

*An: Mind the Gap: Tracing
the Origins of Human
Universals. Ed. P.M.
Kappeler + J. B. Sille.
2010. Pp. 283-298.
New York: Springer.*

Chapter 13 Primate Communication and Human Language: Continuities and Discontinuities

Dorothy L. Cheney and Robert M. Seyfarth

Abstract Here, we review some questions in the production and perception of nonhuman primate vocalizations, focusing on three related issues. First, flexible vocal production separates humans not only from nonhuman primates but also from most other mammals and birds. The rarity of learned, modifiable call production in most mammals suggests that important changes in the mechanisms governing human phonation occurred relatively recently in human evolution, after the divergence of our hominid ancestors from the common ancestors of humans and chimpanzees. Second, while exclusive focus on vocal production reveals clear differences between humans and most other species, a broader examination of call usage and perception paints a more complicated picture, with both similarities and differences between the two groups. Third, an analysis of vocal production and perception draws attention to the very different mechanisms that underlie the behavior of signalers and recipients, even when they are involved in the same communicative event. Nonhuman primates have only a small repertoire of acoustically fixed vocalizations. However, because calls are individually distinctive and each call type is predictably linked to a particular social context, this limited call repertoire can, nonetheless, provide listeners with an open-ended, highly modifiable, and cognitively rich set of meanings, allowing them to construct "narratives" of unseen events. However, although nonhuman primates and other animals seem capable of thinking, as it were, in simple sentences, this ability does not motivate them to speak in sentences. Their knowledge remains largely private. We suggest that long before our ancestors spoke in sentences, they had a language of thought in which they represented the world – and the meaning of call sequences – in terms of actors, actions, and those who are acted upon. The linguistic revolution occurred

D.L. Cheney (✉)

Department of Biology, University of Pennsylvania, Philadelphia, PA, USA
e-mail: cheney@sas.upenn.edu

R.M. Seyfarth

Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA
e-mail: seyfarth@psych.upenn.edu

when our ancestors began to express this tacit knowledge, and to use their cognitive skills in speaking as well as listening.

13.1 Introduction

Darwin (1871) drew attention to a dichotomy in the vocal communication of animals that had perplexed philosophers and naturalists for at least 2,000 years. In marked contrast to human language, he wrote, animal vocalizations appeared to be involuntary expressions of emotion and movement: "When the sensorium is strongly excited, the muscles of the body are generally thrown into violent action; and as a consequence, loud sounds are uttered, . . . although the sounds may be of no use" ([1871] 1981: 83). Two pages later, however, Darwin also acknowledged: "That which distinguishes man from the lower animals is not the understanding of articulate sounds, for, as every one knows, dogs understand many words and sentences. . . . Nor is it the mere capacity of connecting definite sounds with definite ideas; for it is certain that some parrots, which have been taught to speak, connect unerringly words with things, and persons with events" ([1871] 1981: 85).

As a contemporary example of the dichotomy first identified by Darwin, consider Rico, a border collie dog who learned the names of more than 200 different toys (Kaminski et al. 2004). Rico was able to learn and remember the names of new toys by process of exclusion, or "fast mapping," and – like small children – used gaze and attention to guide word learning. But of course, Rico never learned to *say* any of the words he learned. In this respect, his limited vocal production and extensive comprehension are similar to those of human-trained sea lions (Schusterman et al. 2002) and dolphins (Herman et al. 1993).

The dichotomy between call production and perception is puzzling because listeners are also signalers. The same animal that has no problem understanding a word like "ball" is completely unable to articulate it. Indeed, while there are fundamental differences between animal communication and human language in the context of call production, in the context of call perception, these differences seem much less obvious. So why do not animals articulate their mental representations of the world and other individuals more often, and in more contexts? Why does their knowledge remain largely private?

Here, we review some questions in the production and perception of nonhuman primate vocalizations, focusing in particular on the communication of free-ranging baboons (*Papio hamadryas* spp.) (for a more detailed discussion, see Cheney and Seyfarth 2007). For those interested in comparing animal vocal communication with human language, three conclusions emerge. First, flexible vocal production separates humans not only from nonhuman primates but also from most other mammals and birds. The rarity of learned, modifiable call production throughout most of the Class *Mammalia* suggests that important changes in the mechanisms governing human phonation occurred relatively recently in human evolution, after the divergence of our hominid ancestors from the common ancestors of humans and

chimpanzees. Second, while an exclusive focus on vocal production reveals clear differences between humans and most other species, a broader examination of call usage and perception paints a more complicated picture, with both similarities and differences between the two groups. Third, an analysis of vocal production and perception draws attention to the very different mechanisms that underlie the behavior of signalers and recipients, even when they are involved in the same communicative event. Nonhuman primates have only a small repertoire of acoustically fixed vocalizations. However, because calls are individually distinctive and each call type is predictably linked to a particular social context, this limited call repertoire can nonetheless provide listeners with an open-ended, highly modifiable, and cognitively rich set of meanings. Listeners can acquire a huge number of messages from a finite number of signals.

