View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Warwick Research Archives Portal Repository

Manuscript version: Author's Accepted Manuscript
The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:
http://wrap.warwick.ac.uk/125579

How to cite:
Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Chestnut-crowned babbler calls are composed of meaningless shared building blocks

Sabrina Engesser [a,b] *, Jennifer L. Holub [c], Louis G. O'Neill [d,e], Andrew F. Russell [e] † &
Simon W. Townsend [a,b,f] †

[a] Department of Comparative Linguistics, University of Zurich, 8032 Zurich, Switzerland
[b] Center for the Interdisciplinary Study of Language Evolution, University of Zurich, 8032 Zurich, Switzerland
[c] Fowlers Gap Arid Zone Research Station, School of Biological, Earth & Environmental Sciences, University of New South Wales, NSW 2052, Australia
[d] Department of Biological Sciences, Macquarie University, NSW 2109, Australia
[e] Centre for Ecology & Conservation, College of Life & Environmental Sciences, University of Exeter, Penryn, Cornwall TR10 9FE, United Kingdom
[f] Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom

† These authors share last authorship

* Corresponding author
Email: sabrina.engesser@outlook.com, phone: +41 (0)44 634 0223

Short title: Building blocks of babbler multi-element calls

Keywords: Language evolution, phonology, combinatoriality, vocal communication, habituation-discrimination

Abstract

A core component of human language is its combinatorial sound system: meaningful signals are built from different combinations of meaningless sounds. Investigating whether non-human communication systems are also combinatorial is hampered by difficulties in identifying the extent to which vocalizations are constructed from shared, meaningless building blocks. Here we present a novel approach to circumvent this difficulty and show that a pair of functionally distinct chestnut-crowned babbler (*Pomatostomus ruficeps*) vocalizations can be decomposed into perceptibly distinct, meaningless entities that are shared across the two calls. Specifically, by focusing on the acoustic distinctiveness of sound elements using a habituation-discrimination paradigm on wild-caught babblers under standardized aviary conditions, we show that two multi-element calls are composed of perceptibly distinct sounds that are reused in different arrangements across the two calls. Furthermore, and critically, we show that none of the five constituent elements elicits functionally relevant responses in receivers, indicating that the constituent sounds do not carry the meaning of the call; so are contextually meaningless. Our work, which allows combinatorial systems in animals to be more easily identified, suggests that animals can produce functionally distinct calls that are built in a way superficially reminiscent of the way that humans produce morphemes and words. The results reported lend credence to the recent idea that language's combinatorial system may have been preceded by a superficial stage where signalers neither needed to be cognitively aware of the combinatorial strategy in place, nor of its building blocks.

Significance statement

Word generation in human language is fundamentally based on the ability to use a finite set of meaningless sounds in different combinations across contexts. Investigating whether animals share this basic capacity has been hampered by difficulties in identifying the extent to which animal vocalizations can be decomposed into smaller meaningless, yet shared sounds. Using a novel implementation of habituation-discrimination experiments, we show for the first time that a pair of functionally distinct chestnut-crowned babbler (*Pomatostomus ruficeps*) vocalizations are composed of perceptibly distinct, contextually meaningless sounds that are shared across the different calls. We conclude that the individual sounds represent building blocks that generate meaning when combined in a particular way, akin to word formation in human language.

\body

Introduction

A universal feature of human language is its combinatorial structure: a finite set of perceptibly distinct, meaningless sounds (building blocks) can be productively recombined to create a theoretically limitless set of meaningful signals [1]. One way to elucidate candidate origins and/or early forms of the combinatorial feature of language is to test for analogues in the basic process that underpins combinatoriality in the vocalizations of non-human animals [2]. While animals are clearly able to communicate using combinatorial vocal signals [3-8], whether they use meaningless sound elements in different arrangements to generate new meaning is contentious [9, 10]. This contention stems from two sources. First, from ambiguous associations between sound arrangements and meaning: for example, although animal songs are often composed of smaller sound units in different arrangements, precise arrangements are not known to underpin context-specific, or 'propositional', meaning [10-12]. Second, it also stems from difficulties of identifying whether functionally distinct vocalizations can be comprised of a recombinatorial system of shared meaningless sounds (i.e. building blocks) [13-16].

The traditional approach used to deconstruct the building blocks of the combinatorial sound system of human language is through the analysis of minimal pairs: pairs of semantically distinct words that differ in a *single* meaningless sound element, for example 'lap' versus 'tap' [9, 17]. The elements that differ in minimal pairs, in this case /t/ and /l/, are semantically meaningless, but are what serve to differentiate the meaning encoded in the two words. By extension, /t/ and /l/ must each represent distinct, meaning-contrasting sounds. This minimal pairs approach is feasible in human language because its sound elements are present in a plethora of permutations, such that each one used, and the role it plays in differentiating meaning, can be contrasted systematically with others in the repertoire [18]. However, this approach becomes unfeasible for communication systems where different sounds are not productively recombined and occur in prohibitively few combinations to allow direct contrasts of the impacts of single sounds on meaning to be made. Given that the productive usage of different sounds is likely a derived language-specific trait and is not a known feature of animal communication [13], an alternative method is required to test whether functionally distinct vocal signals are built from recombinations of shared sounds that are meaningless in isolation – the hallmark of combinatoriality in human language.

