

Convergent Evolution of Vocal Cooperation without Convergent Evolution of Brain Size

Jeremy I. Borjon^{a, b} Asif A. Ghazanfar^{a, b, c}

^aPrinceton Neuroscience Institute and Departments of ^bPsychology and ^cEcology and Evolutionary Biology, Princeton University, Princeton, N.J., USA

Key Words

Vocal communication · Turn taking · Vocalizations · Common marmosets · Humans

Abstract

One pragmatic underlying successful vocal communication is the ability to take turns. Taking turns – a form of cooperation – facilitates the transmission of signals by reducing the amount of their overlap. This allows vocalizations to be better heard. Until recently, non-human primates were not thought of as particularly cooperative, especially in the vocal domain. We recently demonstrated that common marmosets (*Callithrix jacchus*), a small New World primate species, take turns when they exchange vocalizations with both related and unrelated conspecifics. As the common marmoset is distantly related to humans (and there is no documented evidence that Old World primates exhibit vocal turn taking), we argue that this ability arose as an instance of convergent evolution, and is part of a suite of prosocial behavioral tendencies. Such behaviors seem to be, at least in part, the outcome of the cooperative breeding strategy adopted by both humans and marmosets. Importantly, this suite of shared behaviors occurs without correspondence in encephalization. Marmoset vocal turn taking demonstrates that a large brain size and

complex cognitive machinery is not needed for vocal cooperation to occur. Consistent with this idea, the temporal structure of marmoset vocal exchanges can be described in terms of coupled oscillator dynamics, similar to quantitative descriptions of human conversations. We propose a simple neural circuit mechanism that may account for these dynamics and, at its core, involves vocalization-induced reductions of arousal. Such a mechanism may underlie the evolution of vocal turn taking in both marmoset monkeys and humans.

© 2014 S. Karger AG, Basel

Introduction

Humans orchestrate behavior through spoken language, gestures and, more often than not, a combination of speech and gestures. The evolutionary origins for this uniquely human form of communication are mysterious for many reasons. Primarily, there are few ways to reconstruct what ancestral vocalizations sounded like and how they were used. Moreover, soft tissues such as the brain and parts of the vocal apparatus do not fossilize. Relatedly, the dynamic social behavior that may have driven the evolution of more sophisticated communication is also difficult to reconstruct from the archaeological record. That

said, we can infer from indirect evidence that early hominids were quite social. For example, evidence of hominid fire use dates to at least one million years ago [Berna et al., 2012] and evidence for a group of humans using fire in the form of a hearth has recently been pegged to at least 300,000 years ago [Shahack-Gross et al., 2014]. Whether our hominid ancestors gestured or spoke to one another (or did both) in order to light the hearth is an open question. Yet necessary and foundational to any instance of successful cooperative communication – no matter the modality or sophistication – is the capacity to take turns.

Cooperative Communication: Turn Taking in Humans and Marmoset Monkeys

One way to enhance signal quality during communication is to prevent interference through taking turns. By pausing after transmission, a sender allows signals from other individuals to transpire and be heard before another signal is emitted. The elimination of overlap via turn taking increases the likelihood of the signal being heard accurately. As a consequence, an exchange of signals between two or more individuals has a structure. A successful instance of human vocal turn taking, for example, would involve person 1 speaking while person 2 attends, followed by a response from person 2, be it a statement or an indication for person 1 to continue speaking. The imposition of such structure during communication is a key feature of human social development [Jasnow and Feldstein, 1986; Jaffe et al., 2001]. Mothers will use pauses to establish rhythmicity in vocal, gestural and gaze-driven interactions with their infant [Jaffe et al., 2001]. The establishment of this rhythmicity has been demonstrated to result in measurable gains in word learning [Yu and Smith, 2012]. Thus, we can envision the imposition of structure as a scaffold upon which a developing child can build mature communication strategies. Of course, not all conversations in humans adhere to a strict turn taking rule. Face-to-face conversations between familiar individuals tend to be rapid speech exchanges with a substantial amount of overlap [O’Conaill et al., 1993]. However, between unfamiliar individuals or in unfamiliar contexts, or when individuals are out of view of one another, conversational exchanges solely via speech signals become more regulated, falling back onto basic turn taking behavior [O’Conaill et al., 1993; Sellen, 1995]. Thus, in general and throughout human development, turn taking is a foundational principle upon which more flexible communication can occur.

Given its central importance in everyday human social interactions, it is natural to ask how vocal turn taking, a form of cooperation, evolved. It has been argued that human cooperative vocal communication is unique and, essentially, evolved in three steps (put forth most cogently by Tomasello [2008], but see also Hewes [1973] and Rizzolatti and Arbib [1998] for similar scenarios). First, ape-like ancestors used manual gestures to point and direct the attention of others. Second, later ancestors with prosocial tendencies used manual gestures to mediate shared intentionality. Finally, and most mysteriously, a transition from primarily gestural to primarily vocal forms of cooperative communication came about, perhaps in order to express shared intentionality more efficiently. Typically, no primates other than humans are thought to exhibit cooperative vocal communication, the implication being that communication via turn taking requires a big brain and complex cognitive mechanisms.

Conversely, we hypothesize that vocal turn taking could have evolved through a voluble and prosocial ancestor without the prior scaffolding of manual gestures or big brains. To test this hypothesis, we studied the vocal exchanges of a small New World primate found in South America: the common marmoset (*Callithrix jacchus*) [Takahashi et al., 2013]. Marmosets are part of the Callitrichinae subfamily of the Cebidae family of New World primates. The common marmoset is approximately 20 cm tall, weighs an average of 400 g, and exhibits no sexual dimorphism in body size (fig. 1a). Marmosets display little evidence of shared intentionality and they do not produce manual gestures. Like humans, they are cooperative breeders and voluble. Marmosets are among the very few primate species that form pair bonds and exhibit biparental and alloparental care of infants [Zahed et al., 2008]. These cooperative care behaviors are thought to scaffold prosocial motivational and cognitive processes, such as attentional biases toward monitoring others, the ability to coordinate actions, increased social tolerance and increased responsiveness to others’ signals [Snowdon and Cronin, 2007; Burkart and van Schaik, 2010]. Besides humans, and perhaps to some extent in bonobos [Hare et al., 2007], this suite of prosocial behaviors is not typically seen in other primate species.

