

Coupled Oscillator Dynamics of Vocal Turn-Taking in Monkeys

Daniel Y. Takahashi,^{1,2,*} Darshana Z. Narayanan,^{1,2} and Asif A. Ghazanfar^{1,2,3,*}

¹Neuroscience Institute

²Department of Psychology

³Department of Ecology and Evolutionary Biology
Princeton University, Princeton, NJ 08540, USA

Summary

Cooperation is central to human communication [1–3]. The foundation of cooperative verbal communication is taking turns to speak. Given the universality of turn-taking [4], it is natural to ask how it evolved. We used marmoset monkeys to explore whether another primate species exhibits cooperative vocal communication by taking turns. Marmosets share with humans a cooperative breeding strategy and volubility. Cooperative care behaviors are thought to scaffold prosocial cognitive processes [5, 6]. Moreover, marmosets and other callitrichid primates are very vocal and readily exchange vocalizations with conspecifics [7–11]. By measuring the natural statistics of marmoset vocal exchanges, we observed that they take turns in extended sequences and show that this vocal turn-taking has as its foundation dynamics characteristic of coupled oscillators—one that is similar to the dynamics proposed for human conversational turn-taking [12]. As marmoset monkeys are on a different branch of the evolutionary tree that led to humans, our data demonstrate convergent evolution of vocal cooperation. Perhaps more importantly, our data offer a plausible alternative scenario to “gestural origin” hypotheses for how human cooperative vocal communication could have evolved.

Results

Vocal turn-taking, across all languages, consists of speech exchanges with gaps of silence and minimal overlaps [4, 13]. For multiple rounds, person 1 speaks while person 2 attends, and then person 1 relinquishes the speaking status to allow person 2 to speak or relinquish her turn. We used common marmoset monkeys (*Callithrix jacchus*) to explore whether another primate species shows a similar pattern of cooperative communication. Marmosets don’t have the same sophisticated semantic and syntactic skills as humans, nor do they display much evidence of shared intentionality, but they do have in common with humans a cooperative breeding strategy and volubility [8, 14]. Cooperative care behaviors are thought to scaffold prosocial cognitive processes [5, 6]. Marmosets and other callitrichid primates also readily exchange vocalizations with conspecifics (whether or not they are related or pair bonded) [7, 9–11], particularly when they are out of visual contact. Is this a simple call-and-response (“antiphonal”) behavior seen in numerous species, or is it a sustained temporal coordination of vocal exchanges as in human conversation?

Marmosets were placed in opposite corners of sound-attenuated room, separated by an acoustically transparent but

visually opaque curtain (Figure 1A). We recorded the vocal exchanges produced by ten marmosets paired in various combinations (five cagemate pairs and 20 noncagemate pairs; none were related to each other). Across 54 sessions, we acquired a total of 1,415 phee calls. Phee calls are the long-distance contact calls used by marmosets, and, although marmosets have a number of distinct vocalizations produced in a number of different contexts [8], 99.9% of calls recorded under our experimental conditions were phee calls. Figure 1B shows the waveform and spectrogram of two marmosets exchanging phee calls. We considered a phee call syllable as part of a single call if the interval between two consecutive syllables was less than 1 s. This criterion is justified by the bimodal structure of the intersyllable interval distribution for single marmosets’ vocal output, in which the 1 s threshold separates the first mode of the distribution (representing the interval between syllables within a single phee call) from the second mode that represents the interval between the offset of the last and onset of the first syllables between two calls (Figure 1C). Figure 1D is representative of what our data structure looks like, illustrating the onset and offset timing of the phee calls of one pair of marmosets recorded across five sessions.

We defined call exchanges as two consecutive phee calls from two individuals separated by no more than 30 s of silence ($n = 322$) [9–11]. Consistent with human conversational exchanges, there were zero overlapping calls (i.e., no interruptions) in these exchanges. One possibility is that individuals may be calling at very low rates (but independently of each other), thus making the probability of overlapping with another’s call unlikely. To test the hypothesis that the lack of interruptions was due to truly dependent vocal interactions, we used a permutation test: we exchanged the time series of vocal output of one individual in dyad and paired it with the time series of another, randomly selected individual from a different dyad (Figure 2A). We found that marmosets wait for the vocal exchange partner to finish calling before responding (permutation test, $p < 0.001$) (Figure 2B). A potential caveat is that perhaps one very vocal marmoset generates the call interruptions seen in the permuted data. To test for this, we reran the permutation test an additional ten times, but left one individual out of the data for each iteration. In all cases, the difference between the real data and permuted data remained ($p < 0.001$ for all iterations).

The consistent waiting period of about 5–6 s upon hearing a call establishes a turn-taking rule. The duration of this waiting period is similar to that reported in previous studies with callitrichids [9, 11]. The mechanism that might generate such reliability in vocal exchange dynamics is not known. One possibility is that an individual might produce phee calls with some planned interval, but when he hears a call from another marmoset, that interval is simply reset (Figure 2C); i.e., his subsequent call is delayed by an interval duration predicted by the calling pattern he would have produced spontaneously if he had not heard another individual’s call. A mechanism akin to this is used in chorusing insects and frogs, in which call coordination is an epiphenomenon related to the resetting of an individual’s signal output upon hearing a neighbor’s signal [15]. We tested this hypothesis by comparing the distribution of intervals between consecutive calls from the same subject (that is, two calls without an intervening reply from another) versus consecutive calls in which the second was in presumptive

*Correspondence: takahashi@princeton.edu (D.Y.T.), asifg@princeton.edu (A.A.G.)