We propose that testing whether individuals perceive sound elements within and across functionally distinct calls as acoustically different or equivalent can also serve to decompose the potential building blocks of an animal's vocal system. Further, this approach can be implemented using established habituation-discrimination paradigms previously applied for speech-sound perception in human infants [19] and to assess the information content of whole calls in animals [20-23]. The utility of this habituation-discrimination approach to unpacking the characteristics of elements within calls is based on recent simulations on the emergence of combinatorial signals that define combinatorial structures using trajectories through acoustic and perceptual space [13, 18, 24]. In such simulations, the distance between points along trajectories of acoustic space reflect confusion probabilities, and hence the perceptual discreetness of sound elements. Accordingly, sound elements that are so close in acoustic parameter space so as to be easily confused are in essence perceptibly equivalent, while those that are more distant and seldom confused are essentially distinct. The advantage of this approach is that by focusing on sound discrimination and sharing within and across functionally distinct calls, comparative work investigating whether animal signals are composed of meaningless, recombinatorial entities (or building blocks) becomes feasible; with the potential to shed important light on the origins of combinatoriality.

Our overall aim is to use this new approach to test whether a pair of structurally similar but functionally distinct vocalizations of the chestnut-crowned babbler (*Pomatostomus ruficeps*; Fig 1A) can be decomposed into perceptibly distinct, contextually meaningless entities that are shared across the two calls – the defining feature of combinatoriality. The two calls of this highly social passerine bird from inland southeastern Australia [25] in question are: bi-element flight calls which are uttered when a bird flies off and which function to coordinate group movement (composed of the elements F_1F_2; Fig 1B); and tri-element prompt calls which are produced by an individual when entering the breeding nest in order to stimulate nestling begging during food provisioning (composed of the elements $P_1P_2P_3$; Fig 1B) [26, 27]. The functional distinction between the two calls is confirmed in playbacks on wild birds in on-site aviaries: flight calls induce greater movement and looking outside the aviary, presumably in response to an anticipated incoming bird, while prompt calls induce an 8-fold increase in the amount of time spent looking at a nest placed inside the aviary, presumably because of the natural association between nests and prompt calls [14]. Further, none of the five elements in the two calls is known to be used as stand-alone calls despite >1000 h of recordings in all known socio-ecological contexts, and all differ significantly from uni-element short-distance contact calls used to maintain contact and spacing during feeding [26]. Finally, previous aviary playback experiments also suggested that the distinct meaning encoded in these two multi-element calls is generated by the specific arrangement of the constituent sound elements [14]. However, what is not known is whether or not the constituent elements within these multi-element calls are: (a) perceptibly distinct within calls; (b) perceptibly equivalent across calls; and (c) contextually meaningless. Each of these three facets is required to resolve whether functionally distinct calls are built from smaller, perceptibly distinct and shared, meaningless sounds.

To test these core components of combinatoriality, we used standardized aviary playbacks on wild-caught chestnut-crowned babblers: (i) to identify which of the five sound elements constituting flight and prompt calls (i.e. F_1, F_2, P_1, P_2, P_3) are perceptibly distinct; (ii) to identify which, if any, are shared across the two calls; and (iii) to investigate whether contextually relevant information is encoded in the individual sound elements. To test element distinction versus equivalence, birds were exposed individually to a habituation-discrimination paradigm (Fig 1C). If two elements (e.g. F_1 & F_2) represent perceptibly distinct sounds, we would expect that, after habituating subjects to a series of repetitions of one element (e.g. F_1), switching to the other element (e.g. F_2) would result in a renewed response, measured by investigating changes in the time subjects spent looking into the direction from which the sounds were broadcast – as is customary in habituation-discrimination approaches [20-23]. On the other hand, a lack of response renewal following the habituation sequence would indicate that the contrasted elements are not discriminated and therefore are perceptibly equivalent sounds. Further, to test whether the five elements constituting flight and prompt calls carry contextually relevant meaning, we analyzed functionally relevant behavioral responses, including vocal responses, during the initial habituation phase of each playback. If elements carry relevant meaning, playbacks of flight call elements would be expected to result in babblers looking outside the aviary more and/or moving around the aviary more (see above [14]), whilst for prompt call elements we would expect an increase in time spent looking at the nest provided (see above [14]).

Results

(a) Are calls built from perceptibly distinct sounds?

We first tested whether flight and prompt calls are each comprised of distinct sounds by playing back habituation-discrimination sequences of F_1-F_2 elements from flight calls, and P_1-P_2, P_2-P_3 and P_1-P_3 elements from prompt calls to up to 12 birds individually (see Methods). In this experiment, habituation-discrimination sequences were played in natural order to avoid expectancy violation (i.e. discrimination performance being inflated through playing back elements in an unnatural order). Receivers habituated to habituation sequences (each composed of 20 element repetitions played back at three-second time intervals): subjects spent a median of 19% (*IQR* = 12,29) of their time looking at the speakers during playbacks of the first two elements in habituation sequences but only 1% (*IQR* = 0,6) of their time doing so during the last two elements of habituation sequences. One-sample Wilcoxon-tests were then used to investigate whether any changes in the proportion of time birds spent looking at the loudspeaker during the end of the habituation phase (last two habituation elements) and the discrimination phase significantly deviated from zero. Values significantly greater than zero indicate that habituation and discrimination elements were perceptibly distinct, while values not significantly different from zero indicate elements were not discriminated (i.e. perceived as equivalent sounds).

For the two flight call elements, the proportion of time receivers looked at the speaker increased 6-fold during the discrimination phase, indicating that birds discriminated F_2 from F_1 ($V = 36$, $P = 0.008$, $N = 11$; Fig 2A). As a consequence, we can conclude that the two elements in bi-element flight calls are perceptibly distinct (i.e. $F_1 \neq F_2$). By contrast, tri-element prompt calls do not appear to be composed of three distinct elements. Within prompt calls, significant 2 to 4-fold increases in the time spent looking at the speaker during the discrimination phase were found when P_2 followed P_1 ($V = 28$, $P = 0.016$, $N = 9$; Fig 2A) and when P_3 followed P_2 ($V = 55$, $P = 0.002$, $N = 10$; Fig 2A). However, there was no significant change in the proportion of time spent looking at the speaker between the end of the habituation phase and the discrimination phase when P_3 followed P_1 ($V = 11$, $P = 0.69$, $N = 10$; Fig 2A). These results suggest that the first and third prompt call elements are perceptibly equivalent, and that both are distinct from the second prompt call element.

