



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2023

Cognitive constraints on vocal combinatoriality in a social bird

Watson, Stuart K ; Mine, Joseph G ; O'Neill, Louis G ; Mueller, Jutta L ; Russell, Andrew F ; Townsend, Simon W

DOI: <https://doi.org/10.1016/j.isci.2023.106977>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-255232>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

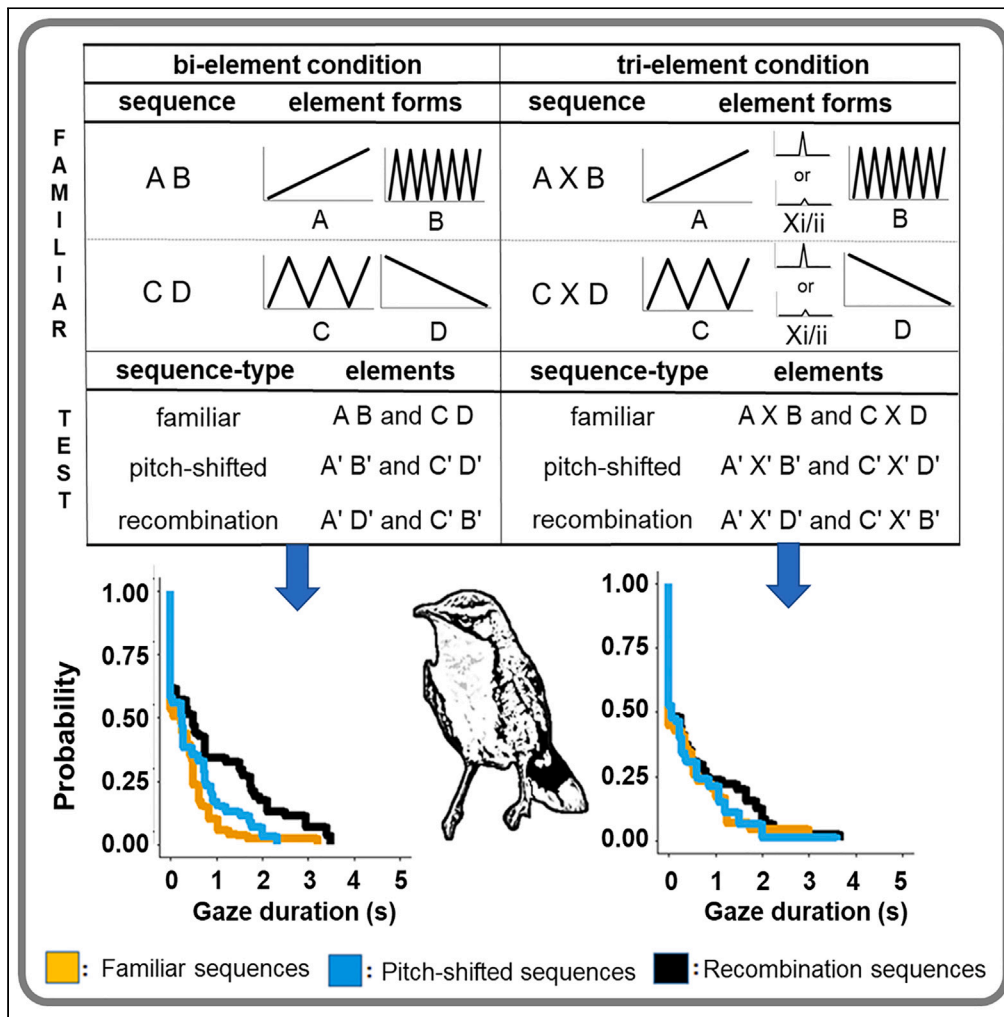
Originally published at:

Watson, Stuart K; Mine, Joseph G; O'Neill, Louis G; Mueller, Jutta L; Russell, Andrew F; Townsend, Simon W (2023). Cognitive constraints on vocal combinatoriality in a social bird. *iScience*, 26(7):106977.

DOI: <https://doi.org/10.1016/j.isci.2023.106977>

Article

Cognitive constraints on vocal combinatoriality in a social bird



Stuart K. Watson,
Joseph G. Mine,
Louis G. O'Neill,
Jutta L. Mueller,
Andrew F. Russell,
Simon W.
Townsend

swatso88@gmail.com

Highlights

The ability to freely recombine a repertoire of distinct elements is key to language

Chestnut-crowned babbler's vocal system is limited to combinations of two distinct sounds

We explored limits of this combinatoriality using an artificial grammar experiment

Data shows babblers can process recombinations of bi- but not tri-element sequences



Article

Cognitive constraints on vocal combinatoriality in a social bird

Stuart K. Watson,^{1,2,3,10,*} Joseph G. Mine,^{1,3,4} Louis G. O'Neill,^{4,5,7} Jutta L. Mueller,⁶ Andrew F. Russell,^{4,6,7,9} and Simon W. Townsend^{1,3,8,9}

SUMMARY

A critical component of language is the ability to recombine sounds into larger structures. Although animals also reuse sound elements across call combinations to generate meaning, examples are generally limited to pairs of distinct elements, even when repertoires contain sufficient sounds to generate hundreds of combinations. This combinatoriality might be constrained by the perceptual-cognitive demands of disambiguating between complex sound sequences that share elements. We test this hypothesis by probing the capacity of chestnut-crowned babblers to process combinations of two versus three distinct acoustic elements. We found babblers responded quicker and for longer toward playbacks of recombined versus familiar bi-element sequences, but no evidence of differential responses toward playbacks of recombined versus familiar tri-element sequences, suggesting a cognitively prohibitive jump in processing demands. We propose that overcoming constraints in the ability to process increasingly complex combinatorial signals was necessary for the productive combinatoriality that is characteristic of language to emerge.

INTRODUCTION

The comparative approach is a powerful tool for examining the evolutionary roots of uniquely human cognitive capacities.¹ The application of this approach to the study of language evolution has been particularly fruitful, identifying many of the cognitive building blocks necessary (but not sufficient) for the emergence of language in a diverse range of non-human animal communication systems.^{2,3} Notably, experiments confirm that birds and mammals can change the meaning of their vocalizations by joining calls together in sequences consistent with basic syntax^{4–7} or by recombining meaningless call elements in different ways to generate alternative meanings in a manner analogous to basic a phonemic system.^{8–10} However, in stark contrast to the complex combinations of words and phonemes in human languages, animals seldom use the same calls or call elements across meaningful sequences of more than two distinct sound units.^{4,11–14} The bi-element structuring of combinatorial calls in animals cannot be due to constraints on vocal production. Many animals have vocal repertoires containing a sufficient number of sounds to theoretically be able to generate hundreds of different sound combinations, since a factorial function underpins the relationship between the number of sounds and the potential number of sound combinations. Indeed, some species of bird can sing hundreds of different songs,¹⁵ but as far as we know, varying a song's composition does not change the semantic meaning.¹⁶ However, even songs that appear to be highly complex can be described and processed as a series of bigrams,¹⁷ and recent evidence suggests that some songbirds may be insensitive to the actual ordering of notes within a song, attending instead to the fine acoustic details of individual notes.¹⁸ This presents a stark contrast with the processing of meaningful call combinations, where the re-ordering of constituent elements seems to render them unrecognisable in species, such as Japanese tits.¹⁹ This difference in how the call types seem to be processed, alongside the fact that songs often contain a myriad of elements whereas meaningful calls typically contain just a few, leads to a tantalizing hypothesis: The cognitive demands of processing sequences with fixed positionality increase according to the number of distinct elements within them, therefore placing a cognitive ceiling on the diversity and complexity of combinatorial calls that can exist within a given system.²⁰

