



Research

Cite this article: Chow CP, Mitchell JF, Miller CT. 2015 Vocal turn-taking in a non-human primate is learned during ontogeny. *Proc. R. Soc. B* **282**: 20150069. <http://dx.doi.org/10.1098/rspb.2015.0069>

Received: 13 January 2015

Accepted: 25 March 2015

Subject Areas:

behaviour, evolution, neuroscience

Keywords:

marmoset, turn-taking, vocal learning

Author for correspondence:

Cory T. Miller

e-mail: corymiller@ucsd.edu

[†]These authors contributed equally to this work.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2015.0069> or via <http://rspb.royalsocietypublishing.org>.

Vocal turn-taking in a non-human primate is learned during ontogeny

Cecilia P. Chow^{1,†}, Jude F. Mitchell^{1,3,†} and Cory T. Miller^{1,2}

¹Cortical Systems and Behavior Laboratory, Department of Psychology, and ²Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA, USA

³Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

Conversational turn-taking is an integral part of language development, as it reflects a confluence of social factors that mitigate communication. Humans coordinate the timing of speech based on the behaviour of another speaker, a behaviour that is learned during infancy. While adults in several primate species engage in vocal turn-taking, the degree to which similar learning processes underlie its development in these non-human species or are unique to language is not clear. We recorded the natural vocal interactions of common marmosets (*Callithrix jacchus*) occurring with both their sibling twins and parents over the first year of life and observed at least two parallels with language development. First, marmoset turn-taking is a learned vocal behaviour. Second, marmoset parents potentially played a direct role in guiding the development of turn-taking by providing feedback to their offspring when errors occurred during vocal interactions similarly to what has been observed in humans. Though species-differences are also evident, these findings suggest that similar learning mechanisms may be implemented in the ontogeny of vocal turn-taking across our Order, a finding that has important implications for our understanding of language evolution.

1. Background

The evolutionary origins of language have been a source of conjecture for many decades [1]. One challenge in reconstructing its evolutionary origins is the limited occurrence of many core linguistic characteristics in extant non-human primate communication systems [2–4]. Although vocal learning has often been considered similarly impoverished amongst our non-human primate cousins [5], this characterization is not entirely accurate. Indeed, robust changes in the acoustic structure of vocal signals learned through sensory feedback, often referred to as sensory-motor learning, is limited among non-human primates [5,6]. However, other aspects of ontogenetic learning, such as call usage and comprehension [7–9], are evident. Language acquisition is not limited to sensory-motor learning, but comprises a myriad of other learning mechanisms [10–12]. Identifying key similarities and differences in vocal learning across a broader range of behaviours would provide a more extensive comparative dataset and, therefore, potentially inform our understanding of language evolution. Vocal turn-taking is one behaviour that occurs both in human [13] and non-human primates [14–17], providing an unique opportunity to directly compare its properties across these closely related species.

Conversations are a fundamental characteristic of human language and reflect the intricate interplay that occurs between social behaviour and communication. The suite of social rules governing the fluid exchange of signals, such as turn-taking, are learned early in development through experience during social interactions [18,19], and exhibit cultural differences [20] in humans. This aspect of vocal learning is distinct from the sensory-motor learning that occurs for signal structure [6,21], but does similarly involve modification through feedback from others [22]. Many non-human primate species engage in an analogous behaviour during natural vocal interactions [14–17]. Vocal turn-taking in non-human primates, however, has not typically been considered in comparative discussions of language learning and evolution. Instead, the overwhelming emphasis has been

placed on sensory-motor learning for signal structure and comparisons with Passerine songbirds [23]. While this aspect of the songbird system has notable analogies to some features of extant language learning, its suitability in reconstructing many aspects of language evolution may be limited. Birds and mammals are only distantly related, separated by over 300 million years of evolution [24]. Like the independent evolution of flight in birds and bats, similarities in vocal learning between songbirds and humans are due to convergent evolution, rather than homologous structures. Non-human primates, therefore, remain pivotal to understanding language evolution.

Here we sought to test whether vocal turn-taking is learned during ontogeny in a species of non-human primate—the common marmoset (*Callithrix jacchus*). Like humans, marmosets never interrupt each other's vocalizations during vocal exchanges [15,17] and will coordinate the timing of their calls relative to conspecifics to avoid environmental noise that may affect the periodicity of these interactions [25]. In fact, the temporal pattern of turn-taking is perceptually salient in this New World primate species, as individuals will cease vocal interactions if a conspecific's response latency is outside a particular period of time [26]. The vocal behaviour in marmosets also exhibits some key differences to human conversations. Notably, marmoset vocal interactions in this context occur over a time scale of several seconds [15,17], while turn-taking occurs within a few hundred milliseconds in human conversations [20], and are limited to a single call type [15,17]. Given the apparent parallels across these species in the social and communicative dynamics of turn-taking, this study sought to examine whether the behaviour is learned in marmosets during development. Similarly to many other non-human primate species [8,27], the acoustic structure of marmoset vocalizations undergoes relatively little change during ontogeny, suggesting that the sensory-motor learning characteristic of songbirds is limited during ontogeny [28], though the occurrence of dialects in adults indicates that the capacity is not entirely absent [29]. Learning in this context of turn-taking would be a function of enacting control over facets of the vocal interaction rather than modifications to the signal structure.

