



# Alarm calls evoke a visual search image of a predator in birds

Toshitaka N. Suzuki<sup>a,b,1</sup>

<sup>a</sup>Center for Ecological Research, Kyoto University, Otsu, 520-2113 Shiga, Japan; and <sup>b</sup>Department of Evolutionary Studies of Biosystems, Graduate University for Advanced Studies, Hayama, 240-0193 Kanagawa, Japan

Edited by Asif A. Ghazanfar, Princeton University, Princeton, NJ, and accepted by Editorial Board Member Marlene Behrmann December 29, 2017 (received for review October 30, 2017)

**One of the core features of human speech is that words cause listeners to retrieve corresponding visual mental images. However, whether vocalizations similarly evoke mental images in animal communication systems is surprisingly unknown. Japanese tits (*Parus minor*) produce specific alarm calls when and only when encountering a predatory snake. Here, I show that simply hearing these calls causes tits to become more visually perceptive to objects resembling snakes. During playback of snake-specific alarm calls, tits approach a wooden stick being moved in a snake-like fashion. However, tits do not respond to the same stick when hearing other call types or if the stick's movement is dissimilar to that of a snake. Thus, before detecting a real snake, tits retrieve its visual image from snake-specific alarm calls and use this to search out snakes. This study provides evidence for a call-evoked visual search image in a nonhuman animal, offering a paradigm to explore the cognitive basis for animal vocal communication in the wild.**

animal communication | cognition | language | referentiality | visual mental image

In human speech, words referring to a specific object (e.g., “moon”) evoke a visual mental image in listeners or readers (1, 2). In cognitive and neural sciences, visual mental imagery can be defined as representations and the accompanying experience of sensory information without a direct external stimulus (3, 4). Thus, at the behavioral level, receivers who retrieve a visual mental image from referential words are expected to enhance their detection of a target object (5, 6). In nonhuman animal communication systems, many key, language-like features have been demonstrated (1, 7), but whether animal signals also evoke mental images in receivers is surprisingly unknown (8–10). Discovering evidence for the retrieval of visual mental images from vocal signals in animals would have fundamental implications for our understanding of animal cognition, neural mechanisms, and the adaptation and evolution of linguistic capabilities (1, 2). Here, using a paradigm, I provide experimental evidence that vocalizations in a nonhuman animal evoke a visual mental image of a referent.

Call-evoked visual mental images may have evolved in animals that produce unique calls to specific objects in their environment (11, 12). One candidate species is the Japanese tit (*Parus minor*), which produces acoustically distinct alarm calls in response to different predators (13). The calls that tits give when and only when encountering predatory snakes are considered “functionally referential” because they elicit specific antisnake behaviors in receivers (14–16) (Fig. 1A). For example, when incubating eggs in the nest, adults respond to snake-specific alarm calls by immediately escaping out of the nest cavity, allowing them to evade attacks from snakes which can invade the cavity (16). When outside of the nest cavity, they respond to snake-specific alarm calls by looking down at the ground nearby their nest tree, which suggests an adaptive behavior to locate approaching snakes (15). In addition to snake-specific alarm calls, tits have evolved another call type used for a wide range of predator types,

including avian and mammalian predators (13) (Fig. 1B). In response to these general alarm calls, receivers approach the sound source and scan the surroundings (17), but do not show any specific behaviors to defend themselves against snakes (15, 16).

Based on these previous studies, I hypothesized that snake-specific alarm calls evoke a visual search image of a snake in tits. A key prediction of this hypothesis is that receivers are primed to detect snakes when hearing snake-specific alarm calls. However, to provide evidence for visual mental imagery, individuals should be primed to detect snakes even in the absence of real snakes, since simply seeing a snake may directly trigger its mental image (3, 4). To account for this possibility, I designed a field experiment to test whether snake-specific alarm calls prime tits to perceive an otherwise indistinct object as a real snake. If the hypothesis is correct, tits are expected to show a specific behavior toward a snake-like object when and only when hearing these calls. Alternatively, if responses to snake-specific alarm calls are not dependent on visual search images, hearing the calls should not specifically improve the detection of a snake-like object, but would simply evoke a stereotyped behavior (e.g., looking down at the ground). I tested the responses of free-living Japanese tits to specific associations between different call types and visual stimuli (snake-like or non-snake-like objects) during their early breeding season.

