MINNESOTA SYMPOSIA ON CHILD PSYCHOLOGY

Development of the Social Brain

Edited by
Jed T. Elison
Maria D. Sera

Volume 39

WILEY

Minnesota Symposia on Child Psychology

Development of the Social Brain

Volume 39

Edited by
Jed T. Elison
Maria D. Sera

WILEY

CHAPTER 3
Marmoset Monkey Vocal Communication: Common Developmental Trajectories With Humans and Possible Mechanisms

ASIF A. GHAZANFAR, DANIEL Y. TAKAHASHI, YISI S. ZHANG AND JEREMY I. BORJON

INTRODUCTION

Like many behaviors (Byrge, Sporns, & Smith, 2014), vocal development is the outcome of interactions between an infant's developing biological system of production (the body and the nervous system) and his experience with caregivers. In humans, such development is marked by a progression: Early vocalizations, like cries, fussing, and cooing sounds, become increasingly complex and speechlike and eventually turn into words (Oller, 2000). In fact, from a purely acoustic perspective, the progression of changes from cries to later vocalizations like cooing and babbling is continuous (Kent & Murray, 1982). At around 6 months, infants enter the babbling period when their vocalizations are most obviously speechlike with the production of consonant-vowel combinations. Thus, early vocalizations act as scaffolding for later, more speechlike vocalizations.

Social feedback from caregivers can be an instrumental driving force that increases the maturation rate of these prelinguistic vocalizations (Gros-Louis, West, & King, 2010). Parental eye contact, contingency, and responsiveness directly impact the quality and vocal pattern of infants' vocalizations (Hsu & Fogel, 2001). Not only is the volubility of infants influenced by social context and past dyadic interactions with caregivers (Franklin et al., 2013; Goldstein, Bornstein, Schwade, Baldwin, & Brandstadter, 2009), but caregivers who preferentially and contingently respond to speechlike infant vocalizations spur the development of more complex vocalizations from those infants (Goldstein, King, & West, 2003; Goldstein & Schwade, 2008). Conversely, parents are more likely to contingently respond to infant vocalizations if they sound more mature or speechlike (Gros-Louis, West, Goldstein, & King, 2006). Taken together, these findings show that contingent parent responses to infants' vocalizations facilitate the latter's development into more mature sounding forms.

To understand the mechanisms underlying these vocal developmental processes and how they evolved, it is important to use an animal model that shares some of the same features. In the following section, we review what we've learned from marmoset monkey vocal development and show they that have a strikingly similar developmental trajectory to that of humans: Marmosets produce babbling-like vocal sequences and the maturation rate of their vocalizations is influenced by parental feedback. We will then provide evidence describing the physiological mechanisms that facilitate this developmental process.

THE MARMOSET MONKEY MODEL SYSTEM

The common marmoset monkey (*Callithrix jacchus*) is a small (300–400 grams, on average), New World species that is native to northeastern Brazil. They live in social groups of ~9 to 15 individuals many of whom are related to each other. Marmosets and other closely related species in the Callitrichid family are quite flexible in their vocal output, especially when compared to Old World primates like macaques and apes. They readily adjust (without any training) the timing of their contact "phee" vocalizations to the timing of conspecific calls (Ghazanfar, Flombaum, Miller, & Hauser, 2001; Ghazanfar, Smith-Rohrberg, Pollen, & Hauser, 2002; Takahashi, Narayanan, & Ghazanfar, 2013), to avoid intermittent background noise (Egnor, Wickelgren, & Hauser, 2007; Roy, Miller, Gottsch, & Wang, 2011), and cooperatively adjust the amplitude of their calls during vocal exchanges in accord with distance from conspecifics (Choi, Takahashi, & Ghazanfar, 2015). Remarkably, marmosets also take turns when they vocalize, exhibiting contingent and repeated exchanges of vocalizations between any two individuals – related or unrelated – for an extended period of time. That is, their behavior is not simply a call-and-response behavior among mates or competitors (Takahashi et al., 2013). This turn-taking behavior by marmosets has the same universal features and coupled oscillator properties as human conversational turn-taking, albeit on a different timescale (Ghazanfar & Takahashi, 2014; Levinson, 2016; Takahashi et al., 2013).

Another noteworthy phenotype of marmoset monkeys is that they are cooperative breeders and typically produce dizygotic twins (Harris et al., 2014). Both parents, as well as older siblings and nonkin, help care for offspring by carrying them and sharing food. Among primates this is very rare: Only humans and members of the Callitrichid family exhibit this reproductive strategy. Thus, in terms of comparative developmental studies among human and nonhuman primates, marmosets are a more compelling analogous species than the phylogenetically closer, but socially dissimilar, Old World apes and monkeys (Elowson, Snowdon, & Lazaro-Perea, 1998). These cooperative breeding behaviors by humans and marmosets lead to prosocial cognitive processes (Burkart et al., 2014; Snowdon & Cronin, 2007), including those related to vocal communication (Borjon & Ghazanfar, 2014). In humans, the vocal behavior of infants stimulates interactions with caregivers, thereby playing a key role in the infants' own development. Thus, in light of their similar infant care strategies and vocal turn-taking behavior, does the vocal behavior of marmoset monkeys develop in a manner similar to that of human vocal behavior (Zuberbühler, 2012)?

BABBLING AND PERINATAL INFLUENCES ON VOCAL OUTPUT

As in humans, variable, spontaneous vocal behaviors are ubiquitous in young marmoset infants (Takahashi et al., 2015). These early vocal behaviors primarily reflect the interplay between the infants' arousal states, sensorimotor coordination, and biomechanical conditions (Zhang & Ghazanfar, 2016). To capture the shape of the developmental trajectory of vocal behavior, one must sample early and densely (Adolph, Robinson, Young, & Gill-Alvarez, 2008), and this is

especially true for marmosets, a species that develops 12 times faster than humans (de Castro Leão, Duarte Dória Neto, & Bernardete Cordeiro de Sousa, 2009; Schultz-Darken, Braun, & Emborg, 2015). Vocalizations must also be recorded in controlled contexts if any sense is to be made of the vocal changes. We recorded infant marmoset vocalizations in two contexts – undirected and directed – starting at postnatal day 1 (P1) and roughly every other day thereafter until they were 2 months of age. In the undirected context, infants were very briefly left alone. In the directed context, infants were in auditory but not visual contact with one of their parents.

