Contagious vocal reaction forms avian vocal exchange: an experimental study in Java sparrows

Sota Kikuchi[+], Noriko Kondo[+], and Hiroki Koda[*]

Graduate School of Arts and Sciences, University of Tokyo, Komaba, Tokyo, 153-8902, Japan

[+]Co-first authors/equal contributions

*Corresponding to HK

HK (ORCID: 0000-0002-0927-3473): hkoda@g.ecc.u-tokyo.ac.jp

Abstract

Vocal exchange or turn-taking - an interaction in which vocalizer alternation occurs at very short intervals to avoid overlapping vocalizations - occurs not only in human conversation but also in animal communications, characterized by temporal regularities in the timing of spontaneous and response vocalizations. This temporal regularity has been observed in social mammals or primates, including humans, but little has been known about vocal exchange of call in birds, a highly vocal animal. Here, we experimentally manipulated a face-to-face setting with Java sparrows to observe their vocal exchange between two birds and analyzed its temporal regularity. The results showed that the temporal regularity of their vocal exchange was significantly different from that of primates. In particular, Java sparrows responded reflexively to the vocalizations of other individuals, but they showed no tendency to await the vocalizations of other individuals. Similar to behavioral synchrony, which is commonly observed in avian species, contagious vocalization possibly shapes avian vocal exchange. These results suggested the need for careful comparative studies of vocal exchange among social animals, considering the differences in the ecological backgrounds of group formation.

Keywords: Vocal exchange, turn-taking, avian call, temporal regularity, behavioural contagion, group formation

1. Introduction

Human conversation is characterized by the phenomenon of vocal turn-taking, in which speakers quickly alternate roles to convey linguistic information to each other (e.g. [1]). This vocal turn-taking involves the maintenance of very short time intervals of about 200 ms between speakers' utterances, while avoiding overlap of utterances between speakers. Interestingly, this feature is independent of linguistic information (e.g., semantic or lexical information) and can be defined simply by the temporal interval between speakers, which can be computed from the onset and offset times of the speech regions for each speaker. Regardless of language and cultural differences, the inter-speech interval in this vocal turn-taking has been confirmed to have a universal temporal regularity [2,3].

Vocal turn-taking, which can be described only by temporal regularity, does not depend on linguistic information, and therefore can be compared to the animals' vocal turn-taking or vocal exchange in the same framework in the studies of animal communication. Indeed, various studies of vocal exchange have been conducted in different species and attempts have been made to compare it with conversational turn-taking in humans [4,5]. Primates, including humans, are animals that have been intensively examined for vocal exchange [4]. In particular, the vocal exchange of *contact calls*, which are commonly observed vocalizations exchanged between individuals in a cohesive social group, has been analyzed in detail to characterize the intervals in the vocal exchange. Similar to human conversation, vocal exchange in many primate species shows a temporal regularity - significant avoidance of vocal overlap and a stable time interval between vocalizations of two individuals (e.g. [6–8]). Additionally, it has been reported that the time interval for repeated vocalizations by the same individual in the absence of response from other individuals (repeated transition) is longer than the time interval for vocalizations by different individuals taking turns (exchange transition) [7,9–14], supporting the similarity in the temporal regularity in the human conversational turn-takings [15,16]. Similarly, temporal regularity of vocal exchange has been also observed in social living non-primate mammals (e.g. rodents [17,18], bats [19], elephants [20,21], meerkats [22], cetaceans

[23–28]). Thus, the temporal regularity consistently observed among humans, nonhuman primates, and non-primate mammals is a common principle and adaptive form of communication in social groups (see conceptual framework of the temporal regularity, see Fig 1).

In contrast, avian vocal exchange has not been as well documented as in mammals. Mammalian vocal communication research focuses on the analysis of a social behavior, such as social relationships and the function of vocalizations within a group. On the other hand, studies on avian vocal communication have focused on "song" interactions between males or between males and females during courtship or territorial defense [29–32]. Although song as a stereotypical sexual display is not suitable for analyzing social communication in groups, call would be appropriate for examining vocal exchange in avian species [33]. Few analyses, however, have yet focused on vocal exchange in call of avian species, as is the case in the maintenance of social cohesion in a mammalian group. In one of the few examples of studies on jungle crows (*Corvus macrorhynchos*) [33], temporal regularity in the vocal exchange of their contact, *'ka call'* - a fixed time exchange interval between different callers - was observed. Similarly, the vocal exchange of calls in common starlings (*Sturnus vulgaris*) [34] or barn owls (*Tyto alba*) [35] also confirmed the avoidance of callers' vocal overlap, indicating temporal regularity in these call exchange. The zebra finches, *Taeniopygia guttata*, a model songbird for the neuroethology and behavioral ecology of their courtship song, has recently improved our understanding of vocal exchange in their calls. Experimental studies that examined the vocal exchange of calls by placing two finches face to face confirmed that call overlap was significantly avoided and call exchange adjustments in vocal timing occurred even in zebra finches as in primates [36–39]. Thus, they argued that this vocal timing regulation, which depends on the timing of other individual calls, is analogous to the primate phenomenon of vocal exchange. Ecologically, these vocal timing regulation in this vocal exchange has been considered a reflection of the social group structure of zebra finches [36], suggesting ecological adaptations in vocal exchange similar to those of mammals.

