

FORMAL COMMENT

Call combinations in birds and the evolution of compositional syntax

Toshitaka N. Suzuki¹*, David Wheatcroft², Michael Griesser³

1 Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Kanagawa, Japan, **2** Department of Animal Ecology, Uppsala University, Uppsala, Sweden, **3** Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

* These authors contributed equally to this work.

* toshi.n.suzuki@gmail.com



Abstract

Syntax is the set of rules for combining words into phrases, providing the basis for the generative power of linguistic expressions. In human language, the principle of compositionality governs how words are combined into a larger unit, the meaning of which depends on both the meanings of the words and the way in which they are combined. This linguistic capability, i.e., compositional syntax, has long been considered a trait unique to human language. Here, we review recent studies on call combinations in a passerine bird, the Japanese tit (*Parus minor*), that provide the first firm evidence for compositional syntax in a nonhuman animal. While it has been suggested that the findings of these studies fail to provide evidence for compositionality in Japanese tits, this criticism is based on misunderstanding of experimental design, misrepresentation of the importance of word order in human syntax, and necessitating linguistic capabilities beyond those given by the standard definition of compositionality. We argue that research on avian call combinations has provided the first steps in elucidating how compositional expressions could have emerged in animal communication systems.

OPEN ACCESS

Citation: Suzuki TN, Wheatcroft D, Griesser M (2018) Call combinations in birds and the evolution of compositional syntax. PLoS Biol 16(8): e2006532. <https://doi.org/10.1371/journal.pbio.2006532>

Received: May 2, 2018

Accepted: July 11, 2018

Published: August 15, 2018

Copyright: © 2018 Suzuki et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Japan Society for the Promotion of Science KAKENHI (grant number 16752305, 18024602, 18033543). received by TNS. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: A, adjective; Conj, conjunctive; N', noun; NP, noun phrase; P, preposition; V, verb; VP, verb phrase.

Provenance: Not commissioned; externally peer reviewed.

Language is an essential human trait, but understanding its evolution remains a challenge, in large part because none of humans' close evolutionary relatives (i.e., the great apes) possess communication systems approaching the complexity and flexibility of language [1]. The communicative power of language depends on multiple co-occurring capabilities, including referentiality, compositionality, syntax, recursivity, and vocal learning [2]. Thus, a profitable approach to gain insights into language evolution is to find and use appropriate analogies of these capabilities in other animals for comparative studies [1,2]. Over the past three decades, both field and laboratory studies have revealed that several linguistic capabilities, such as vocal learning and referentiality, did evolve in a range of animal taxa, providing insights into possible evolutionary drivers of these faculties of language (in a broad sense) [2]. Here, we review recent studies on bird calls demonstrating the first evidence for another key linguistic capability (i.e., 'compositional syntax' [3]) in a nonhuman vocal system [4–6] and discuss recent criticism of the interpretation of these studies.

Syntax is the set of rules for combining words into phrases. In human syntax, combinations of words follow ‘the principle of compositionality’, in which the meaning of a combination depends on both the meaning of its parts and the way in which they are combined [3]. Although combinations of meaningful calls have been documented for nonhuman primates [7,8], they have typically been considered noncompositional, i.e., a structured sequence provides an independent meaning as a whole [3,9]. The lack of evidence for compositional syntax in nonhuman animals makes it difficult to study its origins and evolution using comparative approaches. Assigning meaning to animal signals in the way in which we would assign meaning to human words is a challenge, but the meaning of animal signals could be assessed by observing and associating responses to different types of signals. Thus, three conditions are necessary to demonstrate compositional syntax in nonhuman animals: (i) production of call sequences depends on the context; (ii) response to the whole sequence depends on assessing the meanings of the component calls, rather than simply responding to the whole sequence as a single meaningful unit; and (iii) response to the whole sequence depends on the way in which the component calls are combined.

