

Current Biology

Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys

Highlights

- Development of marmoset contact calls is influenced by contingent parental feedback
- Use of twin infants controlled for genetics, perinatal experience, and growth
- This is the first experimental evidence for vocal production learning in infant monkeys

Authors

Daniel Y. Takahashi, Diana A. Liao,
Asif A. Ghazanfar

Correspondence

takahashiyd@gmail.com (D.Y.T.),
asifg@princeton.edu (A.A.G.)

In Brief

Takahashi et al. show that infant marmoset monkeys are vocal learners. In brief but almost daily sessions, infant marmoset twins were experimentally provided with high or low levels of contingent parental vocal feedback to their vocalizations. More parental feedback accelerated the transition to consistently producing mature contact calls.



Vocal Learning via Social Reinforcement by Infant Marmoset Monkeys

Daniel Y. Takahashi,^{1,2,*} Diana A. Liao,¹ and Asif A. Ghazanfar^{1,2,3,4,*}

¹Princeton Neuroscience Institute

²Department of Psychology

³Department of Ecology and Evolutionary Biology

Princeton University, Princeton, NJ 08544, USA

⁴Lead Contact

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

<http://dx.doi.org/10.1016/j.cub.2017.05.004>

SUMMARY

For over half a century now, primate vocalizations have been thought to undergo little or no experience-dependent acoustic changes during development [1]. If any changes are apparent, then they are routinely (and quite reasonably) attributed to the passive consequences of growth. Indeed, previous experiments on squirrel monkeys and macaque monkeys showed that social isolation [2, 3], deafness [2], cross-fostering [4] and parental absence [5] have little or no effect on vocal development. Here, we explicitly test in marmoset monkeys—a very vocal and cooperatively breeding species [6]—whether the transformation of immature into mature contact calls by infants is influenced by contingent parental vocal feedback. Using a closed-loop design, we experimentally provided more versus less contingent vocal feedback to twin infant marmoset monkeys over their first 2 months of life, the interval during which their contact calls transform from noisy, immature calls to tonal adult-like “phee” calls [7, 8]. Infants who received more contingent feedback had a faster rate of vocal development, producing mature-sounding contact calls earlier than the other twin. The differential rate of vocal development was not linked to genetics, perinatal experience, or body growth; nor did the amount of contingency influence the overall rate of spontaneous vocal production. Thus, we provide the first experimental evidence for production-related vocal learning during the development of a nonhuman primate.

RESULTS

Marmoset monkeys (*Callithrix jacchus*) are a voluble New World species that exhibit a complex system of vocal communication [9]. This system includes vocal turn-taking: two adult marmosets (related or unrelated to each other and out of sight from one another) will exchange extended, phase-locked sequences of contingent phee calls, a type of contact call, acting

in essence like coupled oscillators [10]. Developmentally, infant marmoset monkeys produce long bouts of vocalizations consisting of both immature- and mature-sounding calls [7, 8, 11, 12]. A subset of these calls are immature versions of the phee call [7, 8], and the timing at which these immature calls transform into mature-sounding calls varies widely among infants [7].

One hypothesis to account for this developmental variation is differential feedback from parents. In this scenario, the contingent contact call exchanges that adults use with each other could also be used by parents as they respond to infant calls. This could act as a ratchet for contact call development. For example, studies of naturalistic human infant-parent interactions [13–16], as well as experimental studies [17, 18], reveal that contingent parental responses accelerate the development of infant vocalizations, making them sound more mature (i.e., speech-like). Thus, perhaps marmoset parents that produce contingent vocal responses to infant vocalizations similarly accelerate their development. Indeed, in marmoset monkeys, there is a strong correlation between the amount of contingent vocal feedback from parents and the maturation rate of contact calls [7]. There is no correlation, however, between the overall amount of exposure to parental vocalizations and vocal development [7]. This has led to the suggestion that developing marmoset monkeys—unlike every other nonhuman primate investigated thus far—may be vocal learners [19]. A viable alternative hypothesis is that, instead of an instance of vocal learning, marmoset parents are simply responding more to healthier infants who develop their vocalizations more quickly than others.

We designed an experiment to explicitly test whether or not contingent vocal feedback can increase the rate at which marmoset infants begin producing mature-sounding contact calls. Because marmoset monkeys typically give birth to dizygotic twins [20], we could control for the influence of genetics and the perinatal environment on vocal development [8]. We tested three pairs of twins (six infants) from three different sets of parents. Starting at postnatal day 1 (P1), infants were briefly removed from their home cage and provided different levels of contingent feedback using closed-loop, computer-driven playbacks of parental phee calls (Figures 1A and 1B). One randomly selected twin was provided the best-possible simulated “parent” who provided 100% vocal feedback at an ~1 s delay if the infant produced a low-entropy contact call, i.e., a more mature-sounding call; the other infant was provided a not-so-good “parent” and received vocal feedback for only 10% of

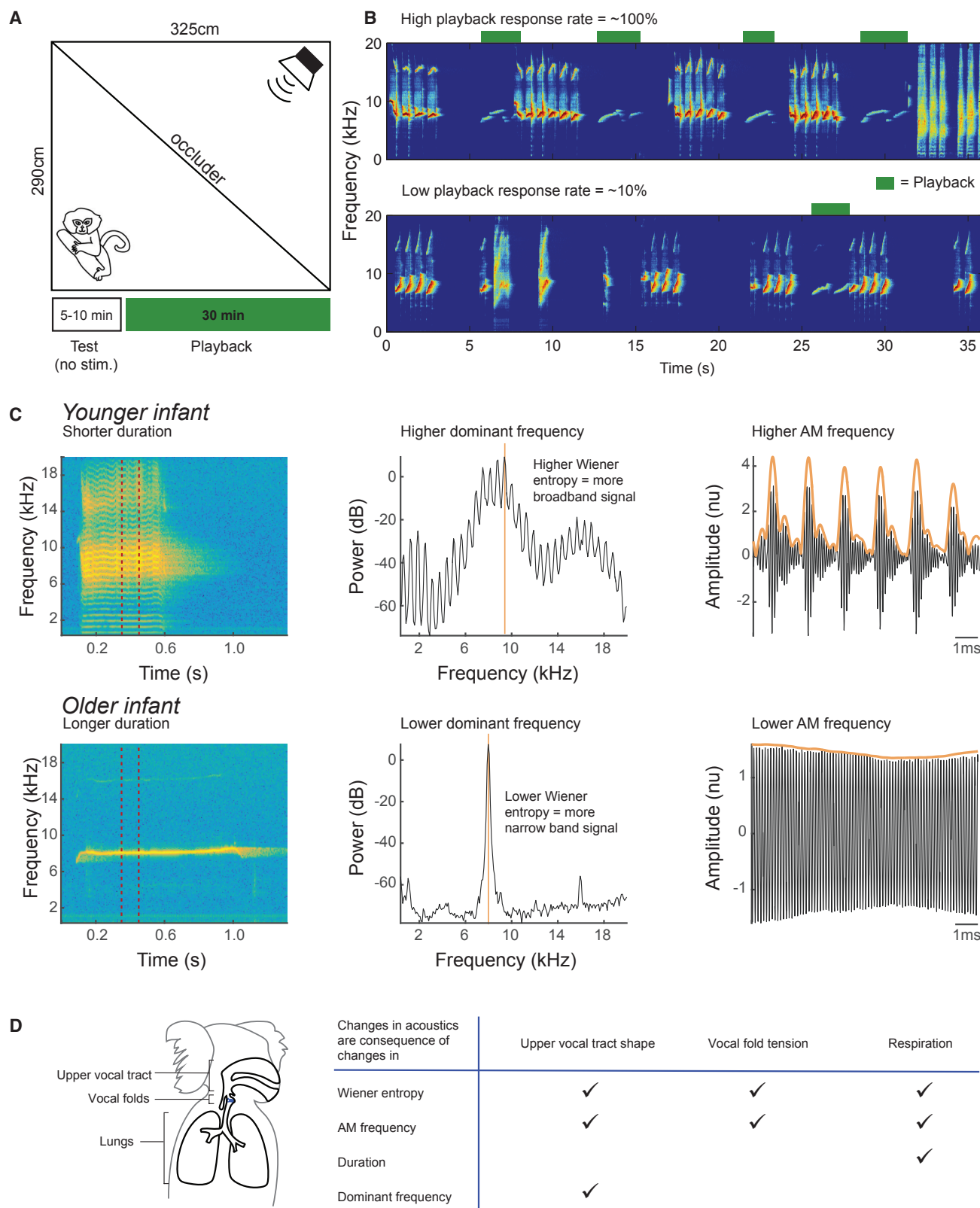


Figure 1. Experimental Design and Potential Acoustic Parameters

(A) Infants were briefly separated from their parents and placed in an acoustically treated testing room. Computer-controlled playbacks were delivered through a speaker. Sessions lasted ~40 min, with the first 5 min (postnatal days 1 to 7) or 10 min used to collect spontaneous vocalizations.

