specific to certain classes of predators (Strushaker 1967, Cheney & Seyfarth 1990), and the five types of agonistic recruitment calls of rhesus macaques (*Macaca mulatta*) designate different classes of opponents in terms of rank and relatedness (Seyfarth *et al.* 1980a; Gouzoules *et al.* 1984). The "seet" calls of yellow warblers (*Setophaga petechia*), similarly, are produced specifically in the presence of brood-parasitic brown-headed cowbirds (*Molothrus ater*), and consistently result in females rushing to sit in their nests (Gill & Sealy 2004). The fundamental message of this work is

that some non-human vocal signals encode precise information regarding specific external objects or events, rather than merely depicting the sender's internal state. This information occurs in the form of statistical regularities in the co-occurrence of certain signals and specific events, so the ability to acquire this information must depend on a cognitive process which involves recognising and measuring these statistical associations in the natural environment (Marler 1961, Seyfarth *et al.* 1980, Cheney & Seyfarth 1988, Townsend *et al.* 2010, Fischer 2013). Thus a more plausible alternative to an affect-conditioning model for the provisioning call of babblers is that this acoustic signal generates specific behavioural responses in receivers as a function of its statistical association with an external event, i.e. the provisioning of chicks within the nest. As such, it may contain a "natural" meaning (Grice 1967, Scarantino 2010), i.e. it carries information that is grounded in natural correlations between events in the environment. This operant definition of the "meaning" of context-specific signals, such as the babbler provisioning call, may represent the most parsimonious explanation for the cognitive process underlying signal reception in non-humans (Bond 2019), and specifically, in chestnut crowned babblers.

Apart from affect-conditioning and associative learning, there is one more candidate cognitive process for explaining the reception of provisioning calls in chestnut crowned babblers that is worth considering in light of the findings presented in this study. This mechanism implies more than just a learned association between an event and an acoustic signal, in fact it entails the conceptual elaboration of meaning in terms of the mental representation of a precise signal referent. However, it does not imply that the referent must be represented as a visual mental image specifically, but rather it remains agnostic as to the precise nature of this semantic elaboration. This is the view espoused by Zuberbuhler, Seyfarth and colleagues (Zuberbuhler *et al.* 1999, Seyfarth *et al.* 2010) and referred to as "conceptual semanticity", as described earlier on in this chapter. With carefully designed experimental procedures, multiple studies have empirically tested this hypothesis and obtained results indicating that receivers attend to a precise, representational meaning, rather than simply the acoustic properties of signals, during vocal communicative events (Zuberbuhler *et al.* 1999, Evans & Evans 2006). For example, in the study on diana monkeys, receivers transferred habituation between signals that differed in their physical features but shared a semantic message, i.e. the presence of a predatory threat.

Crucially, an association-based perspective would predict that, if receivers learned to simply respond to specific stimuli with target responses, they would not transfer habituation between distinct acoustic stimuli.

A very similar experiment to that conducted on diana monkeys, involving manipulation of experience and playback, was also conducted on fowl (*Gallus gallus domesticus*) (Evans & Evans 2006). In this study, hens in the baseline condition were simply exposed to playback of food calls and behavioural responses were recorded in terms of time spent in close frontal inspection of the substrate, a behaviour characteristic of the search for food. In the treatment condition, hens were given a small amount of food (three kernels) before being exposed to playback of food calls. The authors predicted that if hens attended to the meaning of food calls via a mental representation of a feeding opportunity, they should be less likely to exhibit food searching behaviour in response to food calls if they have recently received a small portion of food, as the putative meaning of the playback stimulus is redundant with prior experience. Indeed, hens exhibited significantly less food searching behaviour in the treatment condition compared to the baseline. Thus as in the diana monkey example, the authors concluded that behavioural responses were determined by the putative semantic information contained in the playback stimuli, processed in the form of a mental representation of the meaning of the call, rather than a statistical association between a signal and a particular event.

In neither the diana monkey nor the fowl study is an attempt made to describe the precise nature of the mental representation itself, other than as a mental concept of a specific predatory threat or a feeding opportunity. This open-ended definition of conceptual semanticity may therefore provide an alternative explanation to the findings presented in this study regarding reception of the provisioning call in chestnut crowned babblers. In fact, according to this definition, when hearing provisioning calls babblers may generate an internal representation of the nest that is not tied to its component visual features. For example, babblers may process the meaning of provisioning calls in terms of the function of the nest, i.e. to mediate reproduction via incubation and chick rearing. In this case, they would not be predicted to exhibit preferences of visual interaction between nests differing in their component visual features, unless these visual features were directly linked to the functionality of the different nests. Thus, as conceptual semanticity has offered a plausible explanation for

patterns in behavioural data even where associative learning and affect-conditioning models have fallen short, in multiple cases including an avian species, it remains a valuable alternative hypothesis that may yet be applied successfully to the communication system of chestnut crowned babblers.

With this study, I have aligned with a recently matured and ambitious research objective in the field of comparative communication, i.e. to test for cognitive mechanisms underlying vocal communication that involve the mental representation of signal referents. Specifically, I have drawn from a novel experimental design outlined by Suzuki (2018) aimed at the investigation of visual mental images upon signal reception. This procedure involves the playback of natural acoustic stimuli, accompanied by the opportunity for subjects to interact with carefully selected physical models of signal referents. Following a critique to Suzuki's (2018) paper from Bond (2019), I have implemented additional features into the experimental setup in order to establish with greater certainty the role of visual mental images in signal reception. Specifically, I used physical models of signal referents that differed in their component visual features, so as to tease out more precisely what kind of sensory information might be processed in the form of a visual mental image. Finally, I have applied further modifications to the experimental procedure stemming from personal initiative. These additions, which include the use of artificially generated playback stimuli that exhibit variation in their acoustic similarity to the target context-specific vocal signal, are designed to provide more information on the cueing mechanism resulting in the formation of a visual mental image of signal referents. Specifically, I hypothesised that by relating behavioural responses to acoustic stimuli to the degree of similarity of these artificial stimuli to the target signal, it is possible to distinguish between a graded cueing mechanism and a threshold mechanism for the generation of visual mental images. I consequently propose that further empirical tests of visual mental images should implement both: a) variation in the component visual features of the signal referent models; as well as b) graded variation in the component acoustic features of the playback stimuli.