Recent studies have revealed that call combinations in a bird species, the Japanese tit (*Parus minor*) (Fig 1A), satisfy all three criteria for compositional syntax. Japanese tits have over 10 different notes in their vocal repertoire and use them either singly or in combination with other notes [4]. They produce ‘ABC’ calls when warning conspecifics about predators, while they produce ‘D’ calls when attracting conspecifics [4,5]. Interestingly, tits combine these two calls into ‘ABC-D’ sequences when recruiting conspecifics to mob a stationary predator [4]. Playback experiments [5] showed that tits display different behaviours when hearing ABC calls

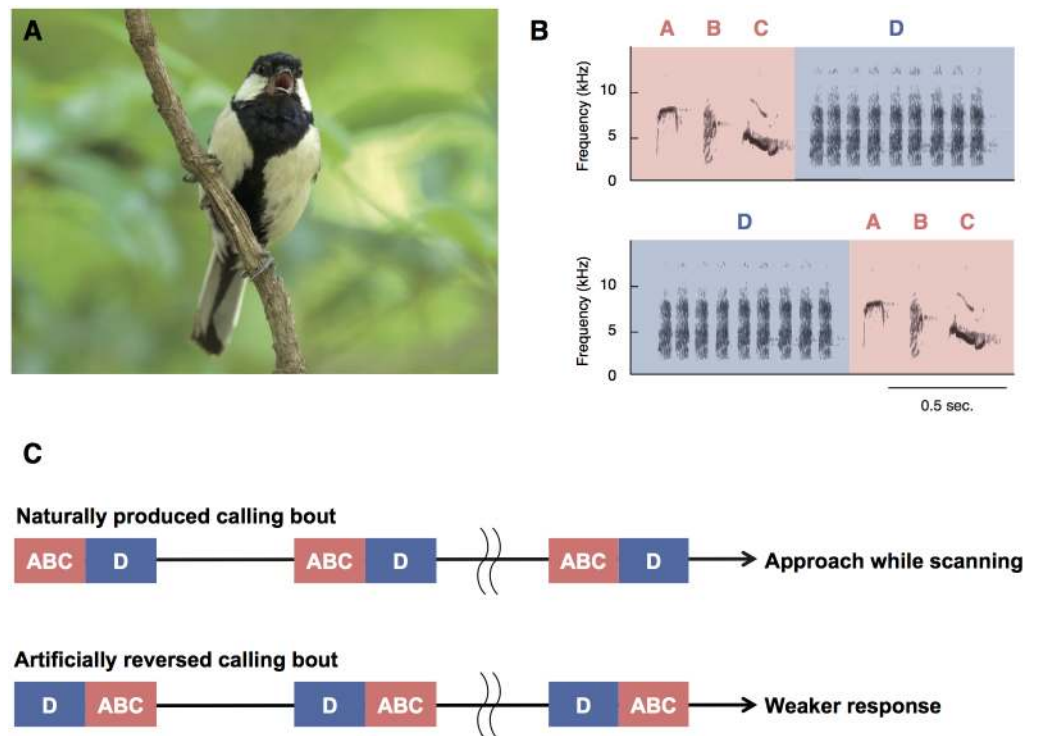


Fig 1. Schematic representation of experimental design. (A) Japanese tit. (B) Sound spectrogram of naturally produced ABC-D and artificially reversed D-ABC sequence. (C) Schematic representation of playback stimulus and responses (Experiment 2 in [5]). A different response to ABC-D and D-ABC sequences indicates that tits recognize call ordering when decoding the sequences. Photo credit: Toshitaka N. Suzuki.