When out of visual contact, marmoset monkeys and other callitrichid primates will participate in vocal exchanges with conspecifics [Ghazanfar et al., 2001, 2002; Miller and Wang, 2006; Chen et al., 2009]. In the laboratory and in the wild, marmosets typically use ‘phee’s, a high-pitched vocalization that can be monosyllabic or multisyllabic, as their contact call (fig. 1b, c) [Bezerra and

Souto, 2008]. A phee call contains information about sex, identity and social group [Norcross and Newman, 1993; Miller et al., 2010]. As for humans, marmoset conversations occur spontaneously with another conspecific regardless of pair-bonding status or relatedness. Marmoset vocal exchanges can last as long as 40 min and have a temporal structure that is strikingly similar to the turn taking rules used by humans [Stivers et al., 2009; Takahashi et al., 2013]. First, there are rarely if ever overlapping calls (i.e. no interruptions and thus, no interference). Second, there is a consistent silent interval between utterances across two individuals (fig. 1b, c). The similarities do not stop there.

In humans, dynamical system models incorporating coupled oscillator-like mechanisms are thought to account for the temporal structure of conversational turn taking and other social interactions [Chapple 1970; Oullier et al., 2008; Schmidt and Morr, 2010]. In the vocal domain, such mechanisms have two basic features: (1) periodic coupling in the timing of utterances across two interacting individuals (fig. 1d) and (2) entrainment, where if the timing of one individual's vocal output quickens or slows, the other's follows suit. The vocal exchanges of marmoset monkeys share both of these features [Takahashi et al., 2013]. Thus, marmoset vocal communication, like human speech communication, can be modeled as loosely coupled oscillators. As a mechanistic description of vocal turn taking, coupled oscillators are advantageous since they are consistent with the data from speech processing that brain oscillations are critical to temporal structure [Giraud and Poeppel, 2012; Hasson et al., 2012] and its evolution [Ghazanfar and Takahashi, 2014]. Further, such oscillations do not necessarily require any higher-order cognitive capacities to operate [Takahashi et al., 2013]. In other words, a coupled oscillator can occur without the involvement of a big brain, something worth considering given the marmoset monkey's small encephalization quotient compared to great apes and humans [Jerison, 1973] (fig. 2).

A schematic of how two marmosets can be coupled to each other via vocalizations is illustrated in figure 3. Consider first, a marmoset producing spontaneous vocalizations by itself with no conspecific responding (fig. 3a). According to the model, preceding a vocalization, this marmoset experiences a drive (perhaps an increase in arousal due to social isolation) to produce a call. As the call is emitted, the sender hears its own signal and this auditory feedback inhibits the drive to continue vocalizing. Once the vocalization ends, the inhibition on the drive to vocalize is released and then, after some refrac-

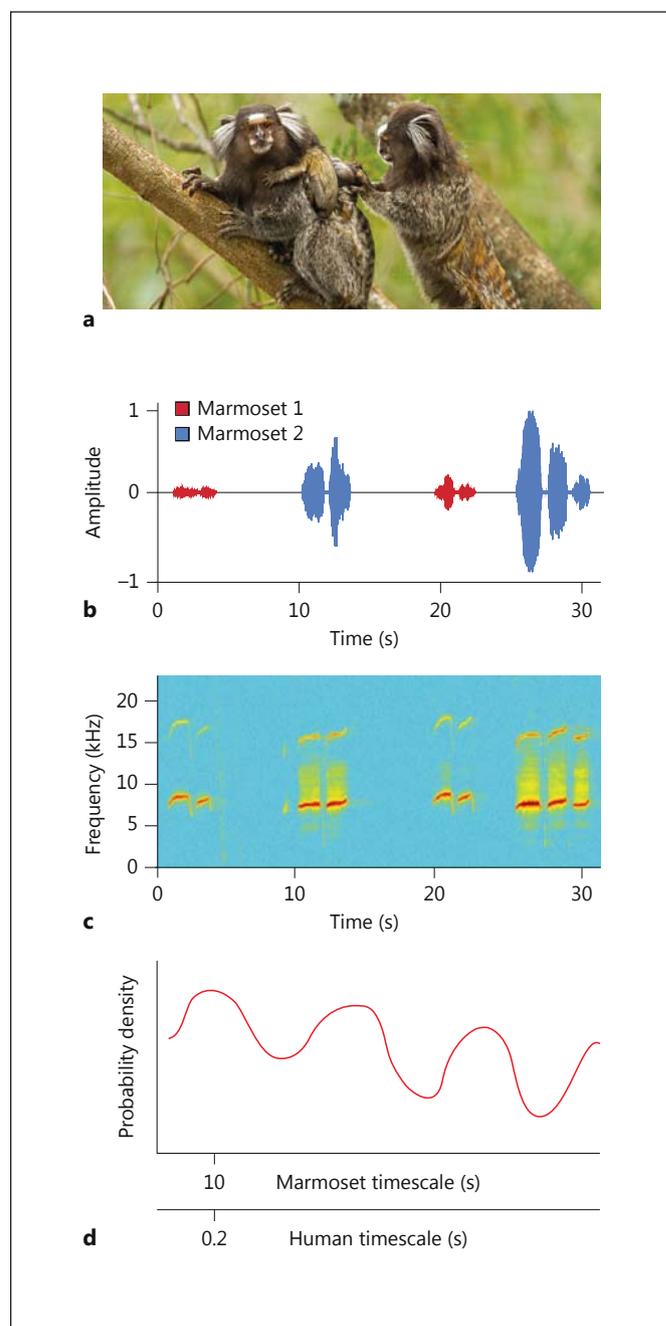


Fig. 1. **a** A pair of common marmosets with two infants. (Copyright, Francesco Veronesi; used under Creative Commons license.) **b, c** Waveform and spectrogram of an example phee call exchange. Used with permission from Takahashi et al. [2013]. **d** Schematic of a cross-correlation plot demonstrating the temporal structure underlying turn taking in marmosets and humans. Note the difference in timescales, whereby the first peak of the cross-correlation is at 10 s for marmosets and 0.2 s for humans.