In general, avian vocal exchange has been comparatively studied within the same framework - functional adaptation to social groups - based on findings in primates. However, we must note the clear ecological differences in group formation between mammals and birds. Mammalian groups, including primates, are really "social groups," in which stable, long-term affinitive relationships based on kinship between mother and offspring form a society, while avian species form temporary, large groups as an anti-predator strategy. Mammals and birds, which have fundamentally different mechanisms for group formation and maintenance, may differ in the role of vocal exchange and the temporal regularities that can be commonly identified in mammals. To our knowledge, however, experimental knowledge of the temporal properties of vocal exchange in avian calls has been extremely limited.

Here, we investigated the temporal regularity of vocal exchanges in Java sparrows (*Lonchura oryzivora*) by experimentally arranging multiple individuals in a face-to-face situation in a soundproof chamber and analyzing the temporal characteristics of their vocal exchanges. Specifically, we measured the time intervals of calls exchanged between different individuals and those repeated by the same individual, and examined the characteristics of their data distribution. We simply predicted that temporal regularity of vocal exchange in avian call would be the equivalent of that in monkeys and humans if their communication is based on the expectation of vocal responses from other individuals, as is the case in the vocal communication in primate and mammalian social groups.

2. Methods

2.1. Subjects

Six adult Java sparrows were used (5 males and 1 female, 2~6 years old). All individuals were individually housed in separate cages (38 cm high x 31 cm wide x 25 cm deep) in the same housing area. They could usually hear and see each other.

2.2. Apparatus

We performed the experiment in a soundproofed chamber (2146 cm H x 1378 cm W x 996 cm D), which placed in a separate area from the housing room. Within the chamber, we designed an experimental space in which individual cages could be placed three-dimensionally (Fig. 2A). There, we placed the two cages in two different locations where the two target birds could audio-visually recognize each other. The horizontal distance between the two cages was 58 cm. We attached two small microphones (CVM-V02C, COMICA), one directly above each of the two cages (about 10 cm) to allow close-up recording of vocalizations in each cage. The microphones were connected via cables to an audio interface (UMC1820 U-PHORIA, Behringer) with a computer (Ubuntu 20.2, CPU: Intel Core 5, RAM 16GB) located outside the soundproof chamber, and two channels were used for recording. Audio signals were recorded at 32-bit quantization/32 kHz sampling rate.

2.3. Procedures

A single session of the experiment consisted of three phases (pre-face-to-face phase, PRE phase; face-to-face phase, F2F phase; and post-face-to-face phase, POST phase) according to the following procedure (Fig. 2A).

First, the animals were exposed to a 15-minute PRE phase; in the PRE phase, only one individual (the prior individual) was introduced into the soundproof room and placed in an individually compartmented area. The purpose of the PRE phase was both to habituate them to the experimental area and to enhance spontaneous vocalizations during the entire experimental session while reducing their stress. Then, the F2F phase followed for 15 minutes. During this F2F phase, the other individual (the facing individual) was introduced into the opposite individual compartment area and faced over the cage. Lastly, a POST phase was set up for 15 minutes. During this POST phase, only the facing individuals were allowed to leave the soundproof room, and the same situation as in the PRE phase was arranged in which only the preceding individuals remained in the chamber. Four sessions were performed for every

combination of two individuals selected from six individuals (15 different pair sets), for a total of 60 sessions (total of 2700 min) of recordings. The order of the pair sets was randomized as much as possible, so that the same individuals did not participate in consecutive experimental sessions.

2.4. Acoustic analysis

2.4.1. Labelling vocal data

Prior to the analysis, all bird vocalizations were labelled (or annotated) in the recorded data of the experimental sessions. Labelling was achieved through three processes: segmentation, classification, and caller identification. All labelling tasks were split and performed by the two independent evaluators (SK, NK), using Raven Pro version 1.6 [40] or Praat version 6.3.10 [41]. Prior to the labelling tasks, the two evaluators checked with each other to agree on the criteria for the segmentation and classification of the vocalization types (Fig 2B).

2.4.1.1. Segmentation

To define intervals in the vocal region, two evaluators independently judged the intervals of vocalizations. More specifically, the onset and offset times of the vocal regions were logged from the recorded data based on the visual inspections of waveform and spectrogram while listening to the audio data.

2.4.1.2. Classifications and caller identifications

The detected vocal regions were classified into three types of vocal types (call, song, and trill). Here, *songs* were classified as "vocal sequences that appear in a stereotyped pattern for an individual," *calls* as "single vocalizations that have no sequential rules and are not songs," and *trills* as "typical sounds that appear as short-duration repetitive pulses".

Caller identification was determined by comparing the intensities of multiple channels of recorded data. The location of the microphone channel with higher intensities was judged as location that were closer to the target individual.