<https://doi.org/10.1371/journal.pbio.2006532.g001>

(moving their heads horizontally as if scanning for danger) and D calls (approaching the sound source). In response to ABC-D sequences, they produce both behaviours. Importantly, tits do not first scan and then approach, as would be predicted if they perceived ABC-D sequences as linear, ordered strings. Instead, they progressively approach the sound source while continuously scanning (Fig 1B). In contrast, playback of artificially reversed sequences ('D-ABC'; Fig 1B) elicited reduced scanning and approaching compared to the response towards natural call sequences (ABC-D). Finally, we addressed the possibility that receivers simply associate the whole ABC-D sequence with a single 'mobbing' meaning rather than assessing the meanings of the constituent calls. We substituted willow tit 'tää' calls, which evoke similar responses in Japanese tits to their own D calls [10], into novel ABC-tää and reversed tää-ABC sequences [6]. We found that Japanese tits exhibit an equivalent response (i.e., approach with scanning) to both natural (ABC-D) and novel (ABC-tää) call sequences. In contrast, when these calls are combined into reversed ordering (tää-ABC), they reduced their responses. Together, these results demonstrate that tits combine different meaningful signals to generate a compound message that depends on the meaning of the elements and the way in which they are combined, i.e., compositional syntax [3]. Thus, we argue that changing the order in which the calls are combined does not produce an alternative compound message but rather a sequence with unclear or ambiguous meaning.

Bolhuis and colleagues [11] argue that these findings do not provide evidence for compositional syntax. They make three main arguments: (i) that tits' lack of responses to D-ABC combinations are a result of such messages being maladaptive; (ii) that linear order plays an unimportant role in human syntax; and (iii) that tit call combinations are not hierarchical or freely productive, as are combinations in human language.

First, they suggest that tits produce weaker responses to reversed (D-ABC) order sequences because only scanning (ABC) before approaching (D), but not the reversed response, would be adaptive. This interpretation, however, is likely to be based on misunderstanding the experimental set-up and tits' behavioural responses. Tits naturally combine ABC and D calls with a 0.5- to 0.15-s interval in between the two calls [5]. This close temporal proximity does not allow tits to display an independent response (scan or approach) to the first call part (ABC or D) before hearing the second call part (D or ABC) (Fig 1B). Moreover, calls were not broadcast in isolation (i.e., a single ABC-D call) but rather in a calling bout, in which tens of sequences were produced with gaps of at least 1.6 s in between. Accordingly, tits do not exhibit a defined sequence of response to each call sequence (e.g., scan and then approach), but exhibit a compound response (approach while scanning) in response to playback of ABC-D calling bouts (Fig 1C). Thus, independent of the adaptiveness of responses, the fact that tits responded differently to natural (ABC-D) and reversed (D-ABC) sequences demonstrated that they perceived the order of call elements when decoding call sequences.

The difference in the tits' responses to naturally produced ABC-D and artificially reversed D-ABC sequences implies that meaning in call sequences is dependent on the order of the lexical elements. However, another argument of Bolhuis and colleagues [11] is that linear order does not play as important a role in human language as is typically thought, and thus, differential responses to ABC-D and D-ABC sequences cannot provide evidence for compositional syntax. This claim is based on the example of the noun phrase 'old men and women'. It can be interpreted as either a collection of old men and women of unspecified age or a collection of old men and old women (Fig 2A). Clearly, structural ambiguity of linguistic expressions can only be observed when more than two morphemes are strung together and are most easily recognized in written language, a highly derived form of language. A phrase with two words usually does have an unambiguous meaning, and in many cases, word order matters when producing a compound meaning. E.g., in English, 'watch out' has an unambiguous meaning,