2. Material and methods

(a) Subjects

A total of 10 infant/juvenile and four adult common marmosets (*C. jacchus*) served as subjects in this experiment from 2010 to 2013. Marmosets were housed in family groups comprising a pair-bonded male and female, as well as one or more generations of offspring. This social organization is consistent with what occurs in wild populations [30]. As marmosets typically give birth to twins, all offspring were a part of a sibling pair. Three of the sibling pairs were from one set of parents, while the other two sibling pairs were from another set of parents. Subjects were housed in the UCSD Cortical Systems and Behavior Laboratory in social groups consisting of pair-bonded mates and their offspring in a colony of approximately 40 animals. We acclimatized the offspring to the procedure from one to three months and started recording them from 4 to 12 months. This time period coincided with at least the three following developmental stages in marmosets. Infancy in marmosets extends from birth until five–six months, followed by a juvenile stage through to approximately 10 months of age and a sub-adult period that occurs until sexual maturity at approximately 18 months [31]. We did not begin experiments until infants freely entered the transport box

and exhibited no sign of duress being separated from the family group. This typically did not occur until four months of age. All experimental procedures followed UCSD IACUC protocol.

(b) Procedure

We transported subjects from their home cage in the colony room to the sound attenuated experimental room using a metal wire transport cage. All recordings took place in a 4×3 m Radio-Frequency Shielded testing room (ETS-Lindgren). The testing room is organized with two rectangular tables positioned at opposite ends of the room 5 m apart with a cloth occluder positioned equidistant between the two tables. Subjects were placed on opposite sides of the room in plastic mesh testing cages ($32 \times 8 \times 46$ cm) in which they were able to climb and freely move. Subjects were always visually occluded from each other during these experiments. A microphone (Sennheiser Model ME-66) was placed 0.3 m directly in front of each subject to record its Phee call vocalizations. All recording sessions lasted 30 min and data were recorded directly to disk for subsequent data analyses.

(c) Test conditions

This study comprised three test conditions that were repeated twice a month on each subject from 4 to 12 months of age: mother–offspring, father–offspring and sibling–sibling. In the latter condition, siblings were always twins from the same parents. The order of these six test sessions was randomized each month.

(d) Data analysis

We recorded vocalizations produced by 14 common marmosets during natural vocal interactions that occurred between visually occluded individuals (10 infants/juveniles, $n = 53\,363$ calls; two fathers, $n = 7436$ calls; two mothers, $n = 15\,267$ calls). Data analysis comprised two stages. The first stage involved identifying vocalizations in the original raw recordings. For this phase, we marked the onset and offset of all vocalizations in the original recordings manually using RAVEN: Interactive Sound Analysis Software (v. 1.3). The timing of these calls provided the basis for analyses of the temporal pattern of interactions in subsequent statistical analyses. The Phee call is the most common call type produced by adults during vocal exchanges in this behavioural context [15,17]. Individual pulses of calls comprising multiple pulses, such as Phee calls, were grouped into single calls if the inter-pulse interval was less than 1000 ms based on previous analyses [29]. In the second stage of analysis, custom written MATLAB (Mathworks v. 2013a) code was then used to extract these vocalizations from the raw recordings and visually identified by trained scorers as one of the four most common call types produced from previous descriptions [28,32]: 'Phee', 'twitter', 'trill', 'trillPhee' or 'other'. The four call types were labelled if and only if the entire call group consisted of that one call type. If there was a mixture of call types, then that call group was labelled as 'other'. These mixed calls are common early in development [28].

In this study, we defined vocal turn-taking as the occurrence of successive, reciprocal vocal exchanges between callers that occur in general social contexts. For all analyses, a response was defined as one subject producing a vocalization within 10 s of the other subject. This duration is based on several previous experiments indicating that calls produced within this time period are perceived as a response, whereas after this period of time it is perceived as the start of a new vocal exchange [26].

(e) Normalized probability plots

To characterize the detailed timing of responses to calls (either appropriate calls or interruptions), we computed histograms that

show the counts of response calls at different delays following the call offset (figures 2 and 4). The histogram of the responses are broken into 0.5 s time bins that are time-locked from the call offset with the first bin having a centre of 0.25 s. The counts in each histogram were normalized by dividing by the average number of counts expected in each time bin given the total number of calls and assuming random timing (i.e. the mean count rate). The average height of the normalized histogram over all delays would thus be one. Peaks above one indicate a response was more likely than the average spontaneous rate while troughs indicate withholding of a response. Probability histograms were also used to quantify the response of the parents to normal and interrupted calls from the juveniles (figure 4). For each infant we labelled which calls interrupted the parent call, with the interval of the parent call defined from the onset to the offset, and in the case of two-pulse Phee calls including the gap between two pulses, thus taking the offset as being the offset of the second pulse. For the analysis of interruptions, we additionally smoothed the probability histograms with a Gaussian kernel with a half-width of 3 s.

(f) Coherence and power spectra analysis

To characterize the degree of reciprocity in vocal interactions, we performed an analysis of the coherence between calling events. Coherence provides a measure that reveals how predictable one sequence of events is from another sequence, broken into different periodicities reflecting the back and forth of vocal exchanges. Each interaction was transformed into a sequence of ones and zeros in discrete time bins of 0.2 s width, where a value of one was used to indicate a period in which a call was being made and zero otherwise (see the electronic supplementary material, figure S3a). Vocal interactions were recorded for 30 min (1800 s) and broken down into 10 segments each of 180 s for analysis. For each month, most pairs of individuals participated in more than a single exchange and in these cases the vocal interactions were concatenated giving more sampled trials (average 20–40 trials per pair). The power spectra of individual series and the spectral coherence between series were computed using multi-taper methods [33]. Smoothing in the frequency domain of 0.025 Hz was used for the 180 s trial duration (giving eight Slepian tapers). Significant peaks in the coherence (as illustrated in the electronic supplementary material, figure S3c for a single session) reflect reciprocal exchanges with a consistent phase relationship. The frequency of the peak in coherence reflects the mean period of these back and forth exchanges, while the narrowness of the peak reflects the consistency of the timing at that period. The height of the peak reflects the strength of the interaction, which depends both on the number of joint events as well as their consistency.