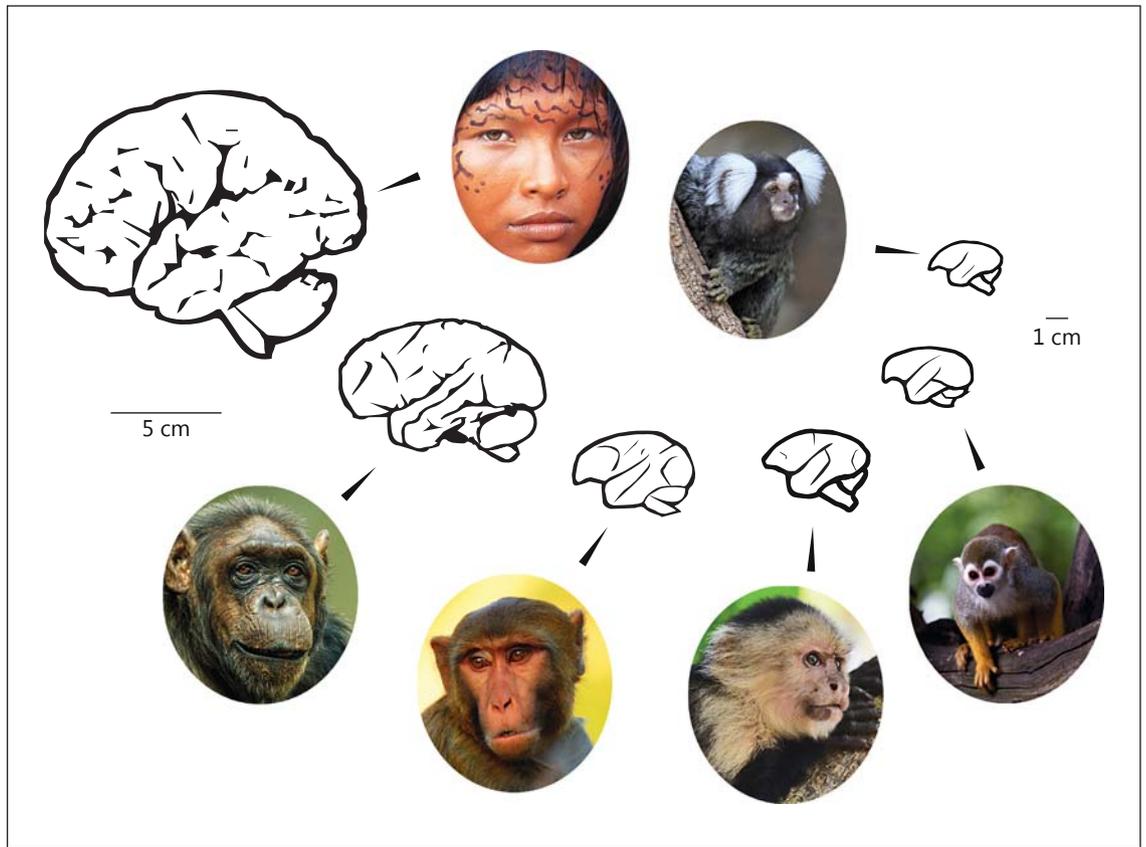


Fig. 2. A schematic demonstrating the variety of brain sizes in the primate order. To date, marmosets and humans are the only two primate species demonstrated to vocally cooperate via an extended series of vocalizations with any other conspecific. (Photographs

used under Creative Commons license; photographers as follows: marmoset by Simon Harris; chimpanzee by S.B.J.; rhesus macaque by Stacey Osburn; capuchin by Dave Spangenburg; squirrel monkey by Tracy Lynn.)

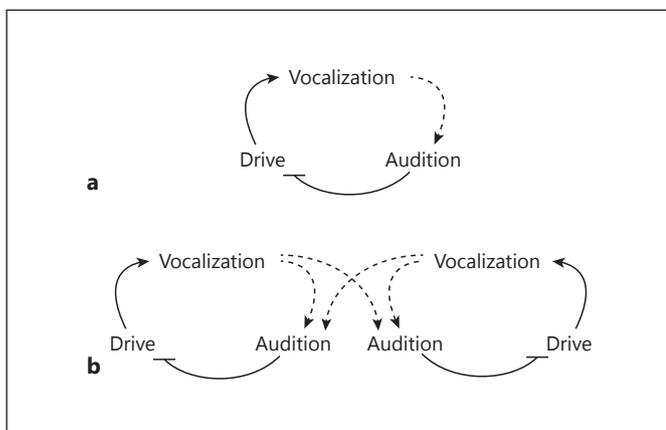


Fig. 3. a An oscillator model of marmoset vocal communication while a marmoset is alone. **b** By introducing a second marmoset, the vocalization of the first can alter the activity of the second, thereby creating a coupled oscillator.

tory period, the marmoset produces another call. Thus, we have a simple oscillatory mechanism. Critically, the inhibitory influence of the auditory input on the drive to vocalize provides a natural mechanism for two marmosets to begin a vocal exchange and to vocally couple with each other and thereby cooperate at a systems level (fig. 3b). A receiver marmoset out of sight from the sender hears the sender's phee call which inhibits its own drive to vocalize, preventing overlapping calls. Once the call ends, the drive to vocalize in the receiver is disinhibited and can hit threshold, whereby another call is initiated and the conversation continues. We now have a coupled oscillator. Since phee calls contain socially salient indexical information about the sender – and turn taking prevents signal degradation caused by overlap – cooperation and social interactions are facilitated and reinforced. Overall, this model is consistent with the affective- and

arousal-based theories of primate communication put forth by Owren et al. [2010] and Rendall et al. [2009].

There is one major difference between human and marmoset monkey vocal exchanges: timescale. Across two individuals, the average human silent interval between utterances is approximately 240 ms, with a very large amount of variance [Stivers et al., 2009]. In marmosets, this interval is much longer, approximately 3–5 s [Takahashi et al., 2013]. This difference in timing can be explained in terms of ‘units of perception’ [Ghazanfar et al., 2001]. Since humans employ words as symbolic units of meaning, our minimal unit of perception in a conversation is on the order of a word or syllable [Chandrasekaran et al., 2009; Lerner et al., 2011]. Marmosets, on the other hand, may use the entire call, which may be multisyllabic and on the order of approximately 3–5 s [Takahashi et al., 2013]. This is consistent with what is known about the units of perception in the contact calling behavior of the closely related cotton-top tamarin [Ghazanfar et al., 2001]. The proposed model in figure 3 can work on either of these time scales.