2.4.1.3. Labelling validation

To check the validity of the labelling, a subset of the data was extracted: 4 sessions of data were selected from 60 sessions of data to include about 1000 vocalizations, and both evaluators independently labelled the data in the F2F phase. The "agreement" of the labelling between the two labelled data (label-SK, label-NK) was examined (For details of the method used to evaluate the agreement rate, see the supplementary materials). Accuracy rate (number of agreements/total number of segmentate vocal region) for classification and caller identification was 95.75% (969/1012 vocalizations). We also examined whether the onset and offset times differed between evaluators, and found no tendency for either evaluator to make biased judgments (i.e., to make judgments statistically significantly earlier or later than the other evaluator).

2.4.2. Definition of the call transition sequences

For the analysis of vocal exchange, here we defined the type of call transition sequences. We defined an "exchange transition" as a case in which a vocalization by one individual is followed by a vocalization by a different individual, and a "repeated transition" as two consecutive vocalizations by the same individual. Note that in exchange transitions, the offset time of the preceding vocalization might be later than the onset time of the following vocalization.

2.5. Statistical procedures

2.5.1. Statistical evaluation of the occurrence rate of vocalizations

First, we compared the frequency of vocalizations across the three conditions, PRE, F2F, and POST, using generalized linear models and model selection methods. We fit a GLMM (error distribution: Poisson distribution, link function: log) with the number of vocalizations of the prior individual per session as the explanatory variable, the condition as the fixed effect term, and the subject and trial as the random intercepts. By comparing the AICs, we estimated which model is better than the null model, which removes the condition from the fixed effects. The

GLMM fitting and AIC calculations were performed with the glmer function of lme4 (v. 1.1-34) [42] and dredge function of MuMIn (v. 1.47.5) [43] in R (v. 4.2.3) [44].

2.5.2. Estimations of interval structures by fitting Gaussian mixture models

The temporal structure of exchange transitions and repeated transitions were statistically evaluated. Specifically, Gaussian mixture models (GMMs) were applied to the time interval data of each transition to examine the distribution structure. The GMM is an attempt to estimate the number of mixtures and parameters (mean and variance), assuming that the data are generated as a mixture of multiple Gaussian distributions consisting of different parameters. Previous studies have shown that 1) the intervals follow logarithms, and 2) the data are generated from two types of Gaussian distributions: a short time interval caused by the association of two vocalization transitions, and a long time interval randomly caused by two independent vocalizations. Our aim of the GMM analysis was to isolate only the distribution of "the time interval when two vocalizations were associated before and after", which is our interest, and we investigated its distribution by fitting the GMM to the data. With the visual inspections of the bimodal distribution of the interval data, we fitted the model with the mixtures of the two components of Gaussian distributions to the log-transformed data to estimate the parameters of the mixture models. Prior to log transformation, we excluded interval cases with negative data, i.e., where the onset time of the subsequent call preceded the offset time of the initial call, and there were only two such cases. One component of estimating the short time interval (“short component”, hereafter) was interpreted as the distribution of "the time interval when two vocalizations were associated before and after” for each of exchange and repeated transitions. For the GMM fitting, we used R’s package `mclust` (v. 6.0.0) [45].

Finally, to statistically compare the interval of the exchange transition and repeat transition, we performed the permutation test as follows. First, we generated values from two normal distributions defined by the mean and variance of the intervals estimated by GMM (two normal distributions for the exchange transition and repeat transition), respectively, and calculated the difference between the generated values. We repeated this calculation 10,000 times to obtain a distribution of 10,000 difference values. We then calculated a 95-percentile

interval consisting of 2.5 and 97.5 percentiles of these values. Statistical differences between the two distributions were discussed by examining whether 0 was included in this interval. We used `rnorm()` function in R to generate data from the normal distribution.

2.5.3. Pseudo turn-taking data by shuffling the vocal data

The short component observed in the exchange transition could be interpreted as a reaction time in the vocal exchange, but we could not rule out the possibility that it might be caused by chance when two individuals vocalize independently at a certain interval of time. To investigate whether the data distributions were generated by chance, we compared the pseudo turn-taking data (generated with reference to [7]) with the actual data of the exchange transition. Pseudo turn-taking data were created by shuffling and combining data from completely different sessions to simulate "virtual" vocal exchanges: time-series data of one individual's vocal onset and offset in the F2F condition and time-series data of another individual's vocalizations in a *different* session. Pseudo turn-taking data were generated by replacing one individual's data from a certain pair combination with one individual's time-series data from the same pair in a *different* session, keeping the pair combination consistent between the actual data and the pseudo turn-taking data (Fig. 2C).

3. Results

3.1. Temporal distributions of call transitions

We recorded 9608 calls during 60 sessions (900 min in each of the three conditions, totaling 2700 min). Comparing the occurrence rate of vocalizations within a session, we found that F2F was the most vocalized condition between PRE, F2F, and POST (Supplementary Figures S2.1. S2.2. condition model's AIC: 4221.3; null model 's AIC: 4486.9).

When we categorized call transitions, we identified 1712 exchange transitions and 3350 repeated transitions. Figure 3A shows the distribution of call offset-onset intervals for exchange

and repeated transitions, respectively. The mean with variance of the estimated short component of GMM are - 0.728 +/- 0.023 and - 0.669 +/- 0.035 for the exchange and repeated transitions, respectively (Fig 3C and 3D, Table S3). This result suggested no statistical difference between the two transitions. Subsequently, using these estimated parameters, we performed a permutation test to obtain the value differences between conditions and calculated the 95th percentile range of the value differences to be [-0.533, 0.403], not biased away from 0 (Supplementary Figure S3).