To quantify how coherence changed over months, parameters of the peak in coherence in the range from 0 to 0.2 Hz were computed for each pair in each month. The peak value was first identified in the range of 0–0.2 Hz. Then the half-width of the peak was computed by decrementing or incrementing from that peak location until half of its height was identified, and the half-width computed as the difference between these values. Then the middle frequency of the peak was computed as a weighted average of the distribution falling within the identified locations of half height. The magnitude of the coherence was taken as the peak value.

3. Results

We recorded 53 363 vocalizations produced by 10 common marmosets during natural vocal interactions over the first year of their respective lives, as well as 22 703 vocalizations

produced by their respective parents. Experiments here focused on naturally occurring vocal exchanges that occur when marmosets are visually occluded from each other and involve the reciprocal exchange of a single call types: Phee calls [17,29]. The two most common errors evident early in the year were frequent interruptions of parents (figure 1a,b) and producing call types inappropriate to the context (figure 1c,d). Interruptions by marmosets in this context may be analogous to co-vocalizing described in human infant–parent conversations [34], while the refinement to only producing Phees may be similar to changes in human infants from vocalic to syllabic sounds during turn-taking development [35]. Adult marmosets rarely interrupt each other during these vocal exchanges [15,17]. Yet young marmosets interrupted callers significantly more than adults until eight months of age (Ranksum test, months 4–7, $p < 0.05$; figure 1a,b). Difference in interruption rates could not be accounted for by the higher frequencies of calling among juveniles (electronic supplementary material, figure S1). Likewise, it was not until the same age (eight to nine months) that marmosets exclusively produced Phee calls during these vocal interactions (Ranksum test, months 4–6 and month 8, $p < 0.05$). Prior to this age, infants produced a myriad of other call types (figure 1c,d; [28,32]) and exhibited marked individual variability (see the electronic supplementary material, figure S2). Interestingly, juvenile marmosets produced fewer errors during vocal interactions with their siblings than with their parents. This was evident for both interruptions (figure 1b; shown in magenta) and proportion of non-Phee calls (figure 1c; shown in magenta). This suggests that the ontogenetic changes for these two errors are not likely a product of a more general developmental or maturational pattern, but reflect a distinct pattern of vocal learning.

(a) Turn-taking in marmosets is a learned vocal behaviour

The development of vocal turn-taking in marmosets exhibited marked differences based on social context, suggesting that the behaviour is learned. Broadly, infant interactions with mothers included turn-taking at a significantly earlier age than with fathers (figure 2). Responses to mothers' calls in early months tended to begin immediately following call offset (figure 2a; peak response at 0.5–1 s), but slowed to a more adult-like distribution comparable to adults in the later months (figure 2b). By contrast, infant marmosets were prone to initiate responses prior to the offset of fathers' calls during early months (figure 2c). In later months (figure 2d), marmoset juveniles tended to withhold their response until the offset of fathers' calls, but emitted their call immediately after this time. Phee calls produced by mothers and fathers in this study were approximately 100 ms different in length suggesting that basic acoustic factors could not account for these differences (two-pulse calls 2.53 ± 0.33 s s.d. and 2.66 ± 0.37 s s.d., one-pulse calls 1.32 ± 0.27 s s.d. and 1.37 ± 0.28 s s.d., for mothers and fathers in this study, respectively).

A key feature of human conversations is reciprocity and its emergent rhythm [20]. Vocal interactions between marmoset parents and their offspring already exhibited significant turn-taking behaviour even in the earliest months measured (four months). We implemented coherence analysis to characterize developmental changes in this feature of marmoset vocal interactions (see Materials and methods and the