It is clear that many species of animals exchange vocalizations, but these usually take the form of a single ‘call-and-response’ as opposed to an extended sequence of vocal interactions. For example, naked mole-rats [Yosida et al., 2007], squirrel monkeys [Masataka and Biben, 1987], Japanese macaques [Sugiura, 1998], large-billed crows [Kondo et al., 2010], bottlenose dolphins [Nakahara and Miyazaki, 2011] and some anurans [Zelick and Narins, 1985; Grafe, 1996] are all capable of simple call-and-response behaviors. However, there are some instances of extended, coordinated vocal exchanges as well. The chorusing behaviors of anurans and insects are the result of competitive coordination between neighboring males in an attempt to attract females [Greenfield, 1994]. Another form of vocal coordination, duetting, occurs between pair-bonded songbirds [Logue et al., 2008] and gibbons [Mitani, 1985]. Like the vocal exchanges of the marmoset monkey, these too indicate that a high-level cognitive capacity is not necessary for vocal turn taking. However, unlike marmosets and humans, these types of vocal coordination occur within the limited contexts of competitive interactions or pair bonds. This inability to flexibly use vocal turn taking across conspecifics, regardless of pair-bonding status or relatedness, stands in stark contrast to the cooperative vocal behavior in marmosets and humans, who frequently initiate and sustain vocal exchanges with any conspecific [Takahashi et al., 2013]. The frequency of vocal interactions with nonrelated and nonpair-bonded individuals may be key to understanding the neural circuitry underlying vocal turn taking.

Why Do Humans and Marmoset Monkeys Exhibit Vocal Turn Taking?

Despite the stark contrasts in brain size, marmoset monkeys and humans can vocally cooperate with any conspecific regardless of pair-bonding status or relatedness. In other primates, vocal cooperation is restricted to certain types of conspecifics and is rarely cooperative. As some 40 million years have passed since the Old World and New World primate lineage split [Steiper and Young, 2006], we argue that vocal cooperation arose as an instance of convergent evolution of behavior, but perhaps through the activation of a shared (homologous) neuronal network. To unpack this argument, we will first examine the mechanisms driving convergent evolution in general, then examine which pressures are shared between marmosets and humans. Finally, we will argue that neuromodulatory changes to a common neural circuit could serve as a viable explanation for the convergent evolution of vocal cooperative behavior in marmosets and humans.

Natural selection acts on the behavior of an individual, not on isolated neural structures [Padberg et al., 2007]. Evolutionary changes to organisms occur as parallel processes, involving the interplay between the environment, an individual’s musculoskeletal structures and the brain [Krubitzer and Kaas, 2005]. In fact, given similar pressures, identical cortical areas can emerge by virtue of the constrained nature of cortical development [Krubitzer and Kaas, 2005; Padberg et al., 2007]. For instance, cortical areas 2 and 5, associated with motor planning and coordination, are very well developed in macaques, an Old World monkey, as well as in *Cebus* monkeys, a New World monkey. In other New World primates, however, areas 2 and 5 are either absent or poorly developed [Padberg et al., 2005]. The reason for this becomes apparent when we consider that, of the New World primates, *Cebus* monkeys are the only species known to use a precision grip – a grip that is common among Old World monkeys and apes (including humans) [Padberg et al., 2007]. Such a grip alters the mechanics of the hand, facilitating object manipulation through an increase in the possible number of digit configurations. In the context of homologous neurodevelopmental trajectories, convergent evolution of biomechanics results in a constrained and predictable change in neural circuitry. Through this process, we see the emergence of identical cortical areas, in this case areas 2 and 5, across two species separated by a common ancestor 40 million years ago. It is possible that similar processes may be at play in marmosets and humans whereby they exhibit homologous neurodevelopmental trajec-

ries that are influenced by convergent features of their socioecological environment – specifically, prosociality and volubility.

Cooperative breeding, a prosocial behavior, is only found in about 3% of mammals [Hrdy, 2005]. Of those mammals, callitrichids are the only non-human primates known to exhibit this strategy [Hrdy, 2005; Burkart et al., 2009]. For marmosets, 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 [Goldizen, 1990; Hrdy, 2005; Burkart and van Schaik, 2010]. Marmoset caregivers actively and frequently provision food for offspring [Yamamoto and Lopes, 2004; Burkart and van Schaik, 2010]. In contrast, infants of other New World monkeys, such as capuchins, rarely receive nonmaternal care, and other group members seldom share food with them. Moreover, marmosets forage while carrying infants, and often compete to carry infants [Santos et al., 1997; Snowdon and Cronin, 2007]. The result of this foraging experience is gains in social learning by infant marmosets. Infant marmosets are highly unlikely to consume novel food items when away from adults unless the infants observed an adult eating the same food [Yamamoto and Lopes, 2004; Voelkl et al., 2006; Vitale and Queyras, 2010]. In contrast, capuchin infants do not exhibit any similar social learning preference: they are just as likely to consume novel foods when they are away or with adults [Fragaszy et al., 1997]. 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]. Experimentally, there is support for the notion that cooperative breeding leads to gains in social cognition [Burkart and van Schaik, 2010]. When compared to their larger, independently breeding sister taxa Cebidae (which include squirrel monkeys and capuchin monkeys), marmosets exhibit greater socio-cognitive performance and much greater evidence of prosociality [Burkart and van Schaik, 2010].

Typically, benefits in socio-cognitive performance are thought to be the result of an increase in the size of the brain, specifically the neocortex [Dunbar, 2009]. This is not the case with the marmoset monkey. Marmoset brains are approximately 4 times smaller than the closely related New World squirrel monkeys and 6 times smaller than those of capuchin monkeys (fig. 2) [Herculano-Houzel et al., 2007]. However, in marmoset monkeys, selective pressures on body size may have led to a more efficient brain. Despite their small overall brain size relative

to other primates, marmosets possess three times as many cortical areas as rats that exhibit a similar body size. To put it another way, a marmoset brain is five times larger than the brain of a rat; the brain of a marmoset is 2.7% of its body size (similar to humans) [Stephan et al., 1980], while the brain of the rat is less than 1% of its own body weight [Donkelaar and Nicholson, 1998]. Thus, these prosocial behaviors and, by extension, cooperative vocal communication must be mediated by the particular form and/or modulation of neural circuits that are not simply a consequence of increasing brain size.