## Results and Discussion

First, I attracted an individual, focal tit by broadcasting snake-specific alarm calls from a speaker hung in a tree. Immediately after the focal tit flew to the tree and perched on a tree branch

### Significance

**In human speech, words often cause listeners to retrieve visual mental images of target objects. In nonhuman animal communication systems, many key, language-like features have been demonstrated, but there is still no evidence that animal signals evoke mental images of objects in receivers. Japanese tits produce specific alarm calls when encountering a predatory snake. Here, I show that simply hearing these calls causes tits to become more visually perceptive to objects resembling snakes (moving sticks). This result indicates that before having detected a real snake, tits retrieve its visual image from snake-specific alarm calls and uses this to search out snakes. This study provides evidence for a call-evoked visual search image in a nonhuman animal.**

Author contributions: T.N.S. designed research, performed research, analyzed data, and wrote the paper.

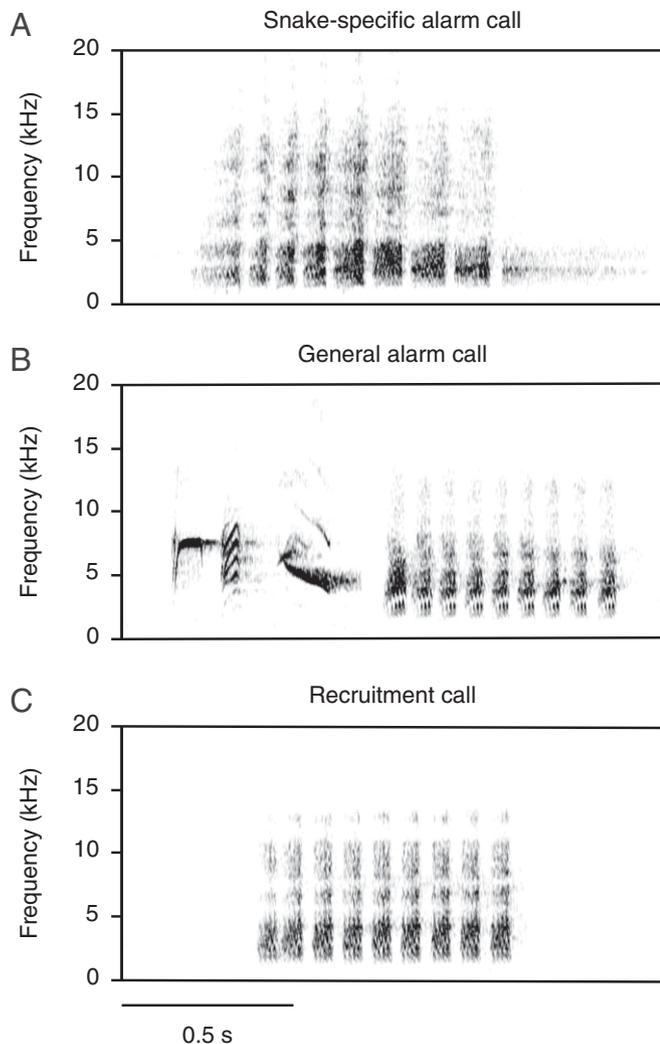
The author declares no conflict of interest.

This article is a PNAS Direct Submission. A.A.G. is a guest editor invited by the Editorial Board.

Published under the PNAS license.

<sup>1</sup>Email: toshi.n.suzuki@gmail.com.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1718884115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1718884115/-DCSupplemental).



**Fig. 1.** Sound spectrograms of Japanese tit vocalizations. (A) Snake-specific alarm call. (B) General alarm call. (C) Recruitment call.

