Can vocal conditioning trigger a semiotic ratchet in marmosets?

Hjalmar K. Turesson and Sidarta Ribeiro*

Brain Institute, Federal University of Rio Grande do Norte, Natal, Brazil

The complexity of human communication has often been taken as evidence that our language reflects a true evolutionary leap, bearing little resemblance to any other animal communication system. The putative uniqueness of the human language poses serious evolutionary and ethological challenges to a rational explanation of human communication. Here we review ethological, anatomical, molecular, and computational results across several species to set boundaries for these challenges. Results from animal behavior, cognitive psychology, neurobiology, and semiotics indicate that human language shares multiple features with other primate communication systems, such as specialized brain circuits for sensorimotor processing, the capability for indexical (pointing) and symbolic (referential) signaling, the importance of shared intentionality for associative learning, affective conditioning and parental scaffolding of vocal production. The most substantial differences lie in the higher human capacity for symbolic compositionality, fast vertical transmission of new symbols across generations, and irreversible accumulation of novel adaptive behaviors (cultural ratchet). We hypothesize that increasingly-complex vocal conditioning of an appropriate animal model may be sufficient to trigger a semiotic ratchet, evidenced by progressive sign complexification, as spontaneous contact calls become indexes, then symbols and finally arguments (strings of symbols). To test this hypothesis, we outline a series of conditioning experiments in the common marmoset (*Callithrix jacchus*). The experiments are designed to probe the limits of vocal communication in a prosocial, highly vocal primate 35 million years far from the human lineage, so as to shed light on the mechanisms of semiotic complexification and cultural transmission, and serve as a naturalistic behavioral setting for the investigation of language disorders.

Keywords: vocal learning, conditioning, operant, marmoset, semiotics, language disorders

Introduction

Since Darwin's (1872) comparative study of emotions, psychological continuity across species has been a dominant view in biology. However, the complexity of human communication has often been taken as evidence that our language reflects a true evolutionary leap, bearing little resemblance to any other animal communication system (Tomasello and Call, 1997; Tomasello, 1999). Scientists have argued bitterly over whether animals possess language or not (Marler, 1970; Griffin, 1984; Savage-Rumbaugh, 1990; Pepperberg et al., 1997; Gelman and Gallistel, 2004; Tomasello et al., 2005). Throughout most of the past century, the putative uniqueness of our language posed serious evolutionary and ethological challenges to a rational explanation of human communication. How did the phonology, syntax and semantics of human language evolve if apes, our closest relatives, appear to be so laconic?

Yet, the increasingly better sampling of animal behavior in the past decades showed that apes and other primates possess vocal referential communication (e.g., Seyfarth et al., 1980), cultural transmission (e.g., Whiten et al., 1999), and even simple combinatorial syntax (e.g., Ouattara et al., 2009). We actually share with certain mammalian and avian species the fundamental sensorimotor mechanisms required for the perceptual and motor aspects of vocal learning (Petkov and Jarvis, 2012), such as the expression in cortico-striatal and cortico-cerebellar circuits of transcription factor FOXP2, which regulates the normal development of human speech (Konopka et al., 2009; Scharff and Petri, 2011). A comprehensive computational analysis of gene expression profiles measured in various brain regions recently allowed for the identification of several functionally and molecularly analogous brain regions for birdsong and human speech (Pfenning et al., 2014). Nevertheless, it remains unclear why have other animals failed to evolve the higher traits of human language.

To shed light on this problem, we begin by reviewing the contributions of semiotics to the problem of vocal communication. Next we review the empirical literature on vocal learning across species, with a focus on the comparison between New and Old World primates. Finally, we present a series of ongoing experiments aimed at probing the semiotic limits of vocal communication in marmosets.

Meaning by Resemblance: Iconic Representations

Semiotics is the logical system developed by Peirce (1958, 1998) to formally describe the communication of an object to an interpretant by way of a sign. In this system, the relationship between sign and object strictly derives from three—and only three—different kinds of mental processes for the establishment of meaning: icon, index or symbol.

If a sign resembles the physical properties of an object, it is said to be an icon of this object. If a sign has spatio-temporal contiguity with an object, it is said to be an index of this object. If a sign represents an object by an entirely arbitrary rule or convention established among interpreters, it is said to be a symbol of that object (Peirce, 1958, 1998). A critical difference between an index and a symbol is that the former requires the simultaneous presence of the object, while the latter often occurs in the complete absence of the object. No other categories exist beyond these three, and by definition, a sign must belong to at least one of them.

Icons are the simplest form of sign, because they convey meaning through the sheer physical similarity with the object, without the need of further abstraction. Arguably, iconic communication would not be simple and perhaps not even possible if parts of the brain did not represent perceptual information in an iconic manner (Ribeiro, 2010). Visual, auditory and somatosensory inputs ascend from the periphery to the central nervous system in a quite organized manner, leading to orderly topographic maps in primary sensory thalamus and neocortex for positions in physical space (Hubel and Wiesel, 1959), sound frequencies in acoustic space (Hind, 1953), and locations along the body surface (Mountcastle, 1957; Mountcastle

et al., 1957; Simons and Woolsey, 1979). As a consequence, object representations in these early processing areas are largely congruent with the objects represented. A well-known example of this feature was provided by the use of 2-deoxyglucose uptake by stimulated neural tissue to imprint a visual grid on the primary visual cortex of macaque monkeys (Tootell et al., 1982). While the experiment revealed the high degree of isomorphism between stimulus and response, it also made explicit that this response is modified by the cortical magnification factor that over-represents the center of the visual field in detriment of the periphery.

Another telling example of the interplay between iconic representations and hardwired filters for ecologically-relevant stimuli can be found in the canary, a seasonal songbird with a characteristic syllable repertoire. The caudomedial nidopallium (NCM) of the canary, an auditory region homologous to the mammalian primary auditory cortex, responds to natural canary whistles according to their frequencies, with low-pitched whistles mapping dorsally in NCM, and gradually more ventral mapping as the pitch increases (Ribeiro et al., 1998). Yet NCM cannot be said to be strictly tonotopic, because artificial stimuli with the same fundamental frequency fail to produce a well-localized response in NCM: computer generated tones activate more diffuse groups of neurons, and guitar notes stripped of their harmonics even more so, without any clear topography. Thus, NCM is "whistletopic" rather than tonotopic, i.e., it carries an orderly representation of the environment that is clearly tuned to the acoustic features of natural stimuli.

If iconic representations in primary cortical areas are more the rule than the exception, examples of iconic communication among non-human animals are not compelling. Inter-specific communication by sheer vocal imitation (Owen-Ashley et al., 2002) is rare; for instance, to date there is no evidence that animals produce onomatopoeic alarm calls to warn against predators. Still, the notion that onomatopoeias played an important role in the early evolution of words is tempting, with old roots in philosophy and linguistics (von Humboldt, 1836; Jakobson and Waugh, 1979; Hinton et al., 1994; Magnus, 2010), and fresh support from studies of infant psychology (Maurer et al., 2006; Imai et al., 2008; Kita et al., 2010) and neuroscience (Osaka et al., 2004; Aziz-Zadeh et al., 2006; Garcia et al., 2014). Judgment of this issue is tied to the relatively recent advent of computer-aided devices to investigate spontaneous behavior of other species in the field. As more studies with better recording techniques are performed, discoveries regarding the natural use of iconic communication may still lead to substantially different conclusions.

Meaning by Spatio-Temporal Contiguity: To Point, to Look, to Indicate

In comparison with icons, indexes constitute a much more flexible type of sign, since the same pointer can be used to mean an endless variety of different objects. The highly informative act of indicating with the finger or gaze is so ubiquitous across different human cultures that one must conclude that its origins are ancient in the *Homo* lineage. Indeed, indexes are far from being exclusive of human communication. A number of studies suggest

that chimps are capable of taking conspecifics gaze direction into account (Call et al., 1998; Hare et al., 2000, 2001). Apes can also follow human pointing, given appropriate methodologies and/or human exposure (Mulcahy and Call, 2009; Lyn et al., 2010; Leavens et al., 2012; Hopkins et al., 2013). Rhesus monkeys follow both human pointing and gaze (Hauser et al., 2007), and even corvids have been shown to follow gaze (Bugnyar et al., 2004; Bugnyar, 2011). The fact that human pointing is followed by domesticated dogs and foxes (Hare et al., 1998, 2002, 2005; Hare and Tomasello, 1999), as well as trained dolphins (Pack and Herman, 2004), suggests that indexical learning can be greatly boosted within a few generations by trait selection, and even within a single generation, by learning from interactions with point users (e.g., human trainers). Altogether, these studies indicate that the capacity for index-based communication is widespread and therefore not a suitable candidate to explain the uniqueness of human language. But it may well be that quantitative differences in the ability to use indexes (Goldin-Meadow, 2007), supported by a large repertoire of arbitrary vocal calls (Ghazanfar et al., 2007), allowed our hominid ancestors to initiate the cultural ratchet that led to contemporary human language.