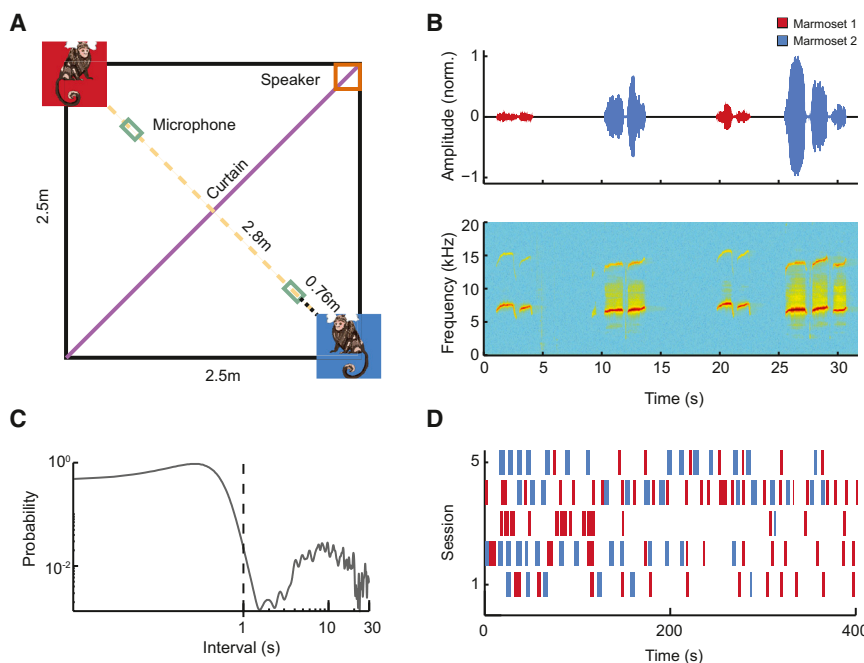


Figure 1. Experimental Setup and Data

(A) Illustration of the experimental setup for the vocal exchange experiment.

(B) Waveform and spectrogram of a sequence of phee call exchanges. x axis, time in seconds; y axis, normalized amplitude (top) and frequency in kHz (bottom). Red and blue waveforms represent calls from marmoset 1 and marmoset 2, respectively.

(C) Distribution of the interval between syllables in log-log scale. x axis, interval in seconds; y axis, probability density of the intervals.

(D) Example of marmoset vocal exchange data. x axis, time in seconds; y axis, session number. The red and blue hashes represent different marmosets. The beginning and end of each hash indicates respectively the onset and offset of a phee call.

reply to another marmoset's call (i.e., during a vocal exchange). This "reset" hypothesis would predict that the interval distributions would be the same. In Figure 2D, the gray line shows that these distributions are, in fact, significantly different ($p < 0.001$, Kolmogorov-Smirnov [K-S] test). The call interval duration of an individual is, on average, significantly shorter (median = 5.63 s) during vocal exchanges than when the same subject produces calls without hearing an intervening call from another individual (median = 11.53 s, p value < 0.001 , Wilcoxon rank-sum test). Thus, a simple reset of vocal production timing upon hearing another's call does not explain these data. An important corollary of this finding is that marmosets distinguish their own calls from those of another; that is, they self-monitor during vocal production as humans do during conversations [16].

A second possibility is that the effect of hearing a phee call simply inhibits the production of one call but without changing the overall structure of the subsequent call sequence. For example, a marmoset may have a motor plan to produce a sequence of three calls, but just prior to producing the second it hears another marmoset's call (Figure 2C). This second call would be suppressed, but the third call would proceed at the originally intended time. We tested this hypothesis by simulating this scenario statistically using a modification of the permutation test. In this test, we exchanged the pairs of callers from different sessions, which generated call exchanges with occasional interruptions (Figure 2B). We then removed the calls that were interrupted (calls with later onsets in the interruptions) and measured the interval distribution between the calls on the simulated sequence (blue line) and compared it to the real interval distribution (green line). If hearing another's call simply suppressed calls in a production sequence, then these interval distributions should be the same. In Figure 2D, the blue line shows that the real vocal exchange interval distribution cannot be explained by an inhibition of this nature ($p < 0.001$, K-S test). These results suggest that marmosets take turns during vocal exchanges, i.e., when the first marmoset produces a call, a second marmoset waits until it finishes and then responds following a specific time interval that

cannot be predicted by the simple mechanisms of resetting (Figure 2D, gray line) or inhibition (Figure 2D, blue line).

