

Animal linguistics: Exploring referentiality and compositionality in bird calls

Toshitaka N. Suzuki 

The Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan

Correspondence

Toshitaka N. Suzuki, The Hakubi Center for Advanced Research, Kyoto University, Yoshida-honmachi, Kyoto 606-8501, Japan.

Email: toshi.n.suzuki@gmail.com

Funding information

Hakubi Project funding; Japan Society for the Promotion of Science, Grant/Award Numbers: 16K18616, 18H05074, 18K14789, 20H03325, 20H05001

Abstract

Establishing the theory of language evolution is an ongoing challenge in science. One profitable approach in this regard is to seek the origins of linguistic capabilities by comparing language with the vocal communication systems of closely related relatives (i.e., the great apes). However, several key capabilities of language appear to be absent in non-human primates, which limits the range of studies, such as direct phylogenetic comparison. A further informative approach lies in identifying convergent features in phylogenetically distant animals and conducting comparative studies. This approach is particularly useful with respect to establishing general rules for the evolution of linguistic capabilities. In this article, I review recent findings on linguistic capabilities in a passerine bird species, the Japanese tit (*Parus minor*). Field experiments have revealed that Japanese tits produce unique alarm calls when encountering predatory snakes, which serve to enhance the visual attention of call receivers with respect to snake-like objects. Moreover, tits often combine discrete types of meaningful calls into fixed-ordered sequences according to an ordering rule, conveying a compositional message to receivers. These findings indicate that two core capabilities of language, namely, referentiality and compositionality, have independently evolved in the avian lineage. I describe how these linguistic capabilities can be examined under field conditions and discuss how such research may contribute to exploring the origins and evolution of language.

KEYWORDS

bird, communication, compositionality, language evolution, referentiality

1 | INTRODUCTION

How did language evolve? This is a long-standing question in science (Fitch, 2010; Hurford, 2014; Tallerman & Gibson, 2012). The generative power of language is based

on semantics and syntax, that is, signals convey independent meanings, and combinations of signals provide compositional messages (Hurford, 2007, 2012). In contrast, animal communication signals have long been considered essentially emotional or motivational in nature, that is, signals are assumed to merely reflect the internal states of signalers and convey neither referential nor compositional information (Hurford, 2007, 2012). Given

Toshitaka N. Suzuki is the recipient of the 22nd Miyadi Award of the Ecological Society of Japan.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Ecological Research* published by John Wiley & Sons Australia, Ltd on behalf of The Ecological Society of Japan

this widely accepted dichotomy, most previous studies on the evolution of language have concentrated on detailed analyses of various linguistic expressions, searching for the minimum set of capabilities (i.e., the faculty of language) required for verbal communication (i.e., the minimalist program; Chomsky, 1965, 1993). However, in the absence of any knowledge of the evolutionary continuity or parallels between language and animal communication systems, the origins and evolution of language remain deep mysteries (Fitch, 2010; Hurford, 2014; Tallerman & Gibson, 2012).

In our ongoing quest to trace the evolution of language, one profitable approach is to seek its origins in the vocal communication of closely related species (the great apes). In this regard, both field and laboratory researches have revealed that humans share several linguistic capabilities, including the associative learning of signs and referents, gestural communication, and use of the direction of another individuals' gaze (i.e., gaze following), with chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) (Genty & Zuberbühler, 2014; Terrace, Petitto, Sanders, & Bever, 1979; Tomasello, Hare, & Lehann, 2007). In contrast, however, numerous other linguistic capabilities, such as learning to produce novel sounds (i.e., audio-vocal learning) and syntax, are apparently lacking in most non-human primates (Hauser, Chomsky, & Fitch, 2002; Hurford, 2012). Thus, direct phylogenetic comparison between humans and the great apes is alone insufficient to gain a complete understanding of the origins and evolution of linguistic capabilities (Hurford, 2012; Suzuki, Wheatcroft, & Griesser, 2019). An alternative fruitful approach lies in identifying convergent cases of linguistic capabilities in phylogenetically distant animals and to perform associated comparative studies (Griesser, Wheatcroft, & Suzuki, 2018; Suzuki, Griesser, & Wheatcroft, 2019; Suzuki, Wheatcroft, & Griesser, 2018). This approach has been applied in studies of the audio-vocal learning of songs in passerine birds (Catchpole & Slater, 2008), and research over the past few decades has revealed remarkable similarities between humans and passerines with respect to the neural mechanisms underlying audio-vocal learning (Berwick, Okanoya, Beckers, & Bolhuis, 2011; Pfenning et al., 2014).

Recent studies on vocal communication in a passerine bird species, the Japanese tit (*Parus minor*; Figure 1), have provided novel insights into instances of convergent linguistic capabilities, among which is referentiality, that is, the ability to convey to receivers, reference to external objects or events using specific signals. This ability had for long been considered unique to humans; however, the findings of field studies conducted over the past four decades have indicated possible evolutionary parallels in



FIGURE 1 The Japanese tit, *Parus minor*. This small passerine has evolved several key capabilities of language, including referentiality and compositionality [Color figure can be viewed at wileyonlinelibrary.com]

several avian and mammalian species (Gill & Bierema, 2013; Suzuki, 2016a; Townsend & Manser, 2013; Zuberbühler, 2009). For example, common ravens (*Corvus corax*) produce so-called “yell” calls on finding carcasses (Heinrich, 1988; Heinrich & Marzluff, 1991), the playback of which has been found to attract conspecifics to the food source (Szipl, Boeckle, Wascher, Spreafico, & Bugnyar, 2015). However, most studies in this field cannot exclude the possibility that signals merely represent a specific motivational state of signalers, rather than denoting external referents (Rendall, Owren, & Ryan, 2009; Wheeler & Fischer, 2012). In this review, I describe studies that have used a novel paradigm to discriminate between these two possibilities in Japanese tits (Suzuki, 2018).