On a summarising note, I suggest that the experimental procedure outlined above, designed to test for visual mental images associated with signal reception, may equally be applied to the study of other forms of mental representation of signal referents. For example, if one is to investigate the mental

representation of conceptual information upon reception of signal referents, this conceptual information may be operationally defined as the behaviourally-relevant function of the referent, rather than its visual features. In this case, an experimental design could be used which offers subjects the chance to interact with models that differ in functional, rather than visual qualities. One prediction of the processing of such conceptual information would be that subjects who generate a mental representation of the function of a signal's referent should be primed to detect model referents that match this function rather than other available model referents. Thus overall, Suzuki's (2018) novel approach has potentially opened the door for multiple avenues of research into the cognitive processes underlying signal reception, other than just visual mental images.

Despite the dearth of available evidence for the mental representation of signal referents in non-human animals, whether in visual or other forms, this research question deserves continued and systematic attention. Linguists, anthropologists and biologists have yet to uncover the various early steps in the evolutionary history of a key behavioural trait which has significantly aided our species in becoming so successful: language (Aitchison 2000). However, it is evident that human language is heavily based on highly referential signals that are associated with elaborate mental representations, both visual and discursive (Marler *et al.* 1992; Premack 1984). Thus even research into the vocal communication of very distantly related species, insofar as it relates to the ecological drivers, the proximate mechanisms and the cognitive underpinnings of complex, representational vocalisations, has the potential to increase our understanding of both vocal communication in general, as well as the evolution of human language.

Chapter 4

Conclusion

In these studies on the acoustic communication of the chestnut crowned babbler (*Pomatostomus ruficeps*), a cooperatively breeding bird of the Australian outback, I have examined several cognitive mechanisms underpinning the reception of complex combinatorial signals. Specifically, I first asked whether individuals possess strict or loose acoustic templates for their vocalisations. I have found that subjects exhibit increasingly appropriate behavioural responses to acoustic signals when the perceived call is more similar to the corresponding acoustic template: this suggests that individuals do not possess strict templates, but are sensitive to the similarity between signals and templates. Next, I tested whether chestnut crowned babblers exhibit evidence for pragmatic inference, i.e. the ability to integrate contextual information upon reception of acoustic signals. I have shown that individuals respond differently to ambiguous stimuli resembling the provisioning call when primed with the correct provisioning call, than when the ambiguous stimuli are presented alone. Therefore, the reception of past acoustic stimuli can be used as contextual information to disambiguate the meaning of future, uncertain signals in this species. Finally, I asked whether the reception of functionally referential signals in *P. ruficeps* is associated with a visual mental image of signal referents. The results indicate that the reception of provisioning calls in babblers does not prime receivers to detect model referents possessing specific component visual features, which suggests that babblers do not form visual mental images of signal referents. Together, these findings strengthen our understanding of the combinatorial rules that underpin the production of babbler vocalisations, as well as the cognitive mechanisms employed by receivers to process combinatorial acoustic signals.

Within the field of vocal communication, we as humans are by no means unique in our ability to generate complex, multi-element vocalisations (Engesser & Townsend 2019). In addition, the capacity to perform mental computations to tie together the elements within a vocalisation, both on the producing and the receiving end, is also not a distinctly human trait, but a widespread characteristic (Fitch 2018, ten Cate 2017). However, the grammatical rules underpinning the production of complex vocalisations across animal species exhibit considerable variation, and are often markedly different from those employed in human

languages (Engesser & Townsend 2019). Some of these rules for the generation of sound combinations are well understood, such as the production of meaning-modifying affixes in Campbell's monkeys (*Cercopithecus campbelli campbelli*) and southern pied babblers (*Turdoides bicolor*) (Ouattara *et al.* 2009, Engesser *et al.* 2018). In other cases, the operations governing sound combination and the generation of meaningful signals remain unclear, as in the case of killer whales (*Orcinus orca*) (Shapiro *et al.* 2011). Furthermore, whereas many of the external features of animal sound combinations have been described in detail, there is much less certainty regarding the cognitive mechanisms and computations that underlie the interpretation of acoustic signals in non-human animals (Wheeler and Fischer 2012).

This study on chestnut crowned babblers provides novel evidence for the grammatical rules and cognitive abilities that underpin the reception of combinatorial signals in an avian species. The first important contribution of this study consists in a deeper understanding of the nature of phonemic contrasts employed by a non-human. Chestnut crowned babblers currently represent the only known species, other than *Homo sapiens*, in which meaningful signals are generated via the rearrangement of meaningless, shared acoustic elements (Engesser *et al.* 2015, Engesser *et al.* 2019). This ability lies at the foundation of the immense generative power of human language: it is the basic principle that underlies the formation of words, by which a limited number of sounds can be organised into an unlimited number of signals. The study of chestnut crowned babbler vocalisations therefore allows a direct assessment of the similarities and differences in human and non-human phoneme use. The first key point that emerges from this study is a substantial difference between the reception of phoneme-based signals in humans and babblers. In humans, vocal signals adhere to strict acoustic templates, such that even minor modifications to existing words result in an alteration or loss of meaning (Yip 2006, Chomsky & Halle 1968). On the other hand, as shown in Chapter 1, a signal in the babbler repertoire may undergo modifications of its phoneme structure and still elicit qualitatively equivalent behavioural responses to the unmodified version. In sum, whereas humans possess strict acoustic templates for phoneme-based signals, chestnut crowned babblers exhibit a loose association between signal and template, determined by a threshold similarity to the correct call.

The phonological layer of acoustic combination present in human speech, by which multiple meaningless acoustic elements are combined like building blocks to generate a vast array of signals, most likely arose via gradual increases in complexity over evolutionary time (Engesser *et al.* 2015, Zuidema & de Boer 2009). In support of this hypothesis, some forms of human language that are recently evolved, such as the Al Sayyid Bedouin sign language, possess a vastly reduced phonological component compared to languages with a longer evolutionary history (Sandler *et al.* 2011, 2014). This would seem to suggest that the evolution of human language passed through a stage where the use of phoneme-like elements was less complex and therefore more restricted in its generative power. Our current knowledge of early forms of human language is limited, and yet a comparative approach can potentially provide insights into the features of vocal communication and sound combination that defined early forms of human speech (Bolhuis *et al.* 2018).