What then is the mechanism that explains the turn-taking dynamics? In the speech literature, dynamic systems

models incorporating oscillator-like mechanisms have been hypothesized to account for conversational turn-taking [12, 17]. We explicitly tested a coupled oscillatory mechanism to explain the temporal structure of marmoset vocal turn-taking. First, we tested for the presence of periodic couplings. The signature of two oscillators coupled in antiphase is the occurrence of events at integer multiples of some fixed period. In this scenario, not only should marmoset 1 produce a sequence of phee calls with a consistent intercall interval, but marmoset 2's phee call replies should be both in between marmoset 1's calls (antiphase) and have a consistent intercall interval with respect to marmoset 1's calls (Figure 3A). To test this, we measured the time interval of marmoset 2's subsequent calls after each of marmoset 1's calls. That is, we measured the interval between marmoset 1's first call and then the marmoset 2's first, second, third... n^{th} calls. We then repeated this for marmoset 1's second call. Using this procedure, we quantified the degree of coupling by calculating the cross-correlation between the call time series of the two marmosets exchanging vocalizations. If the two marmosets are coupled, then the correlation should have significant peaks at regular intervals (Figure 3A)—as seen in human conversational exchanges [18, 19]. If there was no coupling between marmosets, then the correlation would be flat (Figure 3B). To verify the statistical significance of the cross-correlation, we calculated the cross-correlation between the time series of Poisson point processes with the rate of points matching the average call rate in our data.

Our results show that, as predicted for coupled oscillators, the cross-correlation is characterized by the presence of statistically significant periodic peaks, indicating that calls are produced in antiphase with a period of ~ 12 s (Figure 3C). This is identical to what is predicted to occur in human conversations [12, 19] and observed in human conversational turn-taking statistics [18] (though on a much faster time scale; see the Discussion). Moreover, the significant cross-correlation extending until at least 60 s shows that a call influences not only the first response call from another marmoset, but also the dynamics of subsequent call responses. Therefore, the

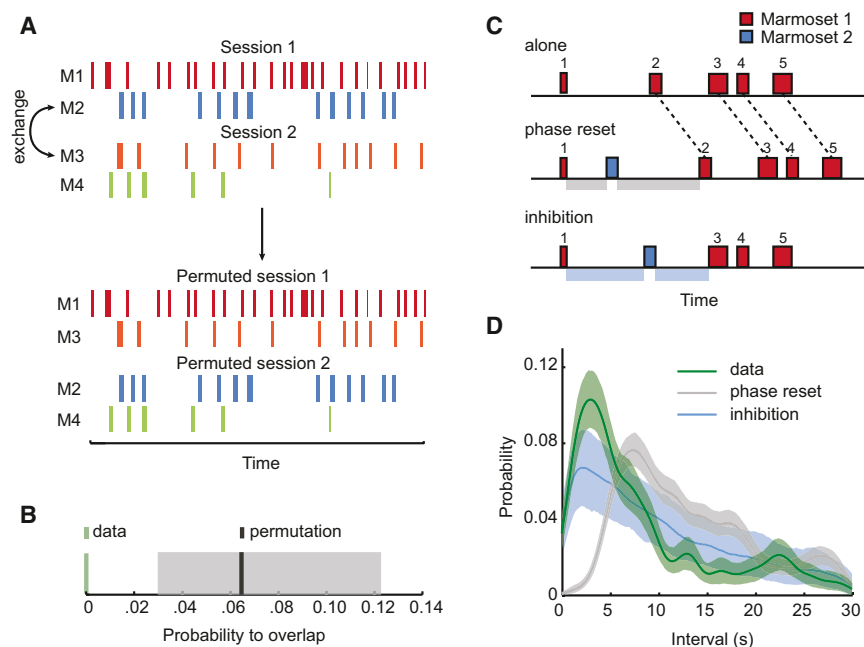


Figure 2. Testing Three Hypotheses for the Dynamics of Marmoset Vocal Exchange

(A) Permutation procedure to test whether the marmosets were producing the call independently of each other or were interacting. x axis, time in seconds. Each hash represents call duration and different colors correspond to different marmosets (M1–M4). (B) Probability that a call overlaps for the data (green line) and permutation test (black line). The gray area represents the 95% confidence interval for the probability of overlap for the permuted data set. (C) Schematic of the phase resetting and inhibitory hypothesis. x axis, time in seconds. Red and blue rectangles represent calls from marmoset 1 and marmoset 2, respectively. The gray and blue shaded regions represent the intervals between vocal exchanges for the resetting and inhibitory hypothesis, respectively. (D) Testing the phase resetting and inhibitory hypothesis. x axis, interval between vocal exchanges in seconds; y axis, the probability density distribution. The shaded regions represent the 95% confidence intervals.

coupling extends to several time scales, consistent with dynamics of a cooperative behavior [20]. One potential caveat is that one marmoset could be acting as an oscillator while the other is not; this would still result in a cross-correlation with periodicity. With a leave-one-out procedure, the blue traces in Figure 3C show that the periodicity is consistently apparent. Moreover, the autocorrelation (where one marmoset's call is correlated with the timing of his own subsequent calls) of each marmoset in a dyad exhibits its own periodicity (Figure 3D), demonstrating that each marmoset is acting as an oscillator.

If marmoset vocal turn-taking dynamics can be modeled as coupled oscillators, then it is expected that the periodicity of call production from one marmoset can be modulated by the periodicity of call production from the other marmoset [21]. That is, the two marmosets should be entrained to each other. To investigate this, we calculated the phase response curve (PRC) of vocal exchanges [22] (Figure 4). We measured three intervals (Figure 4A). First, we measured the duration of the response interval (R) between marmoset 1's call and marmoset 2's response. Second, we measured the interval (PR) between marmoset 1's first call and, after marmoset 2's response, its second call. Third, we measured the interval duration (T_0) of marmoset 1's calls that did not have an intervening call from marmoset 2. This T_0 value was subtracted from PR to give us $PR - T_0$ and will indicate whether marmoset 1's calls are produced more quickly or slowly when responding to marmoset 2. Figure 4B shows three possible outcomes if $PR - T_0$ is plotted against R. In this scenario, a positive slope would indicate that the two marmosets entrain to each other and speed up (or slow down) their calling output depending on the other marmoset, while a negative slope would indicate that when one marmoset speeds up, then the other slows down (or vice versa).