The path to language clearly requires the capacity to generate new meaning by recombining the same sounds across sequences of at least three elements: with just 12 sounds, an order of magnitude more tri-element than bi-element combinations are possible (1320 vs. 132 permutations in each case). However,

¹Department of Comparative Language Science, University of Zurich, Zurich, Switzerland

²Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

³Center for the Interdisciplinary Study of Language Evolution, Zurich, Switzerland

⁴Faculty of Environment, Science and Economy, University of Exeter, Penryn, Cornwall TR10 9FE, UK

⁵Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109 Australia

⁶Institute of Linguistics, University of Vienna, Vienna, Austria

⁷Fowlers Gap Arid Zone Research Station, School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, NSW 2052, Australia

⁸Department of Psychology, University of Warwick, Coventry, UK

⁹These authors contributed equally

¹⁰Lead contact

*Correspondence: swatso88@gmail.com

<https://doi.org/10.1016/j.isci.2023.106977>



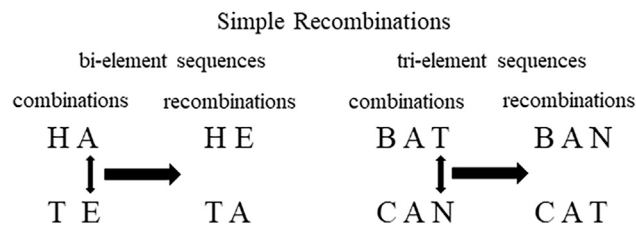


Figure 1. Simple means of generating two new words using reciprocal element recombination in bi- and tri-element sequences

Note that in the bi-element condition, processing the sound only requires recognition of the two elements and which one comes first, whereas in the tri-element condition, the sequence cannot be understood without realization of the additional connection between the first and last elements. It is thus anticipated that processing recombinations of tri-element sequences will be more demanding cognitively.

unlike for bi-element sequences, understanding recombined sequences of three or more elements will more often require the ability to recognize, retain, and process relationships between distinct sounds that are positioned both consecutively and non-consecutively within a sequence.^{21,22} For example, the phonemes /b/, /æ/, /t/, /k/ and /n/ can be combined to make the word *bat*, or recombined to make the words *can*, *ban* or *cat* (Figure 1). In all cases, distinguishing between one meaning and another requires an ability, not only to recognize the constituent elements and their order, but also to process the relationship between the first and last elements. Indeed, infants as young as 10 months old have been found to process non-adjacent relationships between vowels,²³ consonants,²⁴ and fricatives²⁵ within language (see²⁶ for review). Thus, a generative system, which goes beyond simple two-element combinations, is necessarily more cognitively demanding than a more constrained system due to the number of relationships that will often need mapping, and commensurate increases in working-memory requirements.^{22,27} This is also reflected by the fact that the human capacity to process relationships between non-adjacent elements comes online at a later stage in development than the processing of adjacent ones.^{28,29} However, it is worth mentioning that not all non-adjacent elements in a string are necessarily processed as such depending on the layer of representation at work.³⁰

One way of testing the capacity for animals to process recombinations of sound sequences of contrasting structural complexity is to use an “artificial grammar” approach.^{31,32} In such experiments, subjects can be familiarized to artificial sequences of two versus three distinct acoustic elements, for example, and then have their ability to process the associations between constituent elements in each case tested through sequence violations.³³ Using sequences of artificial sounds rather than natural call elements from the subject’s repertoire will invariably be important to remove confounding biases of prior expectation given natural element use.³⁴ In addition, using artificial sequences allows one to better generalize perceptual processing capacity across species than does the use of species-specific elements.³³ Previous studies have used this type of artificial grammar approach to probe the extent to which non-human animals can process the kinds of structures underlying syntactic and phonological structures.^{3,21,22,35,36} Here, we apply this approach in a species which naturally makes use of a limited variety of meaningful sound combinations, to elucidate the cognitive constraints that limit the productive recombination of meaningless call elements across calls, a prerequisite for productive word generation.

Previously, we have provided evidence to suggest that the highly social chestnut-crowned babbler (*Pomatostomus ruficeps*) of outback south-eastern Australia produces sound combinations with properties superficially analogous to phonemic contrasts found in human language.^{8,9} This species does not sing and is not known to be a vocal learner, but possesses a repertoire of at least 18 calls (two of which are combinatorial).³⁷ In addition, just six of these calls are comprised a single element, while the rest are sequences of 2–5 distinct elements, usually in a stereotyped order. Moreover, we have demonstrated that the same call elements can be used in different combinations across functionally distinct contexts. Specifically, movement is associated with the production of a two-sound element “flight” call of the form “A-B” (where A and B are acoustically distinct elements), while nestling provisioning is associated with “prompt” calls which contain the same two elements, but wherein the B element is repeated to form the sequence “B-A-B.” Experiments have confirmed that the A and B elements produced in flight calls are acoustically and perceptually indistinguishable from those found in prompt calls and are meaningless, in that they do not convey context-specific information.^{8,9} Yet, this case of combinatoriality is non-productive—i.e.





| | | bi-element condition | | tri-element condition | |
|--------------------------------------|--|----------------------|---|-----------------------|---|
| | | sequence | element forms | sequence | element forms |
| F A M I L I A R | | AB |  | A X B |  |
| | | CD |  | C X D |  |
| | | sequence-type | elements | sequence-type | elements |
| T E S T | | familiar | A B and C D | familiar | A X B and C X D |
| | | pitch-shifted | A' B' and C' D' | pitch-shifted | A' X' B' and C' X' D' |
| | | recombination | A' D' and C' B' | recombination | A' X' D' and C' X' B' |

Figure 2. Visual representation of each element type (A-D, Xi, Xii) and how they were combined in each sequence type in the familiarization (top) and test phases (bottom)

Familiar elements (e.g. A) could be drawn from any of pitch variants 1–8 of that sound, while pitch-shifted elements (e.g. A') were drawn from variants 9–16. X elements could be of form either Xi or Xii to minimize associations between central and edge elements. Pitch-shifted sequences were to determine whether any effects observed in response to recombination sequences could be explained by mere acoustic novelty rather than learning of sequences. Audio file examples of each sequence type are available at: <https://osf.io/mhgcx/>.

two elements combine to make just two functionally distinct calls, and neither of the sounds is combined in other sequences. Thus, although babblers have a rich call repertoire and can both produce and perceive calls comprising up to five distinct elements, they do not recombine elements across calls of three or more distinct elements. A parsimonious hypothesis is that babblers are constrained to recombine sounds across calls with just two distinct elements because they lack the cognitive capacity to process variation in the composition of sound sequences involving three or more distinct sound elements.²⁰

Here, we tested the capacity of wild-caught chestnut-crowned babblers to process combinations and recombinations of artificial sounds in sequences of two versus three distinct elements in standardised settings (Figure 2). More specifically, we first familiarized chestnut-crowned babblers to sequences of frequency-modulated sine-tone elements involving two or three distinct element types. We then played them sets of three test trials comprising: (i) the same sound sequences to which they were familiarized (“familiar sequences”); (ii) the same sound sequences but pitch-shifted, to test whether the birds can abstract the familiar categorical associations to novel acoustic stimuli (“pitch-shifted sequences”); and (iii) familiar sounds, but in pitch-shifted combinations that also violated the familiar associations between them, to test whether birds can recognize recombinations of the familiar sequences (“recombination sequences”). Throughout the three types of test trial sequence, we measured the subjects’ latency to look toward the source of the sound, as well as the duration of their gaze. If babblers are able to process a given sequence type, we predicted that they would react similarly to familiar and pitch-shifted sequences (given that they are structured according to the same rule). Recombination sequences, on the other hand, should elicit a stronger response because they are unfamiliar recombinations of sounds.^{33,38} Given that chestnut-crowned babblers routinely produce calls comprising both two and three distinct elements, but only use the same elements across sequences of two distinct elements,³⁷ we predicted a capacity to detect recombinations of two-, but not three-element sequences in our experiment.