Indeed, differential neuromodulatory influences on identical neural networks can result in very different behavioral outcomes [Katz, 2011; Marder, 2012]. Take for instance maternal behavior in rats. In female rats, the medial preoptic area (MPOA) is critical for the appearance of maternal behavior, especially when activated by the hormone estrogen [Numan, 1994]. A lesion to the MPOA of a lactating female rat will cease any maternal behavior [Gray and Brooks, 1984]. In the wild and in the laboratory, male rats do not readily exhibit maternal care to their pups; however, following the elimination of testosterone (via castration) and exposure to newborn pups, male rats begin to exhibit maternal care [Bridges et al., 1974]. Importantly, this normally latent male ‘maternal’ behavior is also mediated by the MPOA, as lesions to the MPOA will result in the termination of the maternal behavior [Rosenblatt et al., 1996]. Therefore, the only difference between male and female rats in regards to maternal care may be the presence of testosterone at the MPOA. What this tells us is that, in the course of evolution, it is easier to modify existing neural pathways rather than create new ones. In light of this, we suggest that the variety of ways in which vocalizations are exchanged in primates – from the unrestricted nature of human and marmoset vocal cooperation to the restricted duetting of gibbons and the calls and responses of squirrel monkeys – suggests that a shared neural connectivity homology may exist but is activated in different ways and/or to different degrees. This inconsistent utilization across closely related species perhaps implies variation in neuromodulatory action on the very same or similar circuit [Donaldson and Young, 2008; Katz, 2011; Marder, 2012]. Below, we conclude with a discussion of the possible neuromechanical underpinnings of vocal cooperation.

There is a large overlap in structures considered to underlie social behavior (and vocal behavior, in particular) as well as those underlying motivation and emotional arousal [Cardinal et al., 2002; Syal and Finlay, 2011]. The shared neural architecture enables a pathway in which

social interaction is 'rewarded' via the attachment of motivational value through arousal levels [Syal and Finlay, 2011]. That is, fluctuating arousal levels could be a driving factor in vocal turn taking. As turn taking is inherently rhythmic, it is not unreasonable to consider that this behavior exapted onto currently existing rhythmic mechanisms, such as respiration, heart rate or rhythmic fluctuations in neuroendocrine levels. Changes in these rhythmic mechanisms alter the autonomic nervous system, leading to changes in arousal [Kayaba et al., 2003; Shahid et al., 2012]. While many neuromodulators likely contribute to the production of a vocalization, here is one illustrative example. Orexin is a neuropeptide released by the hypothalamus and involved in the autonomic modulation of breathing and heart rate [Nattie and Li, 2012]. Orexin knockout mice possess a lower basal blood pressure compared to wild types [Kayaba et al., 2003] and intracerebroventricular injection of orexin results in an increase in heart rate and blood pressure [Samson et al., 1999], as well as respiration [Zhang et al., 2005]. Intriguingly, neurotoxic lesions and chemical blocking of regions of the hypothalamus dense in orexin neurons results in a reduction of vocalizations, as well as a lowering of blood pressure and pulse during contextual fear responses [Furlong and Carrive, 2007; Chen et al., 2014]. As orexin is secreted in the brain, there are projections and receptors sensitive to this peptide throughout. In the rat, projections from the arcuate nucleus of the hypothalamus to anterior cingulate cortex (ACC) are known to exhibit orexigenic and anorexigenic phenotypes [Kampe et al., 2009]. The ACC is one particular neural area in which vocal production and arousal intertwine. In cats, rats and monkeys, electrical stimulation of the ACC produces vocalizations [Devinsky et al., 1995; Paus, 2001]. These evoked vocalizations express an internal emotional state [Lewin and Whitty, 1960; Paus, 2001] and, thus, we can consider it one of the 'drives' to vocalize (fig. 3a). Importantly, the human ACC is involved in speech production [Sörös et al., 2006].

Is there evidence for a role of arousal on speech production in humans? Due to the presumption that speech is a purely neocortical-based phenomenon, it is typically argued that human communication is a behavior distanced from the state-governed vocalizations emitted by our non-human primate relatives [Rendall et al., 2009; Owren et al., 2010]. This is a puzzling notion, especially if one takes into consideration the clinical psychology literature demonstrating a relationship between autonomic dysfunction and language production. For example, one characteristic symptom of anxiety disorders and mania

are episodes of rapid speaking, often termed 'pressured speech' [van Kammen and Murphy, 1975; Willson et al., 2005; Pereira et al., 2014]. Manic episodes usually entail a measurable change in autonomic sensitivity [Henry et al., 2010], suggesting that a dysfunctional autonomic system underlies mania [McMeekin, 2002; Latalova et al., 2010; Levy, 2013]. Indeed, those with affective disorders often demonstrate an enlargement of the third ventricle, suggesting dysfunction to the hypothalamic nuclei governing autonomic processes [Bhadoria et al., 2003]. Separate from the clinical literature, speech motor coordination is influenced by fluctuations in the autonomic nervous system in both adults and children [Kleinow and Smith, 2006]. Further, an increase in arousal is often accompanied by measureable changes in vocal prosody [Wiethoff et al., 2008]. Changes in affect are indicated by changes to the fundamental frequency of the voice [McRoberts et al., 1995]. Decelerations in heart rate have also been demonstrated after speech production in typical humans [Peters and Hulstijn, 1984]. Thus, human vocal communication, including speech, is greatly influenced by the autonomic system and fluctuating levels of arousal.

We propose that vocalization-induced reductions of arousal evolved into the mechanism underlying turn taking in marmosets, and perhaps humans as well. In the marmoset monkey, auditory contact through phee calls has been shown to cause a reduction in cortisol levels [Rukstalis and French, 2005]. Unmodulated arousal is known to be damaging and even fatal to primates [Uno et al., 1989]. Thus, modulating arousal levels and cooperative behaviors are intertwined: both producing and hearing vocalizations reduces arousal levels (fig. 3) [Rukstalis and French, 2005]. Consistent with this idea, it has been posited that human conversations play the same role as social grooming seen in apes and monkeys [Dunbar, 1998], suggesting that, in a sense, humans and marmosets have achieved 'grooming at a distance' [Takahashi et al., 2013].