A further linguistic capability of the Japanese tit is compositional syntax, that is, the ability to combine meaning-bearing units into compositional expressions based on certain rules. Although compositional syntax has also been considered a uniquely human trait, tits have been shown to combine discrete types of meaning-bearing calls into fixed-ordered sequences. Moreover, playback experiments have revealed that receivers extract a compositional message from the call sequences using an ordering rule. Herein, I describe how compositional expressions in animals can be distinguished from non-compositional, holistic sequences of meaningless elements (see also Suzuki, Wheatcroft, et al., 2019). I also discuss how these new findings can enhance our understanding of the cognitive mechanisms underlying animal communication and how they can contribute to future comparative studies that seek to establish the origins and evolution of language.

2 | REFERENTIALITY

2.1 | Functionally referential may be emotional

In human speech, words are often used to refer to objects or events (i.e., referents), leading to a triadic relationship among speakers, listeners and referents (Tomasello, 1995, 1999). Human infants develop referential words at 9–12 months of age, contingent on the prior development of several cognitive skills, such as joint attention and audio-vocal learning (Tomasello, 1999). In contrast, animal signals have long been considered expressions of the emotional state of signalers, leading to a simple dyadic relationship between signalers and receivers (Rendall et al., 2009). However, Seyfarth, Cheney, and Marler (1980) challenged this historical assumption by examining the responses of vervet monkeys (*Chlorocebus pygerythrus*) to different types of alarm call. These monkeys produce acoustically discrete alarm calls for different predators, such as leopards, eagles and snakes (Struhsaker, 1967). By examining the response of free-living vervet monkeys to playbacks of the variation in alarm calls, Seyfarth et al. (1980) found that different alarm calls evoke different, presumably adaptive, responses in receivers. Although such specific alarm calls have been described as “functionally referential” (Macedonia & Evans, 1993), it is also claimed that different alarm calls may merely influence receivers’ behavior in the absence of any retrieval of referential information (Rendall et al., 2009). Accordingly, even if monkeys produce different alarm calls for different threats, these sounds could be merely considered expressions of different types of fear, as earlier claimed by Darwin (1871).

2.2 | Enhanced attention and search images

In order to verify whether animal signals are truly referential, it is necessary to determine whether these signals enhance the attention of receivers with respect to target objects, thereby generating a triadic relationship among signalers, receivers and referents. Nevertheless, simply examining the responses of call receivers to target objects is not sufficient to evaluate this possibility, as calls may merely evoke stereotyped behaviors (e.g., a fixed scanning pattern), which may assist in detecting the referents. If this is the case, the enhanced rate of referent detection can be explained in terms of a chain of actions (Bond, 2019).

If alarm calls are truly referential, then a key prediction is that these calls evoke a mental image (or concept)

of predators in the receiver’s mind. In cognitive and neural sciences, the retrieval of mental images is defined as representation and the accompanying experience of sensory information in the absence of a direct external stimulus (Pearson, Naselaris, Holmes, & Kosslyn, 2015). Therefore, to provide evidence for the evocation of visual mental images by acoustic signals, it is necessary to demonstrate that receivers retrieve a mental image even in the absence of the target referent (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Kok, Mostert, & de Lange, 2017; Lee, Kravitz, & Baker, 2012). In this regard, an object that resembles a predator to a certain extent but alone does not evoke a specific behavior, could be used to examine the evocation of visual search images. If alarm calls enhance the visual attention of receivers to predator-like objects prior to having detected a predator, then it would provide evidence that these calls evoke predator-specific search images in receivers (Suzuki, 2019). This paradigm is based on human studies showing that evocation of visual mental images by referential words enables listeners to enhance the detection of otherwise unseen objects (Forder & Lupyan, 2019; Lupyan & Ward, 2013).

2.3 | Alarm calls evoke a search image

Using a combination of snake-like objects and specific alarm calls, Suzuki (2018) examined the possibility that alarm calls evoke certain search images. On encountering a predatory snake, such as a Japanese rat snake (*Elaphe climacophora*), Japanese tits produce acoustically unique alarm calls (Suzuki, 2014), which evoke context-specific anti-snake behaviors in receivers (Suzuki, 2011, 2012a, 2015). For example, when female tits incubate eggs in the nests, they respond to snake-specific alarm calls by immediately fleeing the nest cavity, thereby enabling them to evade attacks from snakes that can invade such cavities (Suzuki, 2015). When outside the cavity, tits respond to snake alarms by scanning the ground near the nesting tree or by looking inside the nest cavity, as if searching for snakes (Suzuki, 2012a). These findings accordingly indicate that the snake-specific alarm calls of Japanese tits do not merely convey the emotional or internal states of signalers (e.g., fear), but may serve to specifically indicate the presence of snakes, similar to the use of human referential words (e.g., “snake”).

Japanese tits were initially attracted by the playback of snake-specific alarm calls. Subsequently, the birds were exposed a wooden stick moved in a snake-like manner using thin string. If tits retrieve a visual mental image of a snake from snake-specific alarm calls, they may use this image to search for a snake and then show a specific

response to the snake-like moving stick. During the playback of snake-specific alarm calls, Japanese tits approached a stick moving snake-like along a tree trunk (Figure 2). However, they did not respond to the same stick when hearing other call types (general alarm calls for non-snake predators or non-alarm, recruitment calls for attracting birds in a non-predatory context). Similarly, tits approached the stick when it was moved in a snake-like manner on the ground in combination with snake alarm calls, but not when combined with general alarm calls. Consequently, stick approaches by tits are not considered to be part of a chain of reactions induced by differences in scanning patterns during different playbacks. In addition, tits did not approach moving sticks when the movement was dissimilar to that of a snake (i.e., swinging on a low shrub). Therefore, on hearing snake-specific alarm calls, tits do not invariably approach any novel objects out of increased curiosity. These results indicate that prior to detecting a real snake, tits retrieve a visual search image from snake-specific alarm calls and use this to search for snakes. Snakes are typically cryptic against the different types of substrates on which they move, such as the ground, tree trunks and branches, and thus the use of visual search images, as opposed to a stereotyped scanning pattern, may contribute to the efficient detection of snakes in complex environments.