RESULTS

Comparison of sequence types within bi- and tri-element conditions

First, we found evidence to suggest that chestnut-crowned babblers can process the recombinations of artificial sequences comprising two distinct elements. During test playbacks of bi-element sound sequences to which they were familiarized, birds took an average of 2.9 s (± 1.79 SD) to look at the speaker and did so for an average of 0.69 s (± 0.83 SD) (See Figure S1, Table S1 for full descriptive statistics). Compared to these familiar sequences, latency and gaze duration changed little in response to playbacks where the familiar bi-element sequence orders were pitch-shifted, but both changed substantially toward playbacks of recombination sequences, in which learned element orders were violated (Figure 3, Table 1). Moreover, for both response latency and gaze duration, the full model (which included playback sequence type as a fixed effect and random effects for individual and group) was a substantially better predictive fit for the data than the null model (which included only random effects, Table 1). For example, as indicated by

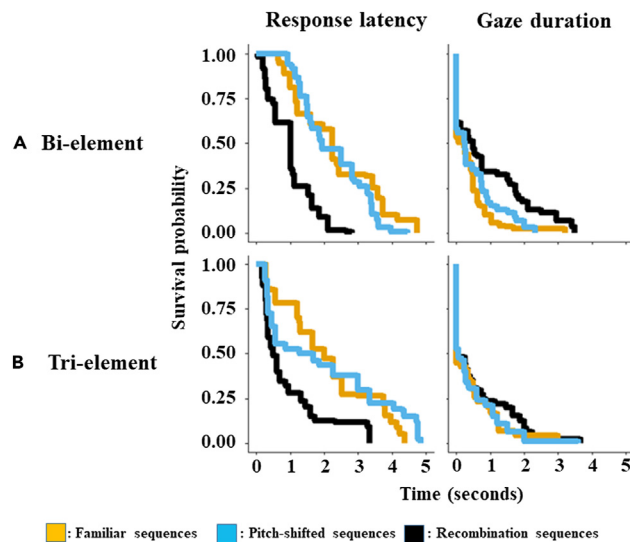


Figure 3. Survival plots for each condition (A: Bi-element, B: Tri-element) and response measure (Top: Latency, bottom: Gaze duration)

Lines indicate the probability that a behavior will start (response latency) or stop (gaze duration) after corresponding amounts of time (x axis). Familiar sequences are identical to those heard during the familiarization phase of the experiment. Pitch-shifted sequences are identical in structure to familiar sequences, but individual elements were pitch-shifted to create acoustic novelty. Recombination sequences are unfamiliar recombinations of familiar sound elements, which are also pitch-shifted. Note that while these plots graphically represent the raw data, they do not control for individual identity and repeated measures, unlike the statistical models described in Table 1. For descriptive plots which average data points by individual, see Figure S1.

the hazard ratios in Table 1, birds responded to recombination sequences on average 1.44× faster and for 1.92× longer than toward pitch-shifted sequences, as well as 1.54× faster and 2.14× longer relative to familiar sequences (Figure 3, Table 1). From these results, we can infer that the birds were readily able to learn the order of the elements in bi-element sequences, and were furthermore sensitive to unfamiliar recombinations of these elements.

By contrast, however, there was no evidence the birds could process recombinations during the tri-element condition. Here, birds (N = 11) took an average of 3.34 s (± 1.91 SD) to respond to the familiarization sequences and did so for 0.50 s (± 0.75 SD) (See Figure S1, Table S1) Here, neither response latencies nor durations reliably differed between sequence-types: responses following recombination sequences were not statistically different from those given in response to playback of familiarization or pitch-shifted sequences (Figure 3B, Table 1). As a consequence, for the tri-element condition, the model including playback trial type as a fixed effect (i.e. familiar, pitch-shifted, recombined) did not explain a reliable amount of variation in either response latency or gaze duration, indicating that sequence type was not a useful predictor for either response measure during tri-element playbacks (Table 1). This lack of difference in reactions to different arrangements of the tri-element sequences suggests that the birds were not able to detect and/or process changes in such call structures.

Comparison of same sequence types between bi-element and tri-element conditions

Finally, to elucidate the reason for the differential responses elicited by different sequence types in the bi-versus tri-element conditions, we compared responses of the same sequence type across the two conditions (e.g. the difference in response to familiarization sequences between bi- and tri-element conditions, see Figure 4). For example, one possibility is that the birds were “confused” by the tri-element condition and became disinterested in the stimuli, but an alternative is that they were able to learn the constituent sounds but not the associations between them sufficiently to process recombinations of those elements. That the latency and duration of responses to familiar sequences were comparable between bi- and tri-element conditions, irrespective of whether or not these sequences were pitch-shifted, suggests that the birds in each condition were similarly familiar with the constituent sounds in the two conditions. However, the response to recombination sequences differed substantially between conditions, with response

Table 1. Outputs for Bayesian survival models comparing behavioral response (latency to look at the speaker and gaze duration) during playbacks of recombination sequences (e.g. AD instead of AB) relative to responses during familiar sequences (e.g. AB) or pitch-shifted variants of familiar sequences

| Condition | Response | wAIC weight | Sequence-contrast | Hazard ratio | Hazard ratio lower 95% bound | Hazard ratio upper 95% bound |
|-------------|------------------|-------------|-------------------------------|--------------|------------------------------|------------------------------|
| Bi-element | Response latency | 0.75 | Recombination vs. pitch-shift | 0.54 | 0.31 | 0.98 |
| | | | Recombination vs. familiar | 0.46 | 0.25 | 0.83 |
| Bi-element | Gaze duration | 0.89 | Recombination vs. pitch-shift | 1.92 | 1.09 | 3.51 |
| | | | Recombination vs. familiar | 2.14 | 1.13 | 4.00 |
| Tri-element | Response latency | 0.09 | Recombination vs. pitch-shift | 0.90 | 0.47 | 1.73 |
| | | | Recombination vs. familiar | 0.88 | 0.48 | 1.66 |
| Tri-element | Gaze duration | 0.21 | Recombination vs. pitch-shift | 1.48 | 0.77 | 2.83 |
| | | | Recombination vs. familiar | 1.44 | 0.73 | 2.69 |

wAIC weight shows the probability that a model has the best predictive fit of those tested. Hazard Ratio indicates the relative rate at which the behavior of interest occurs (first response latency, or cessation of gaze). Hazard ratios that overlap with 1.00 indicate a lack of reliable difference between the sequences.

latency being reduced and gaze duration being extended in bi-element relative to tri-element sequences (Figure 4). These results suggest that the jump from processing recombinations of two versus three distinct elements is cognitively challenging, such that birds were no longer able to keep track of the way in which sequences of three distinct elements were combined and so the critical links between the first and third elements.