3.2. Comparisons with pseudo turn-taking data

In pseudo turn-taking data, we identified 948 exchange transitions and 4168 repeated transitions. Figure 3B shows the distribution of call offset-onset intervals for the exchange transition of the pseudo turn-taking data. Obviously, the short component disappears and no distinct components appear, resulting in a uniform and monotonous distribution. This strongly suggested that the short component identified in the exchange transition originated from the response *call* of the paired individuals to the *call* of the individual.

4. Discussion

Results summary of temporal regularities

Unexpectedly, the results in Java sparrows showed the different temporal regularity from those in vocal exchange/turn-takings observed in humans, primates, and mammals - the short component of the repeated transition is longer than the short component of the exchange transition. Rather, the time distributions between the two transitions were remarkably similar.

Why was no temporal regularity identified in Java sparrows? It may reflect differences in the function of vocal exchange in other animals. The emergence of temporal regularity would be related to the ecological function of vocal exchange. Vocal exchange, which can generally be observed in animal social groups, has been thought to contribute to the maintenance of the spatial cohesiveness of the group. In primates, cetaceans, and mammals, where vocal exchange

is observed, the occurrence of vocal exchange increases when the group is spatially dispersed and the risk of separation from the group increases, such as group movement or restrictions in visibility conditions, or when the distance between individuals is increased. Thus, a function to prevent separation from the group is assumed. Vocal responses from other individuals are thought to function as reciprocal rewards, similar to grooming, in animal social groups. To function as a social reward, the animal would expect a vocal response from the other individual. A situation in which no response is obtained - i.e., expectation violation - would induce repeated vocalizations. As a result, temporal regularity would be observed in animals that typically travel in social groups. Considering the ecological function of vocal exchange in this way, our results suggest that vocal exchange of Java Sparrows does not have the function of maintaining cohesion, and that its social function differs from that of primates and other social groups. Furthermore, it strongly suggests that the expectation of vocalizations by other individuals is absent or extremely weak in Java sparrows.

Sparrows' call served as autonomic reaction to other birds

In contrast, the occurrence rate of calls increased under the F2F condition, clearly indicating that the presence of other individuals increased the rate of calls. Furthermore, the distribution shape of the pseudo-exchange transition generated from the pseudo turn-taking data was significantly different from the actual distribution shape of the exchange transition, as no distinct mode could be observed. These facts indicate that vocal behavior does not occur independently of the presence or behavior of other individuals (various behaviors including vocalization) but depends on the behavior of other individuals and serves as a response to the vocalization of other individuals. The results, together with the fact that the temporal regularity of their vocal exchange differs from that of primates, would suggest two major characteristics of their vocal exchange: 1) they respond reflexively to the vocalizations of other individuals, but 2) their expectations of the vocalizations of other individuals elicited by their own vocalizations are very weak.

What is the ultimate factor making the difference?

The temporal regularity of vocal exchange that had been observed in primate or mammalian groups was notably different from that of the Java sparrows. The present phenomenon, in which behaviors are automatically transmitted between individuals without a strong expectation of reciprocal behavioral responses from other individuals, would suggest a strong involvement of behavioral mechanisms similar to behavioral synchronization commonly confirmed in avian species. Generally, in animals that form extremely large groups, such as bird flock and fish school, behavioral synchronization among individuals is thought to be a key mechanism for maintaining group behavior. In fact, numerous literature has reported various behavioral synchronizations: synchronization of vigilance or scanning behaviors to avoid predators [46,47], synchronization of motion when individuals escape from predators within a flock [48], synchronization of flight angles and speeds between nearby individuals in flight [49], and so on. BOID, the most important discovery in modeling bird flock movements, showed that flock movements emerge through self-organization process based on the local individual-based rules, such as “alignment”, “cohesion”, or “separation”, assuming an automatic behavioral synchronization mechanism in the individual-based rules. More psychologically, automatic behavioral synchronization can be tested as behavioral contagion, although behavioral contagion has also been reported to occur in ravens [50,51] and budgerigars [52]. The behavioral pattern observed in the present vocal exchange may be a form of a time-series behavioral trend resulting from the phenomenon of behavioral synchronization, which is consistently observed in avian species.

Scientific knowledge about turn-taking is also used for applications in the construction of natural conversational systems with humans using large-scale language models. Simulation studies of turn-taking phenomena have also been conducted in recent years to investigate the rules of communicative interaction in animals [7,53]. Of course, in birds, a highly vocal animal, the comparative approach within the above research framework has also been applied, but this approach has been modeled on previous investigations in primates, including humans. The ecological role of vocal exchange needs to be modeled with careful attention to the differences among animals. In particular, a careful approach that explores the evolutionary mechanisms that

shape communication between animals that may differ fundamentally in their herd formation mechanisms is essential, taking into account ecological adaptations.