2.4 | Search images evoked by eavesdropping

Suzuki (2020) subsequently extended the aforementioned studies on Japanese tits to include interspecific

communication. In montane regions of Japan, coal tits (*Periparus ater*) can be found in the same habitats as Japanese tits and often “eavesdrop” on Japanese tit alarm calls (Suzuki, 2016b). Experiments have revealed that similar to Japanese tits, coal tits will also approach snake-like moving sticks in response to hearing Japanese tit snake-specific alarms, whereas they do not approach the same sticks when hearing other call types, or if movement of the stick is dissimilar to that of a snake. These findings accordingly reveal that the retrieval of specific search images from referential calls is not limited to intraspecific communication but can occur in response to interspecific eavesdropping. Recent studies on other species of birds have shown that eavesdropping on the alarm calls of other species is dependent on associative learning between known threats (e.g., visual stimuli from predators or known types of alarm call) and novel sounds (Keen, Cole, Sheehan, & Sheldon, 2020; Magrath, Haff, Fallow, & Radford, 2015; Magrath, Haff, McLachlan, & Iqic, 2015; Potvin, Ratnayake, Radford, & Magrath, 2018). Therefore, it is likely that these birds assign mental images to heterospecific alarm calls through associative learning, although further investigations are necessary to confirm this assumption.

2.5 | Referentiality in other animals?

Although there have been only two studies that have examined the influence of referential calls on visual attention to referents (Suzuki, 2018, 2020), several studies have shown that hearing alarm calls can influence how individuals respond to auditory cues that relate to

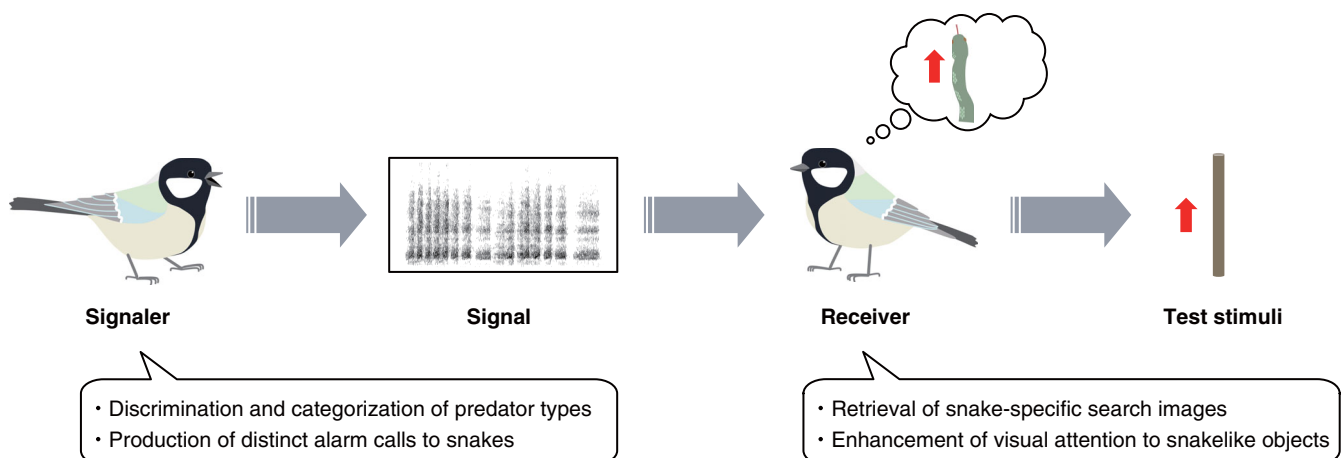


FIGURE 2 The referentiality of a signal can be assessed by examining the attention receivers give to a target referent. By using an object that to a certain extent resembles a predatory snake, but alone does not evoke a response, it can be ascertained whether snake-specific alarm calls evoke predator-specific search images in receivers. If individuals form a snake-specific search image, they may respond to snake-like objects only when hearing snake-specific alarm calls [Color figure can be viewed at wileyonlinelibrary.com]

predators. For example, Diana monkeys (*Cercopithecus diana*) show a reduction in alert responses to predator vocalizations after hearing conspecific alarm calls only if these alarm call types match the predator types (Zuberbühler, Cheney, & Seyfarth, 1999). These findings thus indicate that these monkeys detect referential information related to predator type from specific alarm calls, and thereafter alter their response to subsequent predator vocalizations. This design is comparable to the paradigms used by Suzuki (2018, 2020), but differs in two respects. First, although the monkey study examined the association between two different auditory stimuli, Suzuki (2018, 2020) investigated the association between auditory (calls) and visual (sticks) stimuli, highlighting the importance of the integration of cross-modal information in referential communication. Second, whereas Zuberbühler et al. (1999) used predator vocalizations as the test stimuli, Suzuki (2018, 2020) used an object that bears a certain resemblance to the target referents (i.e., snakes) but alone does not evoke any specific response in receivers. This design allows us to explore the retrieval of mental images, differentiated from the visual (or direct) perception of objects, as visual mental images are defined as the retrieval of mental representations without seeing or awareness of the external referents (Albers et al., 2013; Kok et al., 2017; Lee et al., 2012).

Research over the past four decades has revealed that numerous species of birds and mammals produce specific vocalizations on encountering predators or when finding a source of food (Gill & Bierema, 2013; Suzuki, 2016a; Townsend & Manser, 2013; Zuberbühler, 2009). By adopting a paradigm similar to that used by Suzuki (2018, 2020), it could be determined whether these calls evoke mental states in receivers, thereby generating a triadic relationship among signalers, receivers and referents.

3 | COMPOSITIONALITY

3.1 | Syntax and compositionality

The generative power of language is dependent to a large extent on syntax and compositionality. Syntax is defined as a set of rules whereby words can be combined into well-formed complexes (Hurford, 2012). A combination of words is considered compositional if its overall meaning depends on its elements and the manner in which they are syntactically combined (the principle of compositionality; Partee, ter Meulen, & Wall, 1990; Pelletier, 1994). Studies on animal syntax began with analyses of bird songs. The songs of passerine birds typically consist of multiple sound elements, which are combined according to ordering rules (Catchpole &

Slater, 2008; Podos, Huber, & Taft, 2004). Although bird songs can be structurally complex, their meaning is considered to be simple, with song phrases generally having certain functions, notably facilitating mate attraction, territorial defense, or both (Catchpole & Slater, 2008; Podos et al., 2004). Combinations of sound elements are widely detected in non-human animals, including bats (Bohn, Smarsh, & Smotherman, 2013), mice (Chabout, Sarkar, Dunson, & Jarvis, 2015), mongooses (Fitch, 2012; Jansen, Cant, & Manser, 2012; Rauber, Kranstauber, & Manser, 2020), cetaceans (Payne & McVay, 1971; Mercedo, Herman, & Pack 2005), gibbons (Clarke, Reichard, & Zuberbühler, 2006) and gorillas (Hedwig, Mundry, Robbins, & Boesch, 2015). However, most of these vocal sequences have been considered as holistic sequences, the meanings of which are conveyed by the overall sequences, and consequently, the sequence is not considered as a compositional expression. Thus, in animal studies, the term “syntax” has long been used to signify the rules for combining meaningless elements (Hurford, 2012; Marler, 1998).