DISCUSSION

Both of our behavioral measures showed that only in the bi-element condition did individuals recognize recombinations of familiar sounds as being distinct from the specific artificial sequences to which they were familiarized. Importantly, in the bi-element condition, responses were substantially stronger during recombination sequences than during playbacks of familiar sequences of pitch-shifted elements, suggesting that increased responses cannot be explained by the acoustic novelty of stimuli *per se*. However, we found no evidence that the birds differentiated between familiar, pitch-shifted and recombination sequences in the tri-element condition. We can therefore infer that not only do babblers readily process novel bi-element sequences of artificial sounds and so the predictive relationships between pairs of constituent elements, but that the computational demands of processing recombinations of tri-element sequences appear to be prohibitive. Together, these results suggest that cognitive constraints on processing limit the capacity for chestnut-crowned babblers to recombine sound elements between calls containing three or more distinct elements.

There are several potential reasons why chestnut-crowned babblers did not recognize and/or process recombinations in the tri-element condition. First, they might have been unmotivated or confused by the trial and lost interest during the familiarization phase. We find this explanation unlikely: chestnut-crowned babblers naturally use bi- and tri-element calls in their repertoire,³⁷ and so should not have been preferentially unmotivated to engage in, or been specifically confused by, the tri- versus bi-element familiarization phases. Indeed, the behavioral responses observed were comparable during both the familiar and pitch-shifted trials of bi- and tri-element conditions, suggesting that babblers were equivalently engaged and familiar with the sounds in the two conditions. Alternatively, it might be that the babblers could not process recombinations of tri-element sequences as being “violations” of the sequences experienced during familiarization over the previous day. Further work is required to clarify whether this constraint is genetically imposed or arises through a reduced ability to learn, since babblers do not naturally recombine

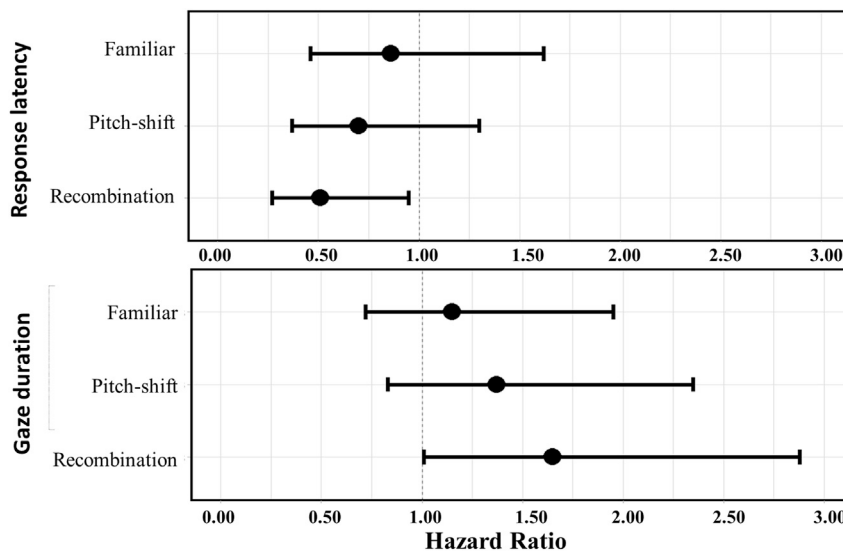


Figure 4. Outputs for between-condition comparisons (bi-element vs. tri-element) of response to sequence types
Top: Response latency. Bottom: Gaze duration. Dots represent mean estimate, lines represent 95% credible intervals. 95% CIs that do not overlap with the horizontal line indicate a robust difference between bi- and tri-element conditions.

elements across tri-element calls.^{8,9} A further possible explanation is that the ecological urgency of B-A-B prompt calls in the babbler system has left them resistant to rearrangements of tri-element sequences of any sounds. This would be an interesting case where broader pattern-recognition capabilities of a species are constrained to the extremely narrow range of ecologically relevant possibilities and might be analogous to data showing that the phonotactic constraints of native English speakers (e.g. the phoneme/ŋ/never occurs in the onset position) interfere with their ability to learn artificial grammars that violate these rules.³⁹ However, we find this unlikely in the case of chestnut-crowned babblers, as if a signal were of such high ecological urgency one would expect it to evolve to be as unambiguous and salient as possible—i.e. not combinatorial and/or not sharing elements with other less urgent calls. Regardless of the mechanism of the constraint at work, the take-home suggestion is that chestnut-crowned babblers appear to be constrained cognitively from processing recombined tri-element sequences, which at least in part, might explain why babblers do not possess a more productive combinatorial vocal system.^{8,9,37}

So what specifically makes the tri-element sequences cognitively demanding to process relative to the bi-element sequences? We identify several alternatives. One straightforward possibility is that limited working memory constrains their ability to mentally represent the entire sequence at once so that the birds simply cannot remember what came at the start of the sequence by the time they get to the end of it. While the birds do have tri-element sequences in their natural repertoire (the “B-A-B” prompt calls), their natural call combinations are composed of very short sounds, whereas our tri-element sequences were 4.5s long in total. Hence, this may be beyond the limits of their working memory, or more simply their motivation to attend to the stimuli for a long time given the lack of ecological relevance. A useful control in further work would therefore be to use stimuli which bear a closer temporal resemblance to their natural calls. However, the birds were able to process the sequences in our bi-element condition which were already 3s. A more likely possibility is that recalling three distinct elements places greater working memory demands than just two, as this requires retrieving three rather than two items from long-term memory and also requires a greater understanding of the positionality of the individual elements. As an illustrative example, the heuristics necessary to recognize the sequence “AXA” (“Contains A” + “A occurs at each edge”) can be relatively simple compared to those necessary for processing “AXB” (“Contains A” + “Contains B” + “A occurs first” + “B occurs last”).^{40,41} This is consistent with artificial grammar studies in humans which find that relationships between distant elements are easier to process when those elements are perceptually similar³⁵ or identical.⁴² Further work with chestnut-crowned babblers could unpack the precise nature of the cognitive constraints in play using additional experimental conditions that probe their ability to process (i) novel tri-element sequences comprising just two distinct elements (i.e. AXA) to determine whether sequences of similar length but simpler composition to those we provided are learnable and

(ii) the same sequences with additional middle-elements (e.g. AXXA, AXXXA and so on) to explore whether simpler compositions of greater length can be learned.