Declarations

Ethics approval

Experiments were reviewed and approved in advance by the Animal Experiment Ethical Committee, the University of Tokyo, Graduate School of Arts and Sciences, (#2022-1 and 2023-8)

Author contributions:

Sota Kikuchi: Conceptualization (equal); data curation (lead); formal analysis (equal); investigation (lead); visualization (lead); writing original draft (equal); writing review and editing (equal).

Noriko Kondo: Conceptualization (equal); formal analysis (equal); investigation (supporting); writing original draft (equal); writing review and editing (equal).

Hiroki Koda: Conceptualization (lead); data curation (supporting); formal analysis (supporting); visualization (supporting); supervision (lead); writing original draft (lead); writing review and editing (lead)

Conflict of interest declaration

We declare we have no competing interests.

Funding

This study was funded by the JSPS KAKENHI (B, #23K25168/22H03914; S, #23H05428) awarded to HK.

Acknowledgements

We are grateful to Riko Goto and Sana Kohmoto for their administrative works.

References

1. Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735. https://doi.org/10.1353/lan.1974.0010

2. Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6:731*. https://doi.org/10.3389/fpsyg.2015.00731

3. Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J. P., Yoon, K. E., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592. https://doi.org/10.1073/pnas.0903616106

4. Levinson, S. C. (2016). Turn-taking in human communication – Origins and implications forlanguage processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

5. Pika, S., Wilkinson, R., Kendrick, K. H., & Vernes, S. C. (2018). Taking turns: Bridging the gap between human and animal communication. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20180598. https://doi.org/10.1098/rspb.2018.0598

6. Levréro, F., Touitou, S., Frédet, J., Nairaud, B., Guéry, J. P., & Lemasson, A. (2019). Social bonding drives vocal exchanges in Bonobos. *Scientific Reports*, *9:711*. https://doi.org/10.1038/s41598-018-36024-9

7. Takahashi, D. Y., Narayanan, D. Z., & Ghazanfar, A. A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Current Biology*, *23*(21), 2162–2168. https://doi.org/10.1016/j.cub.2013.09.005

8. Lemasson, A., Gandon, E., & Hausberger, M. (2010). Attention to elders' voice in non-human primates. *Biology Letters*, *6*(3), 325–328. https://doi.org/10.1098/rsbl.2009.0875

9. Sugiura, H. (1993). Temporal and acoustic correlates in vocal exchange of coo calls in Japanese macaques. *Behaviour*, *124*(3–4), 207–225. https://doi.org/10.1163/156853993X00588

10. Sugiura, H. (2007). Adjustment of temporal call usage during vocal exchange of coo calls in Japanese Macaques. *Ethology*, *113*(6), 528–533. https://doi.org/10.1111/j.1439-0310.2007.01349.x

11. Koda, H. (2004). Flexibility and context-sensitivity during the vocal exchange of coo calls in wild Japanese macaques (Macaca fuscata yakui). *Behaviour*, *141*(10), 1279–1296. https://doi.org/10.1163/1568539042729685

12. Miller, C. T., & Wang, X. (2006). Sensory-motor interactions modulate a primate vocal behavior: Antiphonal calling in common marmosets. *Journal of Comparative Physiology A*, *192*(1), 27–38. https://doi.org/10.1007/s00359-005-0043-z

13. Takahashi, D. Y., Fenley, A. R., & Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1693), 20150370. https://doi.org/10.1098/rstb.2015.0370

14. Yamaguchi, C., Izumi, A., & Nakamura, K. (2009). Temporal rules in vocal exchanges of phees and trills in common marmosets (Callithrix jacchus). *American Journal of Primatology*, *71*(7), 617–622. https://doi.org/10.1002/ajp.20697

15. Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences*, *263*(1), 101–113. https://doi.org/10.1111/j.1749-6632.1975.tb41575.x

16. Bosch, L. T., Oostdijk, N., & Boves, L. (2005). On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, *47*(1–2), 80–86. https://doi.org/10.1016/j.specom.2005.05.009

17. Yosida, S., Kobayasi, K. I., Ikebuchi, M., Ozaki, R., & Okanoya, K. (2007). Antiphonal vocalization of a subterranean rodent, the naked mole-rat (Heterocephalus glaber). *Ethology*, *113*(7), 703–710. https://doi.org/10.1111/j.1439-0310.2007.01371.x

18. Yosida, S., & Okanoya, K. (2009). Naked mole-rat is sensitive to social hierarchy encoded in antiphonal vocalization. *Ethology*, *115*(9), 823–831. https://doi.org/10.1111/j.1439-0310.2009.01677.x

19. Carter, G. G., Skowronski, M. D., Faure, P. A., & Fenton, B. (2008). Antiphonal calling allows individual discrimination in white-winged vampire bats. *Animal Behaviour*, *76*(4), 1343–1355. https://doi.org/10.1016/j.anbehav.2008.04.023

20. Leong, K. M., Ortolani, A., Burks, K. D., Mellen, J. D., & Savage, A. (2003). Quantifying acoustic and temporal characteristics of vocalizations for a group of captive African elephants Loxodonta Africana. *Bioacoustics*, *13*(3), 213–231. https://doi.org/10.1080/09524622.2003.9753499