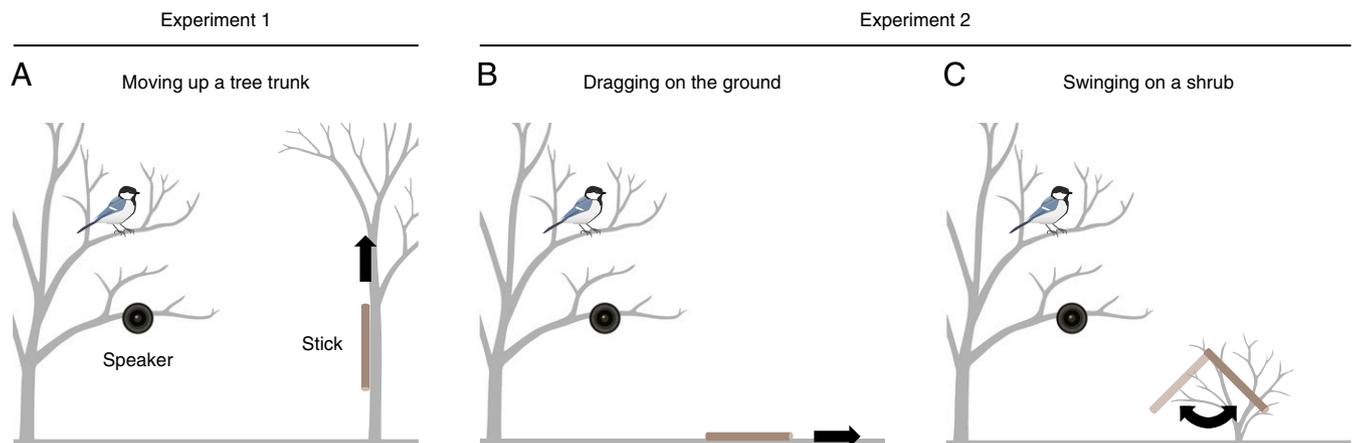
around the speaker, I presented a short stick (18-cm length, 1.5-cm diameter) cut from a dead branch by pulling it up for 1 m

along the tree trunk, using a thin string that I manipulated from a distance (Fig. 2A; Experiment 1). This movement imitates the movement of a snake, as no animals in the study area other than bird-eating snakes (Japanese rat snakes, *Elaphe climacophora*) move up a tree trunk in this way. To control for the possibility that simply perceiving any threatening signal causes tits to respond to the stick movement, I also tested their responses to the stick during playback of general alarm calls. If tits retrieved the visual mental image of a snake from snake-specific alarm calls, then they are expected to exhibit a greater response to the moving stick when hearing snake-specific calls than when hearing general alarm calls.

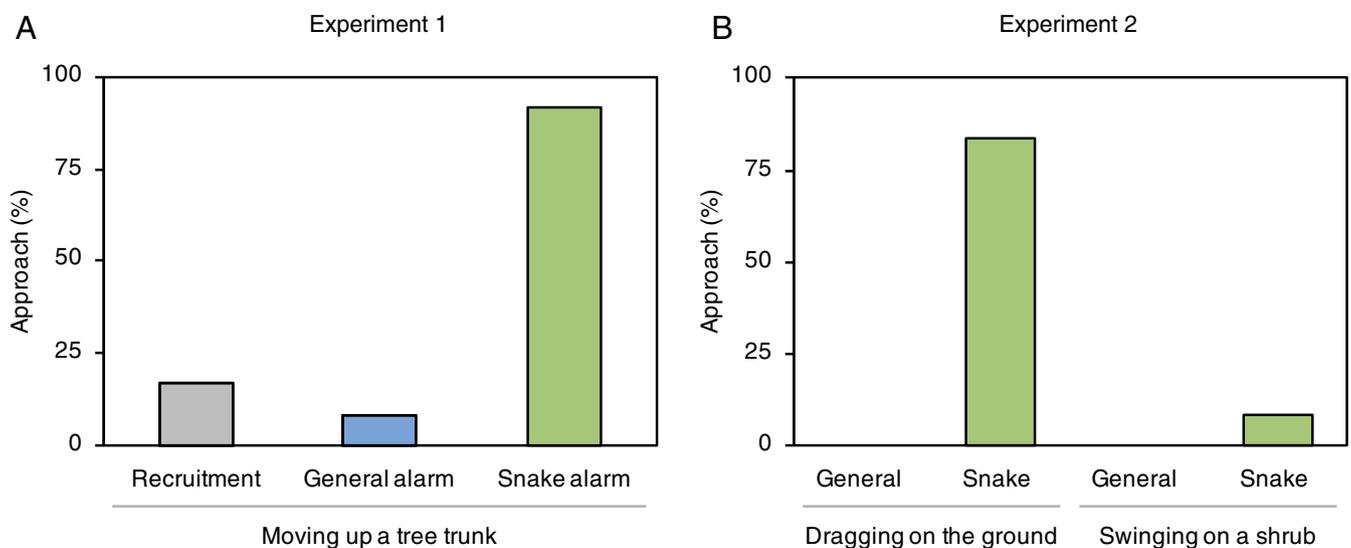
Responses of Japanese tits to the stick differed according to the types of alarm calls they heard (Fig. 3A). When hearing snake-specific alarm calls, focal individuals consistently flew toward and approached within 1 m of the stick moving snake-like up a tree trunk and inspected the surroundings (92%, 11/12 trials; Movie S1). In some cases (58%, 7/12 trials), they also approached within 0.5 m of the stick. In contrast, when hearing general alarm calls, they rarely approached within 1 m of the moving stick (8%, 1/12 trials) (Fisher's exact probability test, 1 m:  $P < 0.001$ ). Thus, in response to snake-specific alarm calls individuals approached the stick, but in response to general alarm calls they did not approach the stick.