Contrary to what is typical for other nonhuman primates (Egnor & Hauser, 2004), we found that marmoset infants exhibit a dramatic change in vocal production in the first few months of postnatal life (Takahashi et al., 2015; Zhang & Ghazanfar, 2016). At P1, vocalizations are more numerous and more variable in their spectrotemporal structure than in later weeks ([Figure 3.1](#)A). The number and variability of calls diminished over 2 months, approaching mature vocal output with exclusive production of phee calls (Takahashi et al., 2013). We measured four acoustic parameters similar to those used for birdsong development (Tchernichovski, Mitra, Lints, & Nottebohm, 2001): duration, dominant frequency, amplitude modulation (AM) frequency, and Wiener entropy (Takahashi et al., 2015). Changes in all four parameters were statistically significant, showing that marmoset vocalizations undergo a transformation in the first 2 months, whereby their utterances lengthen, dominant and AM frequencies decrease, and bandwidths shrink (i.e., entropy decreases). Prior reports of much more subtle developmental changes in monkey vocalizations were frequently attributed to the passive consequences of growth in vocalization-related structures (Egnor & Hauser, 2004). For example, as the vocal folds get bigger and/or the vocal tract gets longer, vocalizations will be produced with acoustics in a lower frequency range (Ghazanfar & Rendall, 2008). Thus, changes in infant marmoset vocalizations – though not subtle – could be attributed solely to physical maturation. To test this, we used body weight as a proxy for overall growth (weight correlates well with vocal apparatus size in monkeys; Fitch, 1997) and tried to predict the pattern of vocal change in the four acoustic parameters using a linear regression. We found that growth alone did not accurately predict acoustic changes in vocal output during marmoset monkey infancy (Takahashi et al., 2015).
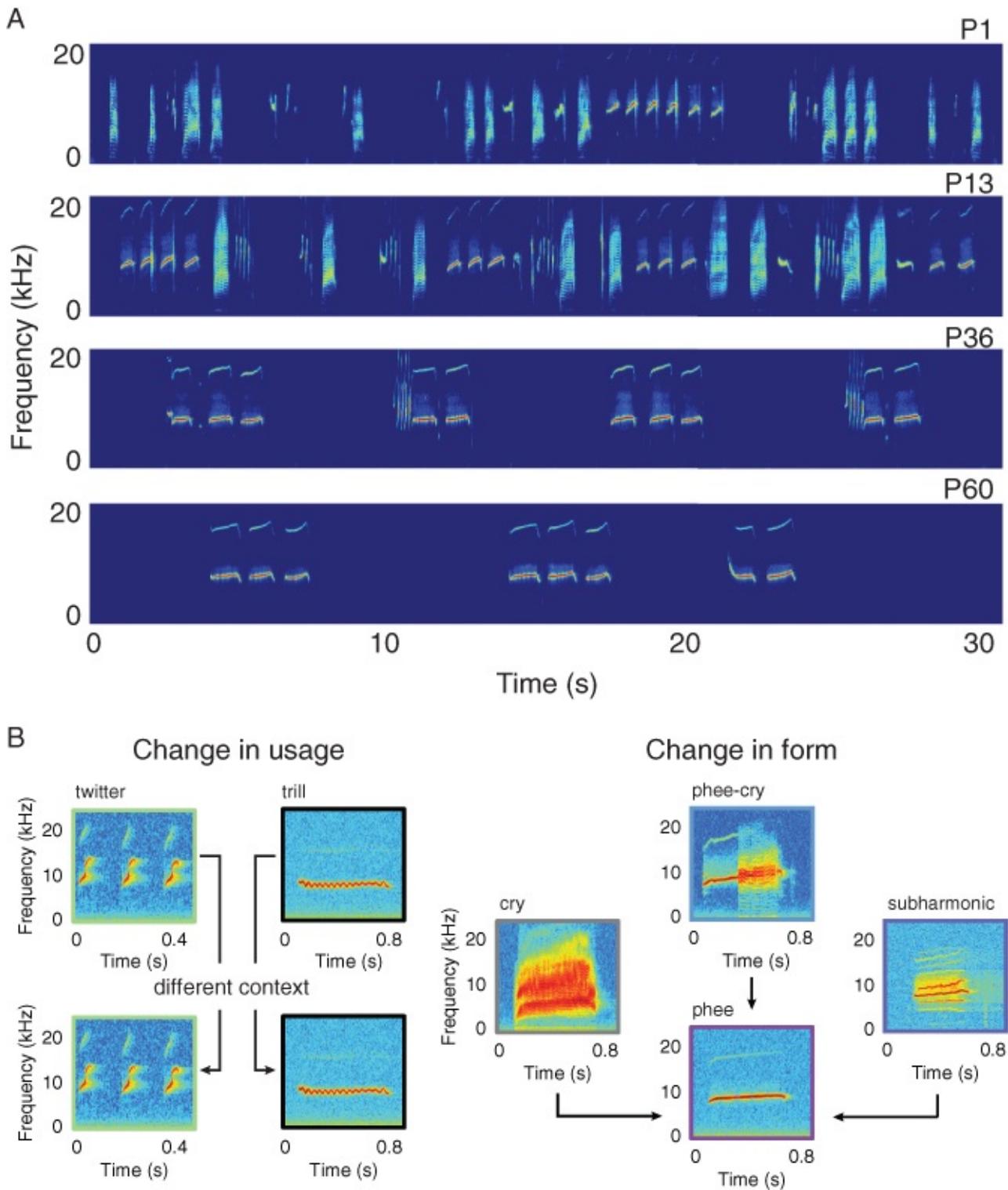
Figure 3.1 Infant marmoset vocalizations undergo dramatic acoustic changes. (A) Vocalizations from one infant. (B) Twitters and trills change usage whereas cries, phee-cries, and subharmonic-phees transition to phee calls.

We next investigated whether a subset of infant marmoset calls served as scaffolding for mature, adult-like calls (Takahashi et al., 2015). Human infant babbling, for instance, is a mixture of vocalizations, a subset of which will be incorporated in adult speech and produced in correct contexts, whereas others are transient (Locke, 1995; Oller, 2000). These transient

syllable structures are also evident in song learning by birds (Tchernichovski et al., 2001). We found that infant marmosets also produce a mixture of mature and immature vocalizations: adult-like calls ("twitters," "trills," and "phees") and immature versions of the contact phee call ("cries," "subharmonic phees," "phee-cries"). By 2 months of age, however, they only produce the phee calls that are appropriate for both the undirected and directed contexts. This suggests that two different vocal learning processes are at work: change in usage (Elowson, Snowdon, & Lazaro-Perea, 1998; Seyfarth & Cheney, 1986) and transformation of immature calls into mature versions (Takahashi et al., 2015; Figure 3.1B). Twitters and trills are produced frequently by marmosets of all ages (Bezerra & Souto, 2008; Pistorio, Vintch, & Wang, 2006), but in adults they are typically produced when in visual contact with conspecifics and not in the undirected context. Thus, twitters and trills undergo change in usage in the first 2 months. That is, by 2 months of age, they stop producing them when they are out of visual contact with a conspecific. By contrast, cries, phee-cries, and subharmonic-phees are only produced by infants (Pistorio et al., 2006; Takahashi et al., 2015). Because these infant-only calls share some features with the phee (e.g., a common duration that is distinct from trills and twitter syllables), they represent immature phees, consistent with vocal transformations observed in preverbal human infants (Kent & Murray, 1982; Scheiner, Hammerschmidt, Jurgens, & Zwirner, 2002) and songbirds (Tchernichovski et al., 2001), but contrasting with prior reports on developing primates (Egnor & Hauser, 2004).