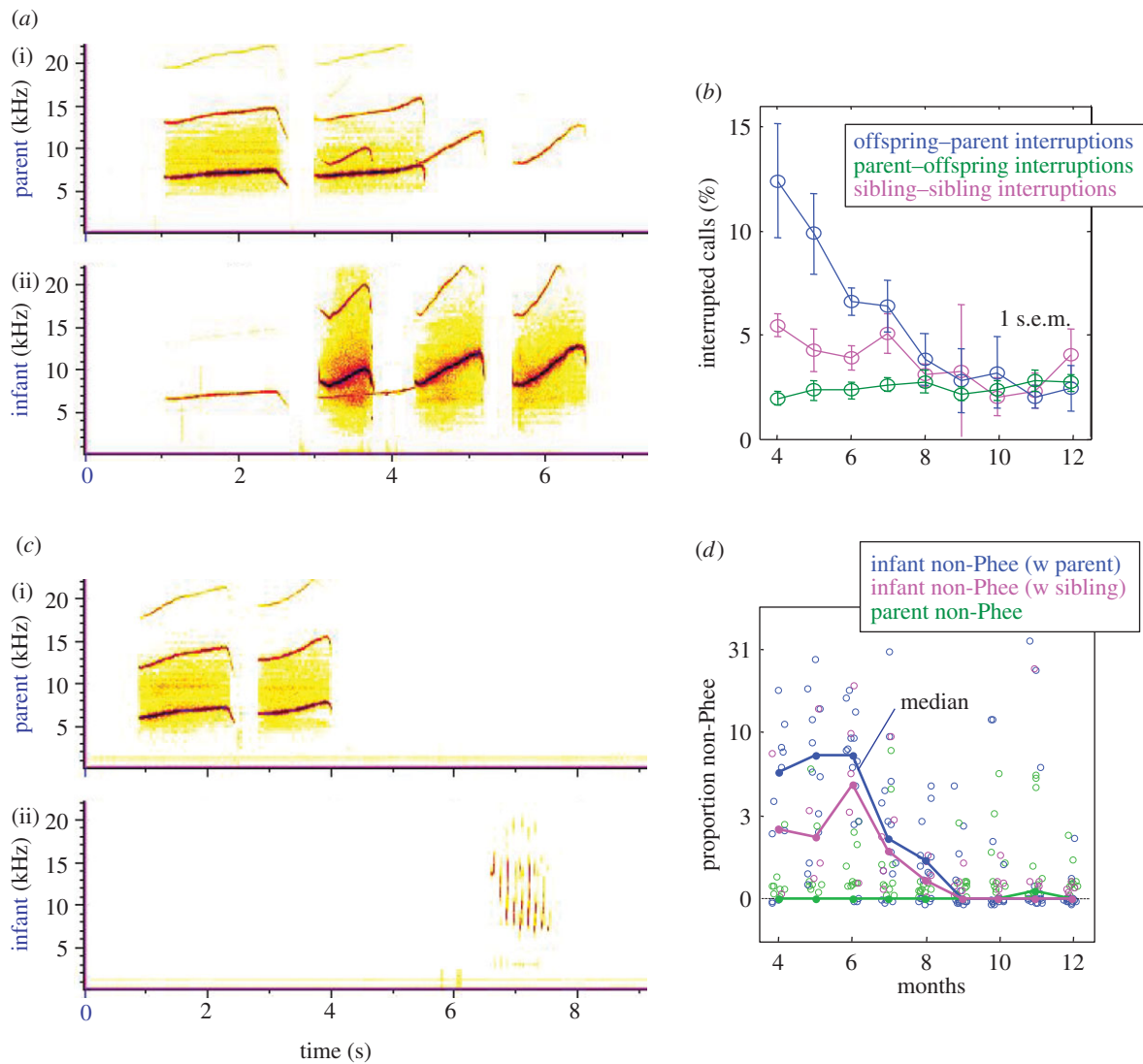


Figure 1. Development of vocal turn-taking errors. (a) A spectrogram shows an example of a vocal exchange in which a juvenile marmoset (ii) interrupts their parent's Phee call (i) by initiating its response prior to the completion of the call. (b) Plots the per cent interrupted calls during parent–offspring interactions and sibling–sibling interactions from 4 to 12 months of age (blue: offspring interruptions of parent calls; green: parent interruptions of offspring calls; pink: sibling interruptions of sibling calls). (c) A spectrogram shows an example of a vocal exchange in which a juvenile marmoset (ii) produces a non-Phee call (twitter) in response to a parent's Phee call (i). The call produced by the juvenile here is a concatenated twitter-Phee produced only in the first six months. (d) Plots the proportion of non-Phee calls produced from 4 to 12 months of age. Individual circles plot data points for each individual, while the solid line plots the median of the population (blue: offspring calls during offspring–parent interactions; green: parent during offspring–parent interactions; pink: sibling calls sibling–sibling interactions).

electronic supplementary material, figure S3). This analysis identifies the frequencies at which the reciprocal exchanges have consistent rhythmic behaviour, as indicated by higher coherence values (shown in colour scale in figure 3*a–c*). The temporal rhythm of vocal interactions with parents showed a general pattern of slowing and narrowing to a tighter, more consistent frequency range from earlier to later months (figure 3*a,b*). The most pronounced difference in vocal turn-taking between mothers and fathers was a weaker coherence for father–offspring vocal exchanges pooled over all months (Ranksum test, $p < 0.05$). There was a significant slowing of the rhythm of interactions that was reflected in the coherence analysis by the overall decrease in the peak frequency of coherence towards slower (lower frequency) periods, with significant decreases for both mother–offspring and father–offspring vocal interactions (figure 3*d*; Ranksum test, $p < 0.05$). These changes in timing did not reflect changes in the parents' spontaneous calling rhythms, which maintained a

constant, slower frequency over development as the juveniles adapted (electronic supplementary material, figure S4). The timing of interactions also became more consistent over the first year for all groups, as reflected by tighter concentration of coherence in a narrower band of frequencies around the peak and measures of the width at half height (figure 3*e*; Ranksum test, $p < 0.05$).

While vocal interactions with both mothers and fathers (light and dark green) showed slowing, those between siblings did not show significant slowing (figure 3*e*; Ranksum test, $p > 0.05$). Additionally, juvenile spontaneous calling did not exhibit significant changes over development for exchanges with their siblings, while they did adapt for their parents (electronic supplementary material, figure S4*c–e*). This suggests that the slowing with parents may be specific to adult vocal interactions and not simply reflect a general developmental trend. In other words, young marmosets learn the rhythms specific to communicating in each social context.