The nature of babbling phoneme use is rudimentary in comparison with human language phonology. The phonemes used in human vocal communication, such as the sound /n/, are re-used productively across a large number of words. By contrast, the combination of meaningless acoustic elements in babblers is non-productive: certain acoustic elements are limited to a subset of the repertoire, rather than being used flexibly across the range of vocal signals. For example, the A and B elements under investigation in this study are present only in the flight and provisioning calls, i.e. roughly 10% of the entire repertoire. The result of this non-productivity in phonemic combination is that most other babbling calls exhibit substantially low acoustic similarity compared to the flight and provisioning calls, as they are composed of different acoustic elements. This between-call dissimilarity may help to explain the finding that babblers exhibit loose, rather than strict, associations between signals and templates. Indeed, an interesting possibility that emerges from these results is that the strictness of the relationship between signals and templates may depend on the productive use of the acoustic elements that make up the signal. When there is no productivity and each call type differs substantially from the others, such as in babbling calls, templates may be relaxed; conversely, when the constituent elements are used productively, such as in human word formation, templates must be strict. This prediction could be tested in the laboratory by conducting a two-step artificial grammar learning (AGL) test on human subjects:

step 1) familiarising subjects with artificial vocabularies which exhibit different levels of productivity, step 2) conducting articulation tests (*sensu* Miller *et al.* 1951) involving the recognition of signals which loosely and strictly adhere to the artificial vocabularies. Subjects familiarised with non-productive vocabularies should recognise loosely adherent stimuli significantly more than those exposed to productive vocabularies. If this prediction were upheld, it would suggest that strict acoustic templates in human vocal communication are only as recent as the emergence of productivity in phoneme use.

The second key finding of this study regards the use of information arising from the context in which signals are emitted. This ability, known as pragmatic inference, has been documented in a number of primate species, yet few studies to date have addressed its presence in birds. Pragmatic inference constitutes one of the most important mechanisms involved in the interpretation of acoustic stimuli, and is fundamental in the case of signals which exhibit a low information content (Smith 1977, Cheney & Seyfarth 1988). Such signals are abundant in the repertoires of many species, and are often the result of constraints on signal production. For example, the vocal repertoires of non-human primates are constrained by both the neurobiological circuits involved in the vocal pathway, as well as the morphology of the vocal apparatus (Hammerschmidt & Fischer 2008, Wheeler & Fischer 2012). In some cases, this results in a limited number of call types being used in a large number of contexts (e.g. Arnold & Zuberbuhler 2013), which reduces the potential for information transfer of the signals themselves (Cheney & Seyfarth 2010). However, despite constraints on signal production, behavioural responses to vocal signals are often flexible and complex. Indeed, the responses to acoustic stimuli of vervet monkeys (*Chlorocebus pygerythrus*), putty-nosed monkeys (*Cercopithecus nictitans*), baboons (*Papio ursinus*) and green monkeys (*Chlorocebus sabaeus*), are strongly affected by prior experience (Cheney & Seyfarth 1988, Arnold & Zuberbuhler 2013, Crockford *et al.* 2007, Price & Fischer 2014). For example, subjects may take into account the existing social relationships, past interactions, or the presence of additional visual or acoustic stimuli, when responding to signals. Therefore, in order to exhibit flexible adaptive responses to signals with a low information content, receivers depend on the ability to extract information not only from the signal, but also from the context in which the signal is given.

Ambiguity in acoustic communication need not arise solely from vocal constraints, but may also be a product of errors in signal production and reception, which are more broadly applicable across the range of taxa that communicate via acoustic signals (Wiley 1994). For example, environmental factors such as wind and background noise degrade the quality of acoustic signals, and therefore reduce the information available in such signals (Dubois & Martens 1984, Brumm & Slabbekoorn 2005). Furthermore, species exhibiting complex combinatorial vocalisations are susceptible to errors in signal production. In particular, where signals are generated via the combination of specific acoustic elements in a given order, such as in human word formation, inconsistencies in element sequence may occur (Frisch & Wright 2002). Chestnut-crowned babblers also utilise such combinatorial signals, and indeed have been observed to produce signals in which the arrangement of acoustic elements differs from that of the natural templates. For example, babblers occasionally combine the “A” and “B” elements used in the flight and provisioning calls into aberrant sequences, such as reverse flight calls “BA” (Crane *et al.* 2016) or unusual 3-element sequences, such as “AAB” and “ABA” (Joseph Mine, unpublished data). Here I have shown that the response to such aberrant combinations is influenced by prior exposure to signals exhibiting the more common, “correct” note sequence. The importance of this cognitive mechanism in the context of acoustic communication has been established in a number of mammalian species (Seyfarth & Cheney 2017), yet this study constitutes one of the first explicit tests of pragmatic inference in an avian species. The most commonly invoked pressure driving the development of this mechanism in mammals is the inflexibility of signal production due to vocal constraints (e.g. Arnold & Zuberbuhler 2013). By contrast, where pragmatic inference occurs in avian species, in which the vocal apparatus is highly specialised and flexible (Suthers *et al.* 2016), the selective pressures on its development are likely to be different. Indeed, as shown by this study, pragmatic inference in chestnut-crowned babblers is useful to circumvent the loss of information arising from ambiguities in the production and reception of signals generated via phonemic contrasts. Thus, further research on pragmatic inference in birds may confirm that although distinct taxa have converged upon this cognitive mechanism to aid in vocal communication, the specific selective pressures driving its emergence are distinct. Finally, the link between pragmatic inference and phonemic

contrasts highlighted here may have important implications for understanding the emergence of phoneme-based signals in human language.

The third and final aim of this thesis was to establish whether, in an avian species, the reception of functionally referential signals is associated with the visual mental image of signal referents. This research question stems from an ongoing debate regarding the cognitive processes involved in signal reception across non-human taxa (Stegmann 2013). Major exponents in this debate have argued that either: a) simple cognitive processes, such as nervous-response conditioning or classic associative learning, govern behavioural responses to even the most context-specific stimuli (Owren *et al.* 2010, Chater & Heyes 1994); or b) complex cognitive processes, such as the processing of semantic information in the form of mental representations, underlie the reception of highly referential signals (Cheney & Seyfarth 1988, Evans & Evans 2006, Zuberbuhler *et al.* 1999, Suzuki 2018). These authors have made important theoretical contributions to the field, such as the affect-conditioning model proposed by Owren & Rendall (1997), or the concept of functionally referential signals first promoted by Marler *et al.* (1992). Furthermore, the empirical evidence available highlights the validity of both perspectives: in certain cases, the link between signals and behavioural responses is mediated simply by variation in motivational states or the statistical associations present in nature, while in others, responses are too flexible to be explained solely by associative learning or affect-conditioning (Wheeler & Fischer 2012). This suggests that although simple cognitive mechanisms in signal reception are relevant to some species, the ability to process signal information in a conceptual form may be more widespread than previously thought (Allen 2013).