Since marmoset monkeys typically give birth to dizygotic twins (Harris et al., 2014), we could also assess the role of perinatal influences on early babbling (Kent & Murray, 1982). We quantified babbling sequences and their development by treating each call type as a "state" and using Markov models to characterize their sequential structure (Zhang & Ghazanfar, 2016). Figure 3.2A shows how babbling sequence change over the course of development. The size of the nodes represents the proportion of the call type and the thickness of the arrow line represents the transition probability. Dashed lines label transition probabilities that dropped below 5% but were significantly greater in the previous week. Overall, babbling sequences exhibited high variability in the first week of life, but increased in stereotypy and monotony over time. However, the trajectory of the transition probabilities that each individual took to reach the stable phee-call state was quite variable. We quantified the individual differences in babbling sequences by calculating the Jensen-Shannon divergence rate (JSDR) between each pair of individuals (Sasahara, Tchernichovski, Takahasi, Suzuki, & Okanoya, 2015). The JSDR measures differences between the transition probabilities among call types weighted by the proportion of each call type. We found that the mean JSDR gradually decreased to near zero in about 8–9 postnatal weeks, reflecting the convergence of all infants ($n = 10$) to the phee-only call state in the undirected context.
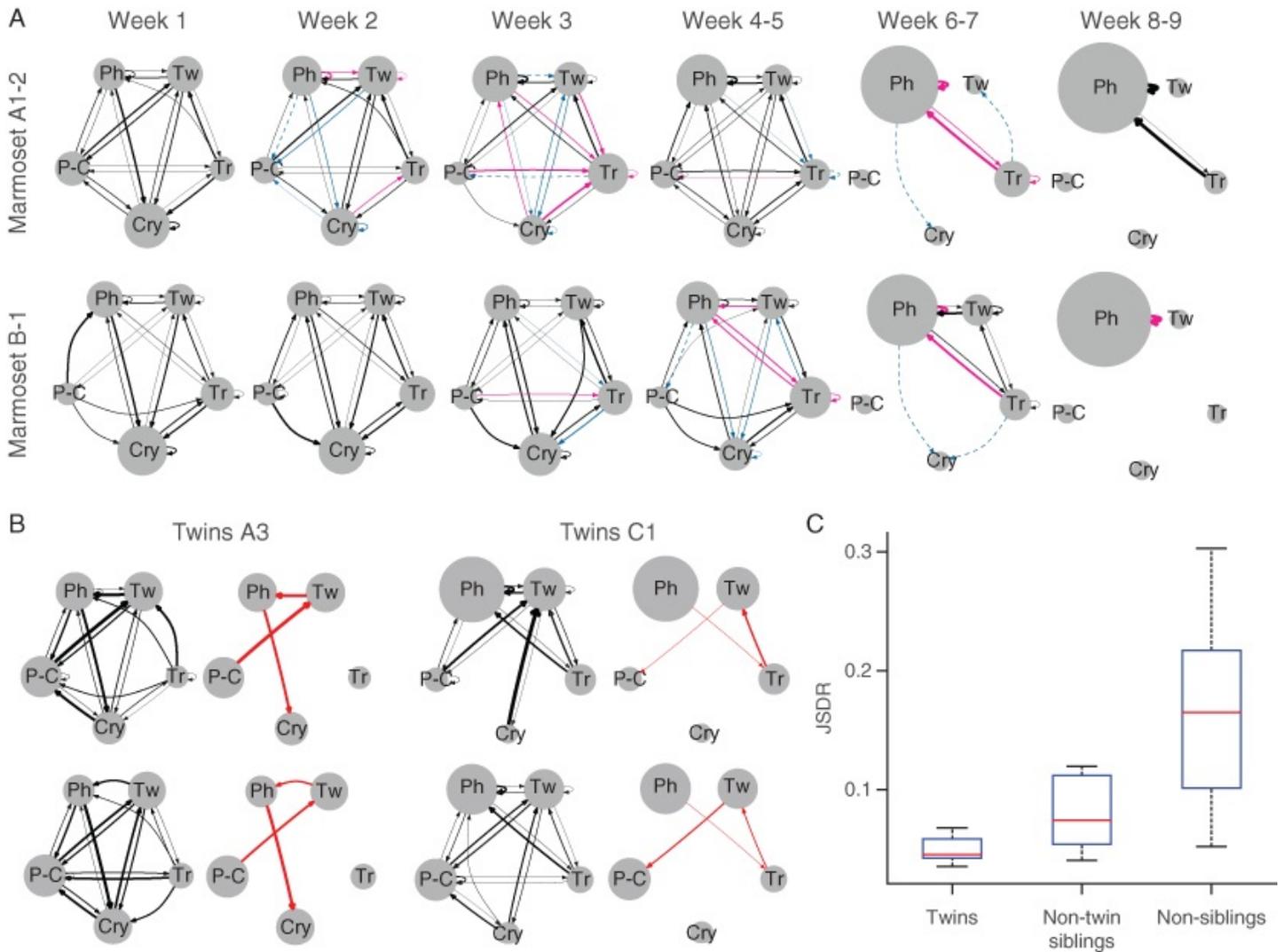
Figure 3.2 Babbling sequences and their similarity among twins. (A) Transition diagrams visualizing vocal sequences from two subjects at different postnatal time points. Each node in the diagram corresponds to a type of call, and the arrows correspond to the transitions between call types. The five most frequently produced call types are: phee (Ph), twitter (Tw), trill (Tr), cry (Cry), and phee-cry (P-C). Node size is proportional to the fraction of the call types, and edge size is proportional to the transition probability between calls. Thin dashed arrows are where transitions dropped below 5% occurrences. (B) Transition diagrams of vocal sequences from the first postnatal week for three sets of twins. Each twin set is arranged in the vertical order with the highlighted most frequent four-call transitions plotted on the right. (C) Comparison of JSDRs in three relationship categories: twins (= 5), nontwin siblings ($n = 12$), and nonsiblings ($n = 28$), p = 3.8e-5, ANOVA.

We then compared babbling sequences produced during the first postnatal week of twins, nontwin siblings, and unrelated infants in order to assess possible perinatal influences on call sequence similarity. Twins are similar genetically and typically share identical perinatal experiences. Since marmoset monkeys produce only dizygotic twins (Harris et al., 2014), nontwin siblings are just as genetically similar as twin siblings, but do not share identical perinatal experiences. If the perinatal environment contributes to the structure of babbling sequences, twin calls should be the most similar, followed by siblings and then age-matched

non-siblings. [Figure 3.2](#)B shows the transition probability of three sets of twins. Twins had very similar babbling sequences in the first week of postnatal life. To quantify this, we used JSDR measures again. We found that twins had greater sequence similarity relative to their nontwin siblings and the nonsiblings, whereas those between the nontwin siblings were smaller than the nonsiblings ([Figure 3.2](#)C).

The greater similarity in early babbling sequences among siblings (twins and nontwins) than the nonsiblings and the greater similarity between dizygotic twins versus their nontwin siblings suggest that prenatal factors, including genetics and/or perinatal environment, shape the early vocal sequences. These factors likely interact with each other during the course of individual development (Gottlieb, 1992), shaping both arousal fluctuation patterns as well as biomechanics. For instance, in rodents, levels of arousal are linked to the actions of the estrogen receptor (Garey et al., 2003), and patterns of estrogen receptor expression are influenced by the epigenetics of maternal care (Champagne & Curley, 2008). Similar data are not available for marmoset monkeys, but other studies in this species demonstrate that early pre- and postnatal environmental events (including quality of parental care) influence patterns of infant behavior (Pryce, Aubert, Maier, Pearce, & Fuchs, 2011) (including vocalizations; Dettling, Feldon, & Pryce, 2002) and gene expression in the brain (Law, Pei, Feldon, Pryce, & Harrison, 2009). Moreover, changes in maternal arousal levels can influence offspring both in the womb and through physical contact postnatally. For example, the cardiorespiratory dynamics of human infants will entrain to their mother's dynamics when they are laying on her body (Van Puyvelde et al., 2015) and during face-to-face communication involving vocalizations (Feldman, Magori-Cohen, Galili, Singer, & Louzoun, 2011). This maternal effect on human infants is present up to 2–3 months of age and is thought to be a continuation of similar maternal influences prenatally. Indeed, very young infants have difficulty self-regulating arousal levels; that regulation comes from parents (Moriceau & Sullivan, 2005). Similarly, in marmoset twins, common contact with parents (via carrying and vocal interactions; Takahashi et al., 2015) during the first week of postnatal life may influence their arousal levels simultaneously to generate similar patterns of fluctuations. Moreover, both prenatal and postnatal influences on infant arousal levels may come from parental vocal output that reflects their own respiration and autonomic state.