Figure 4C shows the PRC and the scatter plot of R versus $PR - T_0$ ($n = 134$). Green dots represent real data; however, because the sampling is not uniform across the 0–30 s response interval, a bias correction method was implemented [22]. The bias correction data are represented by the purple dots. After controlling for this bias, there are two modes of

dynamics in marmoset vocal exchanges. The presence of two modes is expected given that the median intercall interval of a single marmoset calling (T_0) is 12.41 s (Figure 4C, vertical dotted line). We verified its statistical significance using a robust regression algorithm: one from 0 to 12.41 s and the other from 12.41 to 30 s. We observed that there is a positive correlation in the 0 to 12.41 s interval (t test, $p < 0.001$ for both the intercept and coefficient). Moreover, the $PR - T_0$ crosses the zero value in the y axis, indicating that if R (the response from marmoset 2) is short (or long), vocal production by marmoset 1 will speed up (or slow down). For response intervals beyond 12 s, there is no correlation between R and $PR - T_0$ (t test, $p = 0.34$ for the intercept and $p = 0.51$ for the coefficient). Thus, the likelihood of a marmoset producing a “response” to another's call decreases with intervals exceeding ~ 12 s; at this time, it is ambiguous whether marmoset 2's vocalizations are produced spontaneously or in response to marmoset 1's call.

When two oscillators are coupled, they can eventually reach an equilibrium state, i.e., the phase delays between the oscillatory events become stationary. Therefore, if the vocal exchange dynamics is oscillatory and achieves an equilibrium state, we expect that the value of the R when $PR - T_0 = 0$ (i.e., when the phase doesn't change) should match the median interval time in a vocal exchange. This is indeed the case for our data. The $PR - T_0$ is zero when R is ~ 5 s (Figure 4C) and the median in the interval time in a vocal exchange is ~ 5 s (Figure 4C, vertical dashed line; see also Figure 2D, green line).

Discussion

Individuals in a conversation must reciprocally and continuously coordinate their vocal output with each other; i.e., it is necessary that they cooperate [12, 18, 23]. We show that marmoset monkeys exhibit such cooperative vocal behavior. Dyads engage in turn-taking by showing that (1) they prevent overlapping calls by waiting ~ 5 s before responding, (2) the timing of their calls is periodically coupled, and (3) they entrain

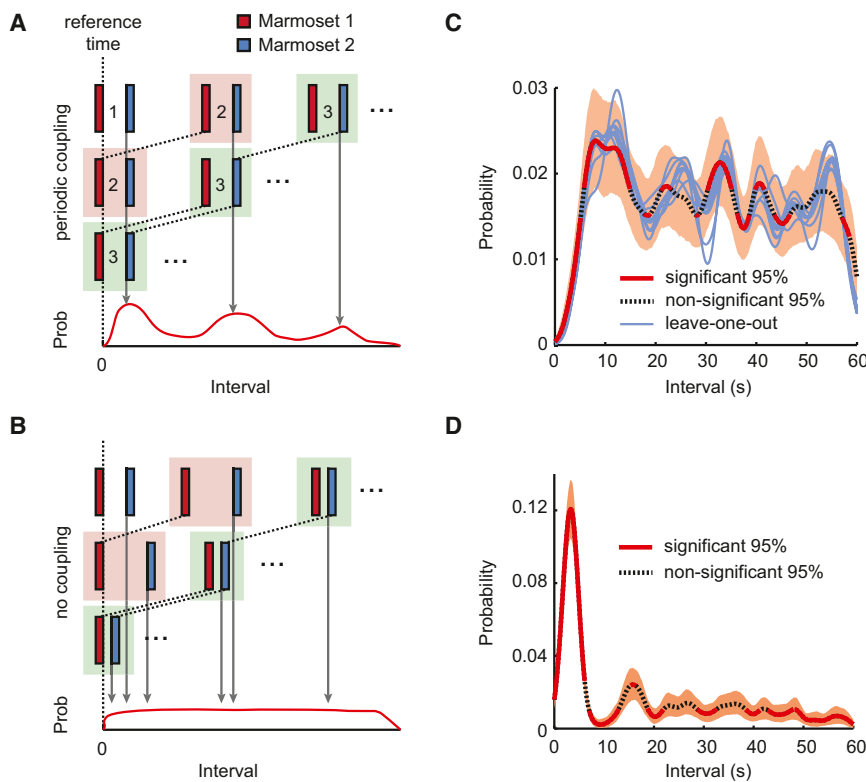


Figure 3. Vocal Exchanges in Marmosets Are Periodically Coupled

(A) Schematic of the calculation of cross-correlation for periodically coupled calls. x axis, time in seconds. Each hash represents different calls. The red and blue colors represent marmoset 1 and marmoset 2, respectively.

(B) Schematic of the calculation of cross-correlation for uncoupled calls. The legend is the same as in (A).