With the current study, I have not added to the available body of evidence suggesting that highly referential signals are processed via the mental representation of signal referents in non-humans. However, I have replicated and potentially improved a recently developed experimental procedure outlined by Suzuki (2018) and Bond (2019) for the investigation of visual mental images associated with signal referents. The explicit study of mental representations is relatively novel within the field of animal communication, and as such, the relevant methodologies and terminologies have not yet achieved a widely accepted form. Inspired by previous work from Zuberbuhler *et al.* (1999), Evans & Evans (2006) and Suzuki (2018), this test of mental representation was carried

out by exposing subjects to both acoustic signals and their presumed referents, and generating clear predictions regarding the presence or absence of mental representations in processing and responding to such signals. Due to the paucity of replication, the available studies on mental representation of signal referents may not as of yet constitute definitive evidence of visual or conceptual processing in non-humans. However, only via replicating this and other experimental designs can the debate be successfully resolved. Therefore, given the youth of this research topic and the controversy surrounding it, renewed interest in generating, perfecting and replicating relevant experimental procedures is warranted.

As the study of animal acoustic communication progresses, it is becoming increasingly clear that non-human animals are capable of producing vocalisations consisting of multiple parts, constructed and processed according to disparate rules, and used in a variety of biologically relevant functions. Thus it appears that human language, which is without a doubt a highly complex form of vocal communication, is not entirely unique in all of its structural features, and neither in its dependence on specific cognitive processes. The sequences we produce are beyond a doubt highly articulate and deeply informative. Unlike any other species, humans combine phonemes productively to produce an astonishing array of words, and then combine these words into hierarchically-structured phrases to maximise information output (Hurford 2011). This results in a multi-tiered system of acoustic combination referred to as duality of patterning (Hauser *et al.* 2002). Yet even in other taxa, meaningless elements are combined like building blocks to make signals that have informative value (Zuidema & de Boer 2018, Engesser *et al.* 2015). Likewise, meaningful vocal signals are arranged into sequences whose information content depends on all constituent parts (Engesser *et al.* 2016, Suzuki *et al.* 2016). In many other vocally-communicating species, information is extracted not only from the signals themselves, but also from the context in which the signals are emitted (Crockford *et al.* 2007, Cheney & Seyfarth 1988). Finally, growing evidence suggests that non-human receivers may also attend to the semantic qualities of signals by generating mental representations of signal referents, as is the case in human language (Arnold & Zuberbuhler 1999, Suzuki 2018). Thus overall, although duality of patterning may set us a step or two above other vocal species in terms

of combinatorial complexity, it appears that many of the proximate and ultimate mechanisms of human vocal communication are shared with other animals.

Comparisons between human and non-human vocal complexity may appeal to both the scientific and wider audience, yet these must be approached with caution, as they are subject to multiple inherent flaws. First, the field of animal combinatorial communication is still a young one. Great efforts have been made in the past two decades to reveal the types of combinations used in the vocal repertoires of several species, and to understand how these combinations enrich the function of vocalisations (reviewed in Engesser & Townsend 2019). Discoveries such as the first evidence of phonemic contrasts and ordering rules in non-human animal vocalisations are fresher still, having occurred only in the past five years (Engesser *et al.* 2015, Suzuki *et al.* 2017). So it is likely that beyond the species investigated so far, many more possess combinatorial repertoires whose complexity has not yet been uncovered and evaluated. Second, comparisons between human and non-human vocal complexity are most often carried out on biased terms. Research in comparative communication frequently attempts to find examples of human language structures, rules and computations in other communication systems, and declares our system as superior when these standards are not met (e.g. Fitch & Hauser 2004). Yet it is not at all necessary that vocal communicative complexity should develop along the same lines among multiple species with entirely different evolutionary backgrounds. The debate should not be limited to whether or not other species are capable of doing what we do, because there is in essence no reason to expect that the evolution of vocal communication across all taxa should ultimately lead to our own outcome, i.e. a recursive and hierarchical system based on duality of patterning. Rather a more daunting task, research might focus instead on uncovering unfamiliar mechanisms by which substantial amounts of information may be conveyed acoustically. Although it may require great flexibility of mind and often incur frustration, this is the only way the field of vocal communication may be fully and effectively explored. Should this kind of approach be successfully implemented, we may in fact discover that evolution has produced other extraordinarily complex systems of vocal communication, alongside our own.

Acknowledgements

I thank professor Andrew F. Russell profoundly, for allowing me the opportunity to work on this project under his supervision. Since day one, I have felt extremely lucky to be a part of his team. He has constantly inspired and motivated me, and significantly contributed to the development of my capacities of critical and creative thinking, scientific writing, statistical analysis, data collection, networking and much more.

I thank Dr. Camille Coye who, despite not being my official supervisor, has dedicated vast amounts of her time and energy to me and this project.

Throughout the year, she has always been available, friendly and helpful, suggesting relevant literature, discussing ideas and perspectives for write-up, and constantly offering constructive criticism, logistical support and statistical advice.

I thank Dr. Sabrina Engesser for guidance with video coding, data extraction and analysis, as well as inspiration for scientific writing.

I thank professor Simon Townsend for useful conversations about combinatorial communication.

I thank professor Alex Thornton for extremely helpful feedback on previous drafts of this manuscript.

I thank Andrea Liebl, Lee Rollins, Louis O'Neill, Kiara L'Herpinier and Sabrina Engesser for conducting the field work that made this entire study possible.

References

- Aitchison, J. (2000). *The seeds of speech: Language origin and evolution*. Cambridge University Press.
- Akçay, Ç., Reed, V. A., Campbell, S. E., Templeton, C. N., & Beecher, M. D. (2010). Indirect reciprocity: song sparrows distrust aggressive neighbours based on eavesdropping. *Animal Behaviour*, *80*(6), 1041-1047.
- Akçay, Ç., Wood, W. E., Searcy, W. A., Templeton, C. N., Campbell, S. E., & Beecher, M. D. (2009). Good neighbour, bad neighbour: song sparrows retaliate against aggressive rivals. *Animal Behaviour*, *78*(1), 97-102.
- Allen, C. (2013). Information and uncertainty in meerkats and monkeys. *Animal communication theory*. Cambridge University Press, Cambridge, 319-335.
- Arnold, K., & Zuberbühler, K. (2013). Female putty-nosed monkeys use experimentally altered contextual information to disambiguate the cause of male alarm calls. *PLoS One*, *8*(6), e65660.
- Bolhuis, J. J., Beckers, G. J., Huybregts, M. A., Berwick, R. C., & Everaert, M. B. (2018). The slings and arrows of comparative linguistics. *PLoS biology*, *16*(9), e3000019.
- Bond, A. B. (2019). Searching images and the meaning of alarm calls. *Learning & behavior*, *47*(2), 109-110.
- Bradbury, J. W., & Vehrencamp, L. (1998). *Principles of animal communication*. Sunderland: Sinauer Associates.
- Browning, L. E., Patrick, S. C., Rollins, L. A., Griffith, S. C., & Russell, A. F. (2012). Kin selection, not group augmentation, predicts helping in an obligate cooperatively breeding bird. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1743), 3861-3869.
- Browning, L. E., Young, C. M., Savage, J. L., Russell, D. J. F., Barclay, H., Griffith, S. C., & Russell, A. F. (2012). Carer provisioning rules in an obligate cooperative breeder: prey type, size and delivery rate. *Behavioral Ecology and Sociobiology*, *66*(12), 1639-1649.
- Brumm, H., & Slabbekoorn, H. (2005). Acoustic communication in noise. *Advances in the Study of Behavior*, *35*, 151-209.
- Burkart, J. M., Roelli, C., & Richiger, R. (2015). When is behavior intentional? The case of proactive prosociality. In *6th EFP meeting*. *Folia Primatologica* (Vol. 86, pp. 253-254).