To confirm the precise make-up of prompt calls, we conducted two further analyses. First, a Friedman test confirmed that there was a significant difference between the extent to which birds discriminated the three contrasted elements in prompt calls ($\chi^2_2 = 10.6$, $P = 0.005$, $N = 7$). Second, post-hoc two-sample Wilcoxon tests were used to compare the differences in the changes in the proportion of time birds spent looking at the speaker during the last two habituation stimuli versus the first two discrimination stimuli across each of the three sets of contrasted elements. These analyses confirmed: (a) that birds did not significantly differ in the extent to which they distinguished P_1 from P_2 versus P_2 from P_3 ($V = 10$, adjusted $P = 0.16$, $N = 9$; P value adjusted for multiple post-hoc testing; Fig 2A); but (b) that responses to P_2 following P_1 and to P_3 following P_2 were both greater than responses to P_3 following P_1 (P_1-P_2 vs. P_1-P_3: $V = 28$, adjusted $P = 0.031$, $N = 7$; P_2-P_3 vs. P_1-P_3: $V = 36$, adjusted P = 0.023, $N = 8$; Fig 2A). Thus, we are confident that the tri-element prompt call is composed of two perceptibly distinct sound types, with $P_1 = P_3$, but P_1 and P_3 to an equal extent $\neq P_2$.

(b) Are perceptibly equivalent sounds shared across calls?

Critical to elucidating whether multi-element calls ostensibly comprise building-blocks is to test whether elements are shared across functionally distinct calls. To investigate whether this is the case for flight and prompt calls, a different set of up to 13 birds received habituation-discrimination sequences comprising combinations of the two flight and three prompt call elements (see Methods). These were F_1 and P_2, F_2 and P_1, F_2 and P_3, $P_{1/3}$ and F_1 – with the elements used as habituation and discrimination stimuli, in this case, alternated because we wished to ensure that any expectancy violation was comparable across contrasts. Again, evidence for habituation during habituation phases was shown, with birds decreasing the percentage of time spent looking at the loudspeaker from a median of 17% ($IQR = 10,30$) to a median of 3% ($IQR = 0,8$) between the beginning and end of the habituation sequences.

Subsequent one-sample Wilcoxon-tests, comparing the change in the proportion of time looking at the speaker between the last two elements of habituation phases and the first two elements of discrimination phases against a null expectation of zero, revealed that the two distinct flight call elements were each perceptually equivalent to at least one of the prompt call elements. In three of the four comparisons, the proportion of time spent looking at the loudspeaker did not significantly increase between the last two stimuli of the habituation phase and the discrimination phase. Specifically, we found F_1 to be perceptually equivalent to P_2 ($V = 18$, $P = 0.58$, $N = 12$; Fig 2B), and F_2 to be perceptibly equivalent to both P_1 ($V = 2$, $P = 0.19$, $N = 10$; Fig 2B) and P_3 ($V = 27$, $P = 0.65$, $N = 9$; Fig 2B). In contrast, the proportion of time birds spent looking at the loudspeaker increased by 4-fold when the prompt call element P_1 or P_3 (which are equivalent, see above) was contrasted with the flight call element F_1; meaning that P_1/P_3 are distinct from F_1 ($V = 55$, $P = 0.002$, $N = 11$; Fig 2B). Thus, these results indicate that bi-element flight calls and tri-element prompt calls both consist of the same two sound types: the first flight and second prompt call elements are perceptibly equivalent (i.e. $F_1 = P_2$), as are the second flight and both first and third prompt call elements (i.e. $F_2 = P_1 = P_3$). In other words, flight and prompt calls comprise the same two building blocks in different combinations.

(c) Do sound elements carry contextual meaning?

In human languages, meaningful signals are built from recombinations of meaningless sounds. To test whether or not the constituent elements of flight and prompt calls carry context-specific meaning, we measured the vocal responses and activity budgets of birds during the first two habituation stimuli of each playback (i.e. H-start, Fig 1C). First, we found no evidence to suggest that playbacks induce birds to respond with either flight or prompt calls: the median number of each call given during the 6 s period of the 82 playbacks included, was zero ($IQR = 0{,}0$). Second, we found no evidence to suggest that birds modify key behaviors in response to the playbacks. For example, we have previously shown that playbacks of flight calls on lone individuals in the aviary environment cause individuals to move around the aviary and to look outside more, while prompt call playbacks cause birds to look more at a nest in an upper corner of the aviary [14]. Here, by contrast, individuals spent little time engaging in behaviors of relevance during the 6 s of each playback analyzed, spending on average: 1.3 s (SD = 1.1) of their time in-movement; 1.3 s (SD = 1.2) looking outside the aviary; and 0.07 s (SD = 0.3) of their time looking at the nest. In addition, the amount of time individuals spent engaged in each of these behaviors was independent of the precise element played (F_1, F_2, P_1, P_2, P_3) (Linear Mixed Model: behavior * element interaction, $\chi^2 = 9.48$, $DF = 8$, $P = 0.30$; Fig 3A) as well as whether or not the elements played were from a flight call (F elements) or a prompt call (P elements) (LMM: behavior * element interaction $\chi^2 = 1.93$, $DF = 2$, $P = 0.38$; Fig 3B). Thus, babblers do not seem to extract contextually meaningful information from the sound elements of the two calls when played back in isolation.