Conclusion

Underlying any mode of cooperative communication is the ability to take turns. Here, we have examined an example of convergent evolution of vocal turn taking in the marmoset monkey and human. As cooperative communication, sociality and arousal levels are fundamentally intertwined, we propose that the vocal turn taking mechanism was scaffolded upon a preexisting arousal-regulation mechanism. In our view, evolutionary pres-

tures leading to an arousal-based coupled oscillatory mechanism for vocal turn taking sets the groundwork for more complex communication to arise by enhancing signal quality. Such a mechanism stands in stark contrast to the traditional argument that cooperative vocal communication arose after being scaffolded onto manual gestures. Of course, there are many ways to cooperatively communicate, but it does not seem likely that our human ancestors progressed linearly from gesture-based communication to vocalized speech [Aboitiz, 2013].

Future Directions

As vocal cooperation may be regulated by arousal, future studies would benefit from simultaneous physiological measurements of the autonomic state during vocal cooperation such as (but not limited to) respiration, heart rate and pupil dilation. Furthermore, since neuromodulators may play a critical role in the differen-

tial emergence of this behavior, cross-species histological maps of neuromodulator-specific receptor density would be invaluable in achieving a comparative perspective – one that would allow us to really test our hypotheses with regard to the differential modulation of homologous circuits. Finally, combined with physiological recordings of neuronal activity, measuring ambient neuromodulator levels during vocal cooperation would be critical in understanding the moment-to-moment fluctuations of neurochemicals and their effect on behavior.

Acknowledgements

We are appreciative of Daniel Takahashi's careful reading of our manuscript as well as his guidance with the coupled oscillator mechanism. We thank Darshana Narayanan, Yayoi Teramoto and Lydia Hoffstaetter for their comments on an early version of the manuscript. This work was supported by the National Science Foundation Graduate Research Fellowship Program (J.I.B.), NIH R01NS054898 (A.A.G.) and the James S. McDonnell Scholar Award (A.A.G.).

References

- Aboitiz F (2013): How did vocal behavior 'take over' the gestural communication system? *Lang Cogn* 5:167–176.
- Berna F, Goldberg P, Horwitz LK, Brink J, Holt S, Bamford M, Chazan M (2012): Microstratigraphic evidence of in situ fire in the Acheulean strata of Wonderwerk Cave, Northern Cape province, South Africa. *Proc Natl Acad Sci USA* 109:E1215–E1220.
- Bezerra BM, Souto A (2008): Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int J Primatol* 29:671–701.
- Bhadoria R, Watson D, Danson P, Ferrier IN, McAllister VI, Moore PB (2003): Enlargement of the third ventricle in affective disorders. *Ind J Psychiatry* 45:147–150.
- Bridges RS, Zarrow MX, Goldman BD, Denenberg VH (1974): A developmental study of maternal responsiveness in the rat. *Physiol Behav* 12:149–151.
- Burkart JM, Hrdy SB, van Schaik CP (2009): Cooperative breeding and human cognitive evolution. *Evol Anthropol Issues News Rev* 18: 175–186.
- Burkart JM, van Schaik CP (2010): Cognitive consequences of cooperative breeding in primates? *Anim Cogn* 13:1–19.
- Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002): Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev* 26:321–352.
- Chandrasekaran C, Trubanova A, Stillitano S, Caplier A, Ghazanfar AA (2009): The natural statistics of audiovisual speech. *PLoS Comput Biol* 5:e1000436.
- Chapple ED (1970): *Culture and Biological Man: Explorations in Behavioral Anthropology*. New York, Holt, Rinehart & Winston.
- Chen HC, Kaplan G, Rogers LJ (2009): Contact calls of common marmosets (*Callithrix jacchus*): influence of age of caller on antiphonal calling and other vocal responses. *Am J Primatol* 71:165–170.
- Chen X, Li S, Kirouac GJ (2014): Blocking of corticotrophin releasing factor receptor-1 during footshock attenuates context fear but not the upregulation of prepro-orexin mRNA in rats. *Pharmacol Biochem Behav* 120:1–6.
- Devinsky O, Morrell MJ, Vogt BA (1995): Contributions of anterior cingulate cortex to behaviour. *Brain* 118:279–306.
- Donaldson ZR, Young LJ (2008): Oxytocin, vasopressin, and the neurogenetics of sociality. *Science* 322:900–904.
- Donkelaar HJ, Nicholson C (1998): *The Central Nervous System of Vertebrates*. Berlin, Springer.
- Dunbar RI (1998): *Grooming, Gossip, and the Evolution of Language*. London, Faber and Faber.
- Dunbar RI (2009): The social brain hypothesis and its implications for social evolution. *Ann Hum Biol* 36:562–572.
- Fragaszy D, Visalberghi E, Galloway A (1997): Infant tufted capuchin monkeys' behaviour with novel foods: opportunism, not selectivity. *Anim Behav* 53:1337–1343.
- Furlong T, Carrive P (2007): Neurotoxic lesions centered on the perifornical hypothalamus abolish the cardiovascular and behavioral responses of conditioned fear to context but not of restraint. *Brain Res* 1128:107–119.
- Ghazanfar AA, Flombaum JI, Miller CT, Hauser MD (2001): The units of perception in the antiphonal calling behavior of cotton-top tamarins (*Saguinus oedipus*): playback experiments with long calls. *J Comp Physiol A* 187:27–35.
- Ghazanfar AA, Smith-Rohrberg D, Pollen AA, Hauser MD (2002): Temporal cues in the antiphonal long-calling behaviour of cottontop tamarins. *Anim Behav* 64:427–438.
- Ghazanfar AA, Takahashi DY (2014): Facial expressions and the evolution of the speech rhythm. *J Cogn Neurosci* 26:1196–1207.
- Giraud AL, Poeppel D (2012): Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci* 15:511–517.
- Goldizen AW (1990): A comparative perspective on the evolution of tamarin and marmoset social systems. *Int J Primatol* 11:63–83.