DEVELOPMENT OF VOCAL TURN-TAKING

Vocal turn-taking is a ubiquitous form of social interaction in our lives. In humans, vocal turn-taking develops during the first year of life (Elias & Broerse, 1996; Hilbrink, Gattis, & Levinson, 2015; Jasnow & Feldstein, 1986; Kajikawa, Amano, & Kondo, 2004; Stern, Jaffe, Beebe, & Bennett, 1975). In the first postnatal months, infants frequently overlap their vocalizations with their parent's utterances. By approximately 9 months of age, they can engage in proto-conversations with adult-like turn-taking dynamics. We investigated the development of vocal turn-taking in marmoset monkeys. Turn-taking in adult marmosets has the same universal features of human conversational turn-taking (albeit on a different timescale) and exhibits the essential features of a coupled oscillator system (Takahashi et al., 2013).

We tested marmoset infants in a directed context in which they could vocally interact with their parents, starting on postnatal day 1 until they were 2 months of age (Takahashi, Fenley, & Ghazanfar, 2016). Because marmoset monkeys develop much faster than humans (de Castro Leão et al., 2009; Schultz-Darken et al., 2015), this time interval represents the equivalent of the first 2 years of human postnatal life. We found that early in postnatal life, marmoset infants overlap their vocalizations with their parents' about 25% of the time (Takahashi et al., 2016), an overlap probability consistent with human 3-month-olds (Hilbrink et al., 2015). This overlap probability decreases to about 15% in marmoset infants by the time they are 2 months of age, also consistent with the near age-equivalent human infant (18-month-olds; Hilbrink et al., 2015). Adult marmosets rarely overlap their vocalizations (Takahashi et al., 2013). Thus, the fact that turn-taking is still developing at 2 months of age in marmosets is analogous to the way that human infants at 18 months of age still exhibit immature vocal interaction dynamics. We also showed that the amount of observed vocalization overlap is larger than what is expected by chance early in postnatal life (Takahashi et al., 2016), similar to what was observed in 13- to 15-week-old human infants (Elias, Hayes, & Broerse, 1986). Thus, marmoset infants undergo the same developmental trajectory for vocal turn-taking as humans and do so during the same life-history stage (Takahashi et al., 2016).

The system of vocal turn-taking between infants and parents is one in which parental responses are contingent upon infant vocalizations. This provides an opportunity for the infant to learn from the parents how to take turns during a vocal exchange. Although it is an intriguing possibility, as far as we know, there is no evidence that parental feedback influences infant turn-taking behavior in humans. We directly tested a possibility of parental influences on turn-taking development in marmosets and found no relationship between the maturation rate of vocal turn-taking and overall frequency of contingent parental responses (Takahashi et al., 2016).

TURN-TAKING AS THE DEVELOPMENTAL SYSTEM UPON WHICH INFANT VOCALIZATIONS ARE LEARNED

Another possibility for learning in this infant-parent vocal system is that contingent parental responses influence infant vocal acoustics. Studies of naturalistic human infant-parent interactions (Bloom, Russell, & Wassenberg, 1987; Gros-Louis, West, & King, 2014; Hsu & Fogel, 2001; Masataka, 1993) as well as experimental studies (Goldstein et al., 2003; Goldstein & Schwade, 2008) reveal that contingent parental responses influence the acoustic structure of subsequent infant vocalizations, making them sound more mature (i.e., speechlike). Along similar lines, we found that subsequent to their parents' vocalizations, marmoset infants increasingly produce longer and more tonal (low entropy) calls over the course of development (Takahashi et al., 2016) (Figure 3.3A). This is yet another parallel with human vocal turn-taking development and consistent with the many ways infants can learn from parents beyond imitation (Tchernichovski & Marcus, 2014). Since we are able to study our marmosets longitudinally, we can track how this real time influence of contingent parental feedback on infant vocal acoustics can impact longer lasting changes in their vocalizations

(Takahashi et al., 2015).
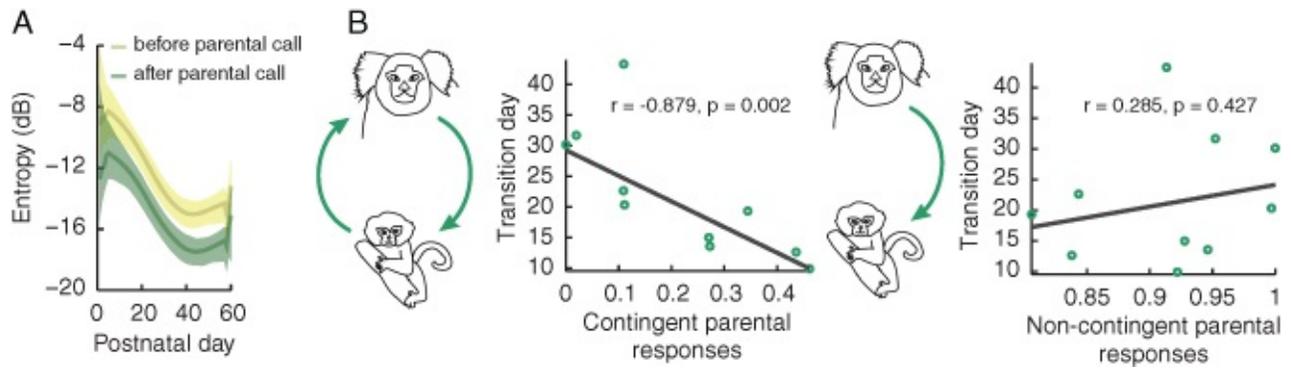
Figure 3.3 Transition from cry to phee is influenced by contingent parental calls. (A) Weighted average entropy of infant calls produced before adult call onset and after adult call offset. The shaded regions indicate the respective 95% confidence intervals. (B) Correlations between the transition day and the proportion of contingent (left) and noncontingent (right) parental responses, respectively.

To assess the effect of parent-infant vocal interactions in marmosets, we quantified their vocal exchanges in the directed context, where infants and their mother or father were in auditory, but not visual, contact, while also measuring the acoustic structure of infant calls (Takahashi et al., 2015). We again used the timeframe of 2 months. During this time interval, immature calls like cries gave way to mature phees calls, but this transition occurred rapidly. For each infant, we used the day on which the phee-cry ratio was 50–50 to mark the infant's transition day. Transitions were typically sharp, but their timing varied substantially across infants (~10 to 40 days). We then investigated if parental responses to infant vocalizations affect the timing of the cries-to-phees transition. This would explain, at least partially, its variability across infants. Parental influence could be via contingent responses and/or simply the number of adult vocalizations the infant has heard. Figure 3.3B (left panel) shows the fraction of infant phees that elicited contingent parental phee responses before the transition day and, remarkably, its significant correlation with the timing of the transition day. Proportions of noncontingent parental calls (91.5% of all calls on average) were not significantly correlated with this timing (Figure 3.3B, right panel). Thus, contingent vocal responses from parents influence the timing of the cries-to-phees transition by reinforcing the production of phee calls (Takahashi et al., 2015).

We addressed two possible caveats to this conclusion. First, it is possible that, through shared genetics, fast transitioning infants are born to more vocally interactive parents. To test this, we correlated the frequency of contingent parental calls and the cry-to-phee transition day for six full-siblings born from the same parents. If shared genetics were driving the result, then there would be no correlation between contingent parental responses and the transition day. We found, however, that there remained a statistically significant correlation (Takahashi et al., 2015). Second, it is possible that the changing patterns of infant calling are due to changes in parental call output. However, we found that neither parent changes their production rates (Takahashi et al., 2015). Based on these analyses, we conclude that the cries-to-phees transition is influenced by contingent responses from parents, not through shared genetics or

changes in parental vocal output.