Although this result supports the hypothesis, there remains the possibility that tits only detect urgency information from both alarm call types (e.g., snake-specific alarm calls convey greater urgency than general alarm calls), thus exhibiting different responses to the stick when hearing different alarm calls. If this is true, tits are expected to approach the stick more often when hearing any alarm call type, than when hearing nonalarm calls. To account for this possibility, I investigated tits' responses to the stick during the playback of recruitment calls (Fig. 1C). Recruitment calls are produced to attract conspecific receivers in nonpredatory contexts, such as when recruiting a mate to the nest or food, and do not evoke any alert behavior (17).

Although general alarm calls do convey more urgency than recruitment calls (17), tits responded similarly to the stick during playbacks of these two call types (Fisher's exact probability test, 1 m:  $P = 1.0$ ; Fig. 3A). However, tits are less likely to approach the stick during playback of recruitment calls (17%, 2/12 trials) compared with the playback of snake-specific alarm calls (Fisher's exact probability test,  $P < 0.001$ ). Thus, tits' responses to the stick are not simply driven by the urgency information encoded



**Fig. 2.** Schematic representation of experimental setup. (A) In Experiment 1, tits were exposed to a stick moving snake-like up a tree trunk immediately after they approached within 2 m above the speaker. (B) In Experiment 2, tits were exposed to another snake-like movement of a stick: dragging on the ground. (C) In Experiment 2, tits were also exposed to non-snake-like movement of a stick: swinging on a low shrub.



**Fig. 3.** Responses of Japanese tits to a stick. (A) Experiment 1. Tits approached a stick moving up a tree trunk during the playback of snake-specific alarm calls, but rarely during the playbacks of general alarm calls and recruitment calls. Sample size:  $n = 36$  trials. Each focal bird was exposed to only one treatment, giving  $n = 12$  trials per treatment. (B) Experiment 2. Tits approached a stick moving on the ground during the playback of snake-specific alarm calls, but not during the playback of general alarm calls. However, they rarely approached the stick when its movement (swinging) was dissimilar to a snake movement regardless of the type of alarm call they heard. Sample size:  $n = 48$  trials. Each bird was exposed to only one treatment, giving  $n = 12$  trials per treatment.

in the calls, but rather, the referential information encoded in the snake-specific alarm calls.

These results demonstrate that snake-specific alarm calls specifically prime tits to respond to a stick moving in a snake-like way, suggesting that they retrieve a visual search image of a snake from these calls. However, another possible explanation is that snake-specific alarm calls cause receivers to approach any object moving in the environment without evoking a specific search image of a snake. According to the hypothesis that snake-specific alarm calls evoke a snake-specific search image, tits are expected to approach a stick when it moves in a snake-like way in a variety of contexts, such as on the ground. In addition, tits should not respond to a stick moving in a pattern that is inconsistent with a snake, such as swinging. To test these possibilities, I conducted an additional experiment to investigate tits' responses to different combinations of alarm calls and stick movement patterns (being moved in a snake-like fashion on the ground or a non-snake-like fashion being swung from a low shrub; Experiment 2; Fig. 2 B and C).