Discussion

Using a novel application of the established habituation-discrimination paradigm, we here demonstrate that a pair of functionally distinct, multi-element calls produced by chestnut-crowned babblers are composed of two perceptibly distinct, contextually meaningless sounds, which are shared across the two vocalizations. Specifically, we show that the first element from bi-element flight calls is distinct from its second element but equivalent to the second element from tri-element prompt calls. Further, the second flight call element is equivalent to the first and third prompt call elements. In addition, none of the individual elements that make up these two calls elicits differential vocal or behavioral responses of relevance in receivers. For example, subjects rarely responded to playbacks with flight or prompt calls, with a total of just nine such calls recorded across the 82 x 6 s playbacks. Moreover, babblers spent little time engaged in behaviors of relevance and the amount of time they did so was not modified by the element played; which would otherwise be expected if the elements encoded flight or prompt call-related information [14]. Together, these results suggest for the first time, that a non-human animal uses meaningless (shared) building blocks in different arrangements to encode distinct meaning.

A core feature of human language is that perceptibly discrete, meaningless sounds are combined in various ways to generate distinct meaning. Testing whether animals use this basic process has been hampered by a focus on minimal pairs as a way to decompose the sound system of a language - that is, identifying building-blocks through a sound's role in differentiating meaning [9, 17]. This approach necessarily requires sounds to occur across a sufficient number of vocalizations to permit meaningful comparisons, which is problematic for largely non-productive communication systems such as those utilized by animals. We demonstrate here that one can identify elements that, in essence, function like building blocks, by rather focusing on the individual perceptibility of sounds used within and across functionally distinct animal calls. We suggest that this novel approach opens up new opportunities to investigate any parallels between animal vocalizations and combinatoriality in human language.

We caution, of course, that any similarities between the combinatorial constructs of animal communication and word generation in human language must be tempered. First, in contrast to the combinatorial structures found in animal communication systems, combinatoriality in human language is hypothetically open-ended, with finite numbers of phonemes used in myriad combinations to generate potentially limitless information. Second, while we have shown previously that at least one element (P_1) appears to be meaning-contrasting [14] and we have shown here that elements across babbler calls (including P_1) can function like building blocks, confirming that shared elements are meaning-differentiating will always be challenging in animals. To mitigate this problem, investigations into whether or not animals use building blocks in their communication systems should limit their comparisons to functionally distinct calls. This will ensure that constituent elements that are shared also play a potential role in generating meaning. Third, the building blocks of babbler calls are separated by silence, whereas in human language, they are not. Whether this is a significant distinction or a likely precursor is yet to be determined.

The acknowledged distinctions between babbler and human combinatoriality notwithstanding, the complexities of human language likely evolved from more rudimentary beginnings. Indeed, recent theoretical work suggests that language's productive combinatorial system was preceded by a superficial stage where the sound elements of signals overlap in their acoustic and perceptual space, but neither needed to be recognized as recombinatorial units nor utilized in a productive way by the system's users [13, 18, 24]. Subsequently, once signalers became aware of their recombinatorial system (i.e. recognize signals as being composed of smaller building blocks), they could evolve strategies (e.g. learning mechanisms) to exploit the combinatorial mechanism productively [13, 18, 24]. We propose that our study provides evidence for such a superficial vocal system by demonstrating bounded, unproductive combinatoriality (i.e. two sounds build only two signals) in babbler vocalizations. Although simple in its structure, this data supports recent hypotheses on human combinatorial systems transitioning from a more rudimentary evolutionary stage (i.e. 'superficial' combinatorial layer) before it fledged into a fully productive combinatorial system [24]. Further experiments are now needed to clarify whether similar, more superficial, combinatorial structures exist in the communication systems of other species and the precise forms they take.

To conclude, our work provides new insights into the potential similarities between animal communication systems and the combinatorial structures of human language, with chestnut-crowned babblers reusing perceptibly distinct elements that are meaningless in isolation, but when used in different arrangements generate distinct meaning. Our study has at least three important implications. First, although we provide novel evidence for ‘superficial’ combinatoriality in non-human animals, we deem it highly improbable that chestnut-crowned babblers are unique amongst animals in their ability to recombine perceptibly distinct and equivalent sounds to generate context-specific calls. Indeed, we are confident that by shifting the empirical focus to an approach that allows combinatorial systems in animals to be more easily identified, additional data in other species will undoubtedly accumulate. Second, whilst species with clearly identifiable internally structured calls, as is the case with chestnut-crowned babblers, represent intuitively more straight-forward test systems, we advocate a more general search for analogues incorporating vocalizations without clear temporal separation as happens to be the case in human language [10]. Either way, further cases are required to provide a coherent understanding of the form of early combinatorial systems, as well as their eco-evolutionary correlates. Finally, using the approach outlined, we believe that comparative work on combinatorial communication in animals will become a significant compliment to game-theoretic modelling [13, 28]; multi-agent simulations [24]; emerging sign language [29]; and communication game work [30] that aim to unpack the evolutionary origins and forms of combinatorial structures and capacities in humans and other animals.

Material and methods

Study species and housing

The study was conducted from July to September 2017 on 25 individuals from 13 different groups of a free-living, color-ringed population of chestnut-crowned babblers, at the Fowlers Gap

Arid Zone Research Station in New South Wales (141°42’E, 31°06’S; for details on the study population and habitat see [25]). Chestnut-crowned babblers are 50 g, group-living, cooperatively breeding passerine birds endemic to inland south-eastern Australia [25], with a known vocal repertoire of at least 18 functionally distinct calls [26]. For experimental procedures, birds were captured and housed in standardized aviaries, and were released back into their original groups after a maximum time of 48 hours (for details on capturing and aviary set-up see [14]). We have confirmed previously that birds are accepted back into their groups without retribution following their temporary absence [31], and in this study measurements of mass following their period in the aviary indicated that birds gained an average of 0.1 g (SD = 2.0) in the aviary. Birds for testing were selected randomly with respect to age and sex, although we never removed the group’s breeding female or individuals with any juvenile plumage (indicating all removed individuals were nutritionally independent and > six months old).