21. Soltis, J., Leong, K., & Savage, A. (2005). African elephant vocal communication I: Antiphonal calling behaviour among affiliated females. *Animal Behaviour*, *70*(3), 579–587. https://doi.org/10.1016/j.anbehav.2004.11.015

22. Demartsev, V., Strandburg-Peshkin, A., Ruffner, M., & Manser, M. (2018). Vocal turn-taking in Meerkat group calling sessions. *Current Biology*, *28*(22), 3661-3666.e3. https://doi.org/10.1016/j.cub.2018.09.065

23. Janik, V. M. (2000). Whistle matching in wild bottlenose dolphins. *Science*, *289*(5483), 1355–1357.

24. Nakahara, F., & Miyazaki, N. (2011). Vocal exchanges of signature whistles in bottlenose dolphins (Tursiops truncatus). *Journal of Ethology*, *29*, 309–320. https://doi.org/10.1007/s10164-010-0259-4

25. Morisaka, T., Yoshida, Y., Akune, Y., Mishima, H., & Nishimoto, S. (2013). Exchange of "signature" calls in captive belugas (Delphinapterus leucas). *Journal of Ethology*, *31*(2), 141–149. https://doi.org/10.1007/s10164-013-0358-0

26. Mishima, Y., Morisaka, T., Ishikawa, M., Karasawa, Y., & Yoshida, Y. (2019). Pulsed call sequences as contact calls in Pacific white-sided dolphins (Lagenorhynchus *obliquidens*). *The Journal of the Acoustical Society of America*, *146*(1), 409–424. https://doi.org/10.1121/1.5116692

27. Caldwell, M. C., & Caldwell, D. K. (1968). Vocalization of Naive Captive Dolphins in Small Groups. *Science*, *159*(3819), 1121–1123. https://doi.org/10.1126/science.159.3819.1121

28. Schulz, T. M., Whitehead, H., Gero, S., & Rendell, L. (2008). Overlapping and matching of codas in vocal interactions between sperm whales: Insights into communication function. *Animal Behaviour*, *76*(6), 1977–1988. https://doi.org/10.1016/j.anbehav.2008.07.032

29. Catchpole, C. K., & Slater, P. J. B. (2008). *Bird Song: Biological Themes and Variations* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511754791

30. Hall, M. L. (2004). A review of hypotheses for the functions of avian duetting. *Behavioral Ecology and Sociobiology*, *55*(5), 415–430. https://doi.org/10.1007/s00265-003-0741-x

31. Catchpole, C. K. (1987). Bird song, sexual selection and female choice. *Trends in Ecology & Evolution*, *2*(4), 94–97. https://doi.org/10.1016/0169-5347(87)90165-0

32. Kroodsma, D. E., & Byers, B. E. (1991). The Function(s) of Bird Song. *American Zoologist*, *31*(2), 318–328. https://doi.org/10.1093/icb/31.2.318

33. Kondo, N., Watanabe, S., & Izawa, E.-I. (2010). A temporal rule in vocal exchange among large-billed crows *Corvus macrorhynchos* in Japan. *Ornithological Science*, *9*(1), 83–91. https://doi.org/10.2326/osj.9.83

34. Chaiken, M. (1990). The ontogeny of antiphonal calling in European starlings. *Developmental Psychobiology*, *23*(3), 233–246. https://doi.org/10.1002/dev.420230304

35. Dreiss, A. N., Ruppli, C. A., Oberli, F., Antoniazza, S., Henry, I., & Roulin, A. (2013). Barn owls do not interrupt their siblings. *Animal Behaviour*, *86*(1), 119–126. https://doi.org/10.1016/j.anbehav.2013.04.019

36. Benichov, J. I., Benezra, S. E., Vallentin, D., Globerson, E., Long, M. A., & Tchernichovski, O. (2016). The forebrain song system mediates predictive call timing in female and male zebra finches. *Current Biology*, *26*(3), 309–318. https://doi.org/10.1016/j.cub.2015.12.037

37. Benichov, J. I., & Vallentin, D. (2020). Inhibition within a premotor circuit controls the timing of vocal turn-taking in zebra finches. *Nature Communications*, *11*, 221. https://doi.org/10.1038/s41467-019-13938-0

38. D'Amelio, P. B., Trost, L., & ter Maat, A. (2017). Vocal exchanges during pair formation and maintenance in the zebra finch (Taeniopygia guttata). *Frontiers in Zoology*, *14*(1), 13. https://doi.org/10.1186/s12983-017-0197-x

39. Ter Maat, A., Trost, L., Sagunsky, H., Seltmann, S., & Gahr, M. (2014). Zebra finch mates use their forebrain song system in unlearned call communication. *PLoS ONE*, *9*(10), e109334. https://doi.org/10.1371/journal.pone.0109334

40. K. Lisa Yang Center for Conservation Bioacoustics. (2019). *Raven Pro: Interactive Sound Analysis Software* (1.6). The Cornell Lab of Ornithology. http://ravensoundsoftware.com