(legend continued on next page)

the low-entropy contact calls it produced (Figure 1B). The use of an ~ 1 s delay is based on data collected under naturalistic conditions showing that parents usually provide vocal responses at around 1 s after an infant vocalization (Figure S1) [21]. The high-versus low-contingent response rates are respectively higher and lower than the average parental contingent response rate during naturalistic infant-parent interactions ($21.35\% \pm 0.17\%$, mean \pm SEM) [7]. Each experimental session lasted 40 min; the first 10 min (5 min in the first week, when the infants were neonates) was used to record the infants' spontaneous vocalizations, and the remaining 30 min was used for playback; the infants were otherwise with their families for the remaining 23 hr and 20 min of each day. The use of maternal versus paternal contact calls was counter-balanced on a session-by-session basis. We used 20 call exemplars from each parent. These conditioning sessions occurred almost every day for 2 months (~ 14 consecutive experiment days + 1 rest day, iterated four times).

We used multiple acoustic measures to quantify developmental changes as a continuous process [7, 22] (Figure 1C). We did this for two reasons. First, it allowed us to measure change without the bias of ethological classifications. Second, it allowed us to see whether some acoustic parameters versus others were more sensitive to contingent parental vocal feedback. The different acoustic parameters can be related to different biological mechanisms or their combination (Figure 1D). For example, an optimal control-based Waddington landscape model of marmoset vocal development revealed that changes in dominant frequency could be completely accounted for by the growth of the vocal tract [23]. The four acoustic measures that we used were duration, dominant frequency, amplitude modulation (AM) frequency, and Wiener entropy, all of which change over the course of development (Figure 1C) [7, 22]: duration increases, dominant frequency decreases, AM frequency decreases, and Wiener entropy decreases as the infant marmosets get older [7].

Because marmoset infants will immediately change (on the order of seconds) their call acoustics to be more mature upon hearing a contingent parental response [21], our acoustic measures were only performed on the spontaneous calls produced by infants during the 5 or 10 min interval at the beginning of every session (Figure 1A). We used multiple linear regression to fit the data with the relevant acoustic measure as the dependent variable and with postnatal day ($n = 193$ days, six infants, 30–33 days/infant), contingency group (high versus low), and twin identity (1, 2, or 3) as the predictors. The inclusion of twin identity allowed us to control for the effect of genetics in the development. All one-way and two-way interactions were included to account for relevant effects. All p values reported below are calculated from the test of nullity of the interaction between post-

natal day and contingency group; we set our alpha level at 0.01. We also report the adjusted R^2 (adj. R^2) of the regression model. The coefficients of the main regression models are reported in the Supplementary Information. Because we wanted to capture the rate of vocal change up until the point at which the infant marmosets produce only mature-sounding contact calls ($>95\%$ phee calls), the regression analyses were done on ages P1 to P35. For all infants ($n = 6$), mature-sounding phee calls were produced almost exclusively after P36, and no effect of contingency group was observed for any of the four acoustic measures after P36 (Figure S2).

We present the Wiener entropy data first, because this measure effectively captures the transformation of noisy, immature (high-entropy) contact calls into tonal, adult-like (low-entropy) calls [7, 8, 21]. For each twin pair, the individual that received more contingent feedback had a faster rate of vocal development, producing mature-sounding (lower-entropy) calls earlier than the other twin. Figure 2A shows that the timing of the transition from immature to mature calls was quicker for the infants that received more contingent feedback (adj. $R^2 = 0.519$, $p = 0.0022$). Figure 2B shows that this pattern held true for each pair: the individual that received more contingent feedback had a steeper rate of vocal development, producing lower-entropy contact calls more quickly than the other twin. Measures of AM frequency revealed a similar pattern. Figure 2C shows that the development of this acoustic parameter was also faster in infants receiving high- versus low-contingent vocal feedback (adj. $R^2 = 0.490$, $p = 0.0068$). Again, this pattern held true for each pair, whereby the individual that received more contingent feedback developed the mature AM frequency more quickly when compared to the other twin (Figure 2D). It is possible that part of the differences in the Wiener entropy and AM frequency development are due to initial differences in vocal behavior exhibited by the infants immediately after birth. To verify this possibility, we tested whether the intercepts of the regression models were different between contingency groups and found no evidence for this (test of nullity for the mean effect of contingency group, Wiener entropy: $p = 0.7851$; AM frequency: $p = 0.0715$).

Because the twins were not identical in their growth rates (Figure 2E), one possibility is that growth accounts for vocal developmental differences. Body weight is a good proxy for overall growth, and weight correlates well with vocal apparatus size in monkeys [24]. We therefore added body weight and its interaction with postnatal day as predictors. The result revealed that differences in weight cannot explain the differential development of the Wiener entropy or AM frequency changes as a function of high versus low contingency (Wiener entropy: adj. $R^2 = 0.494$, $p = 0.0002$; AM frequency: adj. $R^2 = 0.568$, $p = 0.0056$).

(B) Twin infants received either high-contingency playbacks (100%) or low-contingency playbacks (10%). Spectrograms depict when such playbacks (green) were delivered relative to the infant vocalizations. Warmer colors indicate higher values.

(C) Four acoustic parameters change over the course of marmoset vocal development and were measured in the study: duration, dominant frequency, Wiener entropy, and amplitude modulation (AM) frequency. Vertical red dashed lines in spectrograms indicate the time interval used for the analyses in the neighboring panels.

(D) Four acoustic parameters are related to different operations of the vocal apparatus (vocal tract, vocal folds, and lungs/respiration). Wiener entropy and AM frequency changes are associated with changes in the shape of upper vocal tract, vocal fold tension, and respiratory control. Change in dominant frequency is associated with size of upper vocal tract. Change in duration is associated with change in lung capacity and respiratory control.

See also Figure S1.

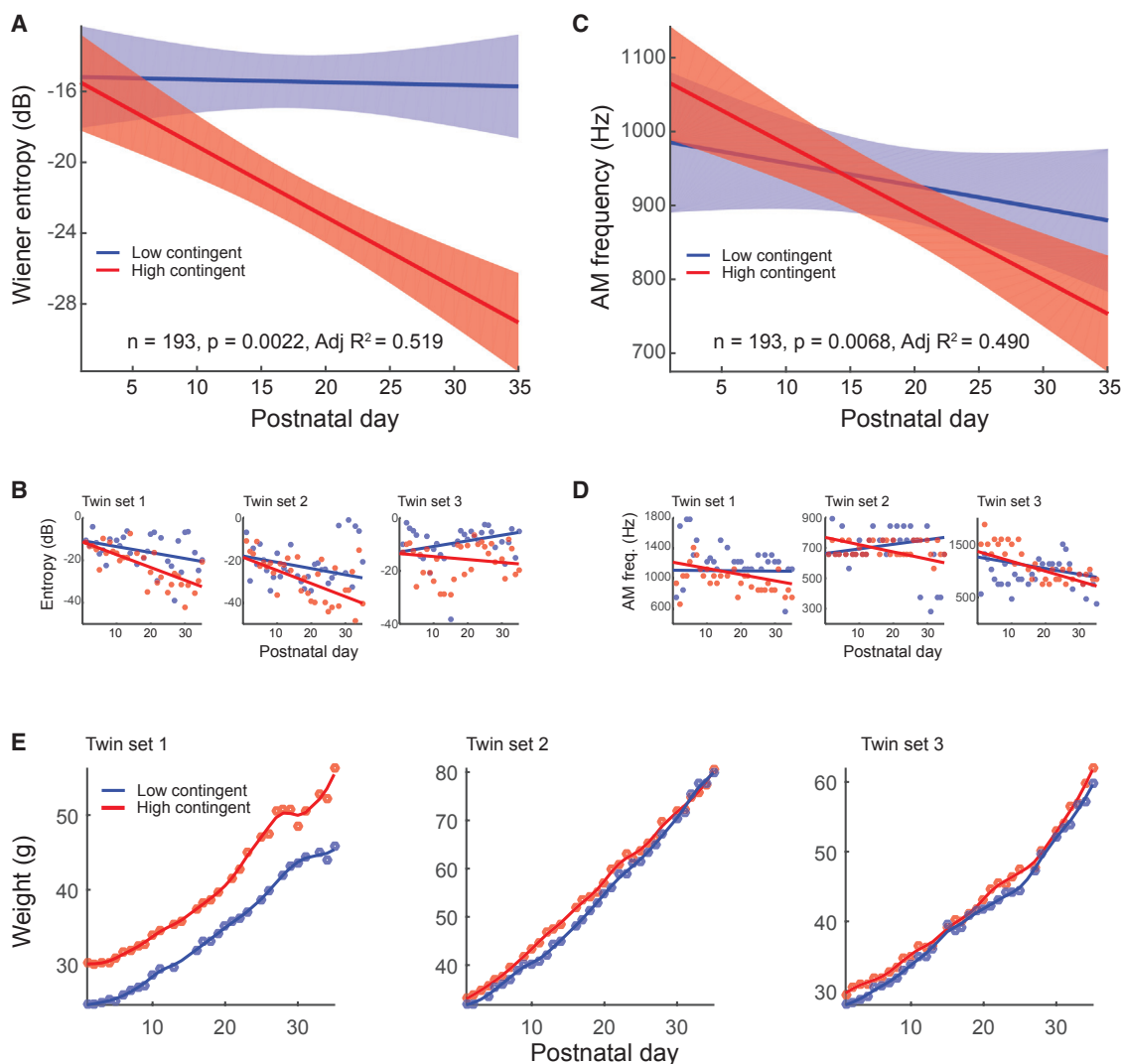


Figure 2. Infants Receiving More Contingent Vocal Feedback Develop Their Vocalizations Faster, and This Change Is Not Related to Growth Differences

(A and B) Wiener entropy (in decibels) changes over postnatal days for high- and low-contingency infants. (A) shows group average; shaded regions indicate 1 SE intervals. (B) shows data for each twin set.