Taken together, these findings suggest that developing marmoset monkeys – unlike every other nonhuman primate investigated thus far – may be vocal learners (Margoliash & Tchernichovski, 2015). However, 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 contingent vocal feedback can increase the rate at which marmoset infants begin producing mature-sounding contact calls (Takahashi, Liao, & Ghazanfar, 2017). Since marmoset monkeys typically give birth to dizygotic twins (Harris et al., 2014), we could again control for the influence of genetics and the perinatal environment on vocal development (Zhang & Ghazanfar, 2016). Starting on P1, infants were provided different levels of contingent feedback using closed-loop, computer-driven playbacks of parental contact calls in almost daily 30-minute sessions for 2 months. One randomly selected twin was given the best possible simulated "parent" who provided 100% contingent vocal feedback; the other infant was provided a not-so-good parent and received only 10% contingent vocal feedback (Figure 3.4A).
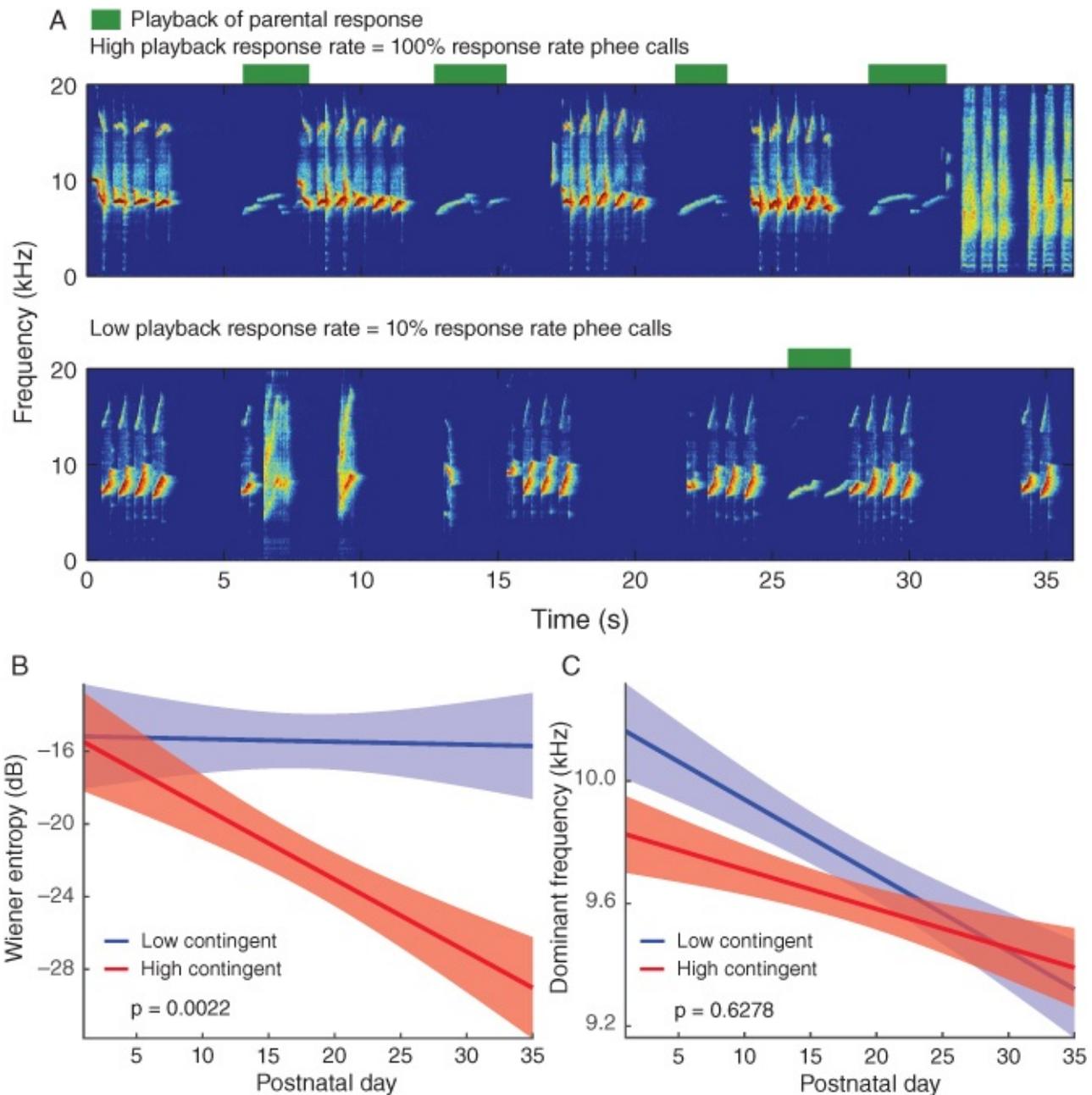
Figure 3.4 Vocal-production learning by infant marmoset monkeys. (A) Twin infants received either high-contingency playbacks (100%) or low contingency playbacks (10%). Spectrograms depict when such playbacks were delivered relative to the infant vocalizations. (B) Wiener entropy (in decibels) changes over postnatal days for high and low contingency infants. (C) Dominant frequency (in kilohertz) changes over postnatal days for high and low contingency infants. Shaded regions indicate 1 standard error intervals.

Our data showed that infant marmoset monkeys who received more contingent feedback learned faster (as measured by entropy; Figure 3.4B). They do this not through imitation but rather through the experience-dependent increase in the control of the vocal apparatus. Calls with high entropy are related to poor muscular control of – and coordination between – respiration and vocal fold tension (Takahashi et al., 2015; Teramoto, Takahashi, Holmes, & Ghazanfar, 2017; Zhang & Ghazanfar, 2016). Thus, more contingent vocal feedback results in faster development of this respiratory and laryngeal control and coordination (Teramoto et al.,

2017). Other acoustic features (e.g., dominant frequency; Figure 3.4C) were unaffected by experience, and this was consistent with the predictions of our integrated framework for marmoset vocal development in which changes in some acoustic features are solely explained by growth (Teramoto et al., 2017).

THE AUTONOMIC NERVOUS SYSTEM AS THE ENGINE FOR VOCAL DEVELOPMENT

Newborn primates are completely dependent upon their caregivers, and the capacity to communicate physiological state is thus particularly important. We wanted to understand how such a system might work. Using electromyography on infant marmoset monkeys, we measured autonomic nervous system (ANS) changes (via heart rate) and respiration while they vocalized in brief social isolation (Zhang & Ghazanfar, 2016). We tested the hypothesis that variable sequences of vocalizations (quantified as Markov chains; Figure 3.2) occur via fluctuations of the ANS and its influence on respiration and (indirectly) the larynx.

As is the case for human infants (MacNeilage, 2008) and songbirds (Sasahara et al., 2015; Tchernichovski et al., 2001), the babbling output of marmoset infants is very rhythmic (Zhang & Ghazanfar, 2016). This rhythmicity suggests that this output is driven by the oscillatory activity of the nervous system. As in developing songbirds (Veit, Aronov, & Fee, 2011), the temporal structure of babbling output by infants is tightly locked to respiration in marmoset monkey infants (Zhang & Ghazanfar, 2016), with respiration providing the power for generating vocal sounds (Ghazanfar & Rendall, 2008). Approximately every second, an utterance phase-locked to the respiratory rhythm is produced. Oddly enough, the time-varying spectral structure of babbling sequences also has a rhythm, but at a rate that is an order of magnitude slower: Spectral entropy, a measure of the noisiness of the sound spectrum (Takahashi et al., 2015; Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000), fluctuates during babbling at a 0.1 Hz frequency (Zhang & Ghazanfar, 2016). This is interesting because the ANS also has a 0.1 Hz rhythm known as the Mayer wave. Thus, in effect, arousal is oscillating at this frequency.