The results of this study have at least two more general implications which we hope will inspire future research. First, they provide a theoretical platform from which to test the conundrum of why combinatorial calls, where communicative function varies with changes in sequence structure, are typically limited to recombinations of just two distinct sound entities (elements or calls),^{4,5,20} whereas animal songs, despite often comprising tens of semantically meaningless sound elements in myriad ways, signal little more than individual presence and current condition.⁴³ We propose that the need to disambiguate meaningful call combinations from one another through holistically processing the type and positionality of individual elements means that the number of distinct elements in a sequence is a significant factor (in addition to e.g. sequence length), which may increase the corresponding cognitive demands. Here, we found that chestnut-crowned babblers were unable to do this with sequences of three distinct elements, which may explain why their natural vocal production system only contains recombinations of two distinct elements. This hypothesis could be further explored by carrying out comparative studies with species known to recombine a greater or fewer number of distinct elements than chestnut-crowned babblers, as well as singing species. Second, by extension, we suggest that increases in processing capacity are an evolutionary prerequisite for increases in the productive power of combinatorial signals, with implications for understanding language evolution. For example, although it has been hypothesised that acoustic constraints in generating new, discriminable sounds were key in promoting the switch to combinatorial structuring during hominin evolution,⁴⁴ it seems likely that overcoming the computational demands associated with processing recombinations of sounds with greater than two entities was an additional fundamental step necessary for productive combinatoriality to emerge. Indeed, it is noteworthy that primates, including squirrel monkeys (*Saimiri sciureus*),⁴⁵ marmosets (*Callithrix jacchus*), and chimpanzees (*Pan troglodytes*) appear capable of processing recombinations of artificial tri-element sound sequences, despite not yet having been demonstrated to use tri-element combinatorial signals in their vocal repertoires.^{33,38,46} The fact that these species demonstrate a more sophisticated capacity for combinatorial processing than chestnut-crowned babblers, a species that makes habitual use of combinatorial calls is both striking and puzzling. It may therefore be that the capacity to process complex combinatorial structures is not a domain-specific cognitive trait, but rather an expression of a more generalized capacity for pattern recognition with applications to other domains, such as social cognition and foraging.^{47,48} This may represent yet another avenue in which increases in social and ecological complexity drive commensurate increases in communicative complexity.^{49–54} One way to shed more light on these important implications is to apply a similar experimental approach to a range of species with combinatorial calls of varying complexity to explore the extent to which the complexity of their production system corresponds with their perceptual capacities. Using standardized “artificial grammar” experiments offers scope for such cross-species comparisons on the coevolution of the capacity to process and produce combinatorial sequences of varying complexity.

Limitations of the study

The stimuli used in our experiment were artificial and very different in both sound and duration from the natural vocal repertoire of the babblers: our acoustic elements lasted 1500ms each, whereas individual babbler vocalisations are typically less than 500ms long.³⁷ It is therefore conceivable that this introduced working memory demands over and above those imposed by the structural properties of our sequences. While the data from our bi-element condition demonstrates this was not an insurmountable hurdle for the birds, it presents a potential 5 confound for their performance in the tri-element condition. Further work may benefit from including additional conditions that use stimuli more acoustically tailored to the features of the study species, to explore the impact of this factor.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS](#)

- Study site and subjects
- Ethical approval
- **METHOD DETAILS**
 - Playback format
 - Acoustic stimuli details
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.106977>.

ACKNOWLEDGMENTS

We thank Simon Griffith, the Dowling family and Keith Leggett for their logistical support at Fowler's Gap. We are grateful to our anonymous reviewers for their feedback on this manuscript.

Funding acknowledgments: S.K.W and S.W.T were funded by the Swiss National Science Foundation (S.K.W & S.W.T: Grant PP00P3_163850). S.K.W was partially funded by NCCR Evolving Language, Swiss National Science Foundation Agreement #51NF40_180888. L.G.O.N. was funded by a joint scholarship from Macquarie University (AUS) and the University of Exeter (UK).

AUTHOR CONTRIBUTIONS

S.K.W, J.G.M., A.F.R. and S.W.T conceptualized the project. S.K.W., J.L.M., L.G.O.N. and A.F.R. carried out bird capture, release and data collection. S.K.W. carried out the statistical analysis. S.W.T. and A.F.R. directed the project. All authors contributed toward interpretation of the data, as well as writing and editing of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community.

Received: April 20, 2022

Revised: November 29, 2022

Accepted: May 24, 2023

Published: May 26, 2023

REFERENCES

1. Zentall, T.R. (2018). The value of research in comparative cognition. *Int. J. Comp. Psychol.* *31*, 1–17.
2. Fedurek, P., and Slocombe, K.E. (2011). Primate vocal communication: a useful tool for understanding human speech and language evolution? *Hum. Biol.* *83*, 153–173. <https://doi.org/10.3378/027.083.0202>.
3. ten Cate, C., and Okanoya, K. (2012). Revisiting the syntactic abilities of non-human animals: natural vocalizations and artificial grammar learning. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *367*, 1984–1994.
4. Coye, C., Townsend, S., and Lemasson, A. (2018). From animal communication to linguistics and back: insight from combinatorial abilities in monkeys and birds. In *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, 187. L.-J. Boë, J. Fagot, and P. Perrier, eds. (Frankfurt am Main: Peter Lang), pp. 196–242.
5. Zuberbühler, K. (2020). Syntax and compositionality in animal communication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* *375*, 20190062. <https://doi.org/10.1098/rstb.2019.0062>.
6. Zuberbühler, K. (2019). Evolutionary roads to syntax. *Anim. Behav.* *151*, 259–265.
7. Townsend, S.W., Engesser, S., Stoll, S., Zuberbühler, K., and Bickel, B. (2018). Compositionality in animals and humans. *PLoS Biol.* *16*, e2006425. <https://doi.org/10.1371/journal.pbio.2006425>.
8. Engesser, S., Holub, J.L., O'Neill, L.G., Russell, A.F., and Townsend, S.W. (2019). Chestnut-crowned babbler calls are composed of meaningless shared building blocks. *Proc. Natl. Acad. Sci. USA* *116*, 19579–19584. <https://doi.org/10.1073/pnas.1819513116>.
9. Engesser, S., Crane, J.M.S., Savage, J.L., Russell, A.F., and Townsend, S.W. (2015). Experimental evidence for phonemic contrasts in a nonhuman vocal system. *PLoS Biol.* *13*, e1002171. <https://doi.org/10.1371/journal.pbio.1002171>.
10. Andrieu, J., Penny, S.G., Bouchet, H., Malaivijitnond, S., Reichard, U.H., and Zuberbühler, K. (2020). White-handed gibbons discriminate context-specific song compositions. *PeerJ* *8*, e9477. <https://doi.org/10.7717/peerj.9477>.
11. Suzuki, T.N., Wheatcroft, D., and Griesser, M. (2017). Wild birds use an ordering rule to decode novel call sequences. *Curr. Biol.* *27*, 2331–2336.e3.
12. Engesser, S., Ridley, A.R., and Townsend, S.W. (2016). Meaningful call combinations