(C and D) AM frequency (in Hz) changes over postnatal days for high- and low-contingency infants. (C) shows group average; (D) shows data for each twin set. (E) Growth of all infants as measured by weight (in g) over postnatal days.

See also [Figures S2](#) and [S3](#).

In contrast to Wiener entropy and AM frequency, [Figures 3A](#) and [3C](#) show that developmental changes in call duration and dominant frequency were not influenced by the amount of contingent feedback (duration: adj. $R^2 = 0.507$, $p = 0.0527$; dominant frequency: adj. $R^2 = 0.557$, $p = 0.6278$). We also didn't find differences in the intercepts between contingency groups (duration: $p = 0.2214$; dominant frequency: $p = 0.0625$). These patterns held true for each twin set ([Figures 3B](#) and [3D](#)). The same result holds when body weight and its interaction with postnatal day are included as predictors in the regression (duration: adj. $R^2 = 0.469$, $p = 0.2697$; dominant frequency: adj. $R^2 = 0.546$, $p = 0.3008$).

Because the different contingency conditions necessarily result in one infant being provided with more parental calls

than the other ([Figure S3](#)), another possibility is that the total amount of playback calls—and not their contingency—can explain the learning effect on the rate of call development. This possibility is not supported by our previous study, which showed that there was a strong correlation between the timing of contact call development and contingency but no correlation with the much greater number of non-contingent parental calls heard by the infant [7]. Nevertheless, we directly tested this hypothesis by including the total number of playback call responses and its interaction with postnatal day as predictors in the regression. The regression result shows that even after controlling for the total number of playback stimuli, the contingency-driven vocal developmental changes remain statistically significant (Wiener entropy: adj. $R^2 = 0.506$, $p = 0.0038$; AM

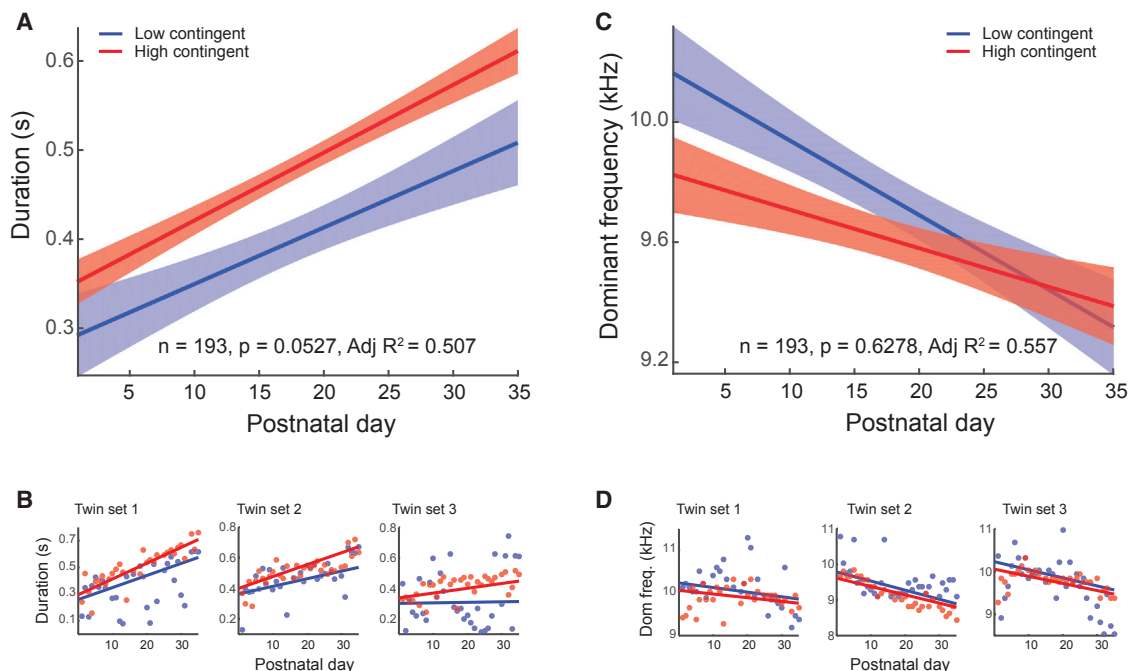


Figure 3. Call Duration and Dominant Frequency Are Not Influenced by the Amount of Vocal Feedback

(A and B) Duration (in s) changes over postnatal days for high- and low-contingency infants. (A) shows group average; shaded regions indicate 1 SE intervals. (B) shows data for each twin set.

(C and D) Dominant frequency (in kHz) changes over postnatal days for high- and low-contingency infants. (C) shows group average; (D) shows data for each twin set.

See also [Figures S2](#) and [S3](#).

frequency: adj. $R^2 = 0.509$, $p = 0.0040$), while the other acoustic parameters remain uninfluenced by the differential feedback (duration: adj. $R^2 = 0.513$, $p = 0.0505$; dominant frequency: adj. $R^2 = 0.544$, $p = 0.3293$).

Given that the difference in the amount of contingency feedback resulted in differential rates of vocal development, we wondered if infants with a faster rate produce more calls in general. A change in the rate of call production could indicate a long-term change in the animal's baseline state, e.g., stress levels [25]. To test this, we fitted a multiple linear regression in which the rate of infant call production (number of calls per minute for each baseline period) was the dependent variable and the predictors were the postnatal day, contingency group, twin set, and all one- and two-way interactions. There was no significant effect of contingency group ($n = 193$, adj. $R^2 = 0.538$, $p = 0.2306$) and no significant interaction between contingency group and postnatal day ($p = 0.6664$) in the first 35 days of postnatal life ([Figure 4A](#)). The same pattern was true later in the second month of life, when infants are producing phee calls exclusively (postnatal days 36 to 60; $n = 130$, adj. $R^2 = 0.145$, contingency group, $p = 0.0919$; interaction between contingency group and postnatal day, $p = 0.7599$) ([Figure 4B](#)). Thus, contingent feedback from parents does not influence the overall rate of vocal production and cannot be linked to long-term differences in the state of the animal.

Our sample size was small for each contingency group ($n = 3$). Thus, it is important to estimate the statistical power of our results as a measure of their robustness. If the power is low, it

can lead to (1) a low probability of finding a true effect, (2) an overestimation of the effect size when the true effect is discovered, and (3) a low positive predictive value (PPV) when an effect is claimed. Naturally, one major source of low power is a small sample size, but sources of high statistical power include experimental designs that control for different confounds like genetics, stimuli, and context (as in the current study) [26]. We calculated the effect size (Cohen's f^2), performed a power analysis, and estimated the PPV of our results. For the regression model with PND, Group, TwinId, and their two-way interactions, the Cohen's f^2 values (with 90% CI) were 0.1841 ([0.0871, 0.3002]), 0.0115 ([0, 0.0365]), 0.1093 ([0.0357, 0.1942]), and 0.0400 ([0, 0.0883]), respectively, for Wiener entropy, AM frequency, duration, and dominant frequency. For $\alpha = 0.01$ (type I error), these effect sizes imply a post hoc achieved power of 0.9985 (Wiener entropy), 0.0929 (AM frequency), 0.9491 (duration), and 0.4546 (dominant frequency).