Present in all mammals, the Mayer wave represents perturbations to the baroreflex (Julien, 2006), the homeostatic mechanism by which blood pressure is maintained via changes in heart rate. This results in an unstable negative feedback control loop that generates self-sustained oscillations at its resonance frequency of 0.1 Hz. How can this ANS rhythm account for the 0.1 Hz entropy fluctuations in infant babbling sequences? The 0.1 Hz ANS rhythm modulates the power of the faster, 1 Hz, respiratory rate (Zhang & Ghazanfar, 2016; Figure 3.5). This creates a scenario in which marmoset infants produce vocalizations at a rate of 1 Hz, but the respiratory power is not constant – it is modulated according to a 0.1 Hz rhythm. Lower respiratory power generates noisy (high entropy, cry-like) vocalizations, while higher respiratory power generates more tonal (low entropy, phee-like) vocalizations (Takahashi et al., 2015; Zhang & Ghazanfar, 2016; Figure 3.5). Thus, as revealed by partial coherence measures of heart rate, respiration, and vocalizations, spectral entropy fluctuates at ~0.1 Hz

because both laryngeal tension and respiratory power are modulated by the ANS rhythm (Zhang & Ghazanfar, 2016).
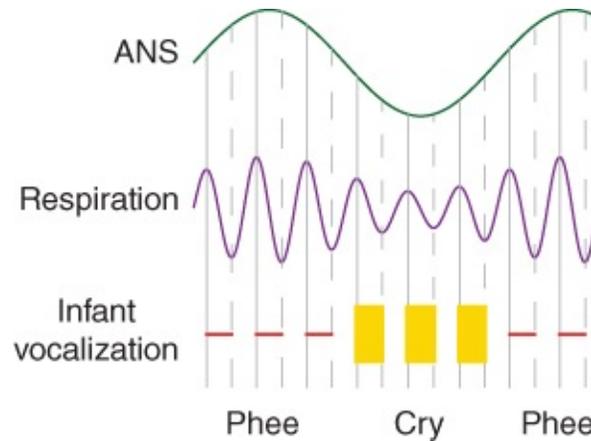
Figure 3.5 Physiological mechanisms of vocal development in marmoset monkeys. Figure shows a schematic illustrating spontaneous vocal production as a function of ANS oscillation and the threshold to vocalize. The continuously produced vocalizations by very young infant marmosets are driven by the natural rhythmic activity of respiration whose power is modulated by the slower, ~0.1 Hz rhythm of the ANS. This consequently changes the quality of the vocalizations so that they fluctuate between high (cry) and low (phee) levels of entropy.

This was the first developmental study in any species (including humans) to quantitatively establish a link between infant vocalizations and physiological states. It demonstrates that infant vocal sequences can potentially encode, and communicate to caregivers, changes in the infant's internal state.

EVOLUTIONARY ORIGINS

The similarities between the developmental trajectories of vocal behavior in humans and marmoset monkeys are striking both in their form and timing (after accounting for the relative rapidity of marmoset development compared to humans). Based on these findings, what can we conclude with regard to how such a developmental system evolved (Borjon & Ghazanfar, 2014; Levinson, 2016)? Typically, with any behavior that two closely related species share, it can be inferred that their last common ancestor also exhibited that behavior. Marmoset monkeys are not very closely related to humans, especially when compared to Old World primates, like chimpanzees or macaque monkeys. If there was evidence that these other primates exhibited similar vocal behaviors, then one could conclude that the species ancestral to both marmosets (and other New World monkeys) and Old World primates (including humans) had the same capacities. Yet, there is no such evidence to date for vocal turn-taking or vocal production learning. Despite suggestions to the contrary (Levinson, 2016), call-and-response behaviors are not the same as turn-taking; they do not exhibit the "coupled" nature of true social interactions observed in marmosets (Takahashi et al., 2013), and in human interactions more generally (De Jaegher, Di Paolo, & Gallagher, 2010; Dumas, de Guzman, Tognoli, & Kelso, 2014; Fogel & Garvey, 2007; Oullier, De Guzman, Jantzen, Lagarde, &

Kelso, 2008). Thus, given the evidence to date, we conclude that vocal turn-taking and vocal learning by marmosets and humans are instances of convergent evolution, possibly as a result of pressures on both species to adopt a cooperative breeding strategy and perhaps through the activation of a shared (homologous) neuronal network (Borjon & Ghazanfar, 2014; Burkart, Hrdy, & van Schaik, 2009).

Cooperative breeding is a prosocial behavior found in only ~3% of mammals (Hrdy, 2005, 2007). Again, among primates, only humans and callitrichids (the primate taxon that includes marmosets) are known to exhibit this strategy (Burkart et al., 2009; Hrdy, 2005). Cooperative breeding occurs when the rearing of infants is greatly reliant on a concerted effort among the breeding female, breeding male, nonbreeding siblings, and occasionally other familiar but unrelated group members (Burkart & van Schaik, 2010; Hrdy, 2005, 2007). In contrast to other monkeys, marmoset caregivers actively and frequently provision food for offspring, and compete with each other for the opportunity to carry offspring (Burkart & van Schaik, 2010; Yamamoto & Lopes, 2004). This cooperative breeding framework, in which nonparents within a social group spontaneously care for offspring other than their own, has been argued to drive uniquely human cognition (Burkart et al., 2009). Vocal turn-taking and its development may thus be specific instances of prosocial behaviors exhibited by humans and marmosets. In this scenario, infants are raised in a social environment in which they may need to compete with others for the attention of caregivers (related or unrelated). Natural selection may have favored individuals who can develop vocal skills (e.g., more mature sounding calls) faster in order to more reliably elicit such care (Zuberbühler, 2012).

CONCLUSIONS

Vocal signals are part of a complex, multidimensional, probabilistic process that includes the physiological states of all the participants (Teramoto et al., 2017; Zeskind, 2013). Our findings demonstrate that infant marmoset monkey calls undergo dramatic changes during the first 2 months of life, both in usage (twitters and trills) and via transformation of cries into mature, adult-like phee calls. The timing of this transition is partly due to maturation, but is also influenced by contingent parental vocal feedback. This is consistent with preverbal vocal development in humans in which two parallel, interactive processes transform infant cries into more mature vocalizations (Byrge et al., 2014; Thelen, 1991). First, natural categories of sounds change as respiratory, laryngeal, and facial components mature. Second, auditory feedback sensitizes infants to certain features of those sounds, and they are modified accordingly. Our findings contrast with previous reports (largely based on squirrel monkeys and macaques) that nonhuman primate vocalizations undergo little or no postnatal change and are impervious to social feedback (Egnor & Hauser, 2004).

From a translational perspective, neurodevelopmental disorders representing two opposite extremes in social communication – autism and Williams syndrome – have been linked to arousal/ANS dysfunction (Bal et al., 2010; Jarvinen & Bellugi, 2013). Thus, to understand early vocal development (and how it may go awry), it is critical to know how arousal/ANS may function in producing individual differences in vocal output. Using marmoset monkeys as a

model system is revealing how rhythmic fluctuations of the ANS may be of critical importance to understanding the early vocal development in humans and other species, acting as the scaffolding upon which vocal development will unfold. These findings are consistent with the increasing recognition that we need to move beyond "imitation" accounts of vocal learning (Syal & Finlay, 2011; Tchernichovski & Marcus, 2014).

ACKNOWLEDGMENTS

REFERENCES

Adolph, K., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, 115, 527–543.

Bal, E., Harden, E., Lamb, D., Van Hecke, A. V., Denver, J. W., & Porges, S. W. (2010). Emotion recognition in children with autism spectrum disorders: Relations to eye gaze and autonomic state. *Journal of Autism and Developmental Disorders*, 40(3), 358–370.