Indexes are very useful to signal to other individuals the presence of conspecifics, predators, preys, or feeding opportunities, and therefore a likely preadaptation of indexical communication is prosociality, i.e., the ability to display behaviors that favor other individuals even in the absence of immediate self-benefit. Prosociality is a widely distributed trait among mammals, from rats (Daniel, 1942; Ben-Ami Bartal, 2011; Márquez et al., 2015) to apes (Horner et al., 2011; von Rohr et al., 2012; Hobaiter and Byrne, 2014; Hobaiter et al., 2014).

Yet, as prosocial as apes can be, they seem to have little capacity for using indexes to share intentions on the execution of collaborative activities (Warneken et al., 2006; but see Leavens et al., 2005). Several researchers have proposed that the key difference between humans and non-humans is shared intentionality, a cornerstone of our exquisite ability to teach and learn, able to create a cultural ratchet that leads to the fast transmission of any useful cultural innovation (reviewed in Tomasello et al., 2005). Joint attention in humans begins at 9–12 months with gaze following, the use of adults for social reference, and the imitation of adult gestures (Bakeman and Adamson, 1984; Moore and Dunham, 1995). The understanding that there are other minds begins as babies start to recognize voluntary behaviors with overt or covert goals (Meltzoff, 1988; Carpenter et al., 1995, 1998). As children develop, adults become outside entities whose attention can be called, both as observers of the child's actions and as producers of behaviors desired by the child. Lack of shared intentionality and deficits in index usage seem to be an important component in autism (Tomasello et al., 2005).

Point following and the partaking of goals appears to be a bottleneck for the learning of indexical signs, and for this reason we designed experiments to quantitatively assess the conditions that allow shared intentionality to arise between pairs of independent vocal agents, either overtly or covertly. These experiments are detailed below, but before they are properly introduced it is necessary to present the quite controversial topic of symbolic communication.

Meaning by Convention: Are We the Symbolic Species?

The search for the key difference between human language and the communication systems in other species led to the proposal that the use of symbols is the distinctive property that sets us apart from the other animals (Deacon, 1998). According to this view, non-human animals are limited to iconic and indexical communication, and therefore utterly incapable of using symbols, the most powerful and flexible sign type. Supposedly the "symbolic species hypothesis" is grounded on the Peircean semiotics, but a closer inspection of semiotics shows that the definition of symbol entirely matches the empirical evidence of referential communication observed in different animal species.

Chimpanzees in captivity can learn to use man-made lexigrams to refer to dozens of different objects and actions (Savage-Rumbaugh et al., 1986; Gillespie-Lynch et al., 2014), effectively expanding their ability to communicate with caregivers and experimenters. However, it has been argued that lexigram use by chimpanzees does not represent true symbolic communication, but rather functional communication based on the instrumental learning of the specific contingencies of the experimental setting (Seidenberg and Petitto, 1987).

Field studies of spontaneous animal communication effectively bypass this concern. Adult vervet monkeys (*Cercopithecus aethiops*), Old World primates from the African savannahs, naturally display three kinds of alarm calls that correspond specifically to the presence of terrestrial, aerial or slithering predators. Upon hearing alarm calls uttered by an adult, other adult monkeys will promptly react to protect themselves, hiding above or below trees in the case of aerial or terrestrial predators, respectively, or moving aside and scanning the ground in the case of snakes. Juvenile vervet monkeys are able to emit the same vocalizations but do so out of context, and therefore do not produce escape reactions in the adults.

Field observation of vervet monkey behavior argues strongly for the symbolic quality of alarm calls among adults (Struhsaker, 1967; Seyfarth et al., 1980). First, the proper context of use of these calls is slowly learned by repeated pairing of alarm call (auditory stimulus) and predator (visual and/or olfactory stimulus), denoting the gradual establishment of a social convention regarding the interpretation of these alarms. Second, experimental playbacks of these alarm calls produce proper escape reactions in the absence of any predator among adults, exemplifying a trademark of symbolic communication, namely that meaning is conveyed in the absence of the object (Seyfarth and Cheney, 1988). Comparable communication systems have been documented in close relatives, Campbell's monkeys (*Cercopithecus campbelli*) and Diana monkeys (*Cercopithecus diana*; Zuberbühler, 2000, 2001). More recently, field research of chimpanzees revealed the use of intentional calls to warn against a predator model (Schel et al., 2013).

Computational simulations of the interactions of artificial creatures representing vocalizing preys and their predators

suggest that the referential code that ascribes specific meaning to each call type arises by chance, through random variations that get to be established and maintained over time. This occurs when the prey-predatory ratio is sufficiently large for the prey population to survive long enough for the establishment and spread of the code (Queiroz and Ribeiro, 2002; Gudwin et al., 2003; Loula et al., 2004; Ribeiro et al., 2007).

Since the original findings in vervet monkeys, the occurrence of referential communication regarding specific predator types (Macedonia and Evans, 1993) has been demonstrated in a variety of non-primate species, including ground squirrels (*Spermophilus* spp.; Owings and Hennessy, 1984), chickens (*Gallus gallus domesticus*; Gyger et al., 1987; Evans and Evans, 1999), prairie dogs (*Cynomys gunnisoni*; Slobodchikoff et al., 1991; Kiriazis and Slobodchikoff, 2006), tree squirrels (*Tamiasciurus hudsonicus* Greene and Meagher, 1998), dwarf mongooses (*Helogale undulata*; Beynon and Rasa, 1989), suricates (*Suricata suricatta*; Manser, 2001; Manser et al., 2001). Similarly, bottlenosed dolphins have been reported to understand human gestures as symbolic representations of body parts (Herman et al., 2001).

Altogether, the empirical and computational results of referential communication among a wide variety of species clearly contend against the notion that humans are the only species to employ symbols. Rather, referential communication in non-human species conform to the notion of dicent symbol in Peircean semiotics, i.e., a symbol that functions "like an index" because its object is "a general interpreted as an existent" (Peirce, 1998). In the semiotic framework, what distinguishes human language from the communication system of other species is our ability to concatenate symbols of symbols of symbols, namely what Peirce labeled "argument."

How Informative is the Structure of Arguments?

Several animal species use sequences of vocalizations to communicate. Notably, birdsong is typically composed of a sequence of phrases, each comprising a string of repeated syllables (Williams and Staples, 1992). In many species, adult males display a stable sequence of vocalizations, with little variation from bout to bout within a season (e.g., canary: Nottebohm et al., 1981) or even across seasons (e.g., zebra-finch: Walton et al., 2012), especially when singing toward a female (Sossinka and Bohner, 1980).

In extreme cases, like in the nightingale, the huge size of the syllabic repertoire leads to a very flexible display (Hultsch et al., 2004). Likewise, mockingbirds have the ability to mimic the vocalizations of other species, resulting in very complex sequences akin to improvisation (Derrickson, 1987). Nevertheless, little or no meaning seems to be ascribed to phrase order, with the clear exception of introductory notes (Price, 1979), which may indicate whether animals are singing directly to a conspecific, or singing alone (Jarvis et al., 1998).

There is no indication that songbirds can shuffle phrases to generate a combinatorial reference code. In fact, this ability seems to be exceedingly rare, uniquely human if not for the (rather limited) example of compositionality in some Old-World
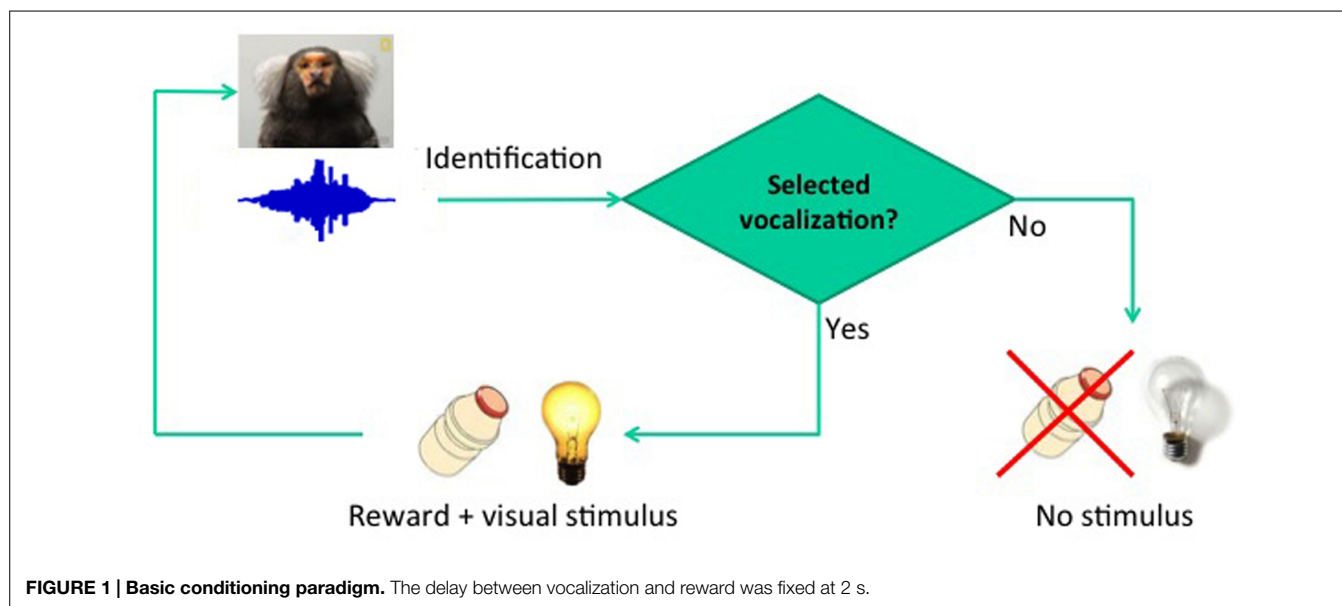
primates, such as the putty-nosed monkey (*Cercopithecus nictitans*), which can employ two vocalizations in a combinatorial manner (Arnold and Zuberbühler, 2006). With the sole exception of suffix usage among Cercopithecines (Ouattara et al., 2009; Coye et al., 2015), and human-tutored apes (Savage-Rumbaugh et al., 1986) and dolphins (Richards et al., 1984), symbolic arguments seem to be exclusively humans.