However, the findings of recent studies have indicated that several animals are able to combine meaningful elements into sequences (Suzuki, Griesser, et al., 2019; Suzuki, Wheatcroft, et al., 2019). Accordingly, it has become necessary to redefine the term “syntax” to correspond with its definition in linguistics. In this regard, Suzuki and Zuberbühler (2019) recently redefined syntax as “a set of principles by which meaning-bearing units can be combined into well-formed complexes,” which matches the definition of human syntax in terms of compositionality and can be applied for the analyses of animal vocal sequences.

3.2 | Compositional syntax in bird calls

Initial support for the occurrence of compositional syntax in animal vocal sequences was provided by the findings of studies on Japanese tits (Suzuki, Wheatcroft, & Griesser, 2016, 2017). Japanese tits produce alert calls (so-called “chicka” calls), which serve to warn conspecifics of the presence of a range of different predators (Suzuki, 2014), whereas they produce acoustically distinct recruitment calls when attracting conspecifics in non-predatory situations (Suzuki et al., 2016). Interestingly, Japanese tits combine these two call types into alert-recruitment call sequences when attracting conspecifics for mobbing predators (Suzuki, 2014). Playback experiments have revealed that tits display different behaviors when hearing alert and recruitment calls, moving their head from side to side, as if scanning for danger, when hearing alert calls,

but approaching the sound source (i.e., a presumed signaler) in response to hearing recruitment calls (Suzuki et al., 2016). In response to alert-recruitment call sequences, tits combine both behaviors, that is, they progressively approach the sound source, while continuously scanning the horizon (Figure 3a). Notably, however, receivers do not appear to produce these two behaviors merely by exhibiting responses to the two meaningful units at the same time, as it has been found that they reduce their response to artificially reversed versions of the same component calls (recruitment-alert call sequences) (Figure 3b). Thus, it is likely that Japanese tits use the same ordering rule (alert-recruitment ordering rule) when combining calls and when decoding call sequences.

3.3 | Decoding novel call sequences

Although Suzuki et al. (2016) have suggested that Japanese tits decode compositional meaning from call sequences using an ordering rule, there remains the possibility that these birds simply reduce their response to reversed call sequences because these sequences are novel and unfamiliar. If this is true, then tits may recognize alert-recruitment call sequences as a holistic message, signifying “mobbing,” rather than by perceiving the meanings of the component calls.

Suzuki et al. (2017) subsequently developed a novel experimental approach to assess this latter possibility. During non-breeding seasons, Japanese tits form mixed-species flocks with willow tits (*Poecile montanus*), within