Bezerra, B. M., & Souto, A. (2008). Structure and usage of the vocal repertoire of *Callithrix jacchus*. *International Journal of Primatology*, 29(3), 671–701. doi:10.1007/s10764-008-9250-0

Bloom, K., Russell, A., & Wassenberg, K. (1987). Turn taking affects the quality of infant vocalizations. *Journal of Child Language*, 14, 211–227.

Borjon, J. I., & Ghazanfar, A. A. (2014). Convergent evolution of vocal cooperation without convergent evolution of brain size. *Brain, Behavior & Evolution*, 84, 93–102.

Burkart, J. M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A.,… Meulman, E. J. (2014). The evolutionary origin of human hyper-cooperation. *Nature Communications*, 5, 4747.

Burkart, J. M., Hrdy, S. B., & van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology*, 18, 175–186.

Burkart, J. M., & van Schaik, C. P. (2010). Cognitive consequences of cooperative breeding in primates? *Animal Cognition*, 13(1), 1–19. doi:10.1007/s10071-009-0263-7

Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain-body-behavior networks. *Trends in Cognitive Sciences*, 18, 395–403.

Champagne, F. A., & Curley, J. P. (2008). Maternal regulation of estrogen receptor α methylation. *Current Opinion in Pharmacology*, 8, 735–739.

Choi, J. Y., Takahashi, D. Y., & Ghazanfar, A. A. (2015). Cooperative vocal control in marmoset monkeys via vocal feedback. *Journal of Neurophysiology*, 114, 274–283.

de Castro Leão, A., Duarte Dória Neto, A., & Bernardete Cordeiro de Sousa, M. (2009). New developmental stages for common marmosets (*Callithrix jacchus*) using mass and age variables obtained by K-means algorithm and self-organizing maps (SOM). *Computers in Biology and Medicine*, 39, 853–859.

De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14, 441–447.

Dettling, A. C., Feldon, J., & Pryce, C. R. (2002). Repeated parental deprivation in the infant common marmoset (*Callithrix jacchus*, primates) and analysis of its effects on early development. *Biological Psychiatry*, 52, 1037–1046.

Dumas, G., de Guzman, G. C., Tognoli, E., & Kelso, J. S. (2014). The human dynamic clamp as a paradigm for social interaction. *Proceedings of the National Academy of Sciences, USA*, *111*, E3726–E3734.

Egnor, S. E. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends in Neurosciences*, 27, 649–654.

Egnor, S. E. R., Wickelgren, J. G., & Hauser, M. D. (2007). Tracking silence: Adjusting vocal production to avoid acoustic interference. *Journal of Comparative Physiology A: Neuroethology Sensory Neural and Behavioral Physiology*, 193(4), 477–483. doi:10.1007/s00359-006-0205-7

Elias, G., & Broerse, J. (1996). Developmental changes in the incidence and likelihood of simultaneous talk during the first two years: A question of function. *Journal of Child Language*, 23, 201–217.

Elias, G., Hayes, A., & Broerse, J. (1986). Maternal control of co-vocalization and inter-speaker silences in mother-infant vocal engagement. *Journal of Child Psychology and Psychiatry*, 27, 409–415.

Elowson, A. M., Snowdon, C. T., & Lazaro-Perea, C. (1998). "Babbling" and social context in infant monkeys: Parallels to human infants. *Trends in Cognitive Sciences*, 2, 31–37.

Elowson, A. M., Snowdon, C. T., & Lazaro-Perea, C. (1998). Infant "babbling" in a non-human primate: Complex vocal sequences with repeated call types. *Behaviour*, 135, 643–664.

Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., & Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development*, 34, 569–577.

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *Journal of the Acoustical Society of America*, 102, 1213–1222.

Fogel, A., & Garvey, A. (2007). Alive communication. *Infant Behavior and Development*, 30, 251–257.

Franklin, B., Warlaumont, A. S., Messinger, D., Bene, E., Nathani Iyer, S., Lee, C.-C., . . . Oller, D. K. (2013). Effects of parental Interaction on infant vocalization rate, variability and vocal type. *Language Learning and Development*, 10(3), 279–296.

Garey, J., Goodwillie, A., Frohlich, J., Morgan, M., Gustafsson, J.-A., Smithies, O.,… Pfaff, D. W. (2003). Genetic contributions to generalized arousal of brain and behavior. *Proceedings of the National Academy of Sciences, USA, 100,* 11019–11022.

Ghazanfar, A. A., Flombaum, J. I., Miller, C. T., & Hauser, M. D. (2001). The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*): Playback experiments with long calls. *Journal of Comparative Physiology. A, Sensory, Neural, and Behavioral Physiology*, 187(1), 27–35.

Ghazanfar, A. A., & Rendall, D. (2008). Evolution of human vocal production. *Current Biology,* 18(11), R457 – R460. doi:10.1016/j.cub.2008.03.030

Ghazanfar, A. A., Smith-Rohrberg, D., Pollen, A. A., & Hauser, M. D. (2002). Temporal cues in the antiphonal long-calling behaviour of cottontop tamarins. *Animal Behaviour*, 64, 427–438.

Ghazanfar, A. A., & Takahashi, D. Y. (2014). The evolution of speech: Vision, rhythm, cooperation. *Trends in Cognitive Science*, 18, 543–553.

Goldstein, M. H., Bornstein, M. H., Schwade, J. A., Baldwin, F., & Brandstadter, R. (2009). The value of vocalizing: Five-month-old infants associate their own non-cry vocalizations with responses from caregivers. *Child Development*, 80(3), 636–644.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, USA,* 100(13), 8030–8035. doi:10.1073/Pnas.1332441100

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, 19(5), 515–523. doi:10.1111/J.1467-9280.2008.02117.X

Gottlieb, G. (1992). *Individual development & evolution: The genesis of novel behavior*. New York, NY: Oxford University Press.

Gros-Louis, J., West, M. J., Goldstein, M. H., & King, A. P. (2006). Mothers provide differential feedback to infants' prelinguistic sounds. *International Journal of Behavioral Development*, 30(6), 509–516. doi:10.1177/0165025406071914

Gros-Louis, J., West, M. J., & King, A. P. (2010). Comparative perspectives on the missing link: Communicative pragmatics. In M. S. Blumberg, J. H. Freeman, & S. R. Robinson (Eds.),

The Oxford handbook of developmental behavioral neuroscience (pp. 684–707). Oxford, UK: Oxford University Press.

Gros-Louis, J., West, M. J., & King, A. P. (2014). Maternal responsiveness and the development of directed vocalizing in social interactions. *Infancy, 19*, 385–408.

Harris, R. A., Tardif, S. D., Vinar, T., Wildman, D. E., Rutherford, J. N., Rogers, J., . . . Aagaard, K. M. (2014). Evolutionary genetics and implications of small size and twinning in callitrichine primates. *Proceedings of the National Academy of Sciences, USA, 111*, 1467–1472.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother-infant interaction. *Frontiers in Psychology*, 6, 1492. doi:10.3389/fpsyg.2015.01492

Hrdy, S. B (2005). Evolutionary context of human development: The cooperative breeding model. In C. S. Carter, L. Ahnert, K. E. Grossmann, S. B. Hrdy, M. E. Lamb, S. W. Porges, & N. Sachser (Eds.), *Attachment and bonding: A new synthesis* (pp. 9–32). Cambridge, MA: MIT Press.

Hrdy, S. B. (2007). Evolutionary context of human development: The cooperative breeding model. In C. A. Salmon & T. K. Shackelford (Eds.), *Family relationships: An evolutionary perspective* (pp. 39–68). New York, NY: Oxford University Press.

Hsu, H.-C., & Fogel, A. (2001). Infant vocal development in a dynamic mother-infant communication system. *Infancy*, 2(1), 87–109.