The notion that the structure of arguments carries important information is tightly linked to graph theory, which deals with networks of interconnected elements, such as successive words during a conversation. Graph theory was initiated by Leonhard Euler in the eighteenth century, and was greatly developed in the twentieth century (Bollobás, 1998) with ever-expanding applications in physics, chemistry, biology and indeed any field in which networks play a key role (Milo et al., 2002; Sporns, 2010). In the context of vocal communication, a graph represents the temporal sequence of vocalizations, with each vocalization represented as a node, and the transition between consecutive vocalizations represented as a directed edge (Ferrer I Cancho and Solé, 2001). The overarching applicability of graph theory to communication suggests that it is particularly well suited to causally bridge very different levels of description of the phenomenon, such as systems neurophysiology and linguistics. If activity within an interconnected network of neurons is very aptly described as a directed graph (Changeux, 1997), so is the sequential production of utterances that characterize vocal communication in general, or human language in particular.

The usefulness of structural speech features to discriminate pathological and non-pathological reports has recently been demonstrated. Specific graph attributes allow the quantitative discrimination of schizophrenic versus manic subjects, even when non-psychotic subjects are included in the sample (Mota et al., 2012, 2014). Similar analyses successfully discriminate patients with Alzheimer's disease from patients with mild cognitive impairment (Bertola et al., 2014). Overall the results indicate that the graph-theoretical analysis of vocalizations is key to the study of language-related deficits in humans, and may have major applicability in animal models of diseases such as autism, such as the marmoset (Shen, 2013). Effective screening for phenotypes of interest may be provided by semiotic experiments based on conditioning (**Figure 1**), designed to track essential quantitative differences in the structure of vocalizations as they become indexes, then dicent symbols, and possibly arguments (**Figure 2**). Before proceeding to the explanation of these experiments and to the justification of the choice of marmosets, it is important to review the known boundaries of vocal learning in primates.

Vocal Flexibility in Non-Human Primates

For symbolic communication to be powerful (i.e., communicate a great number of objects), and/or flexible (i.e., communicate new objects), two aspects must concur. First, the ability to learn to interpret new symbols. Second, the ability to use and produce new symbols for adaptive operation within a given context. While there is plenty of evidence of the former (Savage-Rumbaugh et al., 1978; Savage-Rumbaugh and Romsky, 1989; Gelman and

FIGURE 1 | Basic conditioning paradigm. The delay between vocalization and reward was fixed at 2 s.

Gallistel, 2004), examples of the latter are substantially more limited among non-human primates. Below we briefly review the findings on primate vocal flexibility, understood as the adaptive, cognitively mediated control over vocal production, comprising usage learning, production learning, and acoustic modifications in the presence of noise (Seyfarth and Cheney, 2010).

Vocal Production Learning in Non-Human Primates

Vocal production learning refers to the ability to learn to produce a new vocalization. This can be a modification of a vocalization already in the animal's repertoire, or the production of an entirely new vocalization. Historically, the focus has been on entirely new vocalizations since it is hard to ensure that what appears to be a modified vocalization is not merely a preexisting vocalization that simply was not experimentally observed before (and thus only a matter of usage learning; Janik and Slater, 1997, 2000; Tyack, 2008). Among mammals, imitation of completely novel, often anthropogenic, sounds has only been reported for dolphins, elephants, harbor seals, and humans (Tyack, 2008). However, this narrow way of studying vocal production learning inadvertently excludes many species that at a closer look display modifications of vocalizations beyond the solely motivational or developmental.

One such (subtle) type of vocal modification is the convergence of calls, which takes place when some acoustic features of different individuals' calls converge, often as new social groups form. This has been reported in several species, and among them, primates. In two influential studies Snowdon and Elowson (1999) demonstrated that certain acoustic features of the trill calls produced by pygmy marmosets (*Cebuella pygmaea*)—an intragroup communication call, apparently used for maintaining group cohesion (Snowdon and Hodun, 1981)—converged after different groups were placed in a common acoustic environment
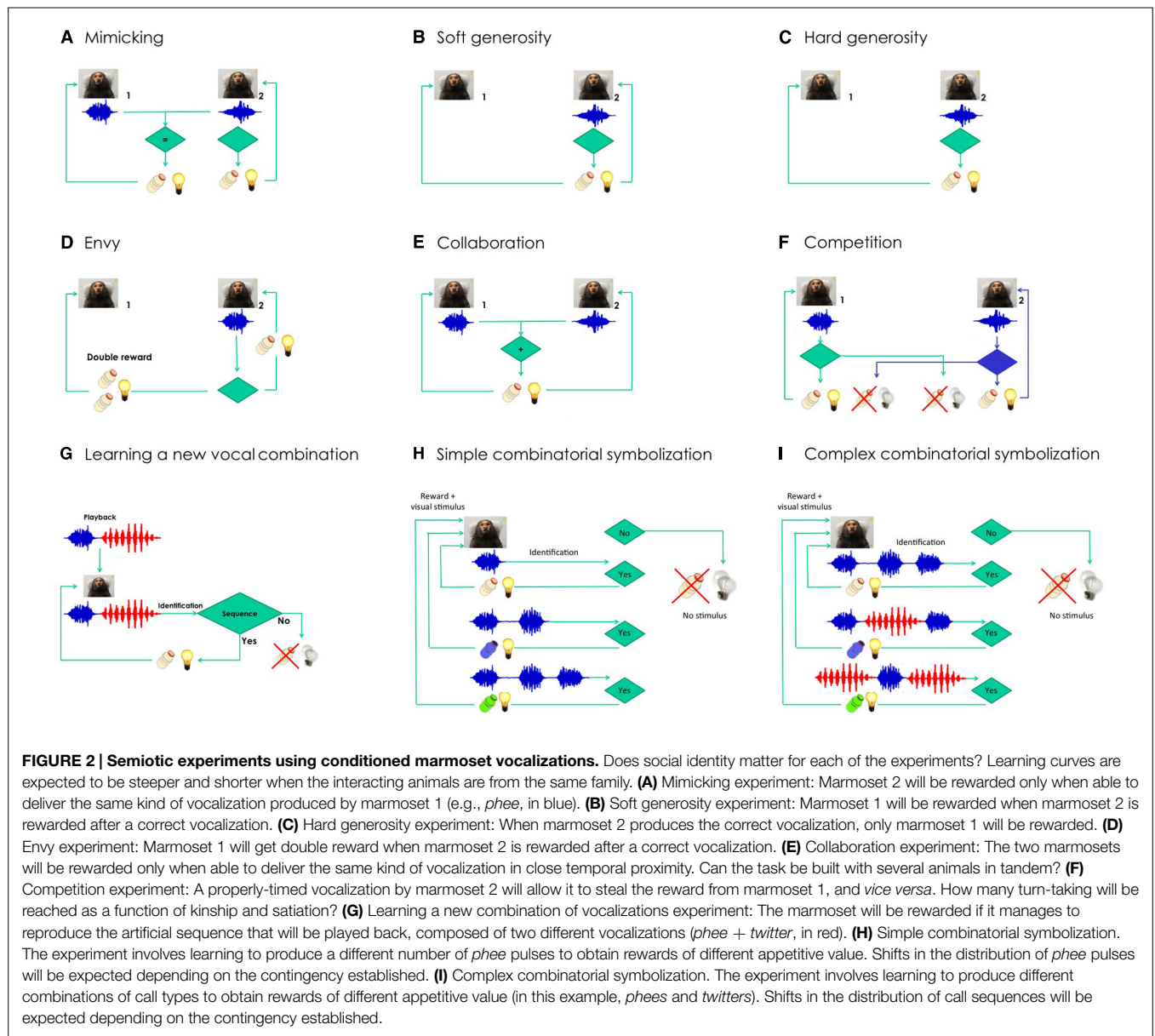
(Elowson and Snowdon, 1994), as well as when individuals were paired (Snowdon and Elowson, 1999). This suggests that pygmy marmosets indeed have some capacity for learning their vocal production, expressed in response to changes in social environments. Similar findings have followed, demonstrating for example the convergence of food grunts after the merging of two chimpanzee groups (Watson et al., 2015), and the convergence of pant hoots (a long distance call) to a version shared within the group (Marshall et al., 1999).