- and compositional processing in the southern pied babbler. *Proc. Natl. Acad. Sci. USA* 113, 5976–5981. <https://doi.org/10.1073/pnas.1600970113>.
13. Suzuki, T.N., and Zuberbühler, K. (2019). Animal syntax. *Curr. Biol.* 29, R669–R671. <https://doi.org/10.1016/j.cub.2019.05.045>.
 14. Suzuki, T.N., Wheatcroft, D., and Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nat. Commun.* 7, 10986. <https://doi.org/10.1038/ncomms10986>.
 15. Devoogd, T.J., Krebs, J.R., Healy, S.D., and Purvis, A. (1993). Relations between song repertoire size and the volume of brain nuclei related to song: comparative evolutionary analyses amongst oscine birds. *Proc. Biol. Sci.* 254, 75–82. <https://doi.org/10.1098/rspb.1993.0129>.
 16. Kroodsmas, D.E., and Byers, B.E. (1991). The function(s) of bird song. *Am. Zool.* 31, 318–328. <https://doi.org/10.1093/icb/31.2.318>.
 17. Katahira, K., Suzuki, K., Okanoya, K., and Okada, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden markov processes. *PLoS One* 6, e24516. <https://doi.org/10.1371/journal.pone.0024516>.
 18. Fishbein, A.R., Fritz, J.B., Idsardi, W.J., and Wilkinson, G.S. (2020). What can animal communication teach us about human language? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375, 20190042. <https://doi.org/10.1098/rstb.2019.0042>.
 19. Suzuki, T.N., Wheatcroft, D., and Griesser, M. (2020). The syntax–semantics interface in animal vocal communication. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 375, 20180405. <https://doi.org/10.1098/rstb.2018.0405>.
 20. Miyagawa, S., and Clarke, E. (2019). Systems underlying human and old world monkey communication: one, two, or infinite. *Front. Psychol.* 10, 1911. <https://doi.org/10.3389/fpsyg.2019.01911>.
 21. Mueller, J.L., Milne, A., and Männel, C. (2018). Non-adjacent auditory sequence learning across development and primate species. *Curr. Opin. Behav. Sci.* 21, 112–119. <https://doi.org/10.1016/j.cobeha.2018.04.002>.
 22. Wilson, B., Spierings, M., Ravnani, A., Mueller, J.L., Mintz, T.H., Wijnen, F., Van Der Kant, A., Smith, K., and Rey, A. (2020). Non-adjacent dependency learning in humans and other animals. *Top. Cogn. Sci.* 12, 843–858.
 23. van Kampen, A., Parmaksiz, G., van de Vijver, R., Höhle, B., Gavarró, A., and Freitas, M.J. (2008). *Language Acquisition and Development: Proceedings of GALA 2007* (Cambridge Scholars Publishing).
 24. Gonzalez-Gomez, N., and Nazzi, T. (2012). Acquisition of nonadjacent phonological dependencies in the native language during the first year of life. *Infancy* 17, 498–524. <https://doi.org/10.1111/j.1532-7078.2011.00104.x>.
 25. Gonzalez-Gomez, N., and Nazzi, T. (2012). In *Phonological feature constraints on the acquisition of phonological dependencies* (Cascadilla Press), pp. 202–212.
 26. Sandoval, M., and Gómez, R.L. (2013). The development of nonadjacent dependency learning in natural and artificial languages. *WIREs Cogn. Sci.* 4, 511–522. <https://doi.org/10.1002/wcs.1244>.
 27. Daneman, M., and Merikle, P.M. (1996). Working memory and language comprehension: a meta-analysis. *Psychon. Bull. Rev.* 3, 422–433. <https://doi.org/10.3758/BF03214546>.
 28. Friederici, A.D., Mueller, J.L., and Oberecker, R. (2011). Precursors to natural grammar learning: preliminary evidence from 4-month-old infants. *PLoS One* 6, e17920. <https://doi.org/10.1371/journal.pone.0017920>.
 29. Teinonen, T., Fellman, V., Näätänen, R., Alku, P., and Huottilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neurosci.* 10, 21. <https://doi.org/10.1186/1471-2202-10-21>.
 30. Halle, M., and Vergnaud, J.R. (2020). On the Framework of autosegmental phonology. In *The Structure of Phonological Representations (Part 1)* (De Gruyter), pp. 65–82. <https://doi.org/10.1515/9783112328088-004>.
 31. Reber, A.S. (1967). Implicit learning of artificial grammars. *J. Verb. Learn. Verb. Behav.* 6, 855–863. [https://doi.org/10.1016/S0022-5371\(67\)80149-X](https://doi.org/10.1016/S0022-5371(67)80149-X).
 32. Beckers, G.J.L., Berwick, R.C., Okanoya, K., and Bolhuis, J.J. (2017). What do animals learn in artificial grammar studies? *Neurosci. Biobehav. Rev.* 81, 238–246. <https://doi.org/10.1016/j.neubiorev.2016.12.021>.
 33. Watson, S.K., Burkart, J.M., Schapiro, S.J., Lambeth, S.P., Mueller, J.L., and Townsend, S.W. (2020). Nonadjacent dependency processing in monkeys, apes, and humans. *Sci. Adv.* 6, eabb0725. <https://doi.org/10.1126/sciadv.abb0725>.
 34. Ravnani, A., Filippi, P., and Tecumseh Fitch, W. (2019). Perceptual tuning influences rule generalization: testing humans with monkey-tailored stimuli. *Iperception.* 10, 2041669519846135.
 35. Wilson, B., Spierings, M., Ravnani, A., Mueller, J.L., Mintz, T.H., Wijnen, F., van der Kant, A., Smith, K., and Rey, A. (2020). Non-adjacent dependency learning in humans and other animals. *Top. Cogn. Sci.* 12, 843–858. <https://doi.org/10.1111/tops.12381>.
 36. Berwick, R.C., Okanoya, K., Beckers, G.J.L., and Bolhuis, J.J. (2011). Songs to syntax: the linguistics of birdsong. *Trends Cogn. Sci.* 15, 113–121. <https://doi.org/10.1016/j.tics.2011.01.002>.
 37. Crane, J.M.S., Savage, J.L., and Russell, A.F. (2016). Diversity and function of vocalisations in the cooperatively breeding Chestnut-crowned Babbler. *Emu - Austral Ornithol.* 116, 241–253. <https://doi.org/10.1071/MU15048>.
 38. Fitch, W.T., and Hauser, M.D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.
 39. Hwang, S., Dell, G., and Fisher, C. (2022). Is the learning of artificial phonotactic rules interfered with by the concurrent experience of English? *Proceedings of the Annual Meeting of the Cognitive Science Society, University of California* 44, 2207–2213.
 40. van Heijningen, C.A.A., Chen, J., van Laatum, I., van der Hulst, B., and ten Cate, C. (2013). Rule learning by zebra finches in an artificial grammar learning task: which rule? *Anim. Cogn.* 16, 165–175.
 41. van Heijningen, C.A.A., De Visser, J., Zuidema, W., and Ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proc. Natl. Acad. Sci. USA* 106, 20538–20543.
 42. Gallagher, G. (2013). Learning the identity effect as an artificial language: bias and generalisation. *Phonology* 30, 253–295. <https://doi.org/10.1017/S0952675713000134>.
 43. Catchpole, C.K., and Slater, P.J.B. (2003). *Bird Song: Biological Themes and Variations* (Cambridge University Press).
 44. Nowak, M.A., and Krakauer, D.C. (1999). The evolution of language. *Proc. Natl. Acad. Sci. USA* 96, 8028–8033. <https://doi.org/10.1073/pnas.96.14.8028>.
 45. Ravnani, A., Sonnweber, R.-S., Stobbe, N., and Fitch, W.T. (2013). Action at a distance: dependency sensitivity in a New World primate. *Biol. Lett.* 9, 20130852.
 46. Leroux, M., and Townsend, S.W. (2020). Call combinations in great apes and the evolution of syntax. *Anim. Behav. Cogn.* 7, 131–139. <https://doi.org/10.26451/abc.07.02.07.2020>.
 47. Pastra, K., and Aloimonos, Y. (2012). The minimalist grammar of action. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 103–117.
 48. Mielke, A., and Carvalho, S. (2022). Chimpanzee play sequences are structured hierarchically as games. Preprint at bioRxiv. <https://doi.org/10.1101/2022.06.14.496075>.
 49. Freeberg, T.M., and Krams, I. (2015). Does social complexity link vocal complexity and cooperation. *J. Ornithol.* 156, 125–132.
 50. Freeberg, T.M. (2006). Social complexity can drive vocal complexity: group size influences vocal information in carolina chickadees. *Psychol. Sci.* 17, 557–561. <https://doi.org/10.1111/j.1467-9280.2006.01743.x>.
 51. Knörnschild, M., Fernandez, A.A., and Nagy, M. (2020). Vocal information and the navigation of social decisions in bats: is social complexity linked to vocal complexity? *Funct. Ecol.* 34, 322–331. <https://doi.org/10.1111/1365-2435.13407>.