(C) Cross-correlation plot for the experimental data. x axis, intervals in seconds; y axis, probability density value. Solid red line represents the cross-correlation values that were significant at 5% level; dotted black lines represent the cross-correlation values that were not significant. Blue lines represent ten interactions in which one marmoset was left out of the analysis.

(D) Autocorrelation plot of the experimental data. The legend is the same as in (C).

to each other such that if one speeds up or slows down their call timing, the other will do so as well. With these features, marmoset vocal turn-taking is very similar to human conversational turn-taking [12, 13], but there are differences as well. One major difference is time scale: gaps between turns for marmosets are several seconds (~3–5 s), whereas gaps for human conversations are usually within hundreds of milliseconds (~250 ms) [4]. The slower time scale for the vocal exchange in marmosets compared to the human may be related to the “units of perception” in each species. For marmosets and other callitrichid species, the minimal unit of communication is likely the whole call, with a duration of ~3–5 s, not each syllable [10], whereas in humans, the word or syllable may be the minimal unit of communication, with much shorter durations [24, 25].

Marmosets, as in human conversations, spontaneously engage in coordinated exchanges with any other conspecific. That marmoset vocal exchanges are periodically coupled and entrained to each other demonstrates that the two individuals are acting like coupled oscillators. Recently, coupled oscillators have been put forth as explanatory models of social interactions in general [20, 26] and conversational turn-taking in particular [12]. These models don’t require higher-order cognitive capacities, have the advantages of parsimony, have support from psychophysical experiments [18, 19], and have consistency with the known neurophysiological processes related to speech [27, 28]. With the current data, we can now link this mechanism as a putative foundation for the evolution of vocal turn-taking. This is not to argue that higher-order cognitive capacities like recognition, prediction, and/or understanding are not important for human conversations [29–31]; no doubt these play important roles guiding our conversations. However, they are not the sole criteria for cooperative communication. Our data, and the relatively small

encephalization quotient of marmoset monkeys relative to the great apes [32], suggest that cooperative communication does not require ape- or human-specific cognitive capacities. Indeed, marmoset vocal turn-taking may represent the kind of foundational scaffold upon which more sophisticated forms of communication were built in humans.

The advantages of extended coordinated vocal interactions may be two-fold: arousal reduction [33] and information extraction [34]. First, for marmoset monkeys that are away from conspecifics, phee calls can act as a form of contact. Such auditory contact reduces stress and arousal levels [35] with its attendant benefits for cognition, physiology and survival [36]. Sustained, coordinated vocal interactions would indicate, in a continuous fashion, that one is being listened to by another. Indeed, in terms of stress reduction, conversational exchanges among humans are thought to play the same role as social grooming does for monkeys and apes [37]. For marmosets, exchanging contact calls may, in effect, be grooming at a distance. Another possibility is that sustained vocal coordination allows for the extraction of important social information. Embedded in their acoustics, marmoset phee calls contain information about gender, identity, social group, and context [38, 39]. It may take a series of exchanges with no interfering and overlapping calls to efficiently and fully extract this information, particularly in noisy forest environments [40]. Turn-taking provides such a mechanism.

Our results show that marmoset engages in vocal turn-taking behavior with conspecifics, whether or not they are pair bonded, related or engaged in mating or territorial encounters. That is, the turn-taking behavior we report is distinct from the coordinated call patterns (e.g., duetting and chorusing) seen in other animals [15, 41], but similar to what is observed in humans (thus, an example of convergent evolution). Moreover, the lack of evidence for such vocal turn-taking in apes motivates, in part, the gestural hypotheses for the origin of human vocal cooperative behavior, whereby cooperation is first exhibited through manual gestures whose cognitive substrates are (somewhat mysteriously) then co-opted by vocalizations [3, 42, 43]. Vocal turn-taking in marmosets supports the notion that cooperative vocal communication could have

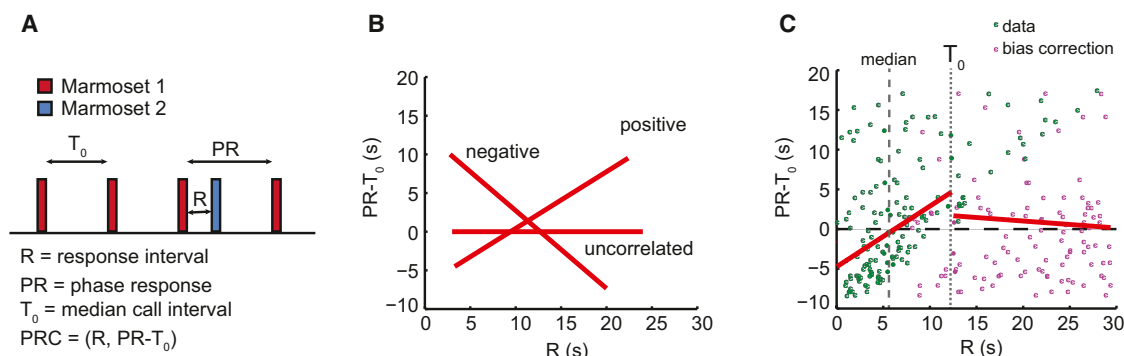


Figure 4. Entrainment of Vocal Exchanges between Marmosets

(A) Schematic of the parameters to calculate the PRC. x axis, time in seconds. The red and blue hashes represent marmoset 1's and marmoset 2's vocalizations, respectively.