Thus, it seems that non-human primates might be capable of vocal production learning, and thus have more vocal flexibility than what was apparent when using the ability to imitate novel vocalizations as a strict requirement (Elowson and Snowdon, 1994; Marshall et al., 1999; Snowdon and Elowson, 1999; Watson et al., 2015). However, the above-mentioned acoustic modifications are subtle, and possibly caused by other factors than learning. Thus, the hard evidence has to be carefully examined, and followed up by better-controlled studies able to rule out the possible confounds. The next section presents some of these confounds.

Vocal Usage in Non-Human Primates

Learning in which context to produce a call is often labeled vocal usage learning and it demonstrates a degree of cognitive control over vocalizations. This kind of vocal flexibility is generally studied with conditioning experiments in which the animal is rewarded for producing calls (any call or a particular type of call). Several studies have presented results suggesting that non-human primates can be conditioned to produce at least some types of vocalizations in response to particular contexts (Sutton et al., 1973; Pierce, 1985; Hihara et al., 2003; Hage et al., 2013), while others have reported failures (Yamaguchi and Myers, 1972; Aitken and Wilson, 1979).

A classic and often cited study offered food reward to three juvenile rhesus monkeys (*Macaca mulatta*) for producing more

FIGURE 2 | Semiotic experiments using conditioned marmoset vocalizations. Does social identity matter for each of the experiments? Learning curves are expected to be steeper and shorter when the interacting animals are from the same family. **(A)** Mimicking experiment: Marmoset 2 will be rewarded only when able to deliver the same kind of vocalization produced by marmoset 1 (e.g., *phee*, in blue). **(B)** Soft generosity experiment: Marmoset 1 will be rewarded when marmoset 2 is rewarded after a correct vocalization. **(C)** Hard generosity experiment: When marmoset 2 produces the correct vocalization, only marmoset 1 will be rewarded. **(D)** Envy experiment: Marmoset 1 will get double reward when marmoset 2 is rewarded after a correct vocalization. **(E)** Collaboration experiment: The two marmosets will be rewarded only when able to deliver the same kind of vocalization in close temporal proximity. Can the task be built with several animals in tandem? **(F)** Competition experiment: A properly-timed vocalization by marmoset 2 will allow it to steal the reward from marmoset 1, and *vice versa*. How many turn-taking will be reached as a function of kinship and satiation? **(G)** Learning a new combination of vocalizations experiment: The marmoset will be rewarded if it manages to reproduce the artificial sequence that will be played back, composed of two different vocalizations (*phee + twitter*, in red). **(H)** Simple combinatorial symbolization. The experiment involves learning to produce a different number of *phee* pulses to obtain rewards of different appetitive value. Shifts in the distribution of *phee* pulses will be expected depending on the contingency established. **(I)** Complex combinatorial symbolization. The experiment involves learning to produce different combinations of call types to obtain rewards of different appetitive value (in this example, *phees* and *twitters*). Shifts in the distribution of call sequences will be expected depending on the contingency established.

and longer calls (Sutton et al., 1973). After 3–4 weeks of training, two of the subjects increased their average number of calls per session, and all three showed an increase in call duration. Pierce (1985) reviewed 12 early attempts to condition the vocalizations of non-human primates and came to the conclusion that they indeed have a considerable amount of control over their own vocal output. However, as in the study from Sutton et al. (1973), several of the results mentioned by Pierce (1985) could possibly be caused by motivational factors instead of learned vocal control.

Criticism: Observational Studies

A major problem with field studies is that they are purely observational, and thus allow for a whole slew of confounding variables to affect the results. For example Arcadi (1996) and Mitani et al. (1999) showed that acoustic features of

the pant-hoot in chimpanzees varies between geographically separated populations. This variation has been interpreted by other authors as an indication that vocal development in chimpanzees involves learning (e.g., Arcadi, 1996). However, careful analysis of a number of environmental factors suggests that the acoustic difference between calls from the two groups does not have to be an effect of learning. For example, the two groups lived in different habitats with different acoustics due to varying density and type of forestation (i.e., primary versus secondary forest), and thus the observed differences in the calls could instead reflect adaptations to different acoustic environments. Further, the amount of interfering noise the pant hoots were subjected to differed according to varying levels of biodiversity in the two habitats. Also the average body size likely differed between the two groups, with corresponding changes in the acoustic features of the calls that depend on body size. Finally, the two groups

were separated by such a large geographical distance that between-group genetic differences could not be excluded (Mitani et al., 1999). Thus, what superficially seemed to imply some form of vocal learning, might only be a consequence of environmental factors.

Criticism: Motivational Factors

A recurrent problem in many of the studies, both on vocal production learning, and vocal usage, is the influence of the motivational or arousal state (Owren et al., 2011; Hage et al., 2013). A number of acoustic features vary with arousal level. For example, call rate, fundamental frequency and call duration all go up with increased arousal (Rendall, 2003). The level of arousal, in turn, can be influenced by a number of factors, including changes in social context and food (Braesicke et al., 2005; De Marco et al., 2011; Machado et al., 2011). Thus, when the animal's context changes this can induce changes in arousal level, and consequently also in some acoustic properties of the calls. To avoid this, utmost care needs be taken to dissociate motivational factors from experimental manipulations and measures.

To date, this has not been systematically done. For example, Sutton et al. (1973) demonstrated that the duration of the *coo* call in rhesus monkeys increased during a conditioning experiment. The *coo* call is produced in multiple contexts, and among them when food is available (Hauser and Marler, 1993). Since call duration increases with arousal, and arousal can increase in the presence of food, it is not unlikely that what appears as the subjects learning to produce longer calls for reward is merely a matter of the subjects learning to associate a particular context with food, and that this increases arousal. Similar alternative explanations can be leveraged to most published studies (see Owren et al., 2011; Hage et al., 2013, for a list of several of these studies).

At present, the most convincing study on vocal conditioning in non-human primates comes from Hage et al. (2013). They trained two rhesus monkeys to produce vocalizations for reward in the presence of a visual cue. One of the two subjects was further trained to produce two different vocalizations, coo and grunt. Which of the two calls was rewarded in a particular trial was indicated by two distinct visual cues. The subject learned to do this and, within the same experimental session, produced nearly exclusively coo calls when the cue indicated so, and grunts when that was indicated. Very few calls were produced in the wrong context. Even though both calls can be considered to be food associated (Hauser and Marler, 1993; Owren and Casale, 1994)—and thus to reflect arousal—the amount and type of reward was equal in both conditions. This means that food-related arousal should also be equal across conditions, and thus it cannot explain the differential calling depending on the visual cue presented. Hage et al. (2013) represent the only clear exception in the literature, which is otherwise confounded by arousal effects.

In summary, in face of the numerous books and review articles written on the evolutionary history of human speech in general and vocal flexibility in non-human primates in particular, it is surprising that better-controlled methods for characterizing the limits of vocal learning in non-human animals have not been developed. Nearly all the positive evidence lack appropriate

controls to rule out confounding effects such as differences in body weight, habitat, and most importantly motivation. Future studies must be designed to effectively dissociate motivational effects from the observable changes in vocal usage and production. The best would be to show that the vocal modifications could be driven in two opposite directions by equally arousing context/social environments. In the next section we present some key methodological improvements in this regard.

If a Marmoset Could Speak, We Should Strive to Understand

A large body of evidence indicates that human language shares many features with communication in other animals, and that the greatest differences lie in the superior human capacity for symbolic compositionality, fast vertical transmission of new symbols, and irreversible accumulation of novel adaptive behaviors that characterizes a cultural ratchet. We hypothesize that increasingly-complex vocal conditioning of an appropriate animal model may be sufficient to trigger a semiotic ratchet, evidenced by progressive sign complexification, as spontaneous contact calls become indexes, then symbols, and finally arguments.