Japanese tits frequently approached within 1 m of a stick dragged on the ground and inspected the surroundings during the playback of snake-specific alarm calls (83%, 10/12 trials). In several cases (25%, 3/12 trials), they also approached within 0.5 m of the stick. However, they never approached the snake-like movement of a stick when hearing general alarm calls (0%, 0/12 trials) (Fisher's exact probability test, 1 m:  $P < 0.0001$ ) (Fig. 3B). Moreover, tits rarely approached a stick that was swung in a non-snake-like motion from a low shrub when hearing either snake-specific (8%, 1/12 trials) or general alarm calls (0%, 0/12 trials) (Fisher's exact probability test, 1 m:  $P = 1.0$ ) (Fig. 3C). In combination with the first experiment, these results demonstrate that tits that heard snake-specific alarm calls did not simply turn their attention to any moving object. Instead they searched for and focused on snake-like objects, regardless of their spatial position. Since snake-specific alarm calls caused tits to specifically approach a stick only if it was moved in a snake-like fashion, these calls did not simply evoke a stereotyped behavior. Thus, I conclude that Japanese tits retrieve a visual search image of a snake from snake-specific alarm calls without needing to see a real snake.

Retrieving a visual search image when hearing calls could provide adaptive benefits to receivers. First, the selective retrieval of a snake search image from snake-specific alarm calls would allow tits to focus on snakes and not on other predators, which should increase the efficiency of finding snakes. Upon detecting a real snake, tits approach the snake closely and hover over it, spreading their wings and tail, which might function in deterring it from predatory attacks (18). In this study, tits approached a snake-like moving stick during the playback of snake-specific alarm calls, but did not exhibit such mobbing behavior, suggesting that they may realize that the stick is not a real snake once they get close enough. Secondly, the retrieval of a snake image from snake-specific alarm calls would allow tits to alter their responses according to context. This is supported by previous studies showing that adult tits respond differently to snake-specific alarm calls depending on whether they are inside or outside of their nest cavities (15, 16). Thus, the need for the efficient detection of external referents and for context-dependent shifts in responses might select for the evolution of communication through visual mental imagery.

This study provides experimental evidence for the retrieval of a visual search image from a specific call type in a nonhuman animal. Using either habituation–dishabituation or expectation of violation paradigms, previous studies have shown that several nonhuman primates and birds associate specific call types with related stimuli (19–22), suggesting that these animals may form mental representations of external objects. However, since most of these studies have investigated the specific associations between two auditory stimuli (i.e., alarm calls and predator vocalizations) (19–21), it has been a long-standing challenge to determine whether specific calls evoke a visual mental image in animals (1, 2). Using an object that somewhat resembles the referent but does not solely evoke a specific behavior, it could be tested whether subjects become more visually perceptive to that object when hearing specific calls. Despite once being considered a uniquely human trait (23), the integration of cross-modal information has now been documented for a wide variety of vertebrates, such as birds (24, 25) and mammals (26–29). While such cross-modal performance has been shown to enhance recognition of individual identity or the emotional state of signalers

(24–29), my results reveal that cross-modal integration of auditory and visual stimuli can be involved in referential communication in animals. In light of this discovery, future research will no doubt begin to uncover the ecological importance of visual mental imagery in animal vocal communication, as well as the developmental and neural mechanisms underlying this cognitive sophistication.

## Materials and Methods

**Study Site and Subjects.** Experiments were conducted on Japanese tits (*P. minor*) in mixed deciduous–coniferous forests in Nagano and Gunma (36°19′–25′ N, 138°27′–39′ E), Japan. In this study site, there were three species of snakes, including Japanese rat snakes, Japanese forest rat snakes (*E. conspicillata*), and tiger keelbacks (*Rhabdophis tigrinus*). Only Japanese rat snakes climb up trees and prey on birds including eggs, young, and adults. Experiments were conducted between May 18 and May 30, 2017 (Experiment 1) and April 22 and May 12, 2014 (Experiment 2), during the early breeding season of tits. Within each experiment, trials were conducted at different sites ( $n = 36$  in Experiment 1,  $n = 48$  in Experiment 2), which were separated by at least 350 m. This distance was enough to ensure that independent data were collected from different individuals, which was confirmed by observations of color-ringed Japanese tits in a previous study (30). Trials took place under calm and dry weather between 0830 and 1600 hours (Japan Standard Time).