We can now estimate the positive prediction value as $PPV = ([1 - \beta] \times R) / ([1 - \beta] \times R + \alpha)$, where $(1 - \beta)$ is the power, α is the type I error control, and R is the odds that a claimed effect is indeed non-null among the effects being claimed [26]. Calculating the pre-study odds R is usually difficult, but we can get an estimate using the literature. Among perturbation studies of non-human primate vocal learning, one study observed a strong effect of social contact [27], and three studies did not [2–4]. Therefore, we can assume a pre-study odds equal to 0.25 and conclude that the PPV for each acoustic parameter is 0.9615 (Wiener entropy), 0.6990 (AM frequency), 0.9596 (duration),

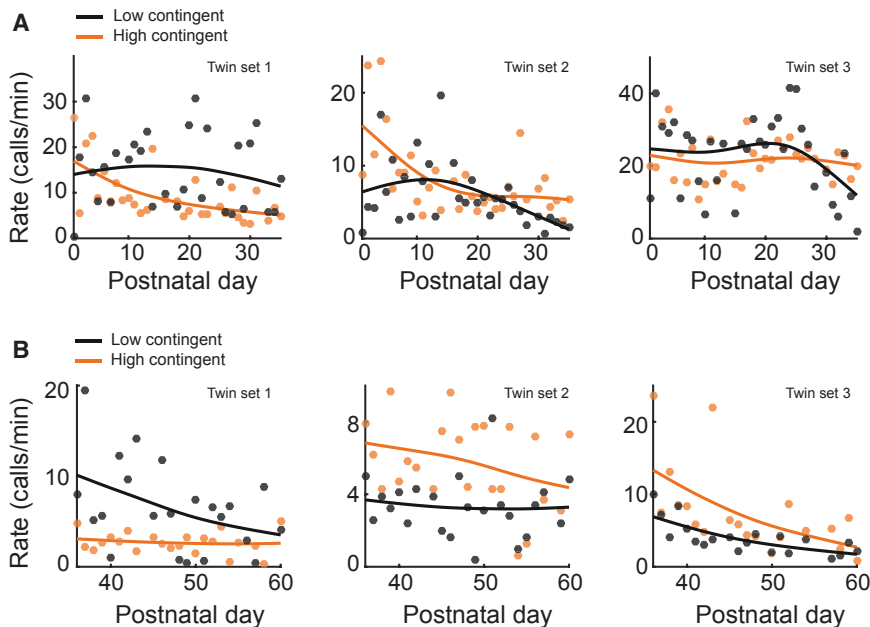


Figure 4. Development of the Rate of Call Production Measured during the Baseline Interval

(A) Rate of call production from postnatal day 1 to postnatal day 35 for each infant. Orange and black lines are cubic spline curves fitted to high- and low-contingency data, respectively.

(B) Rate of call production from postnatal day 36 to postnatal day 60 for each infant. Orange and black lines are cubic spline curves fitted to high- and low-contingency data, respectively.

and 0.9191 (dominant frequency). We conclude that (1) the positive finding for Wiener entropy is highly robust, whereas the finding for AM frequency is less robust; (2) there is not sufficient evidence for the effect of contingency group on duration, but there might be a smaller effect; and (3) there is no evidence for the effect of contingency group on dominant frequency.

Overall, our finding that contingent feedback changes the rate of infant vocal development is statistically robust and is consistent with our previous correlation-based evidence [7].

DISCUSSION

It is generally accepted that there are three varieties of vocal learning: comprehension, usage, and production [28]. Comprehension learning is when an animal learns to respond appropriately—via experience—to vocal signals. For example, infant vervet monkeys learn adaptive responses to alarm calls by watching what adult conspecifics do [29], infant macaques learn via experience to recognize their mother's voice [30], and Diana monkeys learn to respond adaptively to the alarm calls of other species [31]. Usage learning is when an animal learns in which context(s) to produce a pre-existing call in its repertoire. For instance, infant and juvenile vervet monkeys produce adult-like “raptor” alarm calls but do so to the wrong birds; they eventually learn to associate their alarm calling to the bird species [29]. Similarly, infant marmoset monkeys produce some call types in inappropriate contexts [32], and learning the appropriate context is experience dependent [27]. Another instance of usage learning is the vocal conditioning of captive primates: macaques, for example, can be trained to associate their vocalizations with food or context cues [33, 34] and even to arbitrary visual cues [35]. The third variety of vocal learning—production learning—is the experience-dependent change in the acoustic structure of vocalizations [28]. Increasingly, however, the literature has limited the definition of vocal production learning to learning novel vocalizations via imitation (e.g., [36]). Indeed, some have

limited it even further to include only imitation of those vocalizations that involve changes in the laryngeal/syringeal control [37].

Our data show that infant marmoset monkeys exhibit vocal production learning via social reinforcement from parents. They do this not through imitation, but rather through the experience-dependent increase in the control of the vocal

apparatus that allows them to more consistently produce tonal (low-entropy) phee calls. Early in life, infant marmosets produce cries, phee-cries, and subharmonic phees that are immature versions of the phee call (which they also produce, but inconsistently), these immature versions are higher in spectral entropy, dominant frequency, and amplitude modulation frequency, and shorter in duration [7, 8, 21, 23] (Figure 1C). The production of these immature contact calls is related to poor muscular control of—and coordination between—respiration and vocal fold tension [7, 8, 23]. These immature phee calls disappear later in development; they are not produced in any other contexts. The developmental timing of this control and coordination of vocal apparatus elements is what is linked to experience: more contingent vocal feedback results in faster development of respiratory and laryngeal control and coordination [23]. These data support the notion that forms of vocal production learning extend beyond imitation [38, 39].

Marmoset monkeys prove to be exceptional nonhuman primates with regard to vocal learning. Not only do they undergo dramatic changes in the acoustic structure of their vocalizations during development [7, 8], but two recent studies also showed the possibility of vocal learning via experience with parents in this species. In the first study, there was a strong correlation between contingent parental feedback and the rate at which immature versions of the contact phee call transformed into mature versions of the call [7]. In the second study, infant marmoset triplets with limited parental contact (one infant was hand-raised and other two were separated after 3 months) had their vocal behavior as sub-adults compared to their normally reared twin siblings from another pregnancy [27]. The vocal behavior of the triplets with limited parental contact remained immature relative to their normally reared twin siblings by multiple measures, demonstrating that some aspect of parental care is essential for normal vocal development in marmoset monkeys [27]. The present study experimentally demonstrated the specific role of parental vocal feedback in influencing vocal production development. By using

twins, we could reduce the potential confounds of genetics and perinatal experience [8]. Moreover, our analyses showed that neither growth nor the number of calls provided during the play-back sessions influenced the rate of learning. The overall spontaneous call rate of each infant was not influenced by contingency, either. This shows that social reinforcement is acting on a vocal motor learning circuit and not simply changing baseline levels of a behavioral state (e.g., arousal or stress levels).

The social reinforcement-based vocal learning by infant marmoset monkeys is consistent with findings in experimental studies of early (prelinguistic) vocal development in humans [17, 18] and song learning in birds (e.g., cowbirds [40] and zebra finches [41]). In human infants, for instance, immature vocalizations (e.g., cries and cooing sounds, among many others) gradually transform into the consistent, context-dependent production of speech-like babbling output [42, 43]. The speed of this transformation is influenced by contingent parental feedback [13–18]. Another similar development pattern occurs during locomotion learning in humans. Human toddlers alternate between crawling and walking, and only later in development will start to walk consistently with a more adult-like gait [44, 45]. This locomotor transition is also influenced by parental social feedback [46]. Similarly, we found that marmoset infants that received greater contingent vocal feedback than their twin began to consistently produce mature-sounding phee calls earlier in life. Our experimental design used a brief experimental period (35–40 min); for the remaining 23+ hr of the day, the infants were back with their families. Yet this brief experimental manipulation period had a significant effect on infant vocal development. This suggests that the cumulative effect of a socially relevant learning cue (contingency) and repetition can strongly influence the rate of vocal learning.

Given that brief but consistent socially relevant contingent feedback can have a large impact on shaping infant vocal development in humans [17, 18], birds [40, 41], and marmosets (current study), do these animals share a similar learning substrate? A shared neural substrate may involve links between the limbic system and cortico-striatal systems [47]. These two systems are common across vertebrate species. Contingent social responses drive the learning via the cortico-striatal system, and the limbic system (which subsumes the evolutionarily conserved social behavior network [48–50]) drives the motivation to attend to social signals. Vocal learning by infant marmoset monkeys is entirely consistent with this “social motivation-vocal learning” framework [47]. Developing marmoset monkeys also exhibit a pattern of *FoxP2* expression in their cortico-striatal circuit [51] that is analogous to that of songbirds and humans [52], lending credence to the notion that marmoset infants use this circuit to guide their phee call development through social reinforcement, just as birds and humans do. The fact that changes induced by social reinforcement in marmoset infants did not affect all acoustic parameters equally should help investigations of the underlying vocal learning mechanism. The high-contingency condition sped up changes in Wiener entropy and AM frequency but did not influence changes in duration and dominant frequency. In marmoset monkeys, changes in dominant frequency occur simply by a growth-related increase in the length of the vocal tract [23]. Similarly, changes in contact call duration are likely the result of a growth-related increase in lung capacity. By contrast,

our systems-level vocal development landscape framework revealed that the production of tonal versus noisy contact calls (as measured by Wiener entropy) is linked to a learning-related increase in musculature control of the larynx and respiration, as well as the coordination of that control across these two elements of the vocal apparatus [23].