(B) Possible PRC patterns. x axis, response interval (R) in seconds; y axis, $PR - T_0$ in seconds.

(C) PRC for the vocal exchange data. x axis, R in seconds; y axis, $PR - T_0$ in seconds. The value of the median spontaneous call interval duration T_0 is 12.41 s (dotted line). Only sequences of calls of type marmoset 1-marmoset 2 were considered ($n = 134$, green dots). Bias correction points are represented by purple dots. Solid red lines represent the resulting PRC estimate.

evolved in a manner very different than gestural-origins hypotheses predict. In this alternative scenario, existing vocal repertoires could begin to be used in a cooperative, turn-taking manner when prosocial behaviors in general emerged. This occurred in both humans and callitrichid primates when they evolved a cooperative breeding strategy and, as a consequence, more prosocial tendencies such as stronger social bonds, higher tolerance at the group level, cooperative problem solving, and active food sharing [5, 6, 44]. Similar to the “byproduct of selection for temperament” hypothesis put forth to explain the human-like social skills of domestic dogs [45], this raises the possibility that the observation of cooperative vocal turn-taking in marmosets, but not in other nonhuman primates, can be at least in part explained by differences in social motivations that evolved as a byproduct of selection on behaviors related to cooperative breeding [5].

Experimental Procedures

Subjects

The subjects used in this study include ten adult common marmosets (*Callithrix jacchus*) housed at Princeton University. The marmosets were five male-female pairs from different social groups. Animals were fed once daily with standard commercial chow supplemented with fresh fruits and vegetables. The animals had ad libitum access to water. The colony room was maintained at a temperature of approximately 27°C and 50%–60% relative humidity, with a 12 hr light:12 hr dark cycle. Previous to the experiment being conducted, all animals were trained to leave their home cage in a transport box and familiarized with the testing room and testing equipment. All experimental sessions were conducted during daylight hours between 12:00 and 18:00 and were performed in compliance with the guidelines of the Princeton University Institutional Animal Care and Use Committee.

Measuring Vocal Exchanges

We ran each adult marmoset in two experimental conditions: alone and paired. In the alone condition, each marmoset was placed alone in the testing room and the vocalizations were recorded. The alone condition is used in this study only to calculate the distribution of interval between consecutive call syllables and to calculate the median interval between consecutive spontaneous call (T_0 ; see “Phase Response Curve” below). In the paired condition, two animals were placed in the same room and the vocalizations were recorded. All sessions lasted either 15 or 30 min. Each animal was tested only once per day, and subjects were run on the two conditions in randomized order. The experimental room measured 2.5 m × 2.5 m with walls covered in sound attenuating foam. Two tables

(0.66 m in height) were positioned at diagonally opposite corners of the room. The animals were placed—one on each table in the paired condition—in prism-shaped testing boxes made of plexiglas and wire (0.30 m × 0.30 m × 0.35 m). The testing corner was counterbalanced across each monkey and sessions. A speaker was positioned at a third corner equidistant from both testing corners, and pink noise was broadcast at ~45 dB in order to mask occasional noises produced external to the testing room. Digital recorders (ZOOM H4n Handy Recorder) were placed directly in front of each testing box at a distance of 0.76 m. Audio signals were acquired at sampling frequency of 96 kHz. An opaque cloth occluder divided the room in two and prevented the subjects from getting visual cues from each other during the course of the experiment. Each testing box was thoroughly wiped down between each test session to eliminate odors left by previous subjects. For the paired condition, the experimenter ensured that each of the paired marmosets had no visual contact with each other, from the time of removal from the home environment until the end of experiment. Once the subjects were in place, the experimenter turned on both recorders and left the room.

Phee Call Detection

A custom made MATLAB routine automatically detected the onset and offset of any acoustic signal that differed from the background noise at specific frequency range. To detect the differences, we band-passed the entire recording signal between 5 and 8 kHz. This corresponds to the fundamental frequency of marmoset phee calls. We then compared the amplitude of the signal at this frequency band for the periods without call and during a call. A simple threshold was enough to distinguish both periods. Onset-offset gaps longer than one second indicated separate calls, whereas gaps shorter or equal than one second indicated syllables from the same call. After this procedure, we manually verified for each call whether the automatic routine correctly identified single phee calls or combined multiple calls, using the one second separation criteria. Furthermore, we counted the number of syllables of each call, and if there was any mismatch between the automatic onset-offset detection and the call signal observed by the researcher, then we marked the onset and offset of the phee call manually comparing the level of the background noise to the amplitude of the signal. For the paired data set, we had to compare the amplitude of the band-passed signal recorded from the two microphones in the room to determine which of the marmoset was producing a call. When the same call recorded from opposing corners of the room was compared, the amplitude was larger for the microphone at the same corner of the caller. We exhaustively tested and confirmed the accuracy of this procedure.

Statistics of Phee Call Production

The interval distribution between consecutive call syllables, the response time distribution, cross-correlation, and autocorrelation were all estimated using a Gaussian kernel smoothing density estimator implemented in ksdensity routine in MATLAB. The bootstrap confidence interval was

obtained by resampling the call duration data set 1,000 times and then estimating the distributions using the same Gaussian kernel for each resampled data. We plotted the 95% confidence interval for each point by calculating the 97.5% and 2.5% percentile of the resampled estimates. The interval between calls was defined as the time difference between the onset of a call and the offset of a previous call.