An adequate animal model for testing this hypothesis should be amenable to laboratory research, have a rich vocal repertoire, and show prosocial behavior characterized by cooperative signaling and parental scaffolding. There is a continuum across primates regarding the importance of posterior–anterior cortical interactions for the perception and production of vocalizations, by way of the arcuate fasciculus (Rilling et al., 2008). Regions homologous to the Wernicke area for speech perception have been identified in chimpanzees (Gannon et al., 1998), while regions homologous to the Broca area for orofacial control and speech production have been recognized even in the common marmoset (*Callithrix jacchus*; Miller et al., 2010; Simões et al., 2010) a highly vocal New World monkey species split from Old World monkeys around 40 million years ago (Worley et al., 2014).

Marmosets are cooperative breeders (Koenig, 1995), prosocial (Burkart et al., 2007; Burkart and van Schaik, 2013), and capable of attributing intentions to conspecifics (Burkart et al., 2012). Similarly to the development of human speech (Goldstein and Schwade, 2008), the maturation of vocal communication in marmosets depends on parental scaffolding, in the form of contingent vocal (social) feedback that seems to shape the transition to adult vocalizations (Takahashi et al., 2013, 2015). These findings point to marmosets as an adequate animal model for the investigation of the development of indexical, symbolic, and argumental communication.

Below we outline a series of conditioning experiments designed to test the semiotic ratchet hypothesis in marmosets: Is it possible to trigger a cultural ratchet by rewarding specific individual vocalizations, so as to gradually build meaning? The experiments described in **Figure 2** aim to investigate the experimental and possibly natural occurrence of indexes and symbols in marmosets, based on the timing of auditory and visual events, such a gaze orientation, vocalization, and appetitive behavior. The experiments were designed to effectively dissociate motivational

effects from learning-related vocal changes, are suitable for a graph-theoretical analysis of the structure of vocalization sequences, and allow for the investigation of shared intentionality within and across social groups.

The first step is vocal conditioning, using the real-time detection of specific calls from a specific individual to deliver reward (**Figure 1**). Conditioning will be conducted so as to effectively dissociate motivational effects from learning-related vocal changes, first rewarding animals for producing multiple pulses in tandem, and then alternating to the exclusive reward of single-pulse calls. This should lead to substantial variations in the number of pulses produced, providing a direct control for arousal effects.

We will then initiate a series of experiments that take advantage of inter-animal differences in social rank and kinship, comparing results obtained from pairs of animals within and across families. These experiments involve imitation (**Figure 2A**), soft and hard generosity (**Figures 2B,C**), envy (**Figure 2D**), collaboration (**Figure 2E**), competition (**Figure 2F**), and learning a new combination of vocalizations (**Figure 2G**). In all these experiments the role of prosociality will be assessed within and across families. These experiments propose to push the envelope of the complexity of these vocal interactions, measuring the extent of their constraint by social bonds.

We predict that calls will be at first interpreted as prospective indexes of rewards, initially surprising but very reliable, until by sheer repetition the call will transit into a dicent symbol of the reward, to be used voluntarily even in the absence of any appetitive drive (e.g., **Figure 2C**). Experimenting with gradually longer delays between call and reward should also favor symbolic emergence, due to the ever-increasing duration of the time spent in the absence of the object. A further step would be to explore the potential for combinatory semantics in the marmoset, by offering a variety of different rewards for either *phee* calls of different pulse-lengths (**Figure 2H**), or specific combinations of *phee* calls and another call type, the *twitter* (**Figure 2I**).

In summary, we propose that conditioning experiments can provide a fertile empirical framework for the investigation of vocal communication in non-human primates, going beyond perceptual learning to investigate two main directions, namely the emergence of symbolic competence, and the possible capacity for the production of arguments, i.e., symbolic combinatorial

sequences. To test whether a semiotic cultural ratchet was indeed triggered, we will assess whether vocal conditioning becomes faster over time, with an acceleration of cultural transmission across generations. The cultural transmission of correct task execution from parents to infants even without explicit training would be a strong indication that a ratchet has been initiated. Another prediction made by the semiotic ratchet hypothesis is the absence of cultural fallbacks, i.e., once established, a new conditioned communication system should remain stable.

The experiments proposed here also have the potential to serve as a naturalistic behavioral setting for the investigation of language disorders. Structural features of pathological speech in humans, such as the reduced connectivity of the discourse among patients with schizophrenia or bipolar disorder (Mota et al., 2012, 2014), or the reduced density and diameter of speech graphs produced by patients with Alzheimer's disease (Bertola et al., 2014), have great potential as biomarkers for a dense quantitative screening of language deficit phenotypes in transgenic marmosets (Shen, 2013). Animals with autistic phenotypes should display great difficulty in learning tasks with an important social aspect (**Figures 2A–F**) while possibly performing well in solo tasks (**Figures 2G–I**). Ultimately, understanding the processes that generate complex vocal communication in primates may prove crucial to our understanding of the evolution of human language, while at the same time shedding light on the mechanisms underlying its disorders.

References

Aitken, P. G., and Wilson, W. A. (1979). Discriminative vocal conditioning in rhesus monkeys: evidence for volitional control? *Brain Lang.* 8, 227–240. doi: 10.1016/0093-934X(79)90051-8

Arcadi, A. C. (1996). Phrase structure of wild chimpanzee pant hoots: patterns of production and interpopulation variability. *Am. J. Primatol.* 39, 159–178. doi: 10.1002/(SICI)1098-2345(1996)39:3 < 159::AID-AJP2 > 3.0.CO;2-Y

Arnold, K., and Zuberbühler, K. (2006). Language evolution: semantic combinations in primate calls. *Nature* 441, 303. doi: 10.1038/441303a

Aziz-Zadeh, L., Wilson, S. M., Rizzolatti, G., and Iacoboni, M. (2006). Congruent embodied representations for visually presented actions and linguistic phrases describing actions. *Curr. Biol.* 16, 1818–1823. doi: 10.1016/j.cub.2006.07.060

Bakeman, R., and Adamson, L. B. (1984). Coordinating attention to people and objects in mother–infant and peer–infant interaction. *Child Dev.* 55, 1278–1289. doi: 10.2307/1129997

Ben-Ami Bartal, I., Decety, J., and Mason, P. (2011) Empathy and pro-social behavior in rats. *Science* 334, 1427–1430. doi: 10.1126/science.1210789

Bertola, L., Mota, N. B., Copelli, M., Rivero, T., Diniz, B. S., Romano-Silva, M. A., et al. (2014). Graph analysis of verbal fluency test discriminate between patients with Alzheimer's disease, mild cognitive impairment and normal elderly controls. *Front. Aging Neurosci.* 6:185. doi: 10.3389/fnagi.2014.00185

Beynon, P., and Rasa, O. A. E. (1989). Do dwarf mongooses have a language? Warning vocalisations transmit complex information. *S. Afr. J. Sci.* 85, 447–450.

Bollobás, B. (1998). *Modern Graph Theory*. New York: Springer-Verlag.

Braesicke, K., Parkinson, J. A., Reekie, Y., Man, M. S., Hopewell, L., Pears, A., et al. (2005). Autonomic arousal in an appetitive context in primates: a behavioural and neural analysis. *Eur. J. Neurosci.* 21, 1733–1740. doi: 10.1111/j.1460-9568.2005.03987.x

Bugnyar, T. (2011). Knower-guesser differentiation in ravens: others' viewpoints matter. *Proc. Biol. Sci.* 278, 634–640. doi: 10.1098/rspb.2010.1514

Bugnyar, T., Stöwe, M., and Heinrich, B. (2004). Ravens, Corvus corax, follow gaze direction of humans around obstacles. *Proc. Biol. Sci.* 271, 1331–1336. doi: 10.1098/rspb.2004.2738

Burkart, J. M., Fehr, E., Efferson, C., and van Schaik, C. P. (2007). Other-regarding preferences in a non-human primate: common marmosets provision food altruistically. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19762–19766. doi: 10.1073/pnas.0710310104

Burkart, J. M., Kupferberg, A., Glasauer, S., and van Schaik, C. (2012). Even simple forms of social learning rely on intention attribution in marmoset monkeys (*Callithrix jacchus*). *J. Comp. Psychol.* 126, 129–138. doi: 10.1037/a0026025

Burkart, J. M., and van Schaik, C. (2013). Group service in macaques (*Macaca fuscata*), capuchins (*Cebus apella*) and marmosets (*Callithrix jacchus*): a comparative approach to identifying proactive prosocial motivations. *J. Comp. Psychol.* 127, 212–225. doi: 10.1037/a0026392

Call, J., Hare, B. A., and Tomasello, M. (1998). Chimpanzee gaze following in an object-choice task. *Anim. Cogn.* 1, 89–99. doi: 10.1007/s100710050013

Carpenter, M., Akhtar, N., and Tomasello, M. (1998). Fourteen- through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behav. Dev.* 21, 315–330. doi: 10.1016/S0163-6383(98)90009-1

Carpenter, M., Tomasello, M., and Savage-Rumbaugh, S. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees. *Soc. Dev.* 4, 217–237. doi: 10.1111/j.1467-9507.1995.tb00063.x

Changeux, J.-P. (1997). *Neuronal Man: The Biology of Mind*. Princeton, NJ: Princeton University Press.