- Grafe TU (1996): The function of call alternation in the African reed frog (*Hyperolius marmoratus*): precise call timing prevents auditory masking. *Behav Ecol Sociobiol* 38:149–158.
- Gray P, Brooks PJ (1984): Effect of lesion location within the medial preoptic-anterior hypothalamic continuum on maternal and male sexual behaviors in female rats. *Behav Neurosci* 98:703–711.
- Greenfield MD (1994): Synchronous and alternating choruses in insects and anurans: common mechanisms and diverse functions. *Int Comp Biol* 34:605–615.
- Hare B, Melis AP, Woods V, Hastings S, Wrangham R (2007): Tolerance allows bonobos to outperform chimpanzees on a cooperative task. *Curr Biol* 17:619–623.
- Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C (2012): Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci* 16:114–121.
- Henry BL, Minassian A, Paulus MP, Geyer MA, Perry W (2010): Heart rate variability in bipolar mania and schizophrenia. *J Psychiatr Res* 44:168–176.
- Herculano-Houzel S, Collins CE, Wong P, Kaas JH (2007): Cellular scaling rules for primate brains. *Proc Natl Acad Sci USA* 104:3562–3567.
- Hewes GW (1973): Primate communication and the gestural origin of language. *Curr Anthropol* 33:65–84.
- Hrdy SB (2005): Evolutionary context of human development: the cooperative breeding model; in Carter C, Ahnert L, Grossmann K, Hrdy S, Lamb M, Porges S, Sachser N (eds): *Attachment and Bonding: A New Synthesis*. Cambridge, MIT Press, pp 9–32.
- Jaffe J, Beebe B, Feldstein S, Crown CL, Jasnow MD (2001): Rhythms of dialogue in infancy: coordinated timing in development. *Monogr Soc Res Child Dev* 66:1–132.
- Jasnow M, Feldstein S (1986): Adult-like temporal characteristics of mother-infant vocal interactions. *Child Dev* 57:754–761.
- Jerison H (1973): *Evolution of the Brain and Intelligence*. New York, Academic Press.
- Kampe J, Tschöp MH, Hollis JH, Oldfield BJ (2009): An anatomic basis for the communication of hypothalamic, cortical and mesolimbic circuitry in the regulation of energy balance. *Eur J Neurosci* 30:415–430.
- Katz PS (2011): Neural mechanisms underlying the evolvability of behaviour. *Philos Trans R Soc Lond B Biol Sci* 366:2086–2099.
- Kayaba Y, Nakamura A, Kasuya Y, Ohuchi T, Yanagisawa M, Komuro I, Fukuda Y, Kuwaki T (2003): Attenuated defense response and low basal blood pressure in orexin knockout mice. *Am J Physiol Regul Integr Comp Physiol* 285:R581–R593.
- Kleinow J, Smith A (2006): Potential interactions among linguistic, autonomic, and motor factors in speech. *Devel Psychobiol* 48:275–287.
- Kondo N, Watanabe S, Izawa EI (2010): A temporal rule in vocal exchange among large-billed crows (*Corvus macrorhynchos*) in Japan. *Ornithol Sci* 9:83–91.
- Krubitzer L, Kaas J (2005): The evolution of the neocortex in mammals: how is phenotypic diversity generated? *Curr Opin Neurobiol* 15:444–453.
- Latalova K, Prasko J, Diveky T, Grambal A, Kamaradova D, Velartova H, Salinger J, Opavsky J (2010): Autonomic nervous system in euthymic patients with bipolar affective disorder. *Neuro Endocrinol Lett* 31:829–836.
- Lerner Y, Honey CJ, Silbert LJ, Hasson U (2011): Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J Neurosci* 31:2906–2915.
- Levy B (2013): Autonomic nervous system arousal and cognitive functioning in bipolar disorder. *Bipolar Disord* 15:70–79.
- Lewin W, Whitty CWM (1960): Effects of anterior cingulate stimulation in conscious human subjects. *J Neurophysiol* 23:445–447.
- Logue DM, Chalmers C, Gowland AH (2008): The behavioural mechanisms underlying temporal coordination in black-bellied wren duets. *Anim Behav* 75:1803–1808.
- Marder E (2012): Neuromodulation of neuronal circuits: back to the future. *Neuron* 76:1–11.
- Masataka N, Biben M (1987): Temporal rules regulating affiliative vocal exchanges of squirrel monkeys. *Behaviour* 101:311–319.
- McMeeke H (2002): Autonomic peripheral vascular dysregulation and mood disorder. *J Affect Disord* 71:277–279.
- McRoberts GW, Studdert-Kennedy M, Shankweiler DP (1995): The role of fundamental frequency in signaling linguistic stress and affect: evidence for a dissociation. *Percept Psychophys* 57:159–174.
- Miller CT, Mandel K, Wang X (2010): The communicative content of the common marmoset phoe call during antiphonal calling. *Am J Primatol* 72:974–980.
- Miller CT, Wang X (2006): Sensory-motor interactions modulate a primate vocal behavior: antiphonal calling in common marmosets. *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 192:27–38.
- Mitani JC (1985): Gibbon song duets and intergroup spacing. *Behaviour* 92:59–96.
- Nakahara F, Miyazaki N (2011): Vocal exchanges of signature whistles in bottlenose dolphins (*Tursiops truncatus*). *J Ethol* 29:309–320.
- Nattie E, Li A (2012): Respiration and autonomic regulation and orexin. *Progs Brain Res* 198:25–46.
- Norcross JL, Newman JD (1993): Context and gender-specific differences in the acoustic structure of common marmoset (*Callithrix jacchus*) phoe calls. *Am J Primatol* 30:37–54.
- Numan M (1994): A neural circuitry analysis of maternal behavior in the rat. *Acta Paediatr Suppl* 397:19–28.
- O’Conaill B, Whittaker S, Wilbur S (1993): Conversations over video conferences: an evaluation of the spoken aspects of video-mediated communication. *Hum Comput Interact* 8:389–428.
- Oullier O, de Guzman GC, Jantzen KJ, Lagarde J, Kelso JAS (2008): Social coordination dynamics: measuring human bonding. *Soc Neurosci* 3:178–192.
- Owren MJ, Rendall D, Ryan MJ (2010): Redefining animal signaling: influence versus information in communication. *Biol Philos* 25:755–780.
- Padberg J, Disbrow E, Krubitzer L (2005): The organization and connections of anterior and posterior parietal cortex in titi monkeys: do New World monkeys have an area 2? *Cereb Cortex* 15:1938–1963.
- Padberg J, Franca JG, Cooke DF, Soares JGM, Rosa MGP, Fiorani M, Gattass R, Krubitzer L (2007): Parallel evolution of cortical areas involved in skilled hand use. *J Neurosci* 27:10106–10115.
- Paus T (2001): Primate anterior cingulate cortex: where motor control, drive and cognition interface. *Nat Rev Neurosci* 2:417–424.
- Pereira M, Andreatini R, Schwarting RKW, Brenes JC (2014): Amphetamine-induced appetitive 50-kHz calls in rats: a marker of affect in mania? *Psychopharmacology* 231:2567–2577.
- Peters HFM, Hulstijn W (1984): Stuttering and anxiety: the difference between stutterers and nonstutterers in verbal apprehension and physiologic arousal during the anticipation of speech and non-speech tasks. *J Fluency Disord* 9:67–84.
- Rendall D, Owren MJ, Ryan MJ (2009): What do animal signals mean? *Anim Behav* 78:233–240.
- Rizzolatti G, Arbib MA (1998): Language within our grasp. *Trends Neurosci* 21:188–194.
- Rosenblatt JS, Hazelwood S, Poole J (1996): Maternal behavior in male rats: effects of medial preoptic area lesions and presence of maternal aggression. *Horm Behav* 30:201–215.
- Rukstalis M, French JA (2005): Vocal buffering of the stress response: exposure to conspecific vocalizations moderates urinary cortisol excretion in isolated marmosets. *Horm Behav* 47:1–7.
- Samson WK, Gosnell B, Chang JK, Resch ZT, Murphy TC (1999): Cardiovascular regulatory actions of the hypocretins in brain. *Brain Res* 831:248–253.
- Santos CV, French JA, Otta E (1997): Infant carrying behavior in callitrichid primates: *Callithrix* and *Leontopithecus*. *Int J Primatol* 18:889–907.
- Schmidt R, Morr S (2010): Coordination dynamics of natural social interactions. *Int J Sport Psychol* 41:105.
- Sellen A (1995): Remote conversations: the effects of mediating talk with technology. *Hum Comput Int* 10:401–444.