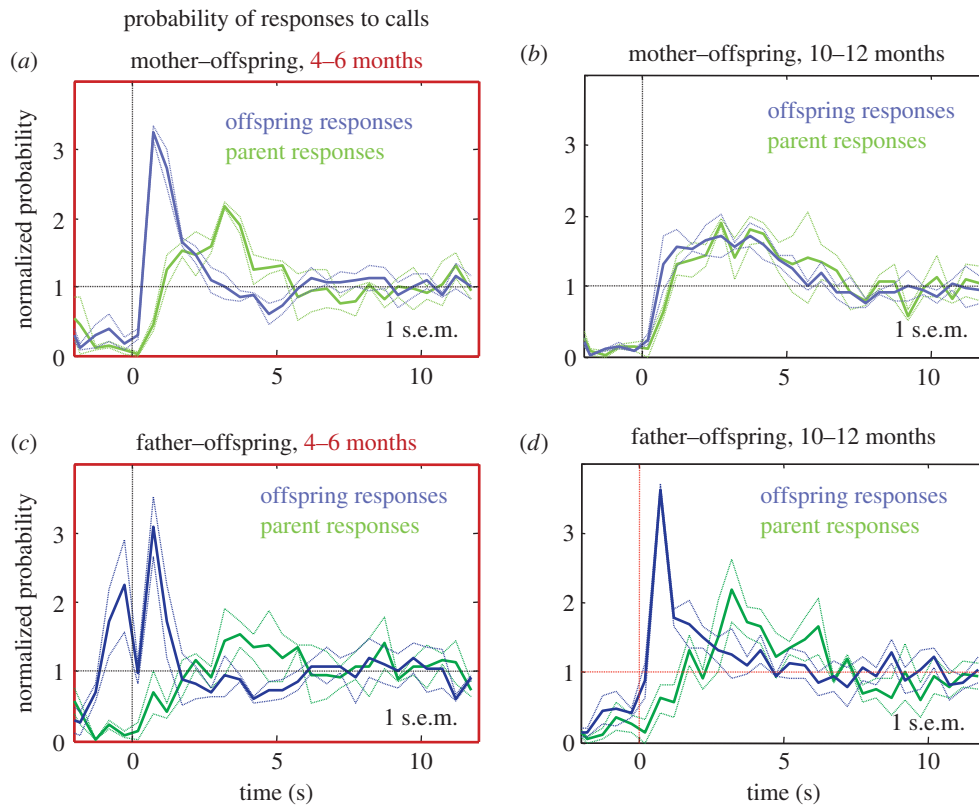


Figure 2. Parent-specific patterns of vocal turn-taking. (a–d) Plots the normalized probability of the response onset for parent–offspring interactions either early (4–6 months) or late (10–12 months) in the year (colour scheme is the same as above). (a) Mother–offspring interactions; early months. (b) Mother–offspring interactions; late months. (c) Father–offspring interactions; early months. (d) Father–offspring interactions; late months.

(b) Parents provide feedback to offspring on behaviour during vocal interactions

Similar to human parents [18,35,36], marmoset parental behaviour provides feedback signals to offspring that could guide learning. When a parent produces a vocal response to their child, it provides a potential positive reinforcement, affirming an interest in continuing the vocal exchange. The absence or delay of a response would, therefore, communicate that the behaviour of the offspring was not appropriate.

We found that parents were significantly less likely to produce a response following an interruption of their own call by their offspring, than during instances when no interruption occurred (Ranksum test, $p < 0.001$). Measuring response probability, we found that although parents' responses to uninterrupted calls were significantly above baseline within the first 4–5 s, response probability following interrupted calls returned to baseline levels only after 7 s. In fact, response probability did not rise above baseline levels during the entire response period (10 s). This effect was evident for both mothers (figure 4a; Ranksum test, $p < 0.01$) and fathers (figure 4b; Ranksum test, $p < 0.01$). Analyses indicated that interruptions effectively ended a vocal exchange. Following an interruption, the probability of producing a Phee call was consistent with what would be expected purely from spontaneous calling (black dashed line, figure 4a,b). Thus, parents effectively ignored offspring following interruptions. By contrast, uninterrupted calls were likely to elicit a response. Thus, the information necessary to reinforce appropriate turn-taking was available to infants, with appropriate responses continuing vocal interactions and interruptions being ignored.

One other error evident in infant marmoset vocal interactions with parents was the production of non-Phee calls.

Marmoset parents also appear to provide feedback during instances when their offspring produce these vocalizations. Analyses indicated that when infants produced call types inappropriate to the specific type of vocal interaction (i.e. not a Phee), parents were significantly more likely to interrupt the offsprings' calls (figure 4c; Ranksum test, $p < 0.01$). This effect was particularly compelling given that non-Phee calls were shorter in duration (mean 0.90 s) than Phee calls (mean 1.84 s; Ranksum test, $p < 0.0001$). This suggests that parents were implementing corrective measures for both of the primary errors evident in juvenile turn-taking and that this may have functioned to guide learning in this communicative context, similarly to what has been observed in human infants [18,35,36].

4. Discussion

Callitrichids (i.e. marmosets and tamarins) are among a small number of non-human primate species that, like humans, pair-bonds and engages in cooperative care of young [30]. These social characteristics have been argued to potentially underlie broader pro-social similarities between humans and marmosets that extend to communicative interactions [15]. As vocal turn-taking involves monitoring the social landscape and coordinating the timing of one's own actions based on others, it has potential to inform not only our understanding of homologies in vocal learning across primates, but the coevolution of communication and cognition [37]. Here we report evidence that the vocal turn-taking in common marmosets is learned during ontogeny and that its development is guided by parents' behaviour.