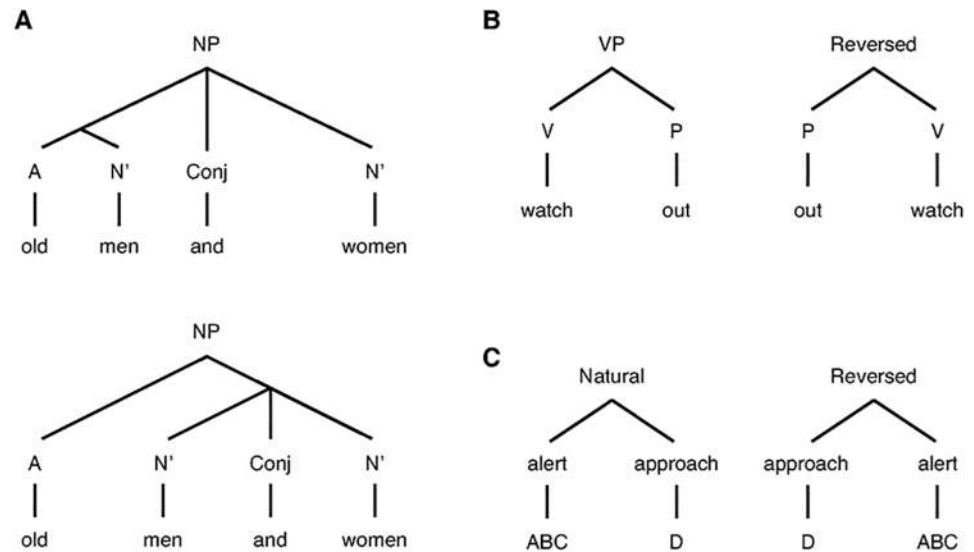


Fig 2. Examples of parse trees for human phrases and Japanese tit call sequences. (A) Ambiguity of meaning can be observed when more than two words are strung together ([old men] and [women]) or [old [men and women]], see [11]. (B) The meaning of phrase depends on syntactic structure. (C) The meaning of Japanese tit sequence also depends on how meaningful calls are combined, indicating that there is also a syntactic structure in which call ordering governs the meaning of sequence. A, adjective; Conj, conjunctive; N', noun; NP, noun phrase; P, preposition; V, verb; VP, verb phrase.

<https://doi.org/10.1371/journal.pbio.2006532.g002>

while the reversed combination ‘out watch’ is difficult to parse and has ambiguous, unclear meaning (Fig 2B). We would predict human responders to respond to the first phrase but not the second precisely because it violates syntactical rules. This is exactly what we observed in Japanese tits: combinations of ABC and D calls evoke compound responses only when they are ordered as ABC-D sequences (Fig 2C).

Finally, Bolhuis and colleagues [11] argue that hierarchical structure and free productivity are essential components of human language that are absent in tit calls. We agree that neither of these components have been demonstrated in tit call combinations, but neither are they part of even Bolhuis and colleagues’ [11] own preferred definition of compositionality in human language. Arguing that these features are not present in Japanese tit call sequences is not suitable to reject the interpretation of our findings. We note that parallels between vocal learning of birdsongs and human language are commonly made [12,13]. An important component of human language learning, unlike birdsong learning, is that meanings are associated to signals [3]. One could argue, as Bolhuis and colleagues [11] do about Japanese tit call combinations, that because this component of human language is absent in birdsong learning, any comparison between the two processes is invalid. Instead, we recognize that previous and ongoing work into birdsong has indeed uncovered deep insight into the neural, genomic, behavioural basis, and evolution of vocal learning more generally, despite the differences between birds and humans. Likewise, we are confident that future research into animal call combinations will provide similar insights into the evolution of compositionality.

In summary, we assert that the comments by Bolhuis and colleagues [11] do not change any interpretation of our results or the conclusion of the studies [4–6]. Our research, and that of others, has shown that call combinations incorporate different meaningful elements, which provides a new model to study the syntax–semantics link in nonhuman species [14]. A growing body of research reveals that combinations of meaningful calls have evolved in non-human animals such as Campbell’s monkeys (*Cercopithecus campbelli*) [8] and pied babblers

(*Turdoides bicolor*) [15], suggesting that compositional syntax may have evolved independently in multiple animal taxa [16]. In addition, recent studies have shown that Japanese tits communicate about predatory threats by using combinations of different linguistic capabilities, such as referential signals and compositional syntax [4,17]. In light of our findings, we believe that animal call sequences can provide an invaluable model system for comparative studies to uncover the origins and evolution of linguistic capabilities that form the basis of syntactic communication.

Author Contributions

Funding acquisition: Toshitaka N. Suzuki.

Writing – original draft: Toshitaka N. Suzuki, David Wheatcroft, Michael Griesser.