**Experimental Stimuli.** Sticks ( $n = 12$ ; 18-cm length, 1.5-cm diameter) were cut from dead branches and were then connected to a thin, black string. Playback stimuli for the three call types were constructed from previous recordings of Japanese tits' vocalizations (13, 14) and were edited by using Adobe Audition 3.0 software (Adobe Corporation). For each playback exemplar, one call was chosen from an individual on the basis of sound quality (i.e., when the signaler was close to the microphone and there was low background noise) and was repeated in a sound file at a rate of five calls per 12 s (one call every 2.4 s), imitating a natural calling rate in Japanese tits (13, 14). Snake-specific alarm calls are composed of a single type of note, whereas general alarm calls can vary subtly with their note composition (13). I used general alarm calls composed of four types of notes (a string of A, B, and C notes followed by 7–10 D notes) since this combination is a typical composition of general alarm calls and recruits individuals to scan for danger (17). Recruitment calls are composed of 7–10 D notes, which recruit individuals to nondangerous situations (17). Previous playback studies have shown that all three of the calls attract Japanese tits to within 2 m of the speaker (15, 17, 31). I filtered out low-frequency (<1 kHz) noise and amplified all of the calls to be played back at a standardized volume (75 dB re: 20  $\mu$ Pa at 1 m from the speaker measured using an SM-325 sound-level meter; AS ONE Corporation). I saved all sound files in .WAV format (16-bit depth, 48.0-kHz sampling rate) to an SD memory card. For each treatment, 12 exemplars were constructed using the calls of 12 individuals that had been recorded with an LS370 parabolic microphone (Fuji Planning Corporation) connected to an R-09HR digital audio recorder (16-bit depth, 48.0-kHz sampling rate; Roland Corporation) (13, 14).

**Experiment 1.** This experiment investigated the responses of tits to a stick moving up a tree trunk during the playback of snake-specific alarm calls, general alarm calls, or recruitment calls. At each experimental site, an AT-SPG50 speaker (Audio-Technica Corporation) was hung from a tree branch at  $1.5 \pm 0.1$  m above the ground. The speaker was connected to an R-09HR digital audio recorder using EXC-12A extension cords (JVC KENWOOD Corporation) to control the playbacks from an observation position  $\sim 10$  m away from the speaker tree. A stick was fixed at a height of  $1.1 \pm 0.2$  m (mean  $\pm$  SD,  $n = 24$ ) on the trunk of a different tree. This tree was carefully chosen with the following criteria: (i) it allowed the stick to be pulled up the tree trunk in a straight line to imitate a snake moving up the trunk, (ii) it allowed the stick to be visible from the position of the speaker, ensuring that an approaching, focal bird was exposed to the stick, and (iii) it allowed the stick to be located at a distance of  $3.2 \pm 0.2$  m (mean  $\pm$  SD,  $n = 36$ ) from the speaker.

Playbacks were started when no birds were visible around the speaker tree and no birds were calling. When a focal tit flew to the speaker tree and approached within 2 m above the speaker where it could see the stick, the movement of the stick was initiated by pulling up the thin string 20 cm and continually repeating this movement five times for 12 s. I then recorded whether the focal tit flew directly toward the moving stick during the first 24 s. Observations were made  $\sim 10$  m from the tree with the stick and the

trials were video-recorded using a GZ-EX350 digital video camera (JVC KENWOOD Corporation). After each trial, I used a tape measure to measure the minimum approach distance of the focal bird to the stick. I confirmed the exact location at which the focal bird made the closest approach by checking the video recording at each experimental site. When the focal tit flew toward the moving stick, it also approached within 1 m of the stick, so I used this distance to determine approach behavior. In consideration for the possibility of observer bias in behavioral data (32), I conducted an interobserver reliability test with a second rater who was naive to the hypothesis but had been accustomed to observing Japanese tit behavior. The second rater observed randomly selected video clips ( $n = 18$ , 50% of the trials) and estimated whether tits approached within 1 m of the stick or not. The results obtained from the author and the second rater showed a high degree of agreement, indicating that there was no observer bias (kappa statistic of 1.0) (32).

The order of trials (snake alarm, general alarm, and recruitment calls) was counterbalanced across treatments so that responses to all treatments were understood under largely similar conditions. In seven trials, the first bird to approach the call playback was from a heterospecific species, such as a coal tit (*Periparus ater*) ( $n = 6$ ) or a willow tit (*Poecile montanus*) ( $n = 1$ ). To account for the possibility that Japanese tits copied the behavior of these other birds, I only used the data from instances where the first individual to approach the speaker was a Japanese tit. Otherwise, I repeated the same treatment at the next site. Unique exemplars of calls ( $n = 12$ ) were used for each trial to completely avoid pseudoreplication (33). Unique exemplars of sticks ( $n = 12$ ) were also used for each trial within each treatment.