- Shahack-Gross R, Berna F, Karkanas P, Lemorini C, Gopher A, Barkai R (2014): Evidence for the repeated use of a central hearth at middle pleistocene (300 ky ago) Qesem Cave, Israel. *J Archaeol Sci* 44:12–21.
- Shahid IZ, Rahman AA, Pilowsky PM (2012): Orexin and central regulation of cardiorespiratory system. *Vitam Horm* 89:159–184.
- Snowdon CT, Cronin KA (2007): Cooperative breeders do cooperate. *Behav Proc* 76:138–141.
- Sörös P, Sokoloff LG, Bose A, McIntosh AR, Graham SJ, Stuss DT (2006): Clustered functional MRI of overt speech production. *Neuroimage* 32:376–387.
- Steiper ME, Young NM (2006): Primate molecular divergence dates. *Mol Phylogenet Evol* 41:384–394.
- Stephan H, Schwedtfeger WK, Baron G (1980): The Brain of the Common Marmoset (*Callithrix jacchus*). Berlin. Springer.
- Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, Hoymann G, Rossano F, de Ruiter JP, Yoon K-E, Levinson SC (2009): Universals and cultural variation in turn-taking in conversation. *Proc Natl Acad Sci USA* 106:10587–10592.
- Sugiura H (1998): Matching of acoustic features during the vocal exchange of coo calls by Japanese macaques. *Anim Behav* 55:673–687.
- Syal S, Finlay BL (2011): Thinking outside the cortex: social motivation in the evolution and development of language. *Dev Sci* 14:417–430.
- Takahashi DY, Narayanan DZ, Ghazanfar AA (2013): Coupled oscillator dynamics of vocal turn-taking in monkeys. *Curr Biol* 23:2162–2168.
- Tomasello M (2008): *Origins of Human Communication*. Cambridge, MIT Press.
- Uno H, Tarara R, Else JG, Suleman MA, Sapolsky RM (1989): Hippocampal damage associated with prolonged and fatal stress in primates. *J Neurosci* 9:1705–1711.
- van Kammen DP, Murphy DL (1975): Attenuation of the euphoriant and activating effects of d- and l-amphetamine by lithium carbonate treatment. *Psychopharmacologia* 44:215–224.
- Vitale A, Queyras A (2010): The response to novel foods in common marmoset (*Callithrix jacchus*): the effects of different social contexts. *Ethology* 103:395–403.
- Voelkl B, Schrauf C, Huber L (2006): Social contact influences the response of infant marmosets towards novel food. *Anim Behav* 72:365–372.
- Wiethoff S, Wildgruber D, Kreifelts B, Becker H, Herbert C, Grodd W, Ethofer T (2008): Cerebral processing of emotional prosody – influence of acoustic parameters and arousal. *Neuroimage* 39:885–893.
- Willson MC, Bell EC, Dave S, Asghar SJ, McGrath BM, Silverstone PH (2005): Valproate attenuates dextroamphetamine-induced subjective changes more than lithium. *Eur Neuropsychopharmacol* 15:633–639.
- Yamamoto ME, Lopes FA (2004): Effect of removal from the family group on feeding behavior by captive *Callithrix jacchus*. *Int J Primatol* 25:489–500.
- Yosida S, Kobayasi KI, Ikebuchi M, Ozaki R, Okanoya K (2007): Antiphonal vocalization of a subterranean rodent, the naked mole-rat (*Heterocephalus glaber*). *Ethology* 113:703–710.
- Yu C, Smith LB (2012): Embodied attention and word learning by toddlers. *Cognition* 125:244–262.
- Zahed SR, Prudom SL, Snowdon CT, Ziegler TE (2008): Male parenting and response to infant stimuli in the common marmoset (*Callithrix jacchus*). *Am J Primatol* 70:84–92.
- Zelick R, Narins PM (1985): Characterization of the advertisement call oscillator in the frog *Eleutherodactylus coqui*. *J Comp Physiol A* 156:223–229.
- Zhang W, Fukuda Y, Kuwaki T (2005): Respiratory and cardiovascular actions of orexin-A in mice. *Neurosci Lett* 385:131–136.