- Burt, J. M., Campbell, S. E., & Beecher, M. D. (2001). Song type matching as threat: a test using interactive playback. *Animal Behaviour*, 62(6), 1163-1170.
- Catchpole, C. K., & Slater, P. J. (2003). *Bird song: biological themes and variations*. Cambridge university press.
- Chater, N., & Heyes, C. (1994). Animal concepts: Content and discontent. *Mind & Language*.
- Cheney, D. L., & Seyfarth, R. M. (1988). Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Animal Behaviour*, 36(2), 477-486.
- Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world: Inside the mind of another species*. University of Chicago Press.
- Chomsky, N., & Halle, M. (1968). The sound pattern of English.
- Clark, C. J. (2009). Courtship dives of Anna's hummingbird offer insights into flight performance limits. *Proceedings of the Royal Society B: Biological Sciences*, 276(1670), 3047-3052.
- Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 507-522.
- Cole, R. A. (2016). *Perception and production of fluent speech*. Routledge.
- Cornwallis, C. K., Botero, C. A., Rubenstein, D. R., Downing, P. A., West, S. A., & Griffin, A. S. (2017). Cooperation facilitates the colonization of harsh environments. *Nature ecology & evolution*, 1(3), 0057.
- Coye, C., Townsend, S., & Lemasson, A. (2018). From Animal Communication to Linguistics and Back: Insight from Combinatorial Abilities in Monkeys and Birds. *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, 187.
- Crane, J. M., Pick, J. L., Tribe, A. J., Vincze, E., Hatchwell, B. J., & Russell, A. F. (2015). Chestnut-crowned babblers show affinity for calls of removed group members: a dual playback without expectancy violation. *Animal Behaviour*, 104, 51-57.
- Crane, J. M., Savage, J. L., & Russell, A. F. (2016). Diversity and function of vocalisations in the cooperatively breeding Chestnut-crowned Babbler. *Emu-Austral Ornithology*, 116(3), 241-253.

- Crockford, C., Wittig, R. M., & Zuberbühler, K. (2017). Vocalizing in chimpanzees is influenced by social-cognitive processes. *Science advances*, 3(11), e1701742.
- Crockford, C., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2007). Baboons eavesdrop to deduce mating opportunities. *Animal Behaviour*, 73(5), 885-890.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, 134(2), 222.
- de Beeck, H. P. O., Wagemans, J., & Vogels, R. (2008). The representation of perceived shape similarity and its role for category learning in monkeys: A modeling study. *Vision Research*, 48(4), 598-610.
- DeGange, A. R., Fitzpatrick, J. W., Layne, J. N., & Woolfenden, G. E. (1989). Acorn harvesting by Florida scrub jays. *Ecology*, 70(2), 348-356.
- Dubois, A., & Martens, J. (1984). A case of possible vocal convergence between frogs and a bird in Himalayan torrents. *Journal für Ornithologie*, 125(4), 455-463.
- Dunbar, R. I. (2003). The origin and subsequent evolution of language. *Studies in the evolution of language*, 3, 219-234.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & psychophysics*, 67(2), 224-238.
- Ellis, J. M., Langen, T. A., & Berg, E. C. (2009). Signalling for food and sex? Begging by reproductive female white-throated magpie-jays. *Animal behaviour*, 78(3), 615-623.
- Emlen, S. T. (1982). The evolution of helping. I. An ecological constraints model. *The American Naturalist*, 119(1), 29-39.
- Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *Wiley Interdisciplinary Reviews: Cognitive Science*, e1493.
- Engesser, S., Crane, J. M., Savage, J. L., Russell, A. F., & Townsend, S. W. (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., & Townsend, S. W. (2019). Chestnut-crowned babbler calls are composed of meaningless

- shared building blocks. *Proceedings of the National Academy of Sciences*, 116(39), 19579-19584.
- Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and compositional processing in the southern pied babbler. *Proceedings of the National Academy of Sciences*, 113(21), 5976-5981.
 - Engesser, S., Ridley, A. R., & Townsend, S. W. (2017). Element repetition rates encode functionally distinct information in pied babbler 'clucks' and 'purrs'. *Animal cognition*, 20(5), 953-960.
 - Evans, C. S. (1997). Referential signals. In *Communication* (pp. 99-143). Springer, Boston, MA.
 - Evans, C. S., & Evans, L. (1999). Chicken food calls are functionally referential. *Animal behaviour*, 58(2), 307-319.
 - Evans, C. S., & Evans, L. (2006). Representational signalling in birds. *Biology letters*, 3(1), 8-11.
 - Fabre-Thorpe, M. (2003). Visual categorization: accessing abstraction in non-human primates. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1435), 1215-1223.
 - Fischer, J. (2013). Information, inference and meaning in primate vocal behaviour. In *Animal Communication Theory: Information and Influence*. Cambridge University Press.
 - Fischer, J., Hammerschmidt, K., Cheney, D. L., & Seyfarth, R. M. (2001). Acoustic features of female chacma baboon barks. *Ethology*, 107(1), 33-54.
 - Fischer, J., Metz, M., Cheney, D. L., & Seyfarth, R. M. (2001). Baboon responses to graded bark variants. *Animal Behaviour*, 61(5), 925-931.
 - Fitch, W. T. (2010). *The evolution of language*. Cambridge University Press.
 - Fitch, W. T. (2018). What animals can teach us about human language: the phonological continuity hypothesis. *Current opinion in behavioral sciences*, 21, 68-75.
 - Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, 303(5656), 377-380.
 - Freeberg, T. M. (2006). Social complexity can drive vocal complexity: group size influences vocal information in Carolina chickadees. *Psychological Science*, 17(7), 557-561.