**Experiment 2.** This experiment investigated the responses of tits to two types of stick movements (along the ground and swinging on a shrub) in combination with playbacks of either snake-specific or general alarm calls. As in Experiment 1, an AT-SPG50 speaker was hung from a tree branch at 1.4 m above the ground, and was connected to an R-09HR digital audio recorder using EXC-12A extension cords. For the snake-like movement along the ground, the stick was positioned on the ground at a distance of 2.5 m from the speaker, and the string was extended from the stick to the observation position to control the movement of the stick from a distance. To generate the non-snake-like swinging movement, the stick was attached to a small shrub ( $1.0 \pm 0.2$  m high, mean  $\pm$  SD,  $n = 24$ ) at a distance of 2.5 m from the speaker and the string was extended to the observation position, so that I could swing the stick from a distance.

Playbacks were started when no birds were visible around the speaker tree and no birds were calling. When a tit flew to the speaker tree and approached within 2 m above the speaker where it could see the stick, the movement of the stick (dragging on the ground or swinging on a shrub) was initiated by pulling the thin string 20 cm and continually repeating this movement five times for 12 s. I then recorded whether the focal tit approached within 1 m of the moving stick during the first 24 s. I also measured the minimum distance at which the focal tit approached the stick, as in Experiment 1. Interobserver reliability tests between the author and a second naive rater were also conducted for these analyses using randomly selected video clips ( $n = 12$  for each movement type, 50% of the trials). The results showed a high degree of agreement between the two raters, indicating that there was no observer bias (kappa statistic: 1.0 for both) (32).

Trials were conducted in 12 blocks, with all 4 treatments (2 call types  $\times$  2 movement types) presented in a randomized order. As in Experiment 1, unique call exemplars and unique sticks were used for each block to avoid pseudoreplication (33). When the first bird to approach an alarm call playback was from a heterospecific species ( $n = 40$  trials), I repeated the same treatment at the next site.

**Statistical Analysis.** Fisher's exact probability tests were used to compare the probability of approach to the movement of the stick. All tests were two-tailed, and statistical significance was set at  $\alpha = 0.05$ . When making multiple comparisons, a false discovery rate control was applied to adjust  $P$  values (34). All of the statistical analyses were performed in R (35) using the *stats* package.

**Ethics Statement.** All experiments were performed in accordance with relevant guidelines and regulations. All experimental protocols were approved by the Animal Care and Use Committees at The Graduate University for Advanced Studies, and adhered to the Guidelines for the Use of Animals of the Association for the Study of Animal Behaviour/Animal Behavior Society (36). This research was performed with permission from the Ministry of the Environment and the Forestry Agency of Japan.

**ACKNOWLEDGMENTS.** I am grateful to Craig Barnett, Michael Griesser, and Daizaburo Shizuka for helpful comments on earlier versions of the manuscript, and to Alexis C. Billings and David Wheatcroft for valuable comments on the

manuscript. I also thank the editor and two anonymous reviewers for their helpful suggestions in revising the manuscript. This work was supported by Japan Society for the Promotion of Science KAKENHI Grant 25-3391 and 16752305.