Jarvinen, A. M., & Bellugi, U. (2013). What does Williams syndrome reveal about the determinants of social behavior? *Frontiers in Human Neuroscience*, 7, 321. doi:10.3389/fnhum.2013.00321

Jasnow, M., & Feldstein, S. (1986). Adultlike temporal characteristics of mother-infant vocal interactions. *Child Development*, 57, 754–761.

Julien, C. (2006). The enigma of Mayer waves: Facts and models. *Cardiovascular Research*, 70(1), 12–21. doi:10.1016/j.cardiores.2005.11.008

Kajikawa, S., Amano, S., & Kondo, T. (2004). Speech overlap in Japanese mother-child conversations. *Journal of Child Language*, 31, 215–230.

Kent, R. D., & Murray, A. D. (1982). Acoustic features of infant vocalic utterances at 3, 6, and 9 months. *Journal of the Acoustical Society of America*, 72, 353–365.

Law, A. J., Pei, Q., Feldon, J., Pryce, C. R., and Harrison, P. J. (2009). Gene expression in the anterior cingulate cortex and amygdala of adolescent marmoset monkeys following parental separations in infancy. *International Journal of Neuropsychopharmacology*, 12, 761–772.

Levinson, S. C. (2016). Turn-taking in human communication – origins and implications for

language processing. *Trends in Cognitive Sciences*, 20(1), 6–14. doi:10.1016/j.tics.2015.10.010

Locke, J. L. (1995). *The child's path to spoken language*. Cambridge, MA: Harvard University Press.

MacNeilage, P. F. (2008). *The origin of speech*. Oxford, UK: Oxford University Press.

Margoliash, D., & Tchernichovski, O. (2015). Marmoset kids actually listen. *Science*, 349(6249), 688–689.

Masataka, N. (1993). Effects of contingent and noncontingent maternal stimulation on the vocal behaviour of three- to five-month old Japanese infants. *Journal of Child Lanuage*, 20, 303–312.

Moriceau, S., & Sullivan, R. M. (2005). Neurobiology of infant attachment. *Developmental Psychobiology*, 47, 230–242.

Oller, D. K. (2000). *The emergence of the speech capacity*. Mahwah, NJ: Erlbaum.

Oullier, O., De Guzman, G. C., Jantzen, K. J., Lagarde, J., & Kelso, J. A. S. (2008). Social coordination dynamics: Measuring human bonding. *Social Neuroscience*, 3(2), 178–192. doi:10.1080/17470910701563392

Pistorio, A. L., Vintch, B., & Wang, X. (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *Journal of the Acoustical Society of America*, 120, 1655–1670.

Pryce, C. R., Aubert, Y., Maier, C., Pearce, P. C., & Fuchs, E. (2011). The developmental impact of prenatal stress, prenatal dexamethasone and postnatal social stress on physiology, behaviour and neuroanatomy of primate offspring: Studies in rhesus macaque and common marmoset. *Psychopharmacology*, 214, 33–53.

Roy, S., Miller, C. T., Gottsch, D., & Wang, X. (2011). Vocal control by the common marmoset in the presence of interfering noise. *The Journal of Experimental Biology*, 214(Pt 21), 3619–3629. doi:10.1242/jeb.056101

Sasahara, K., Tchernichovski, O., Takahasi, M., Suzuki, K., & Okanoya, K. (2015). A rhythmic landscape approach to the developmental dynamics of birdsong. *Journal of the Royal Society Interface*, 12, 20150802.

Scheiner, E., Hammerschmidt, K., Jurgens, U., & Zwirner, P. (2002). Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Phonetics*, 16, 509–529.

Schultz-Darken, N., Braun, K. M., & Emborg, M. E. (2015). Neurobehavioral development of common marmoset monkeys. *Developmental Psychobiology*. doi:10.1002/dev.21360

Seyfarth, R. M., & Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour*, 34, 1640–1658.

Snowdon, C. T., & Cronin, K. A. (2007). Cooperative breeders do cooperate. *Behavioural Processes*, 76, 138–141.

Stern, D. N., Jaffe, J., Beebe, B., & Bennett, S. L. (1975). Vocalizing in unison and in alternation: Two modes of communication within the mother-infant dyad. *Annals of the New York Academy of Sciences*, 263, 89–100.

Syal, S., & Finlay, B. L. (2011). Thinking outside the cortex: Social motivation in the evolution and development of language. *Developmental Science*, 14, 417–430.

Takahashi, D. Y., Fenley, A. R., & Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmosets. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370: 20150370.

Takahashi, D. Y., Fenley, A. R., Teramoto, Y., Narayanan, D. Z., Borjon, J. I., Holmes, P., & Ghazanfar, A. A. (2015). The developmental dynamics of marmoset monkey vocal production. *Science*, 349(6249), 734–738. doi:10.1126/science.aab1058

Takahashi, D. Y., Liao, D. A., & Ghazanfar, A. A. (2017). Vocal learning via social reinforcement by infant marmoset monkeys. *Current Biology*, 27, 1844–1852.

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

Tchernichovski, O., & Marcus, G. (2014). Vocal learning beyond imitation: Mechanisms of adaptive vocal development in songbirds and human infants. *Current Opinion in Neurobiology*, 28, 42–47.

Tchernichovski, O., Mitra, P. P., Lints, T., & Nottebohm, F. (2001). Dynamics of the vocal imitation process: How a zebra finch learns its song. *Science*, 291, 2564–2569.

Tchernichovski, O., Nottebohm, F., Ho, C. E., Pesaran, B., & Mitra, P. P. (2000). A procedure for an automated measurement of song similarity. *Animal Behavior*, 59(6), 1167–1176. doi: 10.1006/anbe.1999.1416

Teramoto, Y., Takahashi, D., Holmes, P., & Ghazanfar, A. A. (2017). Vocal development in a Waddington landscape. *eLife*, 6, e20782.

Thelen, E. (1991). Motor aspects of emergent speech: A dynamic approach. In N. A. Krasnegor, D. M. Rumbaugh, & R. L. Schiefelbusch (Eds.), *Biological and behavioral determinants of language development* (pp. 339–362). Hillsdale, NJ: Erlabaum.

Van Puyvelde, M., Loots, G., Meys, J., Neyt, X., Mairesse, O., Simcock, D., & Pattyn, N. (2015). Whose clock makes yours tick? How maternal cardiorespiratory physiology influences

newborns' heart rate variability. *Biological Psychology, 108,* 132–141. doi:10.1016/j.biopsycho.2015.04.001

Veit, L., Aronov, D., & Fee, M. S. (2011). Learning to breathe and sing: Development of respiratory-vocal coordination in young songbirds. *Journal of Neurophysiology, 106,* 1747–1765.

Yamamoto, M. E., & Lopes, F. D. A. (2004). Effect of removal from the family group on feeding behavior by captive *Callithrix jacchus. International Journal of Primatology, 25,* 489–500. doi:10.1023/B:IJOP.0000019164.98756.9c

Zeskind, P. S. (2013). Infant crying and the synchrony of arousal. In E. Altenmüller, S. Schmidt, & E. Zimmermann (Eds.), *The evolution of emotional communication: From sounds in nonhuman mammals to speech and music in man* (pp. 155–174). Oxford, UK: Oxford University Press.

Zhang, Y. S., & Ghazanfar, A. A. (2016). Perinatally influenced autonomic nervous system fluctuations drive infant vocal sequences. *Current Biology, 26,* 1249–1260. doi:10.1016/j.cub.2016.03.023

Zuberbühler, K. (2012). Cooperative breeding and the evolution of vocal flexibility. In M. Tallerman & K. R. Gibson (Eds.), *The Oxford handbook of language evolution* (pp. 71–81). New York, NY: Oxford University Press.