During and between tests, single birds were kept in one of six compartments of a larger aviary (dimensions of each compartment: 2 x 2 x 2.5 m). Each compartment consisted of a babbler nest, perches and natural substrate. The back side of the aviary comprised a metal-mesh of 1 cm^2 allowing the birds a view to the outside, while the sides were opaque metal and the front consisted of one-way Perspex. During daylight, birds were fed 20 mealworms every two to three hours, and water was provided throughout (see also [14] for details on housing conditions). If two birds were removed from a group at the same time, birds were kept in different compartments, but joined into one compartment overnight. During playback experiments, only one test subject remained in the aviary, while any other birds were removed to an accommodation block out of earshot, to prevent interference with the playback.

Playback stimuli and procedure

Flight and prompt calls used for the creation of playback sequences were recorded using Electret EM-400 condenser tie-clip microphones in combination with a Sony IC-UX533 recorder (sampling frequency 44.1 kHz, 24-bit accuracy). Only high-quality vocalizations were chosen, and flight and prompt call elements were extracted and normalized using Adobe Audition CC 2015. Each playback sequence consisted of 20 habituation stimuli (of one element type) and two subsequent discrimination stimuli (of another element type) broadcast at three-second intervals (Fig 1C). All test subjects were only ever exposed to stimuli originating from unfamiliar individuals. Additionally, to account for pseudo-replication and inevitable among-individual variation in element characteristics owing to, for example body size, the 20 elements used in each habituation sequence always originated from at least eight different individuals (average = 12), while the two discrimination stimuli within a sequence always originated from different individuals. Flight and prompt calls are often given by different individuals in quick succession, so babblers are accustomed to hearing flight and prompt call elements from different individuals in the field. Finally, the 20 elements within the habituation sequences and the two elements within the discrimination sequences were randomly ordered, and each playback sequence/track was only used once, resulting in each test subject receiving unique playback sequences.

Each bird was exposed to 4 unique habituation-discrimination sequences with a break of at least 10 minutes between treatments, leading to a maximum of 100 trials across the 25 birds (but see below). Ten minutes was decided as a minimum because we wished to minimize the amount of time that any co-inhabitant of the aviary was removed for during the playback (with a minimum of 10 mins between treatments, this could be reduced to ca. 40 mins) and pilot work suggested that 10 min intervals did not confound habituation effects. In line with this pilot work, we found here that the change in looking response between H-end and H-start was equivalent for the first and last habitation trials both in the within-call element comparisons (paired, two-sample Wilcoxon test: $V = 32$, $P = 0.62$, $N = 12$ individuals) and among-call element comparisons ($V = 42$, $P = 0.85$, $N = 12$ individuals). Playbacks were broadcast with a natural flight and prompt call amplitude of 50 dB at two meters (measured with a Castle GA206 sound level meter, C-weighted) and using a Braven BRV-X loudspeaker. The loudspeaker was placed outside 1 m away and 1 m shifted towards the side of the open, mesh-enclosed part of the aviary compartment, and was concealed by vegetation. This position was chosen because it facilitated our judgment of gaze direction towards the speaker, which is the key data of interest resulting from habituation-discrimination experiments [20-23]. In order to assess the time subjects looked into the direction of the loudspeaker (and engaged in other relevant behaviors), playbacks were video-taped using a Sony HDR-CX240.

Video coding and trial inclusion criteria

Videos were analyzed frame-by-frame and blindly with respect to playback type using Adobe Audition CC 2015, with the following data extracted from each subject: number of flight and prompt calls given; number of hops/flights; and the amount of time spent looking outside, at the nest in the upper corner and at the loudspeaker. Vocalizations, movement and looking outside were easily coded, but quantifying gaze direction towards specific objects is more challenging because birds have relatively laterally-set eyes compared with humans. Nevertheless, all birds have binocular overlap in their vision to allow them to avoid obstacles during flight, interact with conspecifics, obtain food and pinpoint predators [32]. For passerines, binocular overlaps range from 35-51° (N=13 species, including 6 non-tool-using corvids) [33]. Given that babblers are passerines in the same super family as corvids (Corvidea), suggests that they will have binocular overlap of at least 30° and probably closer to the 40° characteristic of corvids. Further, for one such corvid, the common raven (*Corvus corax*, binocular overlap = 43°, [33]) looking direction towards specific objects during habituation-discrimination experiments has been assessed previously using bill orientation [34]. In line with previous work, we here qualify looking at the speaker or the nest by assessing the orientation of the test bird's bill which had to directly point towards the object in question (± 30°, well within the expected field of binocular overlap). Babblers routinely turn their head in order to pinpoint food, conspecifics and predators, and we have substantial experience with gaze direction for each of these stimuli in the aviary setting. Through double-blind scoring of time spent looking at the speaker during the end of habituation (H-end) and discrimination phases of 41 trials (50% of the 82 included), we found substantial inter-scorer agreement (Interclass Correlation Coefficient for two-way model based on absolute agreement and single rater scores ICC = 0.83, P < 0.001, 95% CI = 0.75-0.89) [35].

Out of the 100 potential trials, 82 were included in the analyses. Two trials were not obtained because we released a bird early due to concerns over a loss of appetite and failed to capture H-start of another trial in the camera. Further, in 5 trials, birds failed to look in the direction of the speaker during the habituation phase, a prerequisite of the habituation-discrimination paradigm, and likewise, a further 11 had to be excluded as they looked at the speaker at least as much during H-end as H-start. There was no systematic bias in the habituation stimuli that were excluded, with each of the 5 habituation elements being removed at least twice.