- Frisch, S. A., & Wright, R. (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics*, 30(2), 139-162.
- Garstang, M. (2004). Long-distance, low-frequency elephant communication. *Journal of Comparative Physiology A*, 190(10), 791-805.
- Gifford III, G. W., Hauser, M. D., & Cohen, Y. E. (2003). Discrimination of functionally referential calls by laboratory-housed rhesus macaques: implications for neuroethological studies. *Brain, behavior and evolution*, 61(4), 213-224.
- Gill, S. A., & Bierema, A. M. K. (2013). On the meaning of alarm calls: a review of functional reference in avian alarm calling. *Ethology*, 119(6), 449-461.
- Gill, S. A., & Sealy, S. G. (2004). Functional reference in an alarm signal given during nest defence: seet calls of yellow warblers denote brood-parasitic brown-headed cowbirds. *Behavioral Ecology and Sociobiology*, 56(1), 71-80.
- Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, 20(11), 818-829.
- Gouzoules, H., Gouzoules, S., & Ashley, J. (1995). Representational signaling in non-human primate vocal communication. In *Current topics in primate vocal communication* (pp. 235-252). Springer, Boston, MA.
- Gouzoules, H., Gouzoules, S., & Marler, P. (1985). External reference and affective signaling in mammalian vocal communication. In *The development of expressive behavior*(pp. 77-101). Academic Press.
- Gouzoules, S., Gouzoules, H., & Marler, P. (1984). Rhesus monkey (*Macaca mulatta*) screams: representational signalling in the recruitment of agonistic aid. *Animal behaviour*, 32(1), 182-193.
- Gregg, J., Inhoff, A. W., & Connine, C. M. (2019). Re-reconsidering the role of temporal order in spoken word recognition. *Quarterly Journal of Experimental Psychology*, 1747021819849512.
- Grice, H. P. (1957). Meaning en *Philosophical Review*, 66 (3). *Reprinted as*, 213-223.
- Grüter, C., & Farina, W. M. (2009). The honeybee waggle dance: can we follow the steps?. *Trends in Ecology & Evolution*, 24(5), 242-247.

- Hammerschmidt, K., & Fischer, J. (1998). The vocal repertoire of Barbary macaques: a quantitative analysis of a graded signal system. *Ethology*, 104(3), 203-216.
- Hammerschmidt, K., & Fischer, J. Constraints in primate vocal production. The evolution of communicative creativity: From fixed signals to contextual flexibility. Edited by: Griebel U, Oller K. 2008.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598), 1569-1579.
- Hurford, J. R. (2007). *Language in the Light of Evolution: The Origins of Meaning*. Oxford University Press.
- Hurford, J. R. (2011). *The origins of grammar: Language in the light of evolution II*. OUP Oxford.
- Ihaka, R., & Gentleman, R. (1996). R: a language for data analysis and graphics. *Journal of computational and graphical statistics*, 5(3), 299-314.
- Jakobson, R. (2012). *Child language, aphasia and phonological universals* (Vol. 72). Walter de Gruyter.
- Jürgens, U. (1979). Vocalization as an emotional indicator A neuroethological study in the squirrel monkey. *Behaviour*, 88-117.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive psychology*, 29(1), 1-23.
- Keenan, S., Lemasson, A., & Zuberbühler, K. (2013). Graded or discrete? A quantitative analysis of Campbell's monkey alarm calls. *Animal behaviour*, 85(1), 109-118.
- Kight, C. R., McNamara, J. M., Stephens, D. W., & Dall, S. R. (2013). Communication as information use: Insights from statistical decision theory. In *Animal Communication Theory: Information and Influence*. Cambridge University Press.
- Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature neuroscience*, 3(9), 946.
- Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (Pan troglodytes). *Child Development*, 76(1), 291-306.

- Leighton, G. M. (2017). Cooperative breeding influences the number and type of vocalizations in avian lineages. *Proceedings of the Royal Society B: Biological Sciences*, 284(1868), 20171508.
- Lieberman, P. (1984). *The biology and evolution of language*. Harvard University Press.
- Lieberman, P. (2006). *Toward an evolutionary biology of language*. Harvard University Press.
- Liebl, A. L., Nomano, F. Y., Browning, L. E., & Russell, A. F. (2016). Experimental evidence for fully additive care among male carers in the cooperatively breeding chestnut-crowned babbler. *Animal Behaviour*, 115, 47-53.
- Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5(5), 552-563.
- Logothetis, N. K., Pauls, J., Bülthoff, H. H., & Poggio, T. (1994). View-dependent object recognition by monkeys. *Current biology*, 4(5), 401-414.
- Marler, P. (1961). The logical analysis of animal communication. *Journal of Theoretical Biology*, 1(3), 295-317.
- Marler, P. (1980). Primate vocalization: affective or symbolic?. In *Speaking of apes* (pp. 221-229). Springer, Boston, MA.
- Marler, P. (1984). Animal communication: affect or cognition. *Approaches to emotion*, 345-365.
- Marler, P., Evans, C. S., & Hauser, M. D. (1992). Animal signals: motivational, referential, or both. *Nonverbal vocal communication: Comparative and developmental approaches*, 66-86.
- Mashburn, L. M., & Whiteley, M. (2005). Membrane vesicles traffic signals and facilitate group activities in a prokaryote. *Nature*, 437(7057), 422.
- McDonald, P. G., & Wright, J. (2011). Bell miner provisioning calls are more similar among relatives and are used by helpers at the nest to bias their effort towards kin. *Proceedings of the Royal Society B: Biological Sciences*, 278(1723), 3403-3411.
- Meise, K., Keller, C., Cowlshaw, G., & Fischer, J. (2011). Sources of acoustic variation: implications for production specificity and call categorization in chacma baboon (*Papio ursinus*) grunts. *The Journal of the Acoustical Society of America*, 129(3), 1631-1641.