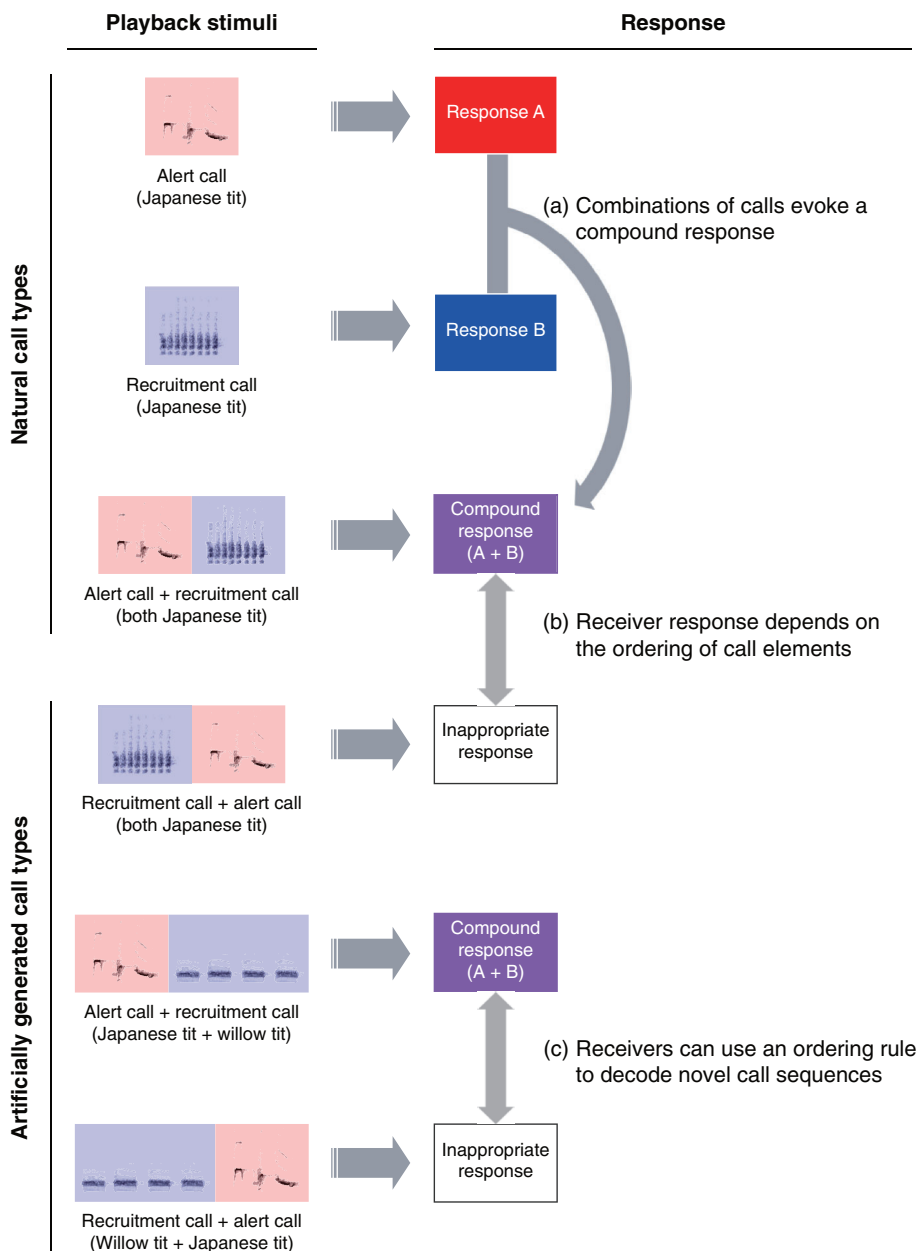


FIGURE 3 The compositionality of call sequences can be assessed by examining the responses of receivers to individual call elements and their combinations (a). According to the definition of syntax and compositionality, it is also necessary to examine the role of call ordering in receivers' responses (b). Furthermore, to rule out the possibility that call sequences provide a single, holistic meaning as a whole, it would be informative to assess the response to artificially generated novel call sequences, such as combinations of calls from two species (c) [Color figure can be viewed at wileyonlinelibrary.com]

which, the Japanese tits approach in response to the recruitment calls of both willow tits and conspecifics to maintain flock cohesion (Suzuki, 2012b, 2012c; Suzuki & Kutsukake, 2017). However, when the recruitment calls of willow tits were artificially shortened to match the length of Japanese tit recruitment calls (while maintaining their natural pitch), Japanese tits did not approach these modified calls (Suzuki et al., 2017), thereby indicating that the responses of Japanese tits are not due to the acoustic similarity of the recruitment calls of willow tits and conspecifics, but rather because they perceive them as two distinct vocalizations with a shared meaning. Therefore, from the perspective of Japanese tit receivers, willow tit recruitment calls are synonymous with their own species' recruitment calls, providing the opportunity to generate artificial novel call sequences composed of conspecific alert calls and heterospecific recruitment calls. In this regard, the findings of playback experiments have revealed that Japanese tits approach both natural and novel (mixed-species) call sequences only when the combinations of call units follow the alert-recruitment ordering, thereby indicating that they use an ordering rule to decode even novel combinations of calls (Figure 3c).

3.4 | Compositional syntax in other animal signals?

An increasing body of evidence indicates that several other animal species may also combine meaning-bearing elements into complex utterances. For example, Campbell's monkeys (*Cercopithecus campbelli*) produce acoustically discrete types of calls ("Krak," "Hok" and "Wak") when perceiving a threat, such as leopards or crowned eagles (Ouattara, Lemasson, & Zuberbühler, 2009a, 2009b), and often combine these calls with a short "oo" sound at the end, producing "Krak-oo," "Hok-oo" and "Wak-oo" vocalizations. Field observations have revealed that two of these sequences ("Krak-oo" and "Hok-oo") are more likely to be produced by monkeys experiencing low-threat situations, such as falling trees or the presence of non-predatory animals, indicating that "oo" may act as a suffix to modify the warning content of alarm calls (Ouattara et al., 2009a, 2009b). Similarly, call combinations have been documented in several species of birds and mammals (Engesser & Townsend, 2019), although in most cases, the findings of these studies could not rule out the possibility that individuals recognize call sequences as a single meaningful unit, but not as a compositional expression (Kuhn, Keenan, Arnold, & Lemasson, 2018; Suzuki, Griesser, et al., 2019; Suzuki, Wheatcroft, et al., 2019). Moreover, even if two meaning-

bearing units are combined, it is possible that the resultant combined call sequences convey an unrelated, third meaning, comparable to idioms in human language. For example, putty-nosed monkeys (*Cercopithecus nictitans*) combine two discrete alarm calls, each of which seemingly denotes a different threat, whereas the call sequences are used to stimulate long-distance group movements (Arnold & Zuberbühler, 2006a, 2006b, 2008, 2012). Further experiments are therefore required to determine whether call combinations in other animals are semantically compositional or whether instead they convey holistic messages.

4 | CONCLUSIONS AND FUTURE DIRECTIONS

4.1 | The family Paridae

In this review, I have described recent findings relating to the linguistic capabilities (referentiality and compositionality) of the Japanese tit. This species belongs to the family Paridae, which worldwide consists of some 55 species of tits, titmice and chickadees (Johansson et al., 2013). In common with Japanese tits, these birds may also use different types of alarm calls for specific predators. For example, "zi" or "seeet" calls are apparently associated with the detection of aerial predators, such as flying raptors (Haftorn, 2000; Zachau & Freeberg, 2012). In addition, snake-specific alarm calls have been recorded in coal tits, willow tits, and varied tits (*Sittiparus varius*) (Suzuki, personal observations). These calls may also convey referential information to receivers. Combinations of calls (or notes) have also been documented in numerous parid species (Krams, Krama, Freeberg, Kullberg, & Lucas, 2012; Lucas & Freeberg, 2007). For example, Carolina chickadees (*Poecile carolinensis*) produce multiple discrete types of notes that are combined to yield a diverse variety of sequences (Freeberg, 2008; Freeberg & Lucas, 2012; Lucas & Freeberg, 2007). The use of different sequences depends on the eliciting context, and therefore may convey different types of information (Lucas & Freeberg, 2007). Moreover, there exists a diverse range in the complexity of vocalizations among different species, thereby providing a model system for comparative studies (Krams et al., 2012).

4.2 | Evolutionary drivers

Species within the family Paridae may provide an ideal opportunity to perform comparative studies for exploring the socioecological factors that drive the evolution of

linguistic capabilities. These birds inhabit a variety of habitats, from forests to savannas, thereby representing an example of adaptive radiation. In addition, they have evolved different systems of sociality, for example, in Oxford (United Kingdom), great tits (*Parus major*) form flocks with a high degree of fission–fusion interactions, whereas blue tits (*Cyanistes caeruleus*) and marsh tits (*Poecile palustris*) form relatively stable groups (Farine, Aplin, Sheldon, & Hoppitt, 2015). In dense forests or in species that form loose flocks, visual contact with other individuals may be limited, thereby favoring individuals that transmit multiple types of information at the same time. In such cases, syntax and compositionality would be selectively advantageous. In contrast, birds living in open habitats or forming cohesive flocks may not need to evolve such complex utterances, as a simple vocalization that attracts the visual attention of other birds might be sufficient to enable the caller to notify other individuals of prevailing circumstances.