Statistical analyses

Element discrimination

Testing whether elements are perceived as dissimilar or equivalent was primarily investigated using a series of one-sample Wilcoxon tests. Specifically, the change in the proportion of time individuals spent looking at the speaker between the discrimination phase (D) and the end of the habituation phase (H-end) was contrasted against a null expectation of zero change (Figs. 1C, 2). The only exception was to further clarify the form of prompt calls. In this case, we additionally used Friedman combined with post-hoc two-sample Wilcoxon tests to test the *differences* in the changes of responses between H-end and D for contrasted pairs of elements (i.e. P_1-P_2 *vs*. P_2-P_3 *vs*. P_1-P_3) - post-hoc *P*-values were adjusted using the Bonferroni-holm method [36]. For all analyses of element discrimination, we used the proportion of time looking at the speaker (rather than absolute time) since the birds were not always in camera view for the entire 6 s H-end and D phases (H-end: mean time in view = 5.9; SD = 0.2, range = 4.8-6.2; D: mean = 6.0, SD = 0.1 range = 5.3-6.4). All statistical analyses were conducted in R (version 3.4.2) - Wilcoxon tests using the “exactRankTest”-package [37], and Friedman tests using the “stats”-package [38].

Element meaning

To investigate whether the five constituent elements of flight and prompt call elements carry contextual meaning, we performed two Linear Mixed effects Models (LMM). In both models, the response term was the amount of time (during the 6 s of H-start for each element, square-root transformed) that individuals were observed: looking outside (not at the speaker); looking at the nest in an upper corner of the aviary; and in-movement (mainly hopping among perches). These behaviors were chosen because we have previously shown in the same aviary set-up that babblers change the duration of each behavior in response to playbacks of flight and prompt calls [14]. It is important to note that the sum percentage of time that individuals engaged in these 3 behaviors averaged just 44%, meaning that individuals could respond to each behavior independently. The term of interest in the first model was the interaction between element type (F_1, F_2, P_1, P_2, P_3) and behavioral response (in-movement, looking-out, looking-nest); while in the second model, we interacted whether or not the element in question was from a flight call (F elements) or a prompt call (P elements) with behavioral response. In both models, time in view was fitted as a covariate and trial identity nested within individual identity were fitted as random intercepts to account for the fact that trials had 3 behavioral responses and that multiple elements were played to the same individual. Model reduction were not performed for either model as in both cases the key result is the interaction between element and behavior. The above two models were fitted in R using the "lme4" package, and the full model with and without the interaction of interest were compared using log-likelihood ratio tests to determine the significance of the interaction term [38, 39].

Acknowledgements: We thank Simon Griffith, Keith Leggett and the Dowling family for logistical support at Fowlers Gap; Kiara L'Herpiniere, Joseph England and Jennifer Page for help with fieldwork; Steven Moran, Volker Dellwo and Stuart Watson for discussions, and three anonymous reviewers for their constructive feedback. The research was approved by the ethics committee of the University of Exeter (Application number 2018/2301).

Data accessibility: All data to reproduce the work is provided as supplementary material.

References

1. Hockett CF (1960) The Origin of Speech. *Sci Am* 203:88-111.
2. Hauser MD, Chomsky N, Fitch WT (2002) The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298(5598):1569-1579.
3. Hurford J (2007) *The origins of meaning* (Oxford University Press, Oxford).
4. Arnold K, Zuberbühler K (2006) Language evolution: Semantic combinations in primate calls. *Nature* 441(7091):303.
5. Zuberbühler K (2018) Combinatorial capacities in primates. *Curr Opin Behav Sci* 21:164-169.

6. Ouattara K, Lemasson A, Zuberbühler K (2009) Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proc Natl Acad Sci USA* 106(51):22026-22031.

7. Engesser S, Ridley AR, Townsend SW (2016) Meaningful call combinations and compositional processing in the southern pied babbler. *Proc Natl Acad Sci USA* 113(21):5976-5981.

8. Suzuki TN, Wheatcroft D, Griesser M (2016) Experimental evidence for compositional syntax in bird calls. *Nat Commun* 7:10986.

9. Yip MJ (2006) The search for phonology in other species. *Trends Cogn Sci* 10(10):442-446.

10. Bowling DL, Fitch WT (2015) Do Animal Communication Systems Have Phonemes? *Trends Cogn Sci* 19(10):555-557.

11. Berwick RC, Okanoya K, Beckers GlJL, Bolhuis JJ (2011) Songs to syntax: the linguistics of birdsong. *Trends Cogn Sci* 15(3):113-121.

12. Engesser S, Townsend SW (2019) Combinatoriality in the vocal systems of non-human animals. *WIREs Cogn Sci* e1493.

13. Zuidema W, de Boer B (2009) The evolution of combinatorial phonology. *J Phon* 37(2):125-144.

14. Engesser S, Crane JM, Savage JL, Russell AF, Townsend SW (2015) Experimental Evidence for Phonemic Contrasts in a Nonhuman Vocal System. *PLoS Biol* 13(6):e1002171.

15. Hailman JP, Ficken MS, Ficken RW (1985) The Chick-a-Dee Calls of *Parus atricapillus* - a Recombinant System of Animal Communication Compared with Written-English. *Semiotica* 56(3-4):191-224.

16. Suzuki TN (2013) Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Anim Behav* 87:59-65.

17. Chomsky N, Halle M (1968) *The Sound Pattern of English* (Harper & Row, New York).

18. Zuidema W, de Boer B (2018) The evolution of combinatorial structure in language. *Curr Opin Behav Sci* 21:138-144.

19. Eimas PD, Siqueland ER, Jusczyk P, Vigorito J (1971) Speech Perception in Infants. *Science* 171(3968):303-306.

20. Charlton BD, Ellis WA, McKinnon AJ, Brumm J, Nilsson K, Fitch WT (2011) Perception of male caller identity in Koalas (*Phascolarctos cinereus*): acoustic analysis and playback experiments. *PLoS ONE* 6(5):e20329.