Coye, C., Ouattara, K., Zuberbuhler, K., and Lemasson, A. (2015). Suffixation influences receivers' behaviour in non-human primates. *Proc. R. Soc. B Biol. Sci.* 282, 20150265. doi: 10.1098/rspb.2015.0265

Daniel, W. J. (1942). Cooperative problem solving in rats. *J. Comp. Psychol.* 34, 361–368.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animal*. London: John Murray.

Deacon, T. W. (1998). *The Symbolic Species: The Co-evolution of Language and the Brain*. New York: W. W. Norton & Company.

De Marco, A., Cozzolino, R., Dessí-Fulgheri, F., and Thierry, B. (2011). Collective arousal when reuniting after temporary separation in *Tonkean macaques*. *Am. J. Phys. Anthropol.* 146, 457–464. doi: 10.1002/ajpa.21606

Derrickson, K. C. (1987). Yearly and situational changes in the estimate of repertoire size in Northern Mockingbirds (*Mimus polyglottos*). *Auk* 104, 198–207.

Elowson, A. M., and Snowdon, C. T. (1994). Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Anim. Behav.* 47, 1267–1277. doi: 10.1006/anbe.1994.1175

Evans, C., and Evans, L. (1999). Chicken food calls are functionally referential. *Anim. Behav.* 58, 307–319. doi: 10.1006/anbe.1999.1143

Ferrer I Cancho, R., and Solé, R. V. (2001). The small world of human language. *Proc. Biol. Sci.* 268, 2261–2265. doi: 10.1098/rspb.2001.1800

Gannon, P. J., Holloway, R. L., Broadfield, D. C., and Braun, A. R. (1998). Asymmetry of chimpanzee planum temporale: humanlike pattern of Wernicke's brain language area homolog. *Science* 279, 220–222. doi: 10.1126/science.279.5348.220

Garcia, R. R., Zamorano, F., and Aboitiz, F. (2014). From imitation to meaning: circuit plasticity and the acquisition of a conventionalized semantics. *Front. Hum. Neurosci.* 8:605. doi: 10.3389/fnhum.2014.00605

Gelman, R., and Gallistel, C. R. (2004). Language and the origin of numerical concepts. *Science* 306, 441–443. doi: 10.1126/science.1105144

Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., and Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Curr. Biol.* 17, 425–430. doi: 10.1016/j.cub.2007.01.029

Gillespie-Lynch, K., Greenfield, P. M., Lyn, H., and Savage-Rumbaugh, S. (2014). Gestural and symbolic development among apes and humans: support for a multimodal theory of language evolution. *Front. Psychol.* 5:1228. doi: 10.3389/fpsyg.2014.01228

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language–and creating language. *Child Dev.* 78, 741–745. doi: 10.1111/j.1467-8624.2007.01029.x

Goldstein, M. H., and Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychol. Sci.* 19, 515–523. doi: 10.1111/j.1467-9280.2008.02117.x

Greene, E., and Meagher, T. (1998). Red squirrels, *Tamiasciurus hudsonicus*, produce predator-class specific alarm calls. *Anim. Behav.* 55, 511–518. doi: 10.1006/anbe.1997.0620

Griffin, D. R. (1984). *Animal Thinking*. Cambridge: Harvard University Press.

Gudwin, R., Loula, A., Ribeiro, S., de Araújo, I., and Queiroz, J. (2003). "A proposal for a synthesis approach of semiotic artificial creatures," in *Recent Developments in Biologically Inspired Computing*, eds L. N. de Castro and F. J. von Zuben (Hershey: Idea Group Publishing), 270–300.

Gyger, M., Marler, P., and Pickert, R. (1987). Semantics of an avian alarm call system: the male domestic fowl, *Gallus domesticus*. *Behaviour* 102, 15–40.

Hage, S. R., Gavrilov, N., and Nieder, A. (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *J. Cogn. Neurosci.* 25, 1692–701. doi: 10.1162/jocn_a_00428

Hare, B., Brown, M., Williamson, C., and Tomasello, M. (2002). The domestication of social cognition in dogs. *Science* 298, 1634–1636. doi: 10.1126/science.1072702

Hare, B., Call, J., and Tomasello, M. (1998). Communication of food location between human and dog (*Canis familiaris*). *Evol. Commun.* 2, 137–159. doi: 10.1075/eoc.2.1.06har

Hare, B., Call, J., and Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Anim. Behav.* 61, 139–151. doi: 10.1006/anbe.2000.1518

Hare, B., Hare, B., Call, J., Call, J., Agnetta, B., Agnetta, B., et al. (2000). Chimpanzees know what conspecifics do and do not see. *Anim. Behav.* 59, 771–785. doi: 10.1006/anbe.1999.1377

Hare, B., Plyusnina, I., Ignacio, N., Schepina, O., Stepika, A., Wrangham, R., et al. (2005). Social cognitive evolution in captive foxes is a correlated by-product of experimental domestication. *Curr. Biol.* 15, 226–230. doi: 10.1016/j.cub.2005.01.040

Hare, B., and Tomasello, M. (1999). Domestic dogs (*Canis familiaris*) use human and conspecific social cues to locate hidden food. *J. Comp. Psychol.* 113, 173–177. doi: 10.1037/0735-7036.113.2.173

Hauser, M. D., Glynn, D., and Wood, J. (2007). Rhesus monkeys correctly read the goal-relevant gestures of a human agent. *Proc. Biol. Sci.* 274, 1913–1918. doi: 10.1098/rspb.2007.0586

Hauser, M. D., and Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behav. Ecol.* 4, 194–205. doi: 10.1093/beheco/4.3.194

Herman, L. M., Matus, D., Herman, E. Y. K., Ivancic, M., and Pack, A. A. (2001). The bottlenosed dolphin's (*Tursiops truncatus*) understanding of gestures as symbolic representations of body parts. *Learn. Behav.* 29, 250–264. doi: 10.3758/BF03192891

Hihara, S., Yamada, H., Iriki, A., and Okanoya, K. (2003). Spontaneous vocal differentiation of coo-calls for tools and food in Japanese monkeys. *Neurosci. Res.* 45, 383–389. doi: 10.1016/S0168-0102(03)00011-7

Hind, J. E. (1953). An electrophysiological determination of tonotopic organization in auditory cortex of cat. *J. Neurophysiol.* 16, 475–489.

Hinton, L., Nichols, J., and Ohala, J. (eds). (1994). *Sound Symbolism*. Cambridge: Cambridge University Press.

Hobaiter, C., and Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Curr. Biol.* 24, 1596–1600. doi: 10.1016/j.cub.2014.05.066

Hobaiter, C., Poisot, T., Zuberbühler, K., Hoppitt, W., and Gruber, T. (2014). Social network analysis shows direct evidence for social transmission of tool use in wild chimpanzees. *PLoS Biol.* 12:e1001960. doi: 10.1371/journal.pbio.1001960

Hopkins, W. D., Russell, J., McIntyre, J., and Leavens, D. A. (2013). Are chimpanzees really so poor at understanding imperative pointing? Some new data and an alternative view of canine and ape social cognition. *PLoS ONE* 8:e79338. doi: 10.1371/journal.pone.0079338

Horner, V., Carter, J. D., Suchak, M., and de Waal, F. B. (2011). Spontaneous prosocial choice by chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13847–13851. doi: 10.1073/pnas.1111088108

Hubel, D. H., and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* 148, 574–591. doi: 10.1113/jphysiol.2009.174151

Hultsch, H., Mundry, R., Kipper, S., and Todt, D. (2004). Long-term persistence of song performance rules in nightingales (*Luscinia megarhynchos*): a longitudinal field study on repertoire size and composition. *Behaviour* 141, 371–390. doi: 10.1163/156853904322981914

Imai, M., Kita, S., Nagumo, M., and Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition* 109, 54–65. doi: 10.1016/j.cognition.2008.07.015

Jakobson, R., and Waugh, L. (1979). *The Sound Shape of Language*. Bloomington: Indiana University Press.

Janik, V., and Slater, P. (2000). The different roles of social learning in vocal communication. *Anim. Behav.* 60, 1–11. doi: 10.1006/anbe. 2000.1410

Janik, V., and Slater, P. B. (1997). Vocal learning in mammals. *Adv. Study Behav.* 26, 59–99. doi: 10.1016/S0065-3454(08)60377-0

Jarvis, E. D., Scharff, C., Grossman, M. R., Ramos, J. A., and Nottebohm, F. (1998). For whom the bird sings: context-dependent gene expression. *Neuron* 21, 775–788. doi: 10.1016/S0896-6273(00)80594-2

Kiriazis, J., and Slobodchikoff, C. N. (2006). Perceptual specificity in the alarm calls of Gunnison's prairie dogs. *Behav. Process.* 73, 29–35. doi: 10.1016/j.beproc.2006.01.015

Kita, S., Kantartzis, K., and Imai, M. (2010). "Children learn sound symbolic words better: evolutionary vestige of sound symbolic protolanguage," in *Smith ADM, Proceedings of the 8th Conference of Evolution of Language*, eds M. Schouwstra, B. de Boer, and K. Smith (London: World Scientific), 206–213.