Predator composition may also be an important factor driving the evolution of avian vocal repertoire. In Asia and southern Europe, there are several species of snakes that depredate on birds (Ha et al., 2020; Sorace et al., 2000), whereas in northern Europe, most snake species are unable to climb trees and therefore do not represent a threat to birds. Consequently, snake-specific alarm calls may not have evolved in northern Europe. Instead, great spotted woodpeckers (*Dendrocopos major*) are one of the major predators of the eggs and nestlings of tits in Europe (Skwarska, Kalinski, Wawrzyniak, & Banbura, 2009), although in Japan, the same species of woodpecker lives in the same habitats as tits but does not attack their nests. Future comparative studies may reveal the socioecological factors that contribute to driving the evolution of referentiality and compositionality in bird calls, thereby providing a unique model for examining the principles and general rules which drive the evolution of linguistic capabilities.

4.3 | Genetic and neural bases

A comparative approach can also be applied to analyze the genetic and neural mechanisms underlying communication. For example, accelerated evolution of early growth response protein 1, a transcription factor that is believed to play a role in vocal communication, learning and memory (Clayton, 2013; Dragunow, 1996; Hara, Kubikova, Hessler, & Jarvis, 2007), has been found in the tit lineage (Laine et al., 2016). A similar positive selection has, however, not been detected in the ground pecker (*Pseudopodoces humilis*), a bird that inhabits open areas (savanna) and seemingly produces less complex

vocalizations (Laine et al., 2016). In humans, forkhead box protein P2 has been shown to play roles in the correct development of language and speech (Enard et al., 2002), and is also associated with song learning in passerines (Haesler et al., 2007; Teramitsu & White, 2006). Although the mechanisms whereby these two genes influence linguistic capabilities remain to be elucidated, further detailed studies may reveal the links between genetic structure, neural mechanisms and vocal communication in parids.

4.4 | Conclusion

Although long considered uniquely human traits, recent studies have revealed that referentiality and compositionality have evolved in the avian lineage and may also be involved in communication systems of a wide range of animals. Several new methodologies, such as object presentation in conjunction with call playbacks or use of mixed-species call sequences, will contribute to enhancing our understanding of how receivers extract information from acoustic signals. Species within the family Paridae have a wide distribution range, including the Northern Hemisphere and Africa, and accordingly represent a valuable group in terms of comparative studies. Detailed investigations on different parid species would advance our understanding of how socioecological factors drive the evolution of linguistic capabilities and their underlying genetic and neural mechanisms.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI grant numbers 16K18616, 18K14789, 18H05074, 20H05001 and 20H03325, and the Hakubi Project funding. I thank two anonymous referees for valuable comments on the manuscript.

ORCID

Toshitaka N. Suzuki  <https://orcid.org/0000-0001-6405-7653>

REFERENCES

- Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*, 1427–1431.
- Arnold, K., & Zuberbühler, K. (2006a). Language evolution: Semantic combinations in primate calls. *Nature*, *441*, 303.
- Arnold, K., & Zuberbühler, K. (2006b). The alarm-calling system of adult male putty-nosed monkeys, *Cercopithecus nictitans maritimi*. *Animal Behaviour*, *72*, 643–653.
- Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*, R202–R203.