41. Paul, B., & David, W. (2023). *Praat: Doing phonetics by computer* (6.3.10). https://www.praat.org.

42. Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

43. Bartoń, K. (2023). *MuMIn: Multi-Model Inference* (1.47.5). https://CRAN.R-project.org/package=MuMIn

44. R Core Team. (2023). *R: A Language and Environment for Statistical Computing* (4.2.3). R Foundation for Statistical Computing. https://www.R-project.org/

45. Scrucca, L., Fop, M., Murphy, T., Brendan, & Raftery, A., E. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal*, *8*(1), 289. https://doi.org/10.32614/RJ-2016-021

46. Fernández-Juricic, E., Siller, S., & Kacelnik, A. (2004). Flock density, social foraging, and scanning: An experiment with starlings. *Behavioral Ecology*, *15*(3), 371–379. https://doi.org/10.1093/beheco/arh017

47. Evans, M. H. R., Lihou, K. L., & Rands, S. A. (2018). Black-headed gulls synchronise their activity with their nearest neighbours. *Scientific Reports*, *8*(1), 9978. https://doi.org/10.1038/s41598-018-28378-x

48. Hemelrijk, C. K., Van Zuidam, L., & Hildenbrandt, H. (2015). What underlies waves of agitation in starling flocks. *Behavioral Ecology and Sociobiology*, *69*(5), 755–764. https://doi.org/10.1007/s00265-015-1891-3

49. Bajec, I. L., & Heppner, F. H. (2009). Organized flight in birds. *Animal Behaviour*, *78*(4), 777–789. https://doi.org/10.1016/j.anbehav.2009.07.007

50. Wenig, K., Pacher, L., & Bugnyar, T. (2022). Testing the contagious nature of allopreening: Bystander ravens are affected by conspecifics' affiliative interactions. *Animal Behaviour*, *184*, 71–80. https://doi.org/10.1016/j.anbehav.2021.12.009

51. Wenig, K., Boucherie, P. H., & Bugnyar, T. (2021). Early evidence for emotional play contagion in juvenile ravens. *Animal Cognition*, *24*(4), 717–729. https://doi.org/10.1007/s10071-020-01466-0

52. Miller, M. L., Gallup, A. C., Vogel, A. R., Vicario, S. M., & Clark, A. B. (2012). Evidence for contagious behaviors in budgerigars (Melopsittacus undulatus): An observational study of yawning and stretching. *Behavioural Processes*, *89*(3), 264–270. https://doi.org/10.1016/j.beproc.2011.12.012

53. Grijseels, D. M., Fairbank, D. A., & Miller, C. T. (2023). A model of marmoset monkey vocal turn-taking. bioRxiv, 2023.10.07.561358. https://doi.org/10.1101/2023.10.07.561358

Figures

Figure 1. Conceptual representation of the mechanism underlying the emergence of temporal regularity in vocal exchange/turn-taking of various social animals. When animals are engaged in a reciprocal vocal exchange (A, top), the individual expects a "response" from the other individual and therefore waits for a response. After the expected response time has elapsed, the individual repeats the vocalization (A, bottom). Consequently, the inter-call interval of repeated transition, t_r, is longer than those of exchange transition, t_e (A, B).

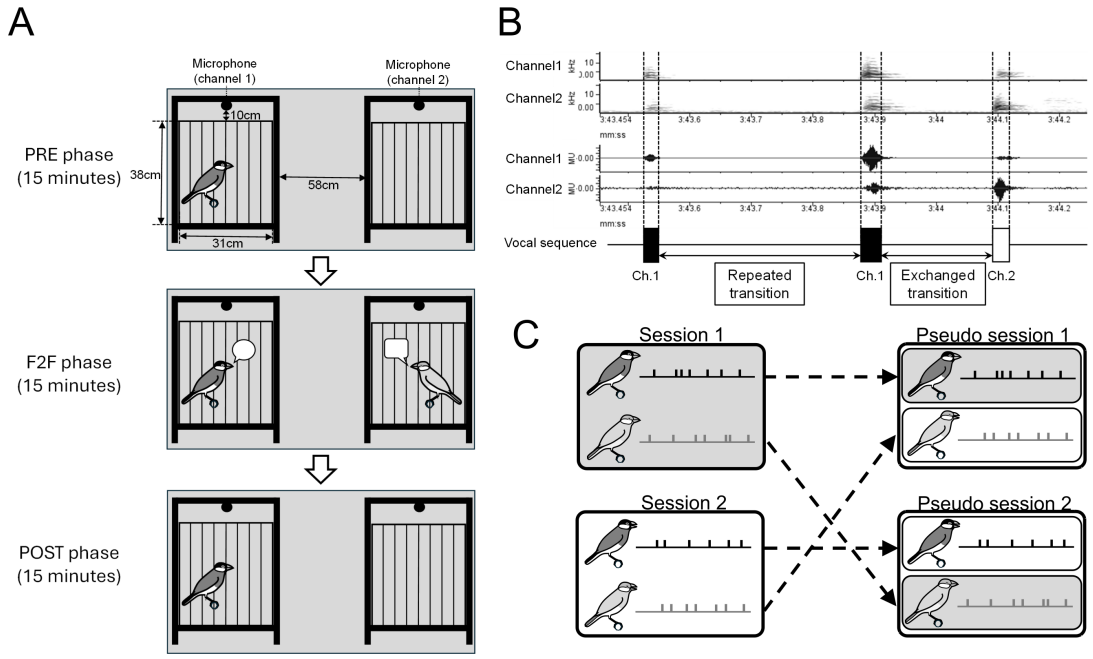

Figure 2. Schematic illustrations of (A) experimental settings, (B) labelling procedure of calls, and (C) the generation process of shuffled session data. (A) The experimental flow of the three phases, along with cage dimensions and spatial arrangement. (B) The sound spectrogram (top), waveform (middle), and segmental call regions labelled by evaluators (bottom), obtained from a 2-channel recording. (C) Workflow to generate pseudo turn-taking data (right). The originally recorded F2F data were replaced in different recording sessions, keeping the individual pair relationship consistency.