Koenig, A. (1995). Group size, composition and reproductive success in wild common marmosets (*Callithrix jacchus*). *Am. J. Primatol.* 35, 311–317. doi: 10.1002/ajp.1350350407

Konopka, G., Bomar, J. M., Winden, K., Coppola, G., Jonsson, Z. O., Gao, F., et al. (2009). Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* 462, 213–218. doi: 10.1038/nature08549

Leavens, D. A., Ely, J., Hopkins, W. D., and Bard, K. A. (2012). Effects of cage mesh on pointing: hand shapes in chimpanzees (*Pan troglodytes*). *Anim. Cogn.* 15, 437–441. doi: 10.1007/s10071-011-0466-6

Leavens, D. A., Russell, J. L., and Hopkins, W. D. (2005). Intentionality as measured in the persistence and elaboration of communication by chimpanzees (*Pan troglodytes*). *Child Dev.* 76, 291–306. doi: 10.1111/j.1467-8624.2005.00845.x

Loula, A., Gudwin, R., and Quieroz, J. (2004). "Symbolic communication in artificial creatures: an experiment in artificial life," in *Proceedings of the 17th Brazilian Symposium on Artificial Intelligence—SIBA (Lecture Notes Computer Science 3171)*, eds A. Bazzan and S. Labidi (São Luis: Springer), 336–345.

Lyn, H., Russell, J. L., and Hopkins, W. D. (2010). The impact of environment on the comprehension of declarative communication in apes. *Psychol. Sci.* 21, 360–365. doi: 10.1177/0956797610362218

Macedonia, J. M., and Evans, C. S. (1993). Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology* 93, 177–197. doi: 10.1111/j.1439-0310.1993.tb00988.x

Machado, C. J., Bliss-Moreau, E., Platt, M. L., and Amaral, D. G. (2011). Social and nonsocial content differentially modulates visual attention and autonomic arousal in rhesus macaques. *PLoS ONE* 6:e26598. doi: 10.1371/journal.pone.: 10.1037/0012-0026598

Magnus, M. (2010). *Gods in the Word: Archetypes in the Consonants*. CreateSpace Independent Publishing Platform.

Manser, M. B. (2001). The acoustic structure of suricates' alarm calls varies with predator type and the level of response urgency. *Proc. Biol. Sci.* 268, 2315–2324. doi: 10.1098/rspb.2001.1773

Manser, M. B., Bell, M. B., and Fletcher, L. B. (2001). The information that receivers extract from alarm calls in suricates. *Proc. Biol. Sci.* 268, 2485–2491. doi: 10.1098/rspb.2001.1772

Marler, P. (1970). Birdsong and speech development: could there be parallels? *Am. Sci.* 58, 669–673.

Márquez, C., Rennie, S. M., Costa, D. F., and Moita, M. A. (2015). Prosocial choice in rats depends on food-seeking behavior displayed by recipients. *Curr. Biol.* 25, 1736–1745. doi: 10.1016/j.cub.2015.05.018

Marshall, A., Wrangham, R., and Arcadi, A. (1999). Does learning affect the structure of vocalizations in chimpanzees? *Anim. Behav.* 58, 825–830. doi: 10.1006/anbe.1999.1219

Maurer, D., Pathman, T., and Mondloch, C. J. (2006). The shape of boubas: sound-shape correspondences in toddlers and adults. *Dev. Sci.* 9, 316–322. doi: 10.1111/j.1467-7687.2006.00495.x

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: long-term memory for novel acts and multiple stimuli. *Dev. Psychol.* 24, 470–476. doi: 10.1037/0012-1649.24.4.470

Miller, C. T., Dimauro, A., Pistorio, A., Hendry, S., and Wang, X. (2010). Vocalization induced CFos expression in marmoset cortex. *Front. Integr. Neurosci.* 4:128. doi: 10.3389/fnint.2010.00128

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827. doi: 10.1126/science.298.5594.824

Mitani, J. C., Hunley, K. L., and Murdoch, M. E. (1999). Geographic variation in the calls of wild chimpanzees: a reassessment. *Am. J. Primatol.* 47, 133–151. doi: 10.1002/(SICI)1098-2345(1999)47:2 < 133::AID-AJP4 > 3.0.CO;2-I

Moore, C., and Dunham, P. J. (1995). *Joint Attention: Its Origins and Role in Development*. New York: Lawrence Erlbaum.

Mota, N. B., Furtado, R., Maia, P. P. C., Copelli, M., and Ribeiro, S. (2014). Graph analysis of dream reports is especially informative about psychosis. *Sci. Rep.* 4, 3691. doi: 10.1038/srep03691

Mota, N. B., Vasconcelos, N. A. P., Lemos, N., Pieretti, A. C., Kinouchi, O., Cecchi, G. A., et al. (2012). Speech graphs provide a quantitative measure of thought disorder in psychosis. *PLoS ONE* 7:e34928. doi: 10.1371/journal.pone. 0034928

Mountcastle, V. B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophysiol.* 20, 408–434.

Mountcastle, V. B., Davies, P. W., and Berman, A. L. (1957). Response properties of neurons of cat's somatic sensory cortex to peripheral stimuli. *J. Neurophysiol.* 20, 374–407.

Mulcahy, N. J., and Call, J. (2009). The performance of bonobos (*Pan paniscus*), chimpanzees (*Pan troglodytes*), and orangutans (*Pongo pygmaeus*) in two versions of an object-choice task. *J. Comp. Psychol.* 123, 304–309. doi: 10.1037/a0016222

Nottebohm, F., Kasparian, S., and Pandazis, C. (1981). Brain space for a learned task. *Brain Res.* 213, 99–109. doi: 10.1016/0006-8993(81)91250-6

Osaka, N., Osaka, M., Morishita, M., Kondo, H., and Fukuyama, H. (2004). A word expressing affective pain activates the anterior cingulate cortex in the human brain: an fMRI study. *Behav. Brain Res.* 153, 123–127. doi: 10.1016/j.bbr.2003.11.013

Ouattara, K., Lemasson, A., and Zuberbühler, K. (2009). Campbell's monkeys use affixation to alter call meaning. *PLoS ONE* 4:e7808. doi: 10.1371/journal. pone.0007808

Owen-Ashley, N. T., Schoech, S. J., and Mumme, R. L. (2002). Context-specific response of Florida scrub-jay pairs to Northern Mockingbird vocal mimicry. *Condor* 104, 858–865.

Owings, D. H., and Hennessy, D. F. (1984). "The importance of variation in sciurid visual and vocal communication," in *The Biology of Ground Dwelling Squirrels*, eds J. O. Murie and G. R. Michener (Lincoln: University of Nebraska Press), 171–200.

Owren, M. J., Amoss, R. T., and Rendall, D. (2011). Two organizing principles of vocal production: implications for nonhuman and human primates. *Am. J. Primatol.* 73, 530–544. doi: 10.1002/ajp.20913

Owren, M. J., and Casale, T. M. (1994). Variations in fundamental frequency peak position in Japanese macaque (*Macaca fuscata*) coo calls. *J. Comp. Psychol.* 108, 291–297. doi: 10.1037/0735-7036.108.3.291

Pack, A. A., and Herman, L. M. (2004). Dolphins (*Tursiops truncatus*) comprehend the referent of both static and dynamic human gazing and pointing in an object choice task. *J. Comp. Psychol.* 118, 160–171. doi: 10.1037/0735-7036.118.2.160

Peirce, C. S. (1958). *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Peirce, C. S. (1998). *The Essential Peirce: Selected Philosophical Writings*. Bloomington: Indiana University Press.

Pepperberg, I. M., Willner, M. R., and Gravitz, L. B. (1997). Development of Piagetian object permanence in a grey parrot (*Psittacus erithacus*). *J. Comp. Psychol.* 111, 63–75. doi: 10.1037/0735-7036.111.1.63

Petkov, C. I., and Jarvis, E. D. (2012). Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* 4:12. doi: 10.3389/fnevo.2012.00012

Pfenning, A. R., Hara, E., Whitney, O., Rivas, M. V., Wang, R., Roulhac, P. L., et al. (2014). Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346, 1256846. doi: 10.1126/science.1256846

Pierce, J. D. (1985). A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26, 202–213. doi: 10.1007/BF02382019

Price, P. H. (1979). Developmental determinants of structure in zebra finch song. *J. Comp. Physiol. Psychol.* 93, 260–277. doi: 10.1037/h0077553

Queiroz, J., and Ribeiro, S. (2002). "The biological substrate of icons, indexes and symbols in animal communication," in *The Peirce Seminar Papers*, Vol. 5, ed. M. Shapiro (Oxford: Berghahn Books), 69–78.