- Arnold, K., & Zuberbühler, K. (2012). Call combinations in monkeys: Compositional or idiomatic expressions? *Brain and Language*, *120*, 303–309.
- Berwick, R. C., Okanoya, K., Beckers, G. J., & Bolhuis, J. J. (2011). Songs to syntax: The linguistics of birdsong. *Trends in Cognitive Sciences*, *15*, 113–121.
- Bohn, K. M., Smarsh, G. C., & Smotherman, M. (2013). Social context evokes rapid changes in bat song syntax. *Animal Behaviour*, *85*, 1485–1491.
- Bond, A. B. (2019). Searching images and the meaning of alarm calls. *Learning and Behavior*, *47*, 109–110.
- Catchpole, C. K., & Slater, P. J. (2008). *Bird song: Biological themes and variations*. Cambridge, England: Cambridge University Press.
- Chabout, J., Sarkar, A., Dunson, D. B., & Jarvis, E. D. (2015). Male mice song syntax depends on social contexts and influences female preferences. *Frontiers in Behavioral Neuroscience*, *9*, 76.
- Chomsky, N. A. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. A. (1993). A minimalist program for linguistic theory. In K. Hare & K. Keyser (Eds.), *The view from building* (Vol. 20, pp. 1–52). Cambridge, MA: MIT Press.
- Clarke, E., Reichard, U. H., & Zuberbühler, K. (2006). The syntax and meaning of wild gibbon songs. *PLoS One*, *1*, e73.
- Clayton, D. F. (2013). The genomics of memory and learning in songbirds. *Annual Review of Genomics and Human Genetics*, *14*, 45–65.
- Darwin, C. (1871). *The descent of man, and selection in relation to sex*. London, England: John Murray.
- Dragunow, M. (1996). A role for immediate-early transcription factors in learning and memory. *Behavior Genetics*, *26*, 293–299.
- Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., ... Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*, 869–872.
- Engesser, S., & Townsend, S. W. (2019). Combinatoriality in the vocal systems of nonhuman animals. *WIREs Cognitive Science*, *10*, e1493.
- Farine, D. R., Aplin, L. M., Sheldon, B. C., & Hoppitt, W. (2015). Interspecific social networks promote information transmission in wild songbirds. *Proceedings of the Royal Society B: Biological Sciences*, *282*, 20142804.
- Fitch, W. T. (2010). *The evolution of language*. Cambridge, England: Cambridge University Press.
- Fitch, W. T. (2012). Segmental structure in banded mongoose calls. *BMC Biology*, *10*, 98.
- Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*, 1105–1123.
- Freeberg, T. M. (2008). Complexity in the chick-a-dee call of Carolina chickadees (*Poecile carolinensis*): Associations of context and signaler behavior to call structure. *Auk*, *125*, 896–907.
- Freeberg, T. M., & Lucas, J. R. (2012). Information theoretical approaches to chick-a-dee calls of Carolina chickadees (*Poecile carolinensis*). *Journal of Comparative Psychology*, *126*, 68–81.
- Genty, E., & Zuberbühler, K. (2014). Spatial reference in a bonobo gesture. *Current Biology*, *14*, 1596–1600.
- Gill, S. A., & Bierema, A. M. K. (2013). On the meaning of alarm calls: A review of functional reference in avian alarm calling. *Ethology*, *119*, 449–461.
- Griessler, M., Wheatcroft, D., & Suzuki, T. N. (2018). From bird calls to human language: Exploring the evolutionary drivers of compositional syntax. *Current Opinion in Behavioral Sciences*, *21*, 6–12.
- Ha, J., Lee, K., Yang, E., Kim, W., Song, H., Hwang, I., ... Jablonski, P. (2020). Experimental study of alarm calls of the oriental tit (*Parus minor*) toward different predators and reactions they induce in nestlings. *Ethology*, *126*, 610–619.
- Haesler, S., Rochefort, C., Georgi, B., Licznarski, P., Osten, P., & Scharff, C. (2007). Incomplete and inaccurate vocal imitation after knockdown of *FoxP2* in songbird basal ganglia nucleus area X. *PLoS Biology*, *5*, e321.
- Haftorn, S. (2000). Contexts and possible functions of alarm calling in the willow tit, *Parus montanus*; the principle of ‘better safe than sorry’. *Behaviour*, *137*, 437–449.
- Hara, E., Kubikova, L., Hessler, N. A., & Jarvis, E. D. (2007). Role of the midbrain dopaminergic system in modulation of vocal brain activation by social context. *European Journal of Neuroscience*, *25*, 3406–3416.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.
- Hedwig, D., Mundry, R., Robbins, M. M., & Boesch, C. (2015). Contextual correlates of syntactic variation in mountain and western gorilla close-distance vocalizations: Indications for lexical or phonological syntax? *Animal Cognition*, *18*, 423–435.
- Heinrich, B. (1988). Winter foraging at carcasses by three sympatric corvids, with emphasis on recruitment by the raven, *Corvus corax*. *Behavioral Ecology and Sociobiology*, *23*, 141–156.
- Heinrich, B., & Marzluff, J. (1991). Do common ravens yell because they want to attract others? *Behavioral Ecology and Sociobiology*, *28*, 13–21.
- Hurford, J. R. (2007). *The origins of meaning: Language in the light of evolution*. Oxford, England: Oxford University Press.
- Hurford, J. R. (2012). *The origins of grammar: Language in the light of evolution II*. Oxford, England: Oxford University Press.
- Hurford, J. R. (2014). *The origins of language: A slim guide*. Oxford, England: Oxford University Press.
- Jansen, D. A. W. A. M., Cant, M. A., & Manser, M. B. (2012). Segmental concatenation of individual signatures and context cues in banded mongoose (*Mungos mungo*) close calls. *BMC Biology*, *10*, 97.
- Johansson, U. S., Ekman, J., Bowie, R. C. K., Halvarsson, P., Ohlson, J. I., Price, T. D., & Ericson, P. G. P. (2013). A complete multilocus species phylogeny of the tits and chickadees (Aves: Paridae). *Molecular Phylogenetics and Evolution*, *69*, 852–860.
- Keen, S. C., Cole, E. F., Sheehan, M. J., & Sheldon, B. C. (2020). Social learning of acoustic anti-predator cues occurs between wild bird species. *Proceedings of the Royal Society B: Biological Sciences*, *287*, 20192513.
- Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 10473–10478.
- Krams, I., Krama, T., Freeberg, T. M., Kullberg, C., & Lucas, J. R. (2012). Linking social complexity and vocal complexity: A parid perspective. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1879–1891.