- Miller, G. A., Heise, G. A., & Lichten, W. (1951). The intelligibility of speech as a function of the context of the test materials. *Journal of experimental psychology*, 41(5), 329.
- Morton, E. S. (1975). Ecological sources of selection on avian sounds. *The American Naturalist*, 109(965), 17-34.
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *The American Naturalist*, 111(981), 855-869.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47(2), 204-238.
- Noveck, I. A., Brooke Lea, R., Davidson, G., & O'Brien, D. P. (1991). Human reasoning is both logical and pragmatic. *Intellectica*, 11(1), 81-109.
- Nowak, M. A., & Krakauer, D. C. (1999). The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14), 8028-8033.
- Nowak, M. A., Krakauer, D. C., & Dress, A. (1999). An error limit for the evolution of language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1433), 2131-2136.
- Otter, K. A., Atherton, S. E., & Van Oort, H. (2007). Female food solicitation calling, hunger levels and habitat differences in the black-capped chickadee. *Animal Behaviour*, 74(4), 847-853.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys use affixation to alter call meaning. *PloS one*, 4(11), e7808.
- Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocal signaling. In *Communication* (pp. 299-346). Springer, Boston, MA.
- Owren, M. J., & Rendall, D. (2001). Sound on the rebound: bringing form and function back to the forefront in understanding nonhuman primate vocal signaling. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, 10(2), 58-71.
- Owren, M. J., Rendall, D., & Bachorowski, J. A. (2003). Nonlinguistic vocal communication. *Primate psychology*, 359-394.
- Owren, M. J., Rendall, D., & Ryan, M. J. (2010). Redefining animal signaling: influence versus information in communication. *Biology & Philosophy*, 25(5), 755-780.

- Palombit, R. A., Seyfarth, R. M., & Cheney, D. L. (1997). The adaptive value of 'friendships' to female baboons: experimental and observational evidence. *Animal Behaviour*, 54(3), 599-614.
- Pan, X., & Sakagami, M. (2012). Category representation and generalization in the prefrontal cortex. *European Journal of Neuroscience*, 35(7), 1083-1091.
- Pearson, D., Donkin, C., Tran, S. C., Most, S. B., & Le Pelley, M. E. (2015). Cognitive control and counterproductive oculomotor capture by reward-related stimuli. *Visual Cognition*, 23(1-2), 41-66.
- Pinker, S. (2010). The cognitive niche: Coevolution of intelligence, sociality, and language. *Proceedings of the National Academy of Sciences*, 107(Supplement 2), 8993-8999.
- Pitcher, B. J., Briefer, E. F., Vannoni, E., & McElligott, A. G. (2014). Fallow bucks attend to vocal cues of motivation and fatigue. *Behavioral Ecology*, 25(2), 392-401.
- Plotkin, J., & Nowak, M. (2001). Major transitions in language evolution. *Entropy*, 3(4), 227-246.
- Polansky, L., Kilian, W., & Wittemyer, G. (2015). Elucidating the significance of spatial memory on movement decisions by African savannah elephants using state–space models. *Proceedings of the Royal Society B: Biological Sciences*, 282(1805), 20143042.
- Portelli, D. J., Barclay, H., Russell, D. J., Griffith, S. C., & Russell, A. F. (2009). Social organisation and foraging ecology of the cooperatively breeding Chestnut-crowned Babbler (*Pomatostomus ruficeps*). *Emu-Austral Ornithology*, 109(2), 153-162.
- Premack, D. (1984). Comparing mental representation in human and nonhuman animals. *Social Research*, 51(4), 985.
- Premack, D. (1984). Possible general effects of language training on the chimpanzee. *Human Development*, 27(5-6), 268-281.
- Premack, D., & Premack, A. J. (1983). *The mind of an ape*. New York: Norton.
- Price, T., & Fischer, J. (2014). Meaning attribution in the West African green monkey: influence of call type and context. *Animal cognition*, 17(2), 277-286.

- Price, T., Wadewitz, P., Cheney, D., Seyfarth, R., Hammerschmidt, K., & Fischer, J. (2015). Vervets revisited: A quantitative analysis of alarm call structure and context specificity. *Scientific reports*, 5, 13220.
- Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean?. *Animal Behaviour*, 78(2), 233-240.
- Russell, A. F. (2016). Chestnut-crowned babbler: dealing with climatic adversity and uncertainty in the Australian arid zone. *Cooperative breeding in vertebrates: studies in ecology, evolution and behavior*, 150-164.
- Russell, A. F., Portelli, D. J., Russell, D. J. F., & Barclay, H. (2010). Breeding ecology of the chestnut-crowned babbler: a cooperative breeder in the desert. *Emu-Austral Ornithology*, 110(4), 324-331.
- Saidel, E. (2009). Attributing mental representations to animals. *The philosophy of animal minds*, 35-51.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207-1218.
- Sandler, W., Aronoff, M., Meir, I., & Padden, C. (2011). The gradual emergence of phonological form in a new language. *Natural language & linguistic theory*, 29(2), 503-543.
- Sandler, W., Aronoff, M., Padden, C., Meir, I., Enfield, N., Kockelman, P., & Sidnell, J. (2014). Language emergence: Al-Sayyid bedouin sign language. *Cambridge handbook of linguistic anthropology*, 250-284.
- Scarantino, A. (2010). Animal communication between information and influence. *Animal Behaviour*, 79(6), e1.
- Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PloS one*, 8(10), e76674.
- Scherer, K. R. (1992). Vocal affect expression as symptom, symbol, and appeal. *Nonverbal vocal communication: Comparative and developmental approaches*, 43-60.
- Schnitzler, H. U., & Denzinger, A. (2011). Auditory fovea and Doppler shift compensation: adaptations for flutter detection in echolocating bats using CF-FM signals. *Journal of Comparative Physiology A*, 197(5), 541-559.
- Scott-Phillips, T. C., Cartmill, E. A., Crockford, C., Gärdenfors, P., Gómez, J. C., Luef, E. M., ... & Sievers, C. (2015). Nonhuman primate communication,

- pragmatics, and the origins of language. *Current Anthropology*, 56(1), 000-000.
- Sereno, M. E., Trinath, T., Augath, M., & Logothetis, N. K. (2002). Three-dimensional shape representation in monkey cortex. *Neuron*, 33(4), 635-652.
 - Seyfarth, R. M., & Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annual review of psychology*, 54(1), 145-173.
 - Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain and language*, 115(1), 92-100.
 - Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current opinion in neurobiology*, 28, 5-9.
 - Seyfarth, R. M., & Cheney, D. L. (2017). Precursors to language: Social cognition and pragmatic inference in primates. *Psychonomic bulletin & review*, 24(1), 79-84.
 - Seyfarth, R. M., & Cheney, D. L. (2017). The origin of meaning in animal signals. *Animal Behaviour*, 124, 339-346.
 - Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science*, 210(4471), 801-803.
 - Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Animal Behaviour*, 28(4), 1070-1094.
 - Seyfarth, R. M., Cheney, D. L., Bergman, T., Fischer, J., Zuberbühler, K., & Hammerschmidt, K. (2010). The central importance of information in studies of animal communication. *Animal Behaviour*, 80(1), 3-8.
 - Shapiro, A. D., Tyack, P. L., & Seneff, S. (2011). Comparing call-based versus subunit-based methods for categorizing Norwegian killer whale, *Orcinus orca*, vocalizations. *Animal behaviour*, 81(2), 377-386.
 - Sharp, S. P., McGowan, A., Wood, M. J., & Hatchwell, B. J. (2005). Learned kin recognition cues in a social bird. *Nature*, 434(7037), 1127.
 - Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, 415(6869), 318.
 - Sison, M., & Gerlai, R. (2010). Associative learning in zebrafish (*Danio rerio*) in the plus maze. *Behavioural brain research*, 207(1), 99-104.
 - Skyrms, B. (2010a). The flow of information in signaling games. *Philosophical Studies*, 147(1), 155.