Permutation Test

To test the possibility that the marmosets are behaving independently (not interacting), we applied a permutation test in which we paired the vocal output of two marmosets in the paired condition recorded from different sessions and chosen at random. These randomly paired sessions were the surrogates for the null hypothesis of two marmosets producing vocalizations independently of each other. We repeated the procedure 1,000 times and calculated the 95% and 99.9% confidence intervals for the permuted data set and compared the results with the experimental data.

Autocorrelation and Cross-correlation

We calculated the cross-correlation function between the offset of a call from say marmoset 1 and the onset of the calls from marmoset 2. The cross-correlation is the distribution of the intervals between marmoset 1's call offset and marmoset 2's calls onset that follows, irrespective of the calls being consecutive or not. The autocorrelation is the distribution of the intervals between marmoset 1's call offset and marmoset 1's calls onset that follows, irrespective of the calls being consecutive or not. To test for the significance of the cross-correlation (autocorrelation) values, we calculated the cross-correlation (autocorrelation) between two Poisson point processes with rate matching the average rate of call production in the paired data set and computed the 95% confidence interval. The bootstrap confidence interval for the cross-correlation was obtained by resampling of the intervals between marmoset 1's call offset and the following marmoset 2 calls. The bootstrap confidence interval for autocorrelation was obtained by resampling of the intervals between marmoset 1's call offset and the following marmoset 1 calls. The cross-correlation and autocorrelation were repeated for each resampled data. The bootstrap procedure was repeated 1,000 times to calculate the 95% confidence intervals.

Phase Response Curve

Consider two marmosets, 1 and 2. Assume that marmoset 1 called (call M1a), and then marmoset 2 called back (call M2) and marmoset 1 called again (call M1b). The interval between the offset of call M1a and onset of call M2 represents the response time interval (R). The interval between the offset of call M1a and onset of call M1b represents the phase response (PR). Finally, T_0 is the median time interval between two consecutive calls from marmoset 1, i.e., the median interval duration between calls with no intervening calls from marmoset 2. To calculate T_0 , we used the call intervals from the alone condition experiment. The graph that describes the relationship between R and $PR - T_0$ is called the phase response curve [22]. Because the PRC estimate is biased when the response time interval (R) is not uniformly sampled, we had to correct for the bias. We corrected for the bias in PRC estimates using the bias correction method proposed in [22]. If marmoset 1 and marmoset 2 don't interact, $PR - T_0$ should be zero for all values of response time. To test whether the PRC was significantly different from a constant line, we fitted two regression lines, one for the interval 0–12.41 s and other for 12.41–30 s using a robust fitting algorithm (iteratively reweighted least square with bisquare weighting function, MATLAB robustfit) and tested for the significance of the coefficients of the regression using a t test.

Acknowledgments

We thank Adele Goldberg, Uri Hasson, and Donald Katz for their insightful comments and criticisms of an earlier draft of this manuscript. We also thank Alison Vitale for help with data collection. This work was supported by the Pew Latin American Fellowship (D.Y.T.), the Brazilian Science without Borders Fellowship (D.Y.T.), and the James S. McDonnell Scholar Award (A.A.G.).