1. Hurford JR (2007) *The Origins of Meaning: Language in the Light of Evolution* (Oxford Univ Press, Oxford).
2. Fitch WT (2010) *The Evolution of Language* (Cambridge Univ Press, Cambridge, UK).
3. Kreiman G, Koch C, Fried I (2000) Imagery neurons in the human brain. *Nature* 408: 357–361.
4. Pearson J, Naselaris T, Holmes EA, Kosslyn SM (2015) Mental imagery: Functional mechanisms and clinical applications. *Trends Cogn Sci* 19:590–602.
5. Lupyan G, Ward EJ (2013) Language can boost otherwise unseen objects into visual awareness. *Proc Natl Acad Sci USA* 110:14196–14201.
6. Kok P, Mostert P, de Lange FP (2017) Prior expectations induce prestimulus sensory templates. *Proc Natl Acad Sci USA* 114:10473–10478.
7. Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298:1569–1579.
8. Rendall D, Owren MJ, Ryan MJ (2009) What do animal signals mean? *Anim Behav* 78: 233–240.
9. Wheeler BC, Fischer J (2012) Functionally referential signals: A promising paradigm whose time has passed. *Evol Anthropol* 21:195–205.
10. Suzuki TN (2016) Semantic communication in birds: Evidence from field research over the past two decades. *Ecol Res* 31:307–319.
11. Zuberbühler K (2003) Referential signaling in non-human primates: Cognitive precursors and limitations for the evolution of language. *Adv Stud Behav* 33:265–307.
12. Gill SA, Bierema AM-K (2013) On the meaning of alarm calls: A review of functional reference in avian alarm calling. *Ethology* 119:449–461.
13. Suzuki TN (2014) Communication about predator type by a bird using discrete, graded and combinatorial variation in alarm calls. *Anim Behav* 87:59–65.
14. Suzuki TN (2011) Parental alarm calls warn nestlings about different predatory threats. *Curr Biol* 21:R15–R16.
15. Suzuki TN (2012) Referential mobbing calls elicit different predator-searching behaviours in Japanese great tits. *Anim Behav* 84:53–57.
16. Suzuki TN (2015) Assessment of predation risk through referential communication in incubating birds. *Sci Rep* 5:10239.
17. Suzuki TN, Wheatcroft D, Griesser M (2016) Experimental evidence for compositional syntax in bird calls. *Nat Commun* 7:10986.
18. Suzuki TN, Ueda K (2013) Mobbing calls of Japanese tits signal predator type: Field observations of natural predator encounters. *Wilson J Ornithol* 125:412–415.
19. Seyfarth RM, Cheney DL (1990) The assessment by vervet monkeys of their own and another species' alarm calls. *Anim Behav* 40:754–764.
20. Zuberbühler K, Cheney DL, Seyfarth RM (1999) Conceptual semantics in a nonhuman primate. *J Comp Psychol* 113:33–42.
21. Zuberbühler K (2000) Causal cognition in a non-human primate: Field playback experiments with Diana monkeys. *Cognition* 76:195–207.
22. Evans CS, Evans L (2007) Representational signalling in birds. *Biol Lett* 3:8–11.
23. Ettliger G (1967) Analysis of cross-modal effects and their relationship to language. *Brain Mechanisms Underlying Speech and Language*, eds Darley FL, Millikan CH (Grune & Stratton, New York), pp 53–60.
24. George I, Richard JP, Cousillas H, Hausberger M (2011) No need to talk, I know you: Familiarity influences early multisensory integration in a songbird's brain. *Front Behav Neurosci* 4:193.
25. Kondo N, Izawa E, Watanabe S (2012) Crows cross-modally recognize group members but not non-group members. *Proc Biol Sci* 279:1937–1942.
26. Ghazanfar AA, Logothetis NK (2003) Neuroperception: Facial expressions linked to monkey calls. *Nature* 423:937–938.
27. Proops L, McComb K, Reby D (2009) Cross-modal individual recognition in domestic horses (*Equus caballus*). *Proc Natl Acad Sci USA* 106:947–951.
28. Sliwa J, Duhamel JR, Pascalis O, Wirth S (2011) Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proc Natl Acad Sci USA* 108:1735–1740.
29. Kulahci IG, Drea CM, Rubenstein DI, Ghazanfar AA (2014) Individual recognition through olfactory-auditory matching in lemurs. *Proc Biol Sci* 281:20140071.
30. Suzuki TN (2012) Long-distance calling by the willow tit, *Poecile montanus*, facilitates formation of mixed-species foraging flocks. *Ethology* 118:10–16.
31. Suzuki TN, Wheatcroft D, Griesser M (2017) Wild birds use an ordering rule to decode novel call sequences. *Curr Biol* 27:2331–2336.
32. Kaufman AB, Rosenthal R (2009) Can you believe my eyes? The importance of interobserver reliability statistics in observations of animal behaviour. *Anim Behav* 78: 1487–1491.
33. Kroodsmas DE, Byers BE, Goodale E, Johnson S, Liu WC (2001) Pseudoreplication in playback experiments, revisited a decade later. *Anim Behav* 61:1029–1033.
34. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289–300.
35. R Core Team (2017) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna).
36. Association for the Study of Animal Behaviour/Animal Behavior Society (2017) Guidelines for the treatment of animals in behavioural research and teaching. *Anim Behav* 123:i–ix.