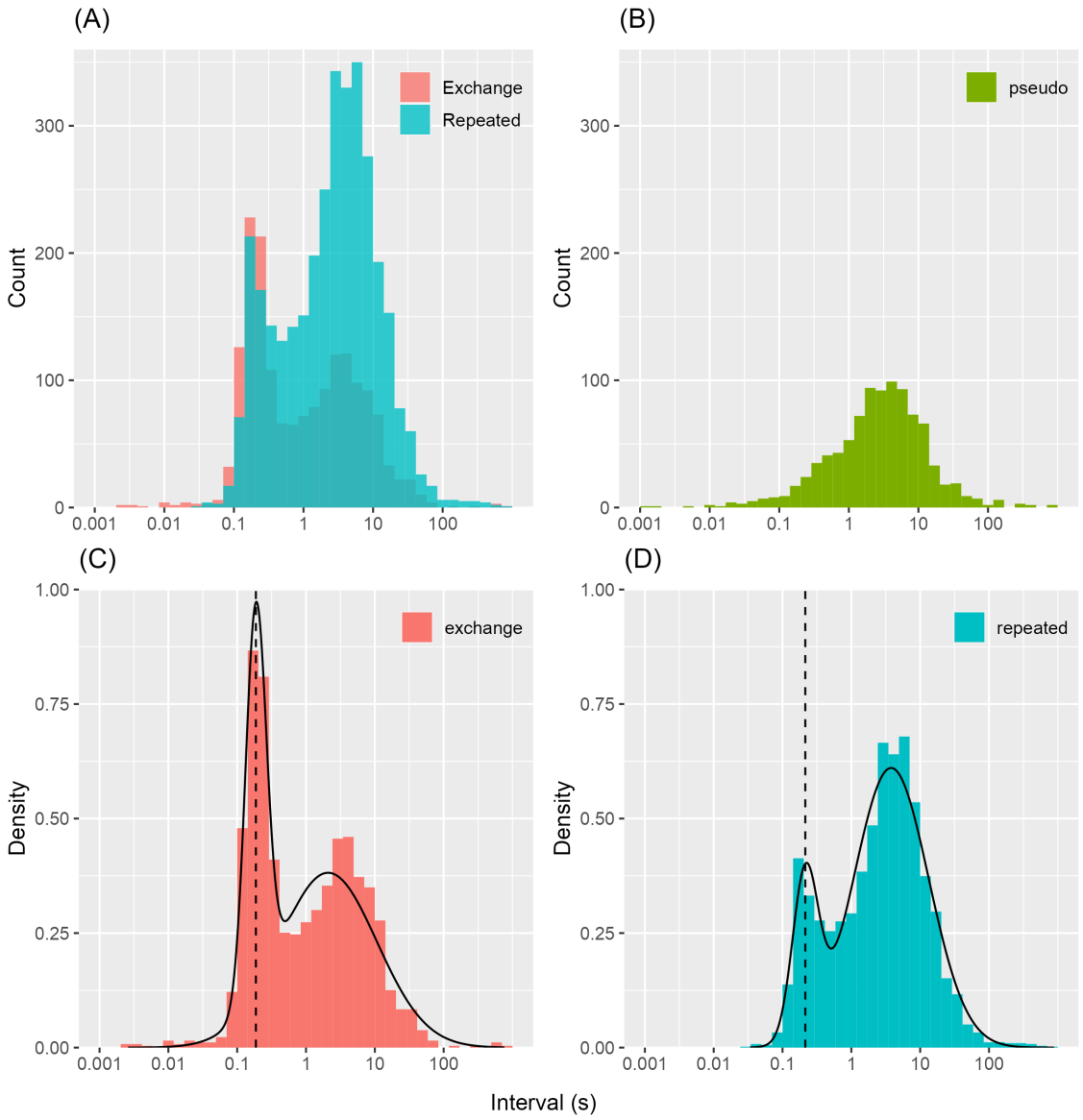

Figure 3. Distributions of inter-call intervals of exchange transitions and repeated transitions. (A) Superposition of two distributions of exchange (red) and repeated (blue) transitions. (B) The distribution of the inter-call intervals of pseudo turn-taking data. (C) Superposition of the histogram and estimated probability density function curve (curve line) fitted by the two modal GMM for the exchange transition and (D) for the repeated transitions. The dashed vertical lines were the positions of the estimated values of mean parameter in the short components of GMMs. All x-axis values are log-transformed.