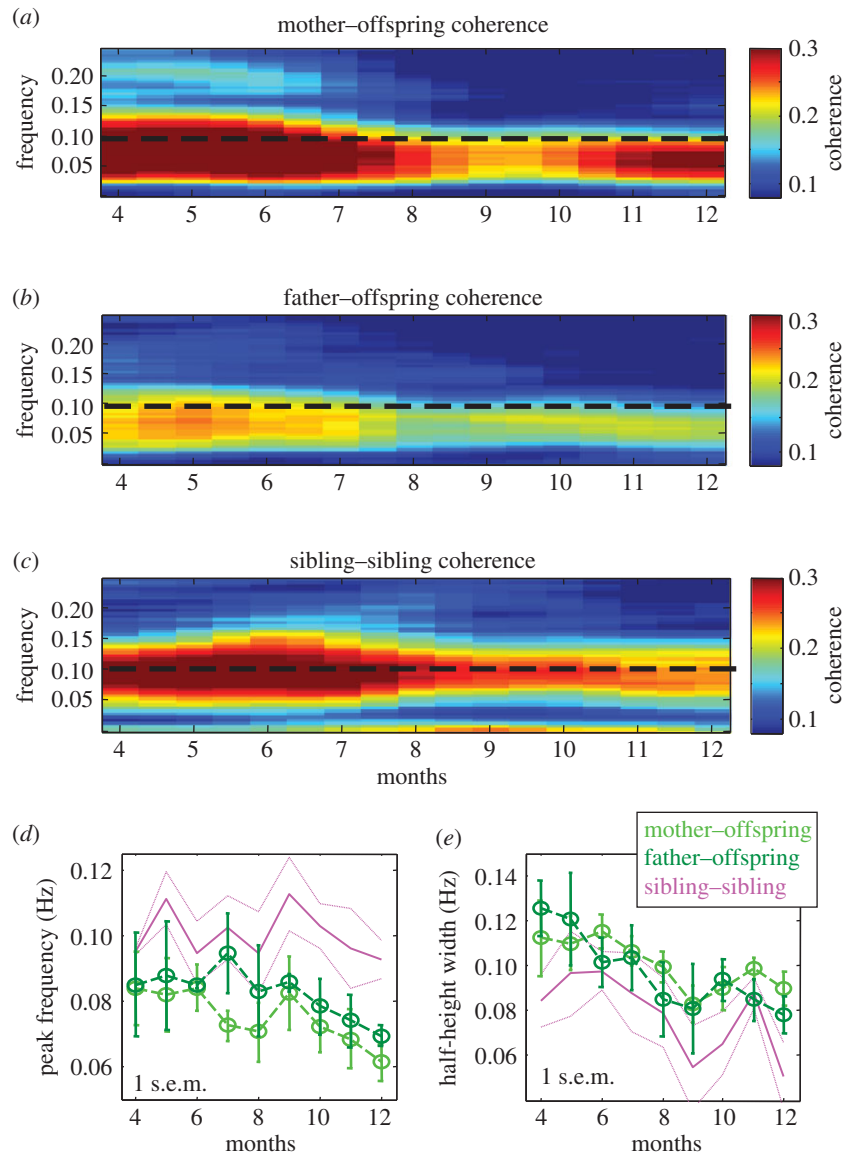


Figure 3. Patterns of vocal reciprocity. Coherence in the timing of vocal responses is plotted in colour scale across months in (a) mother–offspring, (b) father–offspring, and (c) sibling–sibling interactions. The following individual parameters of coherence are plotted from 4 to 12 months for mother–offspring (light green), father–offspring (dark green) and sibling–sibling (magenta) interactions: (d) peak frequency and (e) half height width.

Turn-taking during marmoset vocal interactions is learned during ontogeny. The isolation studies used in songbird research to experimentally test specific factors affecting vocal learning, such as tutoring [38], cannot be performed in primates [39]. Here we interpret evidence of social context-specificity in the development of two aspects of turn-taking as evidence that similar learning mechanisms underlie the development of turn-taking in marmosets. First, errors in vocal interactions (i.e. interruptions and non-Phee calls) were significantly higher in exchanges between young marmosets and their parents than with siblings (figure 1*b,d*). Second, despite only small differences in the duration of Phee calls produced by mothers and fathers, infant marmosets were significantly more likely to interrupt the latter (figure 2*a,c*). In fact, whereas juvenile marmosets adopted an adult-like temporal pattern of exchanges with mothers by 10–12 months (figure 2*b*), the same was not true for vocal interactions with fathers (figure 2*d*). Third, coherence analyses indicated that the rhythm of vocal interactions developed differently based on the social context. While vocal exchanges with mothers and fathers gradually slowed over the first year of life (figure 3*a,b,d*), the frequency

stayed relatively constant during sibling–sibling interactions (figure 3*c,d*). These patterns suggest the ontogenetic changes are not simply due to a general maturational process, but that marmosets learn the relative timing of their vocal responses based on the nuances of the specific social context.

Similar to the ontogeny of human conversational turn-taking [18,35,36], marmoset parents may play an important role in guiding learning of this communication behaviour. Marmoset parents reinforced juveniles by responding to calls that follow the species-typical temporal pattern, whereas interrupted calls elicited no response (figure 4*a,b*). Ceasing to respond following an interrupted call effectively ended the interaction, as calling returned to baseline spontaneous levels, and could have signalled to juveniles that their interruption was an error. Furthermore, when producing non-Phee calls during these vocal interactions, parents were significantly more likely to interrupt their offspring than when Phee calls were produced (figure 4*c*). This behaviour probably functioned as a signal to their offspring that non-Phee calls were not appropriate in the context of vocal exchanges, as Phee calls are the only call type produced by adults in this context [15,17]. Data