Why do humans and marmoset monkeys exhibit similar patterns of vocal development in the early postnatal period? Because some 40 million years have passed since the Old World and New World primate lineages split [53], we suggest that vocal learning arose as a by-product of the convergent evolution of a cooperative breeding system. Cooperative breeding is only found in about 3% of mammals [54]. Of those mammals, marmoset monkeys and others in the callitrichid family are the only nonhuman primates known to exhibit this strategy [54, 55]. For marmosets, the rearing of infants is greatly reliant on a concerted effort among the breeding female, breeding male, non-breeding siblings, and other familiar but unrelated group members. Marmoset caregivers compete to carry infants [56, 57] and frequently provision food to offspring [58, 59]. This cooperative breeding framework, in which non-parents within a social group spontaneously care for offspring other than their own, drives a more general tendency to be prosocial in other domains [60], including vocal communication [6].

How does this lead to vocal production learning in infant marmosets and humans? Care of infants is probably the most important context in which cooperation with unrelated individuals occurs. There is a strong correlation between the amount of infant care provided by others and the reproductive success of a mother [61]. In an environment where caregiver attention is a limited resource and where non-maternal caregivers may have a higher threshold than mothers to provide care, evolution may select for vocal behaviors that help infants attract caregiver attention [62]. A related hypothesis is that human infant vocalizations that sound more speech-like evolved to exploit pre-existing auditory predispositions in adult receivers [63]. The fact that parents of both human and marmoset infants are more likely to give contingent responses to infant vocalizations when those vocalizations sound more adult-like [21, 64] is consistent with this “receiver predisposition” idea [63]. We therefore suggest that the vocal learning mechanism evolved to speed up the production of mature-sounding vocalizations (those that exploit the receiver predispositions) using social feedback because such vocalizations are more likely to elicit caregiver attention.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Marmosets
- METHOD DETAILS
 - Experimental setup
 - Closed loop playbacks
 - Detection of calls and quantification of acoustic parameters

- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Multiple linear regression analysis
 - Linear regression models
 - Effect size and power analysis
- **DATA AND SOFTWARE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cub.2017.05.004>.

AUTHOR CONTRIBUTIONS

D.Y.T.: conceptualization, data collection and curation, data analysis, writing original draft, review and editing. D.A.L.: data analysis, draft review and editing. A.A.G.: conceptualization, supervision, writing original draft, review and editing, funding acquisition.

ACKNOWLEDGMENTS

This work was supported by a Scholar Award from the James S. McDonnell Foundation (A.A.G.) and a National Science Foundation Graduate Research Fellowship (D.A.L.).

Received: February 25, 2017

Revised: April 3, 2017

Accepted: May 3, 2017

Published: May 25, 2017

REFERENCES

1. Egnor, S.E.R., and Hauser, M.D. (2004). A paradox in the evolution of primate vocal learning. *Trends Neurosci.* 27, 649–654.
2. Hammerschmidt, K., Freudenstein, T., and Jürgens, U. (2001). Vocal development in squirrel monkeys. *Behaviour* 138, 1179–1204.
3. Winter, P., Handley, P., Ploog, D., and Schott, D. (1973). Ontogeny of squirrel monkey calls under normal conditions and under acoustic isolation. *Behaviour* 47, 230–239.
4. Owren, M.J., Dieter, J.A., Seyfarth, R.M., and Cheney, D.L. (1992). 'Food' calls produced by adult female Rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques, their normally-raised offspring, and offspring cross-fostered between species. *Behaviour* 120, 218–231.
5. Hammerschmidt, K., Newman, J.D., Champoux, M., and Suomi, S.J. (2000). Changes in rhesus macaque 'coo' vocalizations during early development. *Ethology* 106, 873–886.
6. Borjon, J.I., and Ghazanfar, A.A. (2014). Convergent evolution of vocal cooperation without convergent evolution of brain size. *Brain Behav. Evol.* 84, 93–102.
7. Takahashi, D.Y., Fenley, A.R., Teramoto, Y., Narayanan, D.Z., Borjon, J.I., Holmes, P., and Ghazanfar, A.A. (2015). The developmental dynamics of marmoset monkey vocal production. *Science* 349, 734–738.
8. Zhang, Y.S., and Ghazanfar, A.A. (2016). Perinatally influenced autonomic nervous system fluctuations drive infant vocal sequences. *Curr. Biol.* 26, 1249–1260.
9. Bezerra, B.M., and Souto, A. (2008). Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int. J. Primatol.* 29, 671–701.
10. Takahashi, D.Y., Narayanan, D.Z., and Ghazanfar, A.A. (2013). Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr. Biol.* 23, 2162–2168.
11. Elowson, A.M., Snowdon, C.T., and Lazaro-Perea, C. (1998). Infant 'babbling' in a non-human primate: complex vocal sequences with repeated call types. *Behaviour* 135, 643–664.
12. Pistorio, A.L., Vintch, B., and Wang, X. (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *J. Acoust. Soc. Am.* 120, 1655–1670.
13. Bloom, K., Russell, A., and Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *J. Child Lang.* 14, 211–227.
14. Gros-Louis, J., West, M.J., and King, A.P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy* 19, 385–408.
15. Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to four-month-old Japanese infants. *J. Child Lang.* 20, 303–312.
16. Hsu, H.-C., and Fogel, A. (2001). Infant Vocal Development in a Dynamic Mother-Infant Communication System. *Infancy* 2, 87–109.
17. Goldstein, M.H., King, A.P., and West, M.J. (2003). Social interaction shapes babbling: testing parallels between birdsong and speech. *Proc. Natl. Acad. Sci. USA* 100, 8030–8035.
18. Goldstein, M.H., and Schwade, J.A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol. Sci.* 19, 515–523.
19. Margoliash, D., and Tchernichovski, O. (2015). Marmoset kids actually listen. *Science* 349, 688–689.
20. Harris, R.A., Tardif, S.D., Vinar, T., Wildman, D.E., Rutherford, J.N., Rogers, J., Worley, K.C., and Agaard, K.M. (2014). Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proc. Natl. Acad. Sci. USA* 111, 1467–1472.
21. Takahashi, D.Y., Fenley, A.R., and Ghazanfar, A.A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371, 20150370.
22. Tchernichovski, O., Mitra, P.P., Lints, T., and Nottebohm, F. (2001). Dynamics of the vocal imitation process: how a zebra finch learns its song. *Science* 291, 2564–2569.
23. Teramoto, Y., Takahashi, D.Y., Holmes, P., and Ghazanfar, A.A. (2017). Vocal development in a Waddington landscape. *eLife* 6, e20782.
24. Fitch, W.T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* 102, 1213–1222.
25. Barros, M., de Souza Silva, M.A., Huston, J.P., and Tomaz, C. (2004). Multibehavioral analysis of fear and anxiety before, during, and after experimentally induced predatory stress in *Callithrix penicillata*. *Pharmacol. Biochem. Behav.* 78, 357–367.
26. Button, K.S., Ioannidis, J.P., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S., and Munafò, M.R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14, 365–376.
27. Gülltekin, Y.B., and Hage, S.R. (2017). Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nat. Commun.* 8, 14046.
28. Janik, V.M., and Slater, P.J. (2000). The different roles of social learning in vocal communication. *Anim. Behav.* 60, 1–11.
29. Seyfarth, R.M., and Cheney, D.L. (1986). Vocal development in vervet monkeys. *Anim. Behav.* 34, 1640–1658.
30. Fischer, J. (2004). Emergence of individual recognition in young macaques. *Anim. Behav.* 67, 655–661.
31. Zuberbühler, K. (2000). Interspecies semantic communication in two forest primates. *Proc. Biol. Sci.* 267, 713–718.
32. Elowson, A.M., Snowdon, C.T., and Sweet, C.J. (1992). Ontogeny of trill and J-call vocalizations in the pygmy marmoset, *Cebuella pygmaea*. *Anim. Behav.* 43, 703–715.
33. Hihara, S., Yamada, H., Iriki, A., and Okanoya, K. (2003). Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neurosci. Res.* 45, 383–389.
34. Coudé, G., Ferrari, P.F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S., and Fogassi, L. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PLoS ONE* 6, e26822.
35. Hage, S.R., Gavrilov, N., and Nieder, A. (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *J. Cogn. Neurosci.* 25, 1692–1701.