21. Cheney DL, Seyfarth RM (1988) Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Anim Behav* 36(2):447-486.

22. Fitch WT (2006) Rhesus macaques spontaneously perceive formants in conspecific vocalizations. *J Acoust Soc Am* 120(4):2132-2141.

23. Reby D, Hewison M, Izquierdo M, Pépin D (2008) Red Deer (*Cervus elaphus*) Hinds Discriminate Between the Roars of Their Current Harem-Holder Stag and Those of Neighbouring Stags. *Ethology* 107:954-959.

24. de Boer B, Zuidema W (2010) Multi-Agent Simulations of the Evolution of Combinatorial Phonology. *Adapt Behav* 18(2):141-154.

25. Russell AF (2016) Chestnut-crowned babblers: Dealing with climatic adversity and uncertainty in the Australian arid zone. *Cooperative breeding in vertebrates: studies in ecology, evolution and behavior*, eds Koenig WD, Dickinson JL (Cambridge University Press, Cambridge, MA), pp 150-164.

26. Crane JMS, Savage JL, Russell AF (2016) Diversity and function of vocalisations in the cooperatively breeding Chestnut-crowned Babbler. *Emu* 116(3):241.

27. Young CM, Browning LE, Savage JL, Griffith SC, Russell AF (2013) No evidence for deception over allocation to brood care in a cooperative bird. *Behav Ecol* 24(1):70-81.

28. Nowak MA, Krakauer DC, Dress A (1999) An error limit for the evolution of language. *Proc R Soc B* 266(1433):2131-2136.

29. Sandler W, Aronoff M, Meir I, Padden C (2011) The gradual emergence of phonological form in a new language. *Nat Lang Linguist Th* 29(2):503-543.

30. Verhoef T, Kirby S, de Boer B (2014) Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *J Phon* 43:57-68.

31. Nomano FY, Browning LE, Savage JL, Rollins LA, Griffith SC, Russell AF (2015) Unrelated helpers neither signal contributions nor suffer retribution in chestnut-crowned babblers. *Behav Ecol* 26(4):986-995.

32. Martin GR (2009) What is binocular vision for? A birds' eye view. *J Vis* 9(11):14.11-19.

33. Troscianko J, von Bayern AM, Chappell J, Rutz C, Martin GR (2012) Extreme binocular vision and a straight bill facilitate tool use in New Caledonian crows. *Nat Commun* 3:1110.

34. Reber SA, Boeckle M, Szipl G, Janisch J, Bugnyar T, Fitch WT (2016) Territorial raven pairs are sensitive to structural changes in simulated acoustic displays of conspecifics. *Anim Behav* 116:153-162.

35. Hallgren KA (2012) Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol* 8(1):23-34.

36. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 57:289–300.

37. Hothorn T, Hornik K (2017) exactRankTests: Exact Distributions for Rank and Permutation Tests. R package version 0.8-29. Accessed 01 June 2017.

38. R-Core-Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Accessed 01 June 2017.

39. Bates D, Maechler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1-48.

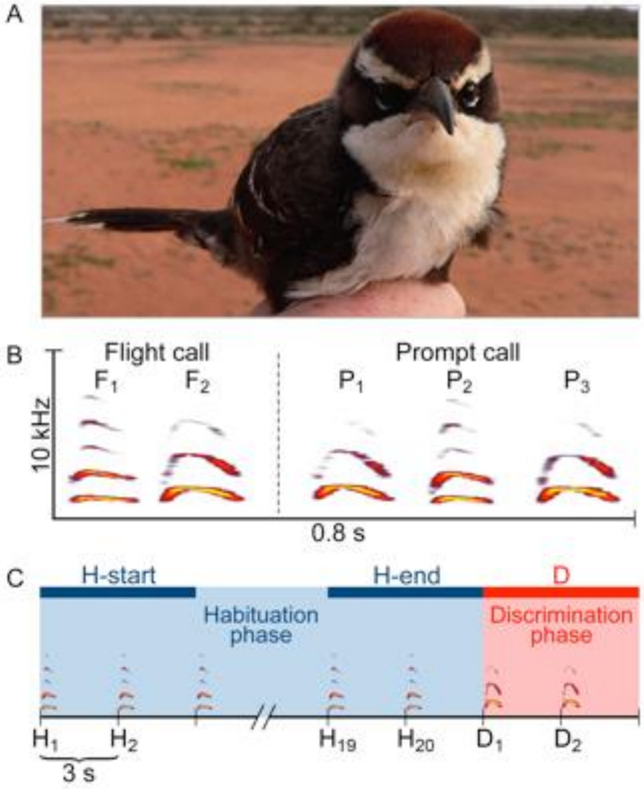
Figure legends

Fig 1. Study design. (A) Chestnut-crowned babbler (credit AF Russell). (B) Spectrogram of a flight and a prompt call, with the flight call being composed of F_1F_2 elements and prompt calls of $P_1P_2P_3$ elements. (C) Schematic overview of the habituation-discrimination experiment. During the habituation phase subjects were accustomed to one element type (from at least 8 different unfamiliar individuals) constituting the habituation stimuli ($H_1 – H_{20}$, e.g. F_1), which was repeated 20 times at three-second intervals. Subsequently, two repetitions of another element type (both from different unfamiliar individuals) constituting the discrimination stimuli ($D_1 – D_2$, e.g. F_2) were broadcast. To assess the discrimination between contrasted elements, the change between the proportion of time subjects looked toward the loudspeaker during the discrimination (D) and the last two habituation stimuli (H-end) was analyzed.