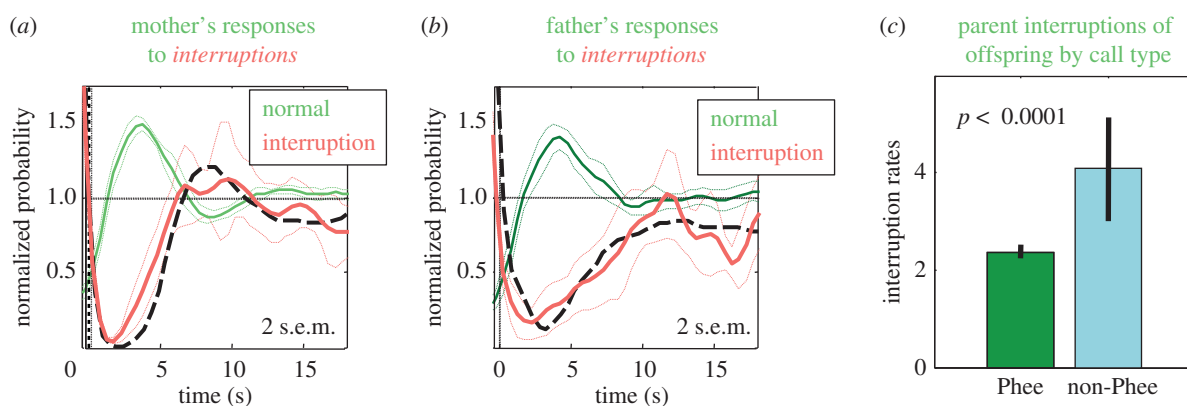


Figure 4. Vocal exchange error correction. The normalized probability of the response over time is plotted for mothers (a) and fathers (b). Green line shows the time course following an uninterrupted call, while the red line shows the time course following interrupted calls. The dashed black line shows the prediction for the response to an interruption when the parents ignore the interruption and respond with the latency expected from their spontaneous rate. (c) Plots the per cent of parents' interruptions for Phee calls and non-Phee calls produced by offspring.

presented here cannot distinguish between the possibility that parents are actively teaching their offspring or simply behaving in a manner that guides social learning. Future experiments will aim to more explicitly test these possibilities.

Communication is not a unitary process. The sensory-motor mechanisms that underlie learning the acoustic structure of words, for example, are distinct from the acquisition of various linguistic and social processes [10–12]. Yet animal models of vocal learning have typically focused on sensory-motor processes underlying changes in the acoustic structure of vocalizations over development [6], an aspect of vocal learning that is limited in non-human primates [7]. This narrow view of vocal learning has diminished the significance of social factors that affect the development of communication systems [21,40] and limited a broader understanding of parallels across primate communication system. A more compelling question may be why similarities with language are evident in communication behaviours, such as turn-taking, but the signals have only rudimentary linguistic properties [2–4]. Given the constraints of the signals themselves, it is

the communicative behaviours and the social interactions they mitigate that may provide the most substantive parallels to language. Studies of how communication behaviours weave into the fabric of primate societies, and its various sophisticated social nuances, may be key to explicating the evolutionary origins of language.

Ethics statement. All research was performed in accordance with local and state laws. These experiments were approved by the UCSD IACUC.

Data availability. The dataset has been made available in the Dryad repository (<http://doi.org/10.5061/dryad.br4rh>).

Acknowledgements. We thank the numerous undergraduate students in the UCSD Cortical Systems and Behavior Laboratory who contributed to performing these experiments over the 3 year period of the study.

Funding statement. This work was supported by NIH R01 DC012087 to C.T.M. and NIH R21 MH104756 to J.M. and C.T.M.

Author contributions. C.C. performed experiments, analysed the data and wrote the paper. J.M. analysed the data and wrote the paper. C.M. designed the experiment, analysed the data and wrote the paper.

References

- Fitch WT. 2010 *The evolution of language*. New York, NY: Cambridge University Press.
- Seyfarth RM, Cheney DL. 2010 Primate vocal communication. In *Primate neuroethology* (eds M Platt, AA Ghazanfar), pp. 84–97. New York, NY: Oxford University Press.
- Seyfarth RM, Cheney DL, Marler P. 1980 Vervet monkey alarm calls: semantic communication in a free-ranging primate. *Anim. Behav.* **28**, 1070–1094. (doi:10.1016/S0003-3472(80)80097-2)
- Arnold K, Zuberbuhler K. 2006 Semantic combinations in primate calls. *Nature* **441**, 303. (doi:10.1038/441303a)
- Janik VM, Slater PJB. 1997 Vocal learning in mammals. *Adv. Study Behav.* **26**, 59–99. (doi:10.1016/S0065-3454(08)60377-0)
- Doupe A, Kuhl P. 1999 Birdsong and human speech: common themes and mechanisms. *Ann. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
- Seyfarth RM, Cheney DL. 1997 Some general features of vocal development in nonhuman primates. In *Social influences on vocal development* (eds CT Snowdon, M Hausberger), pp. 249–273. Cambridge, UK: Cambridge University Press.
- Egnor SER, Hauser MD. 2004 A paradox in the evolution of primate vocal learning. *Trends Neurosci.* **27**, 649–654. (doi:10.1016/j.tins.2004.08.009)
- Seyfarth RM, Cheney DL. 1986 Vocal development in vervet monkeys. *Anim. Behav.* **34**, 1640–1658. (doi:10.1016/S0003-3472(86)80252-4)
- Clark EV. 2009 *First language acquisition*, 2nd edn. Cambridge, UK: Cambridge University Press.
- Bloom P. 2000 *How children learn the meaning of words*. Cambridge, UK: MIT Press.
- Jusczyk PW. 1997 *The discovery of spoken language*. Cambridge, UK: MIT Press.
- Wilson MA, Wilson TP. 2005 An oscillator model of the timing of turn taking. *Psychonomic Bull. Rev.* **12**, 957–968. (doi:10.3758/BF03206432)
- Lemasson A, Glas L, Barbu S, Lacroix A, Guilloux M, Remeuf K, Koda H. 2011 Youngsters do not pay attention to conversational rules: is this so for nonhuman primates. *Sci. Rep.* **1**, 1–4. (doi:10.1038/srep00022)
- Takahashi D, Narayanan D, Ghazanfar AA. 2013 Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* **23**, 2162–2168. (doi:10.1016/j.cub.2013.09.005)
- Symmes D, Biben M. 1988 Conversational vocal exchanges in squirrel monkeys. In *Primate vocal communication* (eds D Todt, P Goedeeking, D Symmes), pp. 123–132. Berlin, Germany: Springer.
- Miller CT, Wang X. 2006 Sensory-motor interactions modulate a primate vocal behavior: antiphonal