36. Fitch, W.T. (2010). *The Evolution of Language* (Cambridge University Press).
37. Petkov, C.I., and Jarvis, E.D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4, 12.
38. Tchernichovski, O., and Marcus, G. (2014). Vocal learning beyond imitation: mechanisms of adaptive vocal development in songbirds and human infants. *Curr. Opin. Neurobiol.* 28, 42–47.
39. Lipkind, D., Marcus, G.F., Bemis, D.K., Sasahara, K., Jacoby, N., Takahasi, M., Suzuki, K., Feher, O., Ravbar, P., Okanoya, K., and Tchernichovski, O. (2013). Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* 498, 104–108.
40. West, M.J., and King, A.P. (1988). Female visual displays affect the development of male song in the cowbird. *Nature* 334, 244–246.
41. Chen, Y., Matheson, L.E., and Sakata, J.T. (2016). Mechanisms underlying the social enhancement of vocal learning in songbirds. *Proceedings of the National Academy of Sciences*, 201522306.
42. Kent, R.D., and Murray, A.D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *J. Acoust. Soc. Am.* 72, 353–365.
43. Scheiner, E., Hammerschmidt, K., Jürgens, U., and Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *J. Voice* 16, 509–529.
44. Adolph, K.E., and Robinson, S.R. (2013). The road to walking: What learning to walk tells us about development. In *Oxford Handbook of Developmental Psychology*, P. Zelazo, ed. (New York: Oxford University Press), pp. 403–443.
45. Adolph, K.E., Vereijken, B., and Shrout, P.E. (2003). What changes in infant walking and why. *Child Dev.* 74, 475–497.
46. Tamis-LeMonda, C.S., Adolph, K.E., Lobo, S.A., Karasik, L.B., Ishak, S., and Dimitropoulou, K.A. (2008). When infants take mothers' advice: 18-month-olds integrate perceptual and social information to guide motor action. *Dev. Psychol.* 44, 734–746.
47. Syl, S., and Finlay, B.L. (2011). Thinking outside the cortex: social motivation in the evolution and development of language. *Dev. Sci.* 14, 417–430.
48. Newman, S.W. (1999). The medial extended amygdala in male reproductive behavior. A node in the mammalian social behavior network. *Ann. N Y Acad. Sci.* 877, 242–257.
49. Goodson, J.L. (2005). The vertebrate social behavior network: evolutionary themes and variations. *Horm. Behav.* 48, 11–22.
50. O'Connell, L.A., and Hofmann, H.A. (2012). Evolution of a vertebrate social decision-making network. *Science* 336, 1154–1157.
51. Kato, M., Okanoya, K., Koike, T., Sasaki, E., Okano, H., Watanabe, S., and Iriki, A. (2014). Human speech- and reading-related genes display partially overlapping expression patterns in the marmoset brain. *Brain Lang.* 133, 26–38.
52. Teramitsu, I., Kudo, L.C., London, S.E., Geschwind, D.H., and White, S.A. (2004). Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *J. Neurosci.* 24, 3152–3163.
53. Steiper, M.E., and Young, N.M. (2006). Primate molecular divergence dates. *Mol. Phylogenet. Evol.* 41, 384–394.
54. Hrdy, S.B. (2005). Evolutionary context of human development: the cooperative breeding model. In *Attachment and Bonding: A New Synthesis*, from the 92nd Dahlem Workshop Report, C. Carter, L. Ahnert, K. Grossmann, S. Hardy, M. Lamb, S. Porges, and N. Sachser, eds. (Cambridge: MIT Press), pp. 9–32.
55. Burkart, J.M., Hrdy, S.B., and van Schaik, C.P. (2009). Cooperative breeding and human cognitive evolution. *Evol. Anthropol.* 18, 175–186.
56. Santos, C.V., French, J.A., and Otta, E. (1997). Infant carrying behavior in callitrichid primates: *Callithrix* and *Leontopithecus*. *Int. J. Primatol.* 18, 889–907.
57. Snowdon, C.T., and Cronin, K.A. (2007). Cooperative breeders do cooperate. *Behav. Processes* 76, 138–141.
58. Burkart, J.M., and van Schaik, C.P. (2010). Cognitive consequences of cooperative breeding in primates? *Anim. Cogn.* 13, 1–19.
59. Yamamoto, M.E., and Lopes, F.A. (2004). Effect of Removal from the Family Group on Feeding Behavior by Captive *Callithrix jacchus*. *Int. J. Primatol.* 25, 489–500.
60. Burkart, J.M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., Huber, J., Isler, K., Kosonen, Z.K., Martins, E., et al. (2014). The evolutionary origin of human hyper-cooperation. *Nat. Commun.* 5, 4747.
61. Ross, C., and MacLarnon, A. (2000). The evolution of non-maternal care in anthropoid primates: a test of the hypotheses. *Folia Primatol. (Basel)* 71, 93–113.
62. Zuberbühler, K. (2012). Cooperative breeding and the evolution of vocal flexibility. In *The Oxford Handbook of Language Evolution*, M. Tallerman, and K.R. Gibson, eds. (New York: Oxford University Press), pp. 71–81.
63. Locke, J.L. (2006). Parental selection of vocal behavior : Crying, cooing, babbling, and the evolution of language. *Hum. Nat.* 17, 155–168.
64. Gros-Louis, J., West, M.J., Goldstein, M.H., and King, A.P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *Int. J. Behav. Dev.* 30, 509–516.
65. Tyler, D.E. (2008). *Robust Statistics: Theory and Methods* (Taylor & Francis).
66. Hair, J.F., Anderson, R.E., Tatham, R.L., and Black, W.C. (1995). *Multivariate Data Analysis*, Third Edition (New York: Macmillan).
67. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Hillsdale, NJ: Lawrence Earlbaum Associates).
68. Olkin, I., and Finn, J.D. (1995). Correlations redux. *Psychol. Bull.* 118, 155–164.
69. Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
<i>Callithrix jacchus</i>	Princeton Neuroscience Institute	N/A
Software and Algorithms		
MATLAB	MathWorks	R2015b
G*Power	http://www.gpower.hhu.de/en.html	v. 3.1.9.2

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Asif A. Ghazanfar (asifg@princeton.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Marmosets

All experiments were approved by, and performed in compliance with, Princeton University Institutional Animal Care and Use Committee and its guidelines. The subjects used in the study were 6 infants (3 twin sets from different parents, 3 males, 3 females, postnatal days 1 to 60), captive common marmosets (*Callithrix jacchus*) housed at Princeton University. The number of subjects used in this study is consistent with, or exceeds, the number of experimental animals used in other studies of nonhuman primate vocal learning (between 1 to 4 animals) [2–4, 27]. The colony room is maintained at a temperature of approximately 27°C and 50%–60% relative humidity, with 12L:12D light cycle. The marmosets live in family groups; all were born in captivity. They had ad libitum access to water and were fed daily with standard commercial chow supplemented with fruits and vegetables. Additional treats were used prior to each session to transfer the adult animals from their home-cage into a transfer cage.

METHOD DETAILS

Experimental setup

Starting from their first postnatal day, we recorded the vocalizations of marmoset monkey infants in two different conditions: test condition, in which the infants were socially isolated (i.e., undirected context) and playback condition, in which the infants received auditory feedback from a closed loop playback system (described in more detail in the next section). Early in life, infants are always carried by their parent. Thus, the parent carrying the infant(s) was first brought from the home cage into a transfer cage. The infant marmoset was then gently separated from the adult and taken to the experiment room where it was placed on a flat piece of foam inside a prism-shaped testing box made of plexiglas (0.30 m × 0.30 m × 0.35 m). The vocalizations we observed were identical in type to those produced when the infant is naturally separated from parents (e.g., when parents push them off or when they transfer them to the other parent for carrying or feeding). The experiment box rested on a table (.76 m in height) in one of two opposing corners of the room. A speaker was placed at the opposing corner, at the same height, and delivered the playback of parental calls. An opaque curtain made of black clothes divided the room to visually occlude the subject from the other corner. A microphone (Sennheiser MKH-416) connected to a computer was placed directly on top of the testing box at a distance of 0.50 m. Audio signals were acquired at a sampling frequency of 96kHz at 16 bit. Most sessions consisted of a 10 min test condition followed by 30 min playback condition. Between postnatal days 1 to 7, however, the test condition lasted for only 5 min to minimize the time very young infants were separated from parents. The total number of sessions for each infant was 52, 53, 56, 56, 53, and 53 (6 infants, 317 sessions, 80,276 utterances during the test period). The number of sessions during the first 35 postnatal days were 31, 30, 33, 33, 33, 33, respectively. The order of the infants in the twin set was counterbalanced. At each session, either the mother or father's calls were played back. The order of which, across sessions, the father or mother's calls were played was counterbalanced between sessions. One of the twin received contingent playback with low probability and the other one received contingent feedback with high probability. The infants were randomly allocated to low or high contingency groups on first postnatal day before any measurement or procedure was performed.