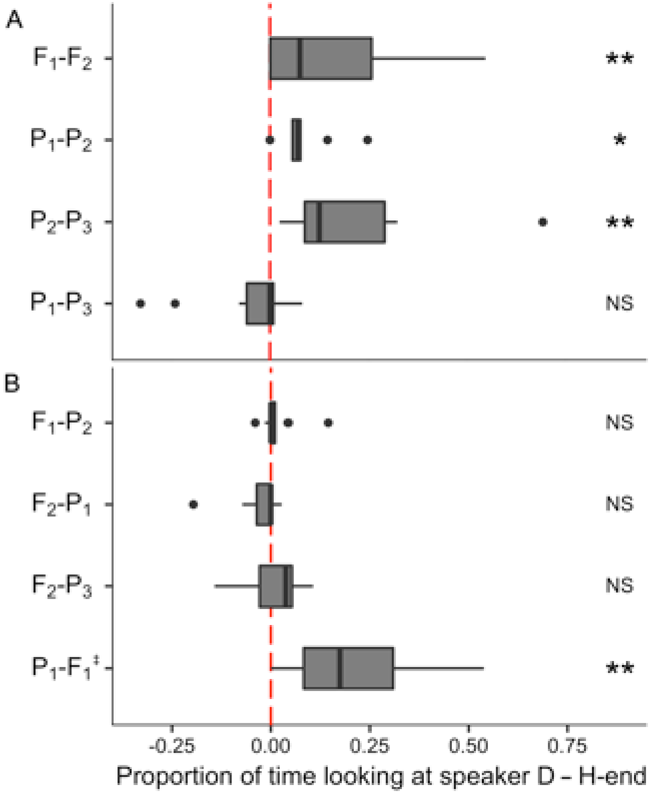
Fig 2. Element discriminations. Results of the habituation-discrimination experiments when contrasting flight and prompt call elements: (A) within flight or prompt calls; and (B) between flight and prompt calls. Figures show the changes in the proportion of time subjects looked at the loudspeaker during the discrimination phase (D) and the end of the habituation phase (H-end) for each element comparison. The dashed vertical (red) line represents the null expectation of no-change. Boxes represent the 25%, 50% and 75% quartiles of the raw data, whiskers extend to 1.5 x inter-quartile ranges, while dots show outliers. Significant changes in the proportion of time spent looking at the loudspeaker between H-end and D are shown with asterisks (* $p < 0.05$, ** $p < 0.01$). In Figure A elements were presented in natural order (as shown), while in B element orders were randomized since no natural order exists in between-call comparisons (‡ denotes that P_1 was alternated with the equivalent sound P_3).

Fig 3. Element meaning. The amount of time individuals spent engaged in behaviors of relevance during H-start when: (A) behavioral responses were considered for each of the 5 element types individually (F_1, F_2, P_1, P_2, P_3); and (B) behavioral responses were considered for flight call (F) elements versus prompt call (P) elements. Shown are the raw data with point sizes indicating the frequency of occurrence at given time values. In Figure A dot shapes (circular or triangular) illustrate the two discriminated sound types (i.e. circular F_1 & P_2; triangular F_2, P_1 & P_3). In Figures A & B red shaded dots illustrate flight call elements and blue shaded dots prompt call elements. Note there is no obvious tendency for different elements to elicit differential behavioral responses. Analyses in each case are based on 246 behavioral responses during the 82 playbacks. In each model, the variance component of the random term 'trial identity' was 0, indicating that the variation in activity budgets within and among trials were equivalent. By contrast, individual identity explained a significant 15% of the residual variance in each model (variance component = 0.04, $P < 0.001$), indicating that some individuals were more active than others. Finally, inclusion of the interaction term of interest in each model raised the AIC by 7 points (Model 1) and 2 points (Model 2), indicating that power of the models were reduced when the interaction terms were included (see text for statistics).

Figure 1

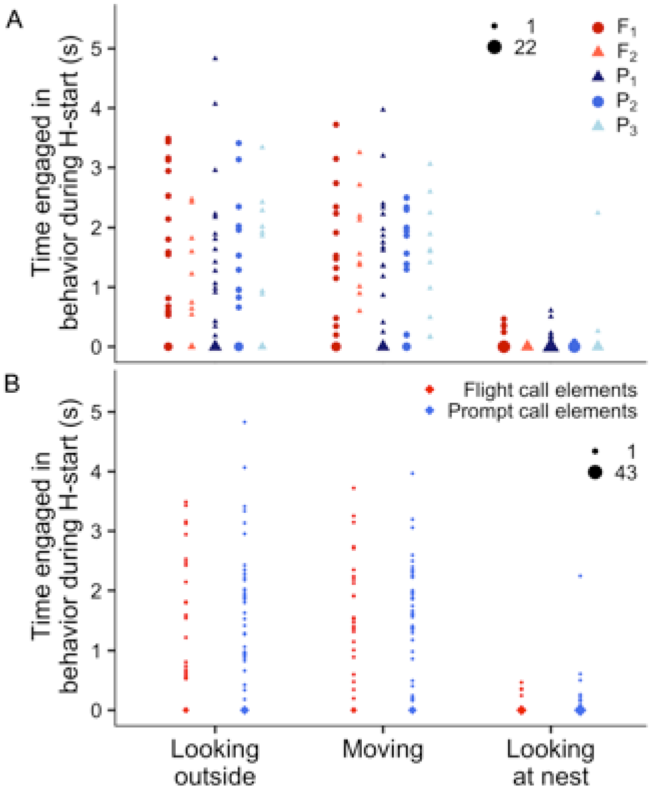

A
B
Flight call
F_1
F_2
Prompt call
P_1
P_2
P_3
10 kHz
0.8 s
C
H-start
H-end
D
Habituation phase
Discrimination phase
H_1
H_2
H_{19}
H_{20}
D_1
D_2
3 s

580 **Figure 2**



581

Figure 3


A
Time engaged in behavior during H-start (s)
0
1
2
3
4
5
1
22
F₁
F₂
P₁
P₂
P₃
B
Time engaged in behavior during H-start (s)
Flight call elements
Prompt call elements
1
43
Looking outside
Moving
Looking at nest