52. Pougault, L., Levréro, F., Leroux, M., Paulet, J., Bombani, P., Dentressangle, F., Deruti, L., Mulot, B., and Lemasson, A. (2022). Social pressure drives “conversational rules” in great apes. *Biol. Rev.* 97, 749–765.
53. Krams, I., Krama, T., Freeberg, T.M., Kullberg, C., and Lucas, J.R. (2012). Linking social complexity and vocal complexity: a parid perspective. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1879–1891. <https://doi.org/10.1098/rstb.2011.0222>.
54. Leroux, M., Chandia, B., Bosshard, A.B., Zuberbühler, K., and Townsend, S.W. (2022). Call combinations in chimpanzees: a social tool? *Behav. Ecol.* 33, 1036–1043. <https://doi.org/10.1093/beheco/arac074>.
55. Rollins, L.A., Browning, L.E., Holleley, C.E., Savage, J.L., Russell, A.F., and Griffith, S.C. (2012). Building genetic networks using relatedness information: a novel approach for the estimation of dispersal and characterization of group structure in social animals. *Mol. Ecol.* 21, 1727–1740. <https://doi.org/10.1111/j.1365-294X.2012.05492.x>.
56. Sorato, E., Gullett, P.R., Griffith, S.C., and Russell, A.F. (2012). Effects of predation risk on foraging behaviour and group size: adaptations in a social cooperative species. *Anim. Behav.* 84, 823–834. <https://doi.org/10.1016/j.anbehav.2012.07.003>.
57. Nomano, F.Y., Browning, L.E., Savage, J.L., Rollins, L.A., Griffith, S.C., and Russell, A.F. (2015). Unrelated helpers neither signal contributions nor suffer retribution in chestnut-crowed babblers. *Behav. Ecol.* 26, 986–995. <https://doi.org/10.1093/beheco/arv023>.
58. Boersma, P., and Weenink, D. (2019). *Praat: Doing Phonetics by Computer*.
59. Friard, O., and Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods Ecol. Evol.* 7, 1325–1330. <https://doi.org/10.1111/2041-210X.12584>.
60. Jahn-Eimermacher, A., Lasarzik, I., and Raber, J. (2011). Statistical analysis of latency outcomes in behavioral experiments. *Behav. Brain Res.* 221, 271–275. <https://doi.org/10.1016/j.bbr.2011.03.007>.
61. R Core Team (2014). *R: A Language and Environment for Statistical Computing*.
62. RStudio Team (2015). *RStudio: Integrated Development for R*.
63. Bürkner, P.C. (2017). brms : an R package for bayesian multilevel models using stan. *J. Stat. Soft.* 80. <https://doi.org/10.18637/jss.v080.i01>.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|--|------------------------|--|
| Deposited data | | |
| Raw data and analysis scripts | Open Science Framework | https://osf.io/mhgcx/ https://doi.org/10.17605/OSF.IO/MHGXCX |
| Software and algorithms | | |
| R statistical programming language | R Project | https://www.r-project.org/ |
| Behavioral Observation Research Interactive Software ('BORIS') | Universita Di Torino | https://www.boris.unito.it/ |

RESOURCE AVAILABILITY

Lead contact

Further information and requests should be directed to and will be fulfilled by the lead contact, Stuart K Watson (swatso88@gmail.com).

Materials availability

This study did not generate any new unique reagents.

Data and code availability

- All raw data used for analysis in this study have been deposited at the following Open Science Framework repository: <https://osf.io/mhgcx/>. This data is publicly available as of the date of publication. DOI is listed in the [key resources table](#).
- All original code has been deposited at the following Open Science Framework repository: <https://osf.io/mhgcx/>. DOI is listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Study site and subjects

The study was conducted at Fowlers Gap Arid Zone Research Station in New South Wales, Australia (141°42'E, 31°06'S). This population has been under long-term investigation since 2004, and details of the study site and population are published elsewhere (e.g.^{55,56}). A total of 24 adults in 10 groups (2–5 per group, comprising 20–44% of the group) were captured in mist nets, selected at random, and transported the few kilometres by vehicle in bird bags to on-site aviary compartments (2 × 2 × 2.5m, see⁹ for further details of housing conditions). It was not possible to discern the sex of the subjects from visual cues. Birds from the same group were housed together, and usually settled and began feeding within 20 min of release into the aviary; birds from different groups were not housed concurrently. Birds were maintained on ~20 mealworms every 3h and water was provided *ad libitum*. Aviary compartments included natural soil substrate, branches and a nest for roosting. All birds were released back into their original groups successfully within 48 h of initial capture, typically gained 1–2 g mass and were accepted back into the group without retribution.⁵⁷

Ethical approval

Ethics approval was provided by UNSW Animal Care and Ethics Committee (06/40A), Macquarie University (2107/025), The University of Exeter, NSW National Parks and Wildlife Service and the Australian Bird and Bat Banding Scheme (3340).

METHOD DETAILS

Playback format

The experiment comprised of a familiarisation phase followed by a test phase; in both cases, stimuli were played on a Braven BRV-X speaker (see below for specific details of stimuli composition in each phase). Starting from at least 1 hour post-settlement in the aviary, birds were provided with 10 familiarisation sessions: 5 on Day 1 and 5 on Day 2, with > 30 min between each session. In each familiarisation session, the birds were played a list of 240 familiarisation sequences (see 'acoustic stimuli' below for details of sequence construction) with 2500ms of silence between each sequence to increase the likelihood they would be perceived as separate. Half of the birds (N = 12) received familiarisations sequences of 2 x 2 distinct elements (i.e., A-B and C-D; 'bi-element condition'; [Figure 2](#)) and the other half received familiarisations sequences of 2 x 3 distinct elements (i.e., A-X-B and C-X-D; 'tri-element condition'; [Figure 2](#)), with former sessions lasting 24 min each and the latter lasting 32 min each. The full lists of sequences used for familiarisation and test phases can be accessed at: <https://osf.io/mhgcx/>.

The test phase was performed on individual birds immediately before release, just after dawn on Day 3 (~45 h after initial capture). To isolate birds for the test phase, birds were removed from the main aviary on nightfall of Day 2 (using a red-light torch) and roosted over-night individually in a covered wooden box (45 × 20 × 20 cm lwh, one side mesh, contained several perches) at room temperature. Test sessions were preceded by a brief 'refamiliarisation' session of 60 different sequences taken from the familiarisation sessions. After two minutes of silence, the experimenter commenced playback of 12 test trials. These 12 trials were comprised of: (i) 4 randomly selected familiar sequences of A-B and C-D (bi-element condition) or A-X-B and C-X-D (tri-element condition); (ii) 4 pitch-shifted sequences of A-X-B and C-X-D, to test whether birds had learned the associations between the elements in familiar sequences, even in novel sound variants; and (iii) 4 recombined sequences, such that D (not B) followed A and C followed B (with an X element between each one in the tri-element condition). See below for full details of how these sequences were defined and constructed. A 5 s response window was left between each bi- or tri-element playback sequence, beginning from the onset of the final sound in the sequence, during which we coded for behavioural responses to the playback trial (latency of first look towards the speaker, and total gaze duration). The test phase was carried out in a featureless 200 × 400 cm room, except for the remotely operated Sony HDR-CX 240E digital camcorder and speaker (volume set at 55 dB to mirror natural call volume) positioned together on a tripod 100 cm from the mesh front of the box containing the bird (the same box they roosted in overnight), inside which the bird could move freely. The speaker was not concealed, so that the source of the sound would be immediately apparent to the subject. The experimenter operated the equipment remotely from outside the room so that they were not visible to the bird.