- Kuhn, J., Keenan, S., Arnold, K., & Lemasson, A. (2018). On the -oo suffix of Campbell's monkeys. *Linguistic Inquiry*, 49, 169–181.
- Laine, V. N., Gossman, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., ... Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, 7, 10474.
- Lee, S.-H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, 59, 4064–4073.
- Lucas, J. R., & Freeberg, T. M. (2007). 'Information' and the chick-a-dee call: Communicating with a complex vocal system. In K. A. Otter (Ed.), *Ecology and behavior of chickadees and titmice: An integrated approach* (pp. 199–213). Oxford, England: Oxford University Press.
- Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 14196–14201.
- Macedonia, J. M., & Evans, C. S. (1993). Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology*, 93, 177–197.
- Magrath, R. D., Haff, T. M., Fallow, P. M., & Radford, A. N. (2015). Eavesdropping on heterospecific alarm calls: From mechanisms to consequences. *Biological Reviews*, 90, 560–586.
- Magrath, R. D., Haff, T. M., McLachlan, J. R., & Iqic, B. (2015). Wild birds learn to eavesdrop on heterospecific alarm calls. *Current Biology*, 25, 2047–2050.
- Marler, P. (1998). Animal communication and human language. In N. G. Jablonski & L. C. Aiello (Eds.), *The origin and diversification of language* (pp. 1–20). San Francisco, CA: California Academy of Sciences.
- Mercedo, I. I. E., Herman, L. M., & Pack, A. A. (2005). Song copying by humpback whales: Themes and variations. *Animal Cognition*, 8, 93–102.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009a). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 22026–22031.
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009b). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE*, 4, e7808.
- Partee, B., ter Meulen, A., & Wall, R. E. (1990). *Mathematical methods in linguistics*. Dordrecht: Kluwer Academic.
- Payne, R. S., & McVay, S. (1971). Songs of humpback whales. *Science*, 173, 587–597.
- Pearson, J., Naselaris, T., Holmes, E. A., & Kosslyn, S. M. (2015). Mental imagery: Functional mechanisms and clinical applications. *Trends in Cognitive Sciences*, 19, 590–602.
- Pelletier, F. J. (1994). The principle of semantic compositionality. *Topoi*, 13, 11–24.
- Pfennig, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., ... Jarvis, E. D. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science*, 346, 1256846.
- Podos, J., Huber, S. K., & Taft, B. (2004). Bird song: The interface of evolution and mechanism. *Annual Review of Ecology, Evolution, and Systematics*, 35, 55–87.
- Potvin, D. A., Ratnayake, C. P., Radford, A. N., & Magrath, R. D. (2018). Birds learn socially to recognize heterospecific alarm calls by acoustic association. *Current Biology*, 28, 2632–2637.
- Rauber, R., Kranstauber, B., & Manser, M. B. (2020). Call order within vocal sequences of meerkats contains temporary contextual and individual information. *BMC Biology*, 18, 119.
- Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour*, 78, 233–240.
- Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Monkey responses to three different alarm calls: Evidence of predator classification and semantic communication. *Science*, 210, 801–803.
- Skwarska, J. A., Kalinski, A., Wawrzyniak, J., & Banbura, J. (2009). Opportunity makes a predator: Great spotted woodpecker predation on tit broods depends on nest box design. *Ornis Fennica*, 86, 109–112.
- Sorace, A., Consiglio, C., Tanda, F., Lanzuisi, E., Cattaneo, A., & Iavicoli, D. (2000). Predation by snakes on eggs and nestlings of great tit *Parus major* and blue tit *P. caeruleus*. *Ibis*, 142, 328–330.
- Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In S. Altmann (Ed.), *Social communication among primates* (pp. 281–324). Chicago, IL: University of Chicago Press.
- Suzuki, T. N. (2011). Parental alarm calls warn nestlings about different predatory threats. *Current Biology*, 21, R15–R16.
- Suzuki, T. N. (2012a). Referential mobbing calls elicit different predator-searching behaviours in Japanese great tits. *Animal Behaviour*, 84, 53–57.
- Suzuki, T. N. (2012b). Long-distance calling by the willow tit, *Poecile montanus*, facilitates formation of mixed-species foraging flocks. *Ethology*, 118, 10–16.
- Suzuki, T. N. (2012c). Calling at a food source: Context-dependent variation in note composition of combinatorial calls in willow tits. *Ornithological Science*, 11, 103–107.
- Suzuki, T. N. (2014). Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Animal Behaviour*, 87, 59–65.
- Suzuki, T. N. (2015). Assessment of predation risk through referential communication in incubating birds. *Scientific Reports*, 5, 10239.
- Suzuki, T. N. (2016a). Semantic communication in birds: Evidence from field research over the past two decades. *Ecological Research*, 31, 307–319.
- Suzuki, T. N. (2016b). Referential calls coordinate multi-species mobbing in a forest bird community. *Journal of Ethology*, 34, 79–84.
- Suzuki, T. N. (2018). Alarm calls evoke a visual search image of a predator in birds. *Proceedings of the National Academy of Sciences of the United States of America*, 115, 1541–1545.
- Suzuki, T. N. (2019). Imagery in wild birds: Retrieval of visual information from referential alarm calls. *Learning and Behavior*, 47, 111–114.
- Suzuki, T. N. (2020). Other species' alarm calls evoke a predator-specific search image in birds. *Current Biology*, 30, 2616–2620.
- Suzuki, T. N., Griesser, M., & Wheatcroft, D. (2019). Syntactic rules in avian vocal sequences as a window into the evolution of compositionality. *Animal Behaviour*, 151, 267–274.

- Suzuki, T. N., & Kutsukake, N. (2017). Foraging intention affects whether willow tits call to attract members of mixed-species flocks. *Royal Society Open Science*, *4*, 170222.
- Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for compositional syntax in bird calls. *Nature Communications*, *7*, 10986.
- Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2017). Wild birds use an ordering rule to decode novel call sequences. *Current Biology*, *27*, 2331–2336.
- Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2018). Call combinations in birds and the evolution of compositional syntax. *PLoS Biology*, *16*, e2006532.
- Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2019). The syntax-semantics interface in animal vocal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*, 20180405.
- Suzuki, T. N., & Zuberbühler, K. (2019). Animal syntax. *Current Biology*, *29*, R669–R671.
- Szipl, G., Boeckle, M., Wascher, C. A. F., Spreafico, M., & Bugnyar, T. (2015). With whom to dine? Ravens' responses to food-associated calls depend on individual characteristics of the caller. *Animal Behaviour*, *99*, 33–42.
- Tallerman, M., & Gibson, K. R. (2012). *The Oxford handbook of language evolution*. Oxford, England: Oxford University Press.
- Teramitsu, I., & White, S. A. (2006). FoxP2 regulation during undirected singing in adult songbirds. *Journal of Neuroscience*, *26*, 7390–7394.
- Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science*, *206*, 891–902.
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, P. J. Dunham, & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.
- Tomasello, M. (1999). The human adaptation for culture. *Annual Review of Anthropology*, *28*, 509–529.
- Tomasello, M., Hare, B., & Lehann, H. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, *52*, 314–320.
- Townsend, S. W., & Manser, M. B. (2013). Functionally referential communication in mammals: The past, present and the future. *Ethology*, *119*, 1–11.
- Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: A promising paradigm whose time has passed. *Evolutionary Anthropology: Issues, News, and Reviews*, *21*, 195–205.
- Zachau, C. E., & Freeberg, T. M. (2012). Chick-a-dee call variation in the context of “flying” avian predator stimuli: A field study of Carolina chickadees (*Poecile carolinensis*). *Behavioral Ecology and Sociobiology*, *66*, 683–690.
- Zuberbühler, K. (2009). Survivor signals: The biology and psychology of animal alarm calling. *Advances in the Study of Behavior*, *40*, 277–322.
- Zuberbühler, K., Cheney, D. L., & Seyfarth, R. M. (1999). Conceptual semantics in a nonhuman primate. *Journal of Comparative Psychology*, *113*, 33–42.

How to cite this article: Suzuki TN. Animal linguistics: Exploring referentiality and compositionality in bird calls. *Ecological Research*. 2021;36:221–231. <https://doi.org/10.1111/1440-1703.12200>