- Skyrms, B. (2010b). *Signals: Evolution, learning, and information*. Oxford University Press.
- Smit, A. B. (1993). Phonologic error distributions in the Iowa-Nebraska articulation norms project: Consonant singletons. *Journal of Speech, Language, and Hearing Research*, 36(3), 533-547.
- Smith, C. L. (2017). Referential signalling in birds: The past, present and future. *Animal Behaviour*, 124, 315-323.
- Smith, W. J. (1977). *The behaviour of communicating: an ethological approach*. Harvard University press.
- Smith, W. J. (1980). *The Behavior of Communicating: An Ethological Approach*. Cambridge,
- Sorato, E., Gullett, P. R., Creasey, M. J., Griffith, S. C., & Russell, A. F. (2015). Plastic territoriality in group-living chestnut-crowned babbler: roles of resource value, holding potential and predation risk. *Animal Behaviour*, 101, 155-168.
- Stegmann, U. (Ed.). (2013). *Animal communication theory: information and influence*. Cambridge University Press.
- Strushaker, T. T. (1967). Auditory communication among vervet monkeys (*Cercopithecus aethiops*) in SA Altmann, ed., *Social Communication among Primates*.
- Suthers, R. A. (2016). *Vertebrate sound production and acoustic communication* (Vol. 53). W. T. Fitch, R. R. Fay, & A. N. Popper (Eds.). Heidelberg: Springer.
- Suzuki, T. N. (2012). Referential mobbing calls elicit different predator-searching behaviours in Japanese great tits. *Animal Behaviour*, 84(1), 53-57.
- Suzuki, T. N. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Animal Behaviour*, 87, 59-65.
- Suzuki, T. N. (2018). Alarm calls evoke a visual search image of a predator in birds. *Proceedings of the National Academy of Sciences*, 115(7), 1541-1545.
- Suzuki, T. N., & Ueda, K. (2013). Mobbing calls of Japanese tits signal predator type: field observations of natural predator encounters. *The Wilson Journal of Ornithology*, 125(2), 412-415.
- Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature communications*, 7, 10986.

- Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2017). Wild birds use an ordering rule to decode novel call sequences. *Current Biology*, 27(15), 2331-2336.
- Tanaka, K., Saito, H. A., Fukada, Y., & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of neurophysiology*, 66(1), 170-189.
- Tembrock, G. (1975). Die Erforschung des tierlichen Stimmausdrucks (Bioakustik). *Biophonetik. Mannheim: Bibliographisches Institut*, 51-67.
- Templeton, C. N., Greene, E., & Davis, K. (2005). Allometry of alarm calls: black-capped chickadees encode information about predator size. *Science*, 308(5730), 1934-1937.
- ten Cate, C. (2017). Assessing the uniqueness of language: Animal grammatical abilities take center stage. *Psychonomic bulletin & review*, 24(1), 91-96.
- Thornton, A., & Boogert, N. J. (2019). Animal Cognition: The Benefits of Remembering. *Current Biology*, 29(9), R324-R327.
- Thornton, A., & McAuliffe, K. (2006). Teaching in wild meerkats. *Science*, 313(5784), 227-229.
- Thornton, A., Clayton, N. S., & Grodzinski, U. (2012). Animal minds: from computation to evolution.
- Toscano, J. C., Anderson, N. D., & McMurray, B. (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic bulletin & review*, 20(5), 981-987.
- Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: the past, present and the future. *Ethology*, 119(1), 1-11.
- Townsend, S. W., Allen, C., & Manser, M. B. (2011). A simple test of vocal individual recognition in wild meerkats. *Biology letters*, 8(2), 179-182.
- Townsend, S. W., Hollén, L. I., & Manser, M. B. (2010). Meerkat close calls encode group-specific signatures, but receivers fail to discriminate. *Animal Behaviour*, 80(1), 133-138.
- Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., ... & Glock, H. J. (2017). Exorcising G. Rice's ghost: an empirical approach to studying intentional communication in animals. *Biological Reviews*, 92(3), 1427-1433.

- Tyack, P. L. (2016). Vocal learning and auditory-vocal feedback. In *Vertebrate sound production and acoustic communication* (pp. 261-295). Springer, Cham.
- Uésugi, K. (1996). The adaptive significance of Batesian mimicry in the swallowtail butterfly, *Papilio polytes* (Insecta, Papilionidae): associative learning in a predator. *Ethology*, *102*(5), 762-775.
- Wagner, H., Takahashi, T., & Konishi, M. (1987). Representation of interaural time difference in the central nucleus of the barn owl's inferior colliculus. *Journal of Neuroscience*, *7*(10), 3105-3116.
- Waters, C. M., & Bassler, B. L. (2005). Quorum sensing: cell-to-cell communication in bacteria. *Annu. Rev. Cell Dev. Biol.*, *21*, 319-346.
- Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: a promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews*, *21*(5), 195-205.
- Wich, S. A., & de Vries, H. (2005). Male monkeys remember which group members have given alarm calls. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1587), 735-740.
- Wiley, H. (2013). Communication as a transfer of information: Measurement, mechanism and meaning. In *Animal Communication Theory: Information and Influence*. Cambridge University Press.
- Wiley, R. H. (1994). Errors, exaggeration, and deception in animal communication. *Behavioral mechanisms in evolutionary ecology*, 157-189.
- Yip, M. J. (2006). The search for phonology in other species. *Trends in cognitive sciences*, *10*(10), 442-446.
- Zuberbühler, K. (2008). Audience effects. *Current Biology*, *18*(5), R189-R190.
- Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, *113*(1), 33.
- Zuidema, W., & De Boer, B. (2009). The evolution of combinatorial phonology. *Journal of Phonetics*, *37*(2), 125-144.
- Zuidema, W., & de Boer, B. (2018). The evolution of combinatorial structure in language. *Current opinion in behavioral sciences*, *21*, 138-144.