Closed loop playbacks

A custom made MATLAB routine based on Psychtoolbox-3 was used to provide playbacks in a closed loop design: it detected infant calls and automatically delivered the parental call responses. The calls produced by infants were captured by a microphone and sent

to a desktop computer where it was saved and processed in real time. For the purpose of closed loop playback, we considered a call any sound with an amplitude large enough to cross and stay above a fixed threshold for more than 2 s (note: immature and mature phoe calls are multi-syllabic), possibly with some silent periods each lasting less than 400 ms. Furthermore, if the ratio between the power spectrum in the 8–10 kHz range and 4–6 kHz was larger than 2:1, then that signal was considered a more mature-sounding contact call. When such a narrow band call was detected, a parental call was played back through a speaker at ~60dB (measured at 0.1 m from the testing box) with ~1 s interval between infant call offset and playback onset. The 1 s interval was chosen based on the distribution of parental response intervals during natural interactions (Figure S1) [21]. The parameters of the playback system were optimized to detect infant calls and deliver playback precisely using data collected from a single infant from a separate gestation that was not included in this study to avoid double dipping.

Detection of calls and quantification of acoustic parameters

The detection and quantification of the acoustic parameters were similar to what have been described previously [7]. To determine the onset and offset of a syllable, a custom made MATLAB routine automatically detected the onset and offset of any signal that differed from the background noise at specific frequency range. To detect the differences, we first bandpass filtered the entire recording signal between 6 and 11 kHz. This corresponds to the frequency region where the infant marmoset calls have the highest power, which is not necessarily the fundamental frequency (F0), *i.e.*, the lowest frequency of the periodic components of the sound. The choice of 6–11 kHz frequency range allowed us to detect 100% of calls. Second, we resampled the signal to 1 kHz sampling rate, applied the Hilbert transform and calculated the absolute value to obtain the amplitude envelope of the signal. The amplitude envelope was further low pass filtered to 50 Hz. A segment of the recording without any call (silent) was chosen as a comparison baseline. The 99th percentile of the amplitude value in the silent period was used as the detection threshold. Sounds with amplitude envelope higher than the threshold were considered a possible vocalization. Finally, to ensure that sounds other than vocalizations were not included, a researcher verified whether each detected sound was a vocalization or not based on the spectrogram and amplitude of the signal.

After detecting the onset and offset of the call syllable, a custom made MATLAB routine calculated the duration, dominant frequency, amplitude modulation (AM) frequency, and Wiener entropy of each syllable. The duration of syllable is the difference between the offset and onset of the sound amplitude that crossed the threshold. To calculate the dominant frequency of a syllable, we first calculated the spectrogram and obtained the frequencies at which the spectrogram had maximum power for each time point. The dominant frequency of a syllable was calculated as the maximum of those frequencies. The spectrogram was calculated using a FFT window of 1024 points, Hanning window, with 50% overlap. The AM frequency was calculated in the following way. First, the signal was bandpass filtered between 6 to 10 kHz and then a Hilbert transform was applied. The absolute value of the resulting signal gives us the amplitude envelope of the modulated signal. The 6–10 kHz frequency range was found to give accurate values for the syllable envelope. Finally, the AM frequency was calculated as the dominant frequency of the amplitude envelope. The Wiener entropy is the logarithm of the ratio between the geometric and arithmetic means of the values of the power spectrum across different frequencies [7, 22]. The Wiener entropy represents how broadband the power spectrum of a signal is. The closer the signal is to white noise, the higher the value of Wiener entropy will be.

QUANTIFICATION AND STATISTICAL ANALYSIS

For all analysis, we adopted the Type I error $\alpha = 0.01$, below which we considered statistically significant.

Multiple linear regression analysis

MATLAB fitlm routine was used to fit a *robust multiple linear regression* to the data. The robust regression is more tolerant against outliers, deviation from normality, heteroscedasticity in the data and is in general superior to ordinary multiple linear regression [65]. We used the bisquare weight function with constant 4.685, which is the default in MATLAB. In Figure 2A,B we fitted the multiple linear regression model

$$\text{Entropy} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId} + e \cdot \text{PND} \cdot \text{Group} + f \cdot \text{PND} \cdot \text{TwinId} + \text{error},$$

where PND is the postnatal day, Group is the contingency group and TwinId is the twin identity. Both Group and TwinId were dummy variables, which was coded as 0 or 1 for each category. Therefore, the regression model is, when written explicitly using the dummy variables,

$$\text{Entropy} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId}_1 + e \cdot \text{TwinId}_2 + f \cdot \text{PND} \cdot \text{Group} + g \cdot \text{PND} \cdot \text{TwinId}_1 + h \cdot \text{PND} \cdot \text{TwinId}_2 + \text{error},$$

where, Group = 0 if it is low contingency and 1 if it is high contingency. TwinId₁ = 1 if the data is from the second twin set and TwinId₂ = 1 if it is from the third twin set. If the data is from the first twin set, both TwinId₁ and TwinId₂ = 0. In this way all the subject identities are coded on the dummy variables and, therefore, in our regression model, we are assuming that each marmoset infant has its own intercept and slope. Moreover, this coding scheme allow us to control for the effect of genetics. For all other regressions, Group and TwinId were always coded in the same way, but for sake of clarity, we will write the regression in the compact form without repeating the dummy variables (Group and TwinId).

The multiple linear model fitted in [Figures 2C](#) and [2D](#) was:

$$\text{AMfreq} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId} + e \cdot \text{PND} \cdot \text{Group} + f \cdot \text{PND} \cdot \text{TwinId} + \text{error}$$

For the regression controlling for weights, we fitted the multiple linear models:

$$\text{Entropy} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId} + e \cdot \text{Weight} + f \cdot \text{PND} \cdot \text{Group} + g \cdot \text{PND} \cdot \text{TwinId} + h \cdot \text{PND} \cdot \text{Weight} + \text{error}$$

$$\text{AMFreq} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId} + e \cdot \text{Weight} + f \cdot \text{PND} \cdot \text{Group} + g \cdot \text{PND} \cdot \text{TwinId} + h \cdot \text{PND} \cdot \text{Weight} + \text{error}$$

The cubic spline curve was fitted to the weights of each infant using MATLAB fitlm. Any missing weight data was interpolated using the values of the cubic spline curve.

For the multiple linear regressions in [Figures 3](#), the models had exactly the same predictors as in the models for [Figure 2](#), but using as dependent variables the duration of syllables and dominant frequency.

To test the effect of total number of stimuli calls on the previous session, we also fitted the multiple linear regressions

$$\text{Dependent} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId} + e \cdot \text{NStim} + f \cdot \text{PND} \cdot \text{Group} + g \cdot \text{PND} \cdot \text{TwinId} + h \cdot \text{PND} \cdot \text{NStim} + \text{error},$$

where the dependent variable was either entropy, AM frequency, duration or dominant frequency and NStim is the number of feed-back stimuli that the infant received in the previous session. Because, the interaction term between PND and NStim was not statistically significant for any of the acoustic parameters, we excluded the interaction term and reduced the model to

$$\text{Dependent} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId} + e \cdot \text{NStim} + f \cdot \text{PND} \cdot \text{Group} + g \cdot \text{PND} \cdot \text{TwinId} + \text{error}.$$

By experimental design, NStim is correlated with Group. To verify if the degree of collinearity of these predictors could cause difficulty in the interpretation of the coefficients, we calculated the variance inflation factor (VIF). Values of VIF larger than 10 are usually considered problematic [66]. We found that the VIF of NStim is 4.7.

The linear regression fitted to the data shown in [Figure 4](#) was

$$\text{CallRate} = a + b \cdot \text{PND} + c \cdot \text{Group} + d \cdot \text{TwinId} + e \cdot \text{PND} \cdot \text{Group} + f \cdot \text{PND} \cdot \text{TwinId} + \text{error},$$

where CallRate is the rate of call production by the infant in test condition.

Linear regression models

We report below the estimated regression coefficients, standard errors, t-values, and p values of the models used to test the effect of interaction between postnatal day and contingency group. All models were tested against the constant model and were significantly different ($p < 0.0001$).