Acoustic stimuli details

Acoustic elements

We generated 6 types of acoustic elements, which we refer to as elements A, B, C, D, Xi and Xii ([Figure 2](#)). For each element type, we generated 16 pitch variants, each of which had identical pitch contours but different starting pitches, which we refer to as e.g. A1, A5, B3, C16, etc. Half of these variants (numbers 1–8) were used in familiarisation sequences, with variant 1 having its onset start at 500Hz and each subsequent variant starting 50Hz higher. Variants 9–16 were used in the pitch-shift and recombination sequences, with variant 9 starting at 1100Hz and each subsequent variant starting 50Hz higher. A gap of 250Hz (850–1100) was inserted between variants 8 and 9 to increase the saliency of the difference between familiarisation and pitch-shift/recombination sequences. All elements within a sequence were played back at the same volume in an otherwise silent room. All elements were well within the babblers' natural call range (<300 to >4000 Hz) with previous research suggesting that babblers can discern natural call elements that differ in frequency of <50 Hz.³⁷ All elements had a duration of 1500ms, with a 10ms volume fade in/out to eliminate sound onset effects. All stimuli were generated using the software Praat.⁵⁸ These elements were designed with the intent that each category is easily acoustically distinguishable from the next, and have been productively applied to examining the sequential processing abilities of humans and primates (32). A sample sound generation script, and all sounds associated with this experiment, can be downloaded from: <https://osf.io/mhgcx/>.

Sequence composition

The birds assigned to the bi-element condition were exposed to sequences comprised of elements A, B, C and D only, while those exposed to the tri-element condition additionally heard Xi and Xii elements. The reason for oscillating between the use of these two X elements was to minimise the likelihood that individuals simply learned another set of associations between the middle (i.e. Xi) and edge elements (i.e. A and B) to process the tri-element sequences, thus necessitating an understanding of the relationship between the non-consecutive first and last elements. In other words, while the relationship between A/B or C/D had total predictive certainty, the predictive accuracy between edge and central elements was relatively low, encouraging reliance on these non-adjacent relations. An additional consideration was that introducing variability to this central element mitigated the possibility that the birds would ignore it and parse the edge elements as adjacent. Within both bi- and tri-element sequences, there was 500ms of silence between each element. Correspondingly, bi-element sequences lasted a total of 3500ms, whereas tri-element sequences lasted 5500ms.

Familiarisation sequences were random assortments of frequency variants 1–8 played during the familiarisation phase (i.e. starting frequencies varying from 500 to 850 Hz). Pitch-shifted sequences were identical to familiarisation sequences in all regards except that they were comprised of variants 9–16 (i.e. starting frequencies from 1100–1450 Hz). Finally, recombined sequences were also pitch-shifted (i.e. from 1100–1450 Hz), but this time the sequences were recombined in the form A(Xi/ii)D and B(Xi/ii)C.

List composition

In the familiarisation phase, sequences were played in a pseudorandom order such that each combination of element variants (e.g., A1-B6, A2-B4, C7-D2, A5-B5 etc; total combinations = 60) appeared four times (240 sequences in the list), but never more than three times in a row. All lists contained the same number (N = 120) of AB sequences and C-D sequences (with Xi/Xii inserted in the tri-element condition).

During the test phase, each bird was exposed to 4 familiar, 4 pitch-shifted and 4 recombined sequences (12 trials total). These were played in pseudorandom order such that no condition was heard more than three times in a row to minimise habituation. Whether a subject heard a pitch-shifted or recombination sequence as the first playback in the list was counter-balanced across subjects, so that a potential effect of the mere novelty of the pitch-variants was controlled for. For familiar and pitch-shifted sequences in the test phase, half were A-B sequences, and the other half were C-D sequences. For the recombined sequences, half were of form A-D and the other half were C-B (with Xi/Xii inserted in the middle for sequences in the tri-element condition). Sequences heard by birds in the bi- and tri-element conditions were identical in the sounds used to comprise them, the only difference being the addition of Xi and Xii in the middle of tri-element sequences.

QUANTIFICATION AND STATISTICAL ANALYSIS

All videos were coded frame-by-frame using BORIS behavioural observation software⁵⁹ by SKW. Specifically, we coded the time and duration of each occasion the bird looked directly towards the speaker during the 5 s response window that followed the onset of the final sound in a playback sequence. A direct look was determined as occasions where the bird's beak was oriented straight-on with the camera (which was positioned at the speaker). This response measure has been validated in previous playback experiments with this species.⁸ While it may be that some birds attended to the speaker at an offset angle, it was not possible to reliably code a standardised offset criterion (e.g. orientations within 30 degrees of the speaker) due to the non-uniform positionality of the birds within the cage during testing. If the bird was already looking towards the camera at the time of onset of the final element in a sequence, or if the bird was vocalising over the playback sounds, the trial was not used for analysis (55 / 252 trials). When not reacting to the experiment, birds were generally resting or moving inside the box. For each trial, we then extracted: (a) the latency of a subject's first look towards the speaker (hereafter 'response latency'); and (b) the total amount of time spent looking directly toward the speaker (hereafter 'gaze duration').

Inter-observer reliability tests were carried out on ~25% of all trials (N = 50, N = 25 for each condition). To this end, a second coder (J.G.M.) was provided with muted clips of the 5 s response windows to ensure that they were blind to both condition and stimulus type (the primary coder, SKW, used unmuted videos). Pearson's correlation analysis suggested an overall high level of agreement between observers for both

response latency (bi-element condition r : 0.937, tri-element condition r : 0.966, overall r : 0.952). and gaze duration (bi-element condition r : 0.804, tri-element condition r : 0.952, overall r : 0.841).

To allow for the inclusion of trials in which there was no response from the subject, we censored this data using survival analyses.⁶⁰ Specifically, we employed Bayesian Cox proportional hazards models.⁶⁰ One full model and one null model were fit for each condition (bi-element and tri-element conditions) and each outcome variable (response latency and gaze duration) for a total of 8 models. The full model included a fixed effect of sequence type (3-level factor: familiarisation, pitch-shifted and recombination sequences) and a random effect of individual identity, whereas the null model contained only the latter. Comparisons between full and null models were carried out using Watanabe-Akaike information criterion (wAIC) weights in order to determine the relative likelihood that a given model provided the best predictive fit for the data. Whether there was a difference between conditions (bi-element / tri-element) in response to the same sequence type was then explored using six further models examining each combination of sequence-type and outcome variable. Because multiple datapoints were used for each subject, each of these models also included random intercepts for each individual.

All models were implemented in R⁶¹ and RStudio⁶² using the package 'brms'.⁶³ Model chain convergence was assessed by inspecting trace plots, rhat values (all equal to 1.00) and effective sample sizes (all over 1000). All data and code used for analysis, as well as markdowns of model outputs, can be accessed at the following repository: <https://osf.io/mhgcx/>.