Writing – review & editing: Toshitaka N. Suzuki, David Wheatcroft, Michael Griesser.

References

1. Fitch WT. The evolution of language. Cambridge: Cambridge University Press; 2010.
2. Hauser MD, Chomsky N, Fitch WT. The faculty of language: what is it, who has it, and how did it evolve? *Science*. 2002; 298: 1569–1579. <https://doi.org/10.1126/science.298.5598.1569> PMID: 12446899
3. Hurford JR. The origins of grammar: Language in the light of evolution II. Oxford: Oxford University Press; 2011.
4. Suzuki TN. Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Anim Behav*. 2014; 87:59–65. <https://doi.org/10.1016/j.anbehav.2013.10.009>
5. Suzuki TN, Wheatcroft D, Griesser M. Experimental evidence for compositional syntax in bird calls. *Nat Commun*. 2016; 7: 10986. <https://doi.org/10.1038/ncomms10986> PMID: 26954097
6. Suzuki TN, Wheatcroft D, Griesser M. Wild birds use an ordering rule to decode novel sequences. *Curr Biol*. 2017; 27: 2331–2336. <https://doi.org/10.1016/j.cub.2017.06.031> PMID: 28756952
7. Arnold K & Zuberbühler K. Language evolution: semantic combinations in primate calls. *Nature* 2006; 441: 303. <https://doi.org/10.1038/441303a> PMID: 16710411
8. Ouattara K, Lemasson A, and Zuberbühler K. Campbell's monkeys concatenate vocalizations into context-specific sequences. *Proc Natl Acad Sci*. 2009; 106: 22026–22031. <https://doi.org/10.1073/pnas.0908118106> PMID: 20007377
9. Schlenker P, Chemla E, Zuberbühler K. What do monkey calls mean? *Trends Cogn Sci*. 2016; 20: 894–904. <https://doi.org/10.1016/j.tics.2016.10.004> PMID: 27836778
10. Suzuki TN, Kutsukake N. Foraging intention affects whether willow tits call to attract members of mixed-species flocks. *R Soc Open Sci*. 2017; 4: 170222. <https://doi.org/10.1098/rsos.170222> PMID: 28680675
11. Bolhuis JJ, Beckers GJL, Huybregts MAC, Berwick RC, Everaert MBH. Meaningful syntactic structure in songbird vocalizations? *PLoS Biol*. 2018; 16: e2005157. <https://doi.org/10.1371/journal.pbio.2005157> PMID: 29864124
12. Bolhuis JJ, Okanoya K, Scharff C. Twitter evolution: converging mechanisms in birdsong and human speech. *Nat Rev Neurosci*. 2010; 11: 747–759. <https://doi.org/10.1038/nrn2931> PMID: 20959859
13. Berwick RC, Okanoya K, Beckers GJL, Bolhuis JJ. Songs to syntax: the linguistics of birdsong. *Trends Cogn Sci*. 2011; 15: 113–121. <https://doi.org/10.1016/j.tics.2011.01.002> PMID: 21296608
14. Griesser M, Wheatcroft D, Suzuki TN. From bird calls to human language: exploring the evolutionary drivers of compositional syntax. *Curr Opin Behav Sci*. 2018; 21: 6–12. <https://doi.org/10.1016/j.cobeha.2017.11.002>
15. Engesser S, Ridley AR, Townsend SW. Meaningful call combinations and compositional processing in the southern pied babbler. *Proc Natl Acad Sci*. 2016; 113: 5976–5981. <https://doi.org/10.1073/pnas.1600970113> PMID: 27155011
16. Townsend SW, Engesser S, Stoll S, Zuberbühler K, Bickel B. (2018) Compositionality in animals and humans. *PLOS Biology* 16(8): e2006425. <https://doi.org/10.1371/journal.pbio.2006425>
17. Suzuki TN. Alarm calls evoke a visual search image of a predator in birds. *Proc Natl Acad Sci*. 2018; 115: 1541–1545. <https://doi.org/10.1073/pnas.1718884115> PMID: 29378940