- calling in common marmosets. *J. Comp. Physiol. A* **192**, 27–38. (doi:10.1007/s00359-005-0043-z)
18. Snow C. 1977 The development of conversation between mothers and babies. *J. Child Lang.* **4**, 1–22. (doi:10.1017/S0305000900000453)
 19. Bruner JS. 1975 The ontogenesis of speech acts. *J. Child Lang.* **2**, 1–19. (doi:10.1017/S0305000900000866)
 20. Stivers T *et al.* 2009 Universals and cultural variation in turn-taking in conversation. *Proc. Natl Acad. Sci. USA* **106**, 10 587–10 592. (doi:10.1073/pnas.0903616106)
 21. Janik VM, Slater PJB. 2000 The different roles of social learning in vocal communication. *Anim. Behav.* **60**, 1–11. (doi:10.1006/anbe.2000.1410)
 22. Gros-Louis J, West MJ, Goldstein MH, King AJ. 2006 Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. Behav. Dev.* **30**, 509–516. (doi:10.1177/0165025406071914)
 23. Bolhuis JJ, Everaert M. 2013 *Birdsong, speech and language*. Cambridge, MA: MIT Press.
 24. Kumar S, Hedges SB. 1998 A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920. (doi:10.1038/31927)
 25. Roy S, Miller CT, Gottsch D, Wang X. 2011 Vocal control by the common marmoset in the presence of interfering noise. *J. Exp. Biol.* **214**, 3619–3629. (doi:10.1242/jeb.056101)
 26. Miller CT, Beck K, Meade B, Wang X. 2009 Antiphonal call timing in marmosets is behaviorally significant: interactive playback experiments. *J. Comp. Physiol. A* **195**, 783–789. (doi:10.1007/s00359-009-0456-1)
 27. Seyfarth RM, Cheney DL. 1999 Production, usage and response in nonhuman primate vocal learning. In *The design of animal communication* (eds MD Hauser, M Konishi), pp. 391–418. Cambridge, MA: MIT Press.
 28. Pistorio A, Vintch B, Wang X. 2006 Acoustic analyses of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* **120**, 1655–1670. (doi:10.1121/1.2225899)
 29. Miller CT, Mandel K, Wang X. 2010 The communicative content of the common marmoset Phee call during antiphonal calling. *Am. J. Primatol.* **72**, 974–980. (doi:10.1002/ajp.20854)
 30. Digby LJ, Barreto CE. 1993 Social organization in a wild population of *Callithrix jacchus*: part 1: group composition and dynamics. *Folia Primatol.* **61**, 123–134. (doi:10.1159/000156739)
 31. Yamamoto M. 1993 From dependence to sexual maturity: the behavioural ontogeny of Callitrichidae. In *Marmosets and tamarins: systematics, behavior and ecology* (ed. AB Rylands), pp. 235–254. London, UK: Oxford University Press.
 32. Bezera BM, Souto A. 2008 Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int. J. Primatol.* **29**, 671–701. (doi:10.1007/s10764-008-9250-0)
 33. Jarvis MR, Mitra B. 2001 Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Comput.* **13**, 717–749. (doi:10.1162/089976601300014312)
 34. Elias G, Hayes A, Broerse J. 1986 Maternal control of co-vocalization and inter-speaker silences in mother-infant vocal engagements. *J. Child Psychol. Psychiatry* **27**, 409–415. (doi:10.1111/j.1469-7610.1986.tb01842.x)
 35. Bloom K. 1988 Quality of adult vocalizations affects the quality of infant vocalizations. *J. Child Lang.* **15**, 469–480. (doi:10.1017/S0305000900012502)
 36. Hsu H, Fogel A. 2001 Infant vocal development in a dynamic mother-infant communication system. *Infancy* **2**, 87–109. (doi:10.1207/S15327078IN0201_6)
 37. Seyfarth RM, Cheney DL. 2014 Evolution of language from social cognition. *Curr. Opin. Neurobiol.* **28**, 5–9. (doi:10.1016/j.conb.2014.04.003)
 38. Marler P. 1970 A comparative approach to vocal learning: song development in white-crowned sparrows. *J. Comp. Physiol. Psychol. Monogr.* **71**, 1–25. (doi:10.1037/h0029144)
 39. Harlow HF, Dodsworth RO, Harlow MK. 1965 Total social isolation in monkeys. *Proc. Natl Acad. Sci. USA* **54**, 90–97. (doi:10.1073/pnas.54.1.90)
 40. Goldstein MH, King AJ, West MJ. 2003 Social interaction shapes babbling: testing parallels between birdsong and speech. *Proc. Natl Acad. Sci. USA* **100**, 8030–8035. (doi:10.1073/pnas.1332441100)