Entropy $\sim 1 + \text{PND} \cdot \text{Group} + \text{PND} \cdot \text{TwinId}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept)	−1.0732	0.2167	−4.9523	1.6461e−06
PND	−0.028009	0.010559	−2.6527	0.0086788
Group	−0.058207	0.21319	−0.27302	0.78514
TwinId_1	−0.73452	0.26191	−2.8044	0.005579
TwinId_2	−0.21587	0.26217	−0.8234	0.41134
PND:Group	−0.032279	0.01041	−3.1009	0.0022313
PND:TwinId_1	−0.0013244	0.012797	−0.10349	0.91768
PND:TwinId_2	0.049178	0.012803	3.8411	0.00016818

AMfreq $\sim 1 + \text{PND} \cdot \text{Group} + \text{PND} \cdot \text{TwinId}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept)	1107.7	59.223	18.703	1.6341e−44
PND	−0.3784	2.8856	−0.13113	0.89581
Group	105.62	58.264	1.8129	0.071474
TwinId_1	−444.54	71.578	−6.2106	3.3977e−09
TwinId_2	152.61	71.648	2.1299	0.034499
PND:Group	−7.7853	2.8448	−2.7366	0.0068124
PND:TwinId_1	3.3999	3.4974	0.97213	0.33226
PND:TwinId_2	−10.507	3.499	−3.003	0.0030433

Duration $\sim 1 + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept)	0.2592	0.02994	8.6573	2.3366e-15
PND	0.009027	0.0014588	6.188	3.8273e-09
Group	0.036144	0.029455	1.2271	0.22135
TwinId_1	0.10384	0.036186	2.8696	0.0045888
TwinId_2	0.042662	0.036222	1.1778	0.24039
PND:Group	0.0028045	0.0014382	1.95	0.052687
PND:TwinId_1	-0.0040556	0.0017681	-2.2938	0.022925
PND:TwinId_2	-0.0086818	0.0017689	-4.908	2.0113e-06

DomFreq $\sim 1 + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept)	10234	94.532	108.26	3.018e-169
PND	-10.81	4.606	-2.3468	0.019992
Group	-174.27	93.001	-1.8738	0.062531
TwinId_1	-450.33	114.25	-3.9415	0.00011473
TwinId_2	-3.813	114.37	-0.033341	0.97344
PND:Group	2.2053	4.541	0.48566	0.62779
PND:TwinId_1	-14.693	5.5825	-2.6321	0.0092037
PND:TwinId_2	-8.08	5.5851	-1.4467	0.14967

Entropy $\sim 1 + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId} + \text{PND} \times \text{Weight}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept)	1.0381	1.2497	0.83068	0.40724
PND	-0.065573	0.026399	-2.4839	0.013895
Group	0.19384	0.26076	0.74334	0.45823
TwinId_1	-0.22484	0.39588	-0.56794	0.57077
TwinId_2	-0.016016	0.28399	-0.056397	0.95509
Weight	-0.079176	0.049991	-1.5838	0.11497
PND:Group	-0.03973	0.010633	-3.7367	0.00024906
PND:TwinId_1	0.0078016	0.026358	0.29598	0.76758
PND:TwinId_2	0.046426	0.01376	3.3741	0.0009041
PND:Weight	0.0019526	0.00069189	2.8221	0.0052986

AMfreq $\sim 1 + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId} + \text{PND} \times \text{Weight}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept)	2509.2	309.91	8.0965	7.7001e-14
PND	30.551	6.5469	4.6665	5.9033e-06
Group	277.62	64.668	4.2931	2.8539e-05
TwinId_1	-121.99	98.176	-1.2426	0.21561
TwinId_2	284.48	70.428	4.0394	7.8784e-05
Weight	-60.478	12.398	-4.8782	2.318e-06
PND:Group	-7.3904	2.6368	-2.8028	0.0056124
PND:TwinId_1	34.435	6.5366	5.2679	3.8493e-07
PND:TwinId_2	-5.6191	3.4123	-1.6467	0.10134
PND:Weight	0.29006	0.17159	1.6905	0.092639

Duration $\sim 1 + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId} + \text{PND} \times \text{Weight}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept) 0.086621 0.17803 0.48656 0.62715
 PND 0.0034503 0.003761 0.91741 0.36013

Group 0.026253 0.037149 0.70669 0.48066
 TwinId_1 0.080725 0.056398 1.4313 0.15404
 TwinId_2 0.024033 0.040458 0.59403 0.55323
 Weight 0.0072414 0.0071219 1.0168 0.3106
 PND:Group 0.0016772 0.0015147 1.1072 0.26965
 PND:TwinId_1 -0.0093137 0.003755 -2.4803 0.01403
 PND:TwinId_2 -0.0084514 0.0019603 -4.3114 2.6477e-05
 PND:Weight 1.4905e-05 9.8569e-05 0.15121 0.87998

DomFreq $\sim 1 + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId} + \text{PND} \times \text{Weight}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept) 10624 560.78 18.946 5.199e-45
 PND -0.4841 11.847 -0.040863 0.96745
 Group -160.03 117.02 -1.3676 0.17312
 TwinId_1 -356.62 177.65 -2.0074 0.046174
 TwinId_2 56.528 127.44 0.44356 0.65788
 Weight -16.719 22.434 -0.74528 0.45706
 PND:Group 4.951 4.7714 1.0376 0.30081
 PND:TwinId_1 -5.6357 11.828 -0.47646 0.63431
 PND:TwinId_2 -8.7727 6.1747 -1.4208 0.15709
 PND:Weight 0.0261 0.31049 0.084062 0.9331

Entropy $\sim 1 + \text{NStim} + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept) -1.0594 0.21981 -4.8195 3.0015e-06
 PND -0.028755 0.010712 -2.6844 0.0079286
 Group 0.12121 0.2956 0.41006 0.68224
 TwinId_1 -0.75217 0.26418 -2.8472 0.0049126
 TwinId_2 -0.2013 0.26559 -0.75793 0.44946
 NStim -0.00163 0.0018645 -0.87424 0.38313
 PND:Group -0.031149 0.010617 -2.9338 0.0037743
 PND:TwinId_1 0.0021952 0.012975 0.16918 0.86584
 PND:TwinId_2 0.050836 0.013021 3.9042 0.00013259

AMfreq $\sim 1 + \text{NStim} + \text{PND} \times \text{Group} + \text{PND} \times \text{TwinId}$

Estimated Coefficients:

Estimate SE tStat pValue

(Intercept) 1065.3 58.295 18.274 3.2489e-43
 PND 1.5478 2.8408 0.54485 0.58652
 Group -95.55 78.395 -1.2188 0.22447
 TwinId_1 -411.68 70.062 -5.8759 1.9362e-08
 TwinId_2 168.22 70.437 2.3883 0.017939
 NStim 1.665 0.49446 3.3673 0.00092433
 PND:Group -8.1971 2.8158 -2.9112 0.0040459
 PND:TwinId_1 0.58617 3.441 0.17035 0.86492
 PND:TwinId_2 -13.907 3.4532 -4.0273 8.244e-05

Duration ~1 + NStim + PND*Group + PND*TwinId

Estimated Coefficients:

Estimate SE tStat pValue

	Estimate	SE	tStat	pValue
(Intercept)	0.25613	0.030179	8.4869	6.9102e-15
PND	0.0091292	0.0014707	6.2074	3.4868e-09
Group	0.0062594	0.040585	0.15423	0.8776
TwinId_1	0.10452	0.036271	2.8816	0.0044266
TwinId_2	0.044412	0.036465	1.2179	0.22481
NStim	0.00024752	0.00025598	0.96693	0.33485
PND:Group	0.0028697	0.0014577	1.9686	0.050497
PND:TwinId_1	-0.0043113	0.0017814	-2.4201	0.016487
PND:TwinId_2	-0.0093599	0.0017877	-5.2356	4.461e-07

DomFreq ~1 + NStim + PND*Group + PND*TwinId

Estimated Coefficients:

Estimate SE tStat pValue

	Estimate	SE	tStat	pValue
(Intercept)	10221	98.247	104.03	2.0214e-165
PND	-10.317	4.7878	-2.1548	0.032478
Group	-424.87	132.12	-3.2157	0.0015373
TwinId_1	-409.35	118.08	-3.4667	0.00065583
TwinId_2	6.0924	118.71	0.051321	0.95913
NStim	1.6855	0.83334	2.0226	0.04456
PND:Group	4.6416	4.7456	0.97809	0.32931
PND:TwinId_1	-18.061	5.7994	-3.1142	0.0021395
PND:TwinId_2	-12.989	5.8199	-2.2318	0.026832

CallRate~PND*Group+PND*TwinId

Estimated Coefficients:

Estimate SE tStat pValue

	Estimate	SE	tStat	pValue
(Intercept)	16.342	1.9319	8.4591	8.0002e-15
PND	-0.26321	0.09413	-2.7962	0.0057162
Group_1	-2.286	1.9006	-1.2028	0.23061
TwinId_2	-5.2229	2.3349	-2.2369	0.02649
TwinId_3	9.6716	2.3372	4.1381	5.3129e-05
PND:Group_1	0.040069	0.092801	0.43177	0.66641
PND:TwinId_2	0.05292	0.11409	0.46386	0.6433
PND:TwinId_3	0.1245	0.11414	1.0908	0.27678

Effect size and power analysis

We calculated the local effect size of the contingency group (Group) for the model in Figure 2 and Figure 3. We used as a measure of the effect size the Cohen's f^2 [67]. To calculate the confidence interval, we used the Olkin and Finn's approximation [68]. The power was calculated using G*Power 3 [69].

DATA AND SOFTWARE AVAILABILITY

Data and MATLAB code used for analysis of Figures 2, 3, and 4 are available in DRYAD Digital Repository: <http://dx.doi.org/10.5061/dryad.76bn8>.