Received: July 17, 2013

Revised: August 30, 2013

Accepted: September 2, 2013

Published: October 17, 2013

References

1. Grice, H.P. (1975). Logic and conversation. In *Syntax and Semantics, Volume 3*, P. Cole and J. Morgan, eds. (New York: Academic Press), pp. 43–58.
2. Levinson, S.C. (2006). On the human interactional engine. In *Roots of Human Sociality*, N.J. Enfield and S.C. Levinson, eds. (New York: Berg Publishers), pp. 39–69.
3. Tomasello, M. (2008). *Origins of Human Communication* (Cambridge: MIT Press).
4. Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J.P., Yoon, K.-E., and Levinson, S.C. (2009). Universals and cultural variation in turn-taking in conversation. *Proc. Natl. Acad. Sci. USA* 106, 10587–10592.
5. Burkart, J.M., Hrdy, S.B., and van Schaik, C.P. (2009). Cooperative breeding and human cognitive evolution. *Evol. Anthropol.* 18, 175–186.
6. Burkart, J.M., and van Schaik, C.P. (2010). Cognitive consequences of cooperative breeding in primates? *Anim. Cogn.* 13, 1–19.
7. Chen, H.C., Kaplan, G., and Rogers, L.J. (2009). Contact calls of common marmosets (*Callithrix jacchus*): influence of age of caller on antiphonal calling and other vocal responses. *Am. J. Primatol.* 71, 165–170.
8. Bezerra, B.M., and Souto, A. (2008). Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int. J. Primatol.* 29, 671–701.
9. Miller, C.T., and Wang, X.Q. (2006). Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 192, 27–38.
10. Ghazanfar, A.A., Flombaum, J.I., Miller, C.T., and Hauser, M.D. (2001). The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*): playback experiments with long calls. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* 187, 27–35.
11. Ghazanfar, A.A., Smith-Rohrberg, D., Pollen, A.A., and Hauser, M.D. (2002). Temporal cues in the antiphonal long-calling behaviour of cotton-top tamarins. *Anim. Behav.* 64, 427–438.
12. Wilson, M., and Wilson, T.P. (2005). An oscillator model of the timing of turn-taking. *Psychon. Bull. Rev.* 12, 957–968.
13. Sacks, H., Schegloff, E.A., and Jefferson, G. (1974). Simplest systematics for organization of turn-taking for conversation. *Language* 50, 696–735.
14. Zahed, S.R., Prudom, S.L., Snowdon, C.T., and Ziegler, T.E. (2008). Male parenting and response to infant stimuli in the common marmoset (*Callithrix jacchus*). *Am. J. Primatol.* 70, 84–92.
15. Greenfield, M.D., Tourtellot, M.K., and Snedden, W.A. (1997). Precedence effects and the evolution of chorusing. *Proc. Biol. Sci.* 264, 1355–1361.
16. Wilson, T.P., and Zimmerman, D.H. (1986). The structure of silence between turns in 2-party conversation. *Discourse Process.* 9, 375–390.
17. Chapple, E.D. (1970). *Culture and Biological Man: Explorations in Behavioral Anthropology* (New York: Holt, Rinehart & Winston).
18. O'Dell, M.L., Nieminen, T., and Lennes, M. (2012). Modeling turn-taking rhythms with oscillators. *Linguist. Ural.* 48, 218–227.
19. Benus, S., Gravano, A., and Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *J. Pragmatics* 43, 3001–3027.
20. Schmidt, R., and Morr, S. (2010). Coordination dynamics of natural social interactions. *Int. J. Sport Psychol.* 41, 105–106.
21. Schultheiss, N.W., Prinz, A.A., and Butera, R.J. (2012). *Phase Response Curves in Neuroscience: Theory, Experiment, and Analysis* (New York: Springer).
22. Phoka, E., Cuntz, H., Roth, A., and Häusser, M. (2010). A new approach for determining phase response curves reveals that Purkinje cells can act as perfect integrators. *PLoS Comput. Biol.* 6, e1000768.
23. Pereira, A.F., Smith, L.B., and Yu, C. (2008). Social coordination in toddler's word learning: interacting systems of perception and action. *Connect. Sci.* 20, 73–89.
24. Chandrasekaran, C., Trubanova, A., Stillitano, S., Caplier, A., and Ghazanfar, A.A. (2009). The natural statistics of audiovisual speech. *PLoS Comput. Biol.* 5, e1000436.
25. Lerner, Y., Honey, C.J., Silbert, L.J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31, 2906–2915.
26. Oullier, O., de Guzman, G.C., Jantzen, K.J., Lagarde, J., and Kelso, J.A.S. (2008). Social coordination dynamics: measuring human bonding. *Soc. Neurosci.* 3, 178–192.

27. Giraud, A.L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517.
28. Hasson, U., Ghazanfar, A.A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* **16**, 114–121.
29. Duncan, S., and Fiske, D.W. (1979). Dynamic patterning in conversation. *Am. Sci.* **67**, 90–98.
30. Cassell, J. (2000). Embodied conversational interface agents. *Commun. ACM* **43**, 70–78.
31. Gravano, A., and Hirschberg, J. (2011). Turn-taking cues in task-oriented dialogue. *Comput. Speech Lang.* **25**, 601–634.
32. Jerison, H.J. (1973). *Evolution of the Brain and Intelligence* (New York: Academic Press).
33. Owren, M.J., Amoss, R.T., and Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *Am. J. Primatol.* **73**, 530–544.
34. Seyfarth, R.M., Cheney, D.L., Bergman, T., Fischer, J., Zuberbuhler, K., and Hammerschmidt, K. (2010). The central importance of information in studies of animal communication. *Anim. Behav.* **80**, 3–8.
35. Rukstalis, M., and French, J.A. (2005). Vocal buffering of the stress response: exposure to conspecific vocalizations moderates urinary cortisol excretion in isolated marmosets. *Horm. Behav.* **47**, 1–7.
36. Uno, H., Tarara, R., Else, J.G., Suleman, M.A., and Sapolsky, R.M. (1989). Hippocampal damage associated with prolonged and fatal stress in primates. *J. Neurosci.* **9**, 1705–1711.
37. Dunbar, R. (1998). *Grooming, Gossip, and the Evolution of Language* (Cambridge: Harvard University Press).
38. Miller, C.T., Mandel, K., and Wang, X. (2010). The communicative content of the common marmoset phee call during antiphonal calling. *Am. J. Primatol.* **72**, 974–980.
39. Norcross, J.L., and Newman, J.D. (1993). Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phee calls. *Am. J. Primatol.* **30**, 37–54.
40. Morrill, R.J., Thomas, A.W., Schiel, N., Souto, A., and Miller, C.T. (2013). The effect of habitat acoustics on common marmoset vocal signal transmission. *Am. J. Primatol.* **75**, 904–916.
41. Hall, M.L. (2004). A review of hypotheses for the functions of avian duetting. *Behav. Ecol. Sociobiol.* **55**, 415–430.
42. Hewes, G.W. (1973). Primate communication and the gestural origin of language. *Curr. Anthropol.* **14**, 5–24.
43. Rizzolatti, G., and Arbib, M.A. (1998). Language within our grasp. *Trends Neurosci.* **21**, 188–194.
44. Snowdon, C.T., and Cronin, K.A. (2007). Cooperative breeders do cooperate. *Behav. Processes* **76**, 138–141.
45. Hare, B., and Tomasello, M. (2005). Human-like social skills in dogs? *Trends Cogn. Sci.* **9**, 439–444.