Rendall, D. (2003). Acoustic correlates of caller identity and affect intensity in the vowel-like grunt vocalizations of baboons. *J. Acoust. Soc. Am.* 113, 3390–3402. doi: 10.1121/1.1568942

Ribeiro, S. (2010). *Song, Sleep, and the Slow Evolution of Thoughts: Studies on Brain Representation*. Lambert Academic Publishing.

Ribeiro, S., Cecchi, G. A., Magnasco, M. O., and Mello, C. V. (1998). Toward a song code: evidence for a syllabic representation in the canary brain. *Neuron* 21, 359–371. doi: 10.1016/S0896-6273(00)80545-0

Ribeiro, S., Loula, A., de Araújo, I., Gudwin, R., and Queiroz, J. (2007). Symbols are not uniquely human. *Biosystems* 90, 263–272. doi: 10.1016/j.biosystems.2006.09.030

Richards, D. G., Wolz, J. P., and Herman, L. M. (1984). Vocal mimicry of computer generated sounds and vocal labeling of objects by a bottlenosed dolphin, *Tursiops truncatus*. *J. Comp. Psychol.* 98, 10–28. doi: 10.1037/0735-7036.98.1.10

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., et al. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nat. Neurosci.* 11, 426–428. doi: 10.1038/nn2072

Savage-Rumbaugh, E. S. (1990). Language acquisition in a nonhuman species: implications for the innateness debate. *Dev. Psychobiol.* 23, 599–620. doi: 10.1002/dev.420230706

Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., and Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan paniscus*). *J. Exp. Psychol. Gen.* 115, 211–235. doi: 10.1037/0096-3445.115.3.211

Savage-Rumbaugh, E. S., and Romsky, M. A. (1989). "Symbol acquisition and use by *Pan troglodytes*, *Pan paniscus*, *Homo sapiens*," in *Understanding Chimpanzees*, eds P. G. Heltne and L. A. Marquardt (Cambridge, MA: Harvard University Press), 266–295.

Savage-Rumbaugh, E. S., Rumbaugh, D. M., and Boysen, S. (1978). Symbolic communication between two chimpanzees (*Pan troglodytes*). *Science* 201, 641–644. doi: 10.1126/science.675251

Scharff, C., and Petri, J. (2011). Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 2124–2140. doi: 10.1098/rstb.2011.0001

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., and Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PLoS ONE* 8:e76674. doi: 10.1371/journal.pone.0076674

Seidenberg, M. S., and Petitto, L. A. (1987). Communication, symbolic communication, and language in child and chimpanzee: comment on Savage-Rumbaugh, McDonald, Sevcik, Hopkins, and Rupert (1986). *J. Exp. Psychol. Gen.* 116, 279–287. doi: 10.1037/0096-3445.116.3.279

Seyfarth, R. M., and Cheney, D. L. (1988). Empirical tests of reciprocity theory: problems in assessment. *Ethol. Sociobiol.* 9, 181–187. doi: 10.1016/0162-3095(88)90020-9

Seyfarth, R. M., and Cheney, D. L. (2010). Production, usage, and comprehension in animal vocalizations. *Brain Lang.* 115, 92–100. doi: 10.1016/j.bandl.2009.10.003

Seyfarth, R. M., Cheney, D. L., and Marler, P. (1980). Monkey responses to three different alarm calls: evidence of predator classification and semantic communication. *Science* 210, 801–803. doi: 10.1126/science.7433999

Shen, H. (2013). Precision gene editing paves way for transgenic monkeys. *Nature* 503, 14–15. doi: 10.1038/503014a

Simões, C. S., Vianney, P. V. R., de Moura, M. M., Freire, M. A. M., Mello, L. E., Sameshima, K., et al. (2010). Activation of frontal neocortical areas by vocal production in marmosets. *Front. Integr. Neurosci.* 4:123. doi: 10.3389/fnint.2010.00123

Simons, D. J., and Woolsey, T. A. (1979). Functional organization in mouse barrel cortex. *Brain Res.* 165, 327–332. doi: 10.1016/0006-8993(79)90564-X

Slobodchikoff, C. N., Kiriazis, J., Fischer, C., and Creef, E. (1991). Semantic information distinguishing individual predators in the alarm calls of Gunnison's prairie dogs. *Anim. Behav.* 42, 713–719. doi: 10.1016/S0003-3472(05)80117-4

Snowdon, C. T., and Elowson, A. M. (1999). Pygmy marmosets modify call structure when paired. *Ethology* 105, 893–908. doi: 10.1046/j.1439-0310.1999.00483.x

Snowdon, C. T., and Hodun, A. (1981). Acoustic adaptation in pygmy marmoset contact calls: locational cues vary with distances between conspecifics. *Behav. Ecol. Sociobiol.* 9, 295–300. doi: 10.1007/BF00299886

Sossinka, R., and Bohner, J. (1980). Song types in the zebra finch (*Poephila guttata castanotis*). *Z. Tierpsychol.* 53, 123–132.

Sporns, O. (2010). *Networks of the Brain*, 1st Edn. London: The MIT Press.

Struhsaker, T. T. (1967). Social structure among vervet monkeys (*Cercopithecus aethiops*). *Behaviour* 29, 6–121. doi: 10.1163/156853967X00073

Sutton, D., Larson, C., Taylor, E. M., and Lindeman, R. C. (1973). Vocalization in rhesus monkeys: conditionability. *Brain Res.* 52, 225–231. doi: 10.1016/0006-8993(73)90660-4

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

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

Tomasello, M. (1999). The human adaptation for culture. *Annu. Rev. Anthropol.* 28, 509–529.

Tomasello, M., and Call, J. (1997). *Primate Cognition*. Oxford: Oxford University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; discussion 691–735. doi: 10.1017/S0140525X05000129

Tootell, R. B., Silverman, M. S., Switkes, E., and De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science* 218, 902–904. doi: 10.1126/science.7134981

Tyack, P. L. (2008). Convergence of calls as animals form social bonds, active compensation for noisy communication channels, and the evolution of vocal learning in mammals. *J. Comp. Psychol.* 122, 319–331. doi: 10.1037/a0013087

von Humboldt, W. (1836). *The Heterogeneity of Language and its Influence on the Intellectual Development of Mankind New edition: On Language. On the Diversity of Human Language Construction and Its Influence on the Mental Development of the Human Species*. New York: Cambridge University Press. [2nd Revised Edn, 1999].

von Rohr, C. R., Koski, S. E., Burkart, J. M., Caws, C., Fraser, O. N., Ziltener A., et al. (2012). Impartial third-party interventions in Captive chimpanzees: A reflection of community concern. *PLoS ONE* 7: e32494. doi: 10.1371/journal.pone.0032494

Walton, C., Pariser, E., and Nottebohm, F. (2012). The zebra finch paradox: song is little changed, but number of neurons doubles. *J. Neurosci.* 32, 761–774. doi: 10.1523/JNEUROSCI.3434-11.2012

Warneken, F., Chen, F., and Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Dev.* 77, 640–663. doi: 10.1111/j.1467-8624.2006.00895.x

Watson, S. K., Townsend, S. W., Schel, A. M., Wilke, C., Wallace, E. K., Cheng, L., et al. (2015). Vocal learning in the functionally referential food grunts of chimpanzees. *Curr. Biol.* 25, 1–5. doi: 10.1016/j.cub.2014.12.032

Whiten, A., Goodall, W. C., McGrew, T., Nishida, V., Reynolds, V., Sugiyama, C., et al. (1999). Cultures in chimpanzees. *Nature* 399, 682–685. doi: 10.1038/21415

Williams, H., and Staples, K. (1992). Syllable chunking in zebra finch (*Taeniopygia guttata*) song. *J. Comp. Psychol.* 106, 278–286. doi: 10.1037/0735-7036.106.3.278

Worley, K. C., Warren, W. C., Rogers, J., Locke, D., Muzny, D. M., Mardis, E. R., et al. (2014). The common marmoset genome provides insight into primate biology and evolution. *Nat. Genet.* 46, 850–857. doi: 10.1038/ng.3042

Yamaguchi, S. I., and Myers, R. E. (1972). Failure of discriminative vocal conditioning in Rhesus monkey. *Brain Res.* 37, 109–114. doi: 10.1016/0006-8993(72)90350-2

Zuberbühler, K. (2000). Local variation in semantic knowledge in wild Diana monkey groups. *Anim. Behav.* 59, 917–927.

Zuberbühler, K. (2001). Predator-specific alarm calls in Campbell's monkeys, *Cercopithecus campbelli*. *Behav. Ecol. Sociobiol.* 50, 414–422. doi: 10.1007/s002650100383