13.2 Call Production and Development

Monkeys and apes have a small repertoire of calls that show relatively little modification in their acoustic properties during development. The development of vocal production is largely unaffected by variations in auditory experience or rearing (see Hammerschmidt and Fischer 2008 for a recent review). For example, or male baboons produce loud two-syllable *alarm wahoos* when they encounter lions give during competitive contests with other males, but the two types of wahoo differ according to a number of acoustic measures (Fischer et al. 2002). Similarly, the *alarm* barks given by females and juveniles are acoustically similar to the *contact* barks that baboons give when they become lost or separated from their companions. Again, however, there are subtle acoustic differences between the two bark types that allow them to be distinguished by ear (Fischer et al. 2001a).

Unlike predator alarm calls, which depend in a fairly simple way on the type of predator or the degree of danger, the vocalizations given by animals during social interactions are elicited by a more complex array of factors that may include both the immediate social context and the history of interactions between the particular individuals involved. Baboon grunts offer an example.

Grunts are the baboons' most common vocalization. They are individually distinctive, given in a variety of non-aggressive circumstances, and seem to comprise two acoustically graded types (Owren et al. 1997; Rendall 2003). The *move* grunt is typically given in the context of group movement. Like similar calls given by other primate species, *move* grunts function to alert other individuals to the signaler's intentions to travel in a particular direction. By contrast, the *infant* grunt is given during many sorts of friendly interactions and functions as a signal of benign intent. *Infant* grunts are most commonly given in the context of infant handling, but they also occur during grooming, other friendly behavior, and reconciliation. The two grunt subtypes thus differ in the specificity of the stimuli that elicit them. *Move* grunts are linked to a specific context. By comparison, *infant*

grunts are given in a wide variety of friendly situations. Whereas *infant* grunts function to signal benign intent toward one specific individual, *move* grunts broadcast the signaler's intentions to many individuals. *Move* and *infant* grunts exemplify the variation found in the baboon vocal repertoire, and indeed in the vocal repertoires of many other primates. Some calls are tightly linked to a relatively narrow context, whereas others are used in a wider variety of circumstances. Some calls are directed at a specific individual, whereas others are more widely broadcast.

The acoustic intergradation in many nonhuman primate vocalizations may be caused by gradation in the caller's arousal or emotional state. Analyses of baboons' *contact* and *alarm* barks, *contest* and *alarm* wahoos, and *move* and *infant* grunts all suggest that acoustic variation is consistent with variation in the caller's emotions (Fischer et al. 2001a, 2002, 2004; Rendall 2003, see also Jürgens 1995). In each of these cases, however, graded calls whose production may be determined largely by the signaler's emotions are nonetheless perceived as discretely different vocalizations (Rendall et al. 1999). And once listeners have recognized that each of these discretely different signals is predictably linked to a particular event, calls have the potential to convey to listeners a meaning that goes far beyond information about the signaler's emotional state.

In marked contrast to children, who learn to produce and comprehend thousands of new words during their first three years of life, monkeys and apes rarely modify their vocal repertoires by adding new sounds. Although some primates make subtle modifications in their vocalizations as a result of experience (Hauser 1992; Elowson and Snowdon 1994; Mitani and Brandt 1994; Crockford et al. 2004) and can modify the loudness of their calls through auditory feedback (Hage et al. 2006), a baboon in Kenya produces more or less the same sounds in the same contexts as a baboon in Botswana. This conclusion follows not only from research on many primate species (Seyfarth and Cheney 1997a) but also from a cross-fostering experiment involving two closely related species of macaques. In this experiment, juveniles that had been fostered into a group of macaques from another species continued to produce their own species' calls, despite being physically capable of producing their adoptive species' calls (Owren et al. 1993). Monkeys seem genetically predisposed to give particular calls in particular contexts.

This is not to say that nonhuman primate call production is involuntary. In both the field and the laboratory, nonhuman primates can control their vocalization and choose to either vocalize or remain silent. After behaving aggressively toward a subordinate, for example, a baboon may give a "reconciliatory" grunt to her opponent or she may not (Cheney et al. 1995b; Cheney and Seyfarth 1997). Similarly, when capuchin monkeys (*Cebus capucinus*) find food, they may call or remain silent (Gros-Louis 2004). Even in highly emotional circumstances like encounters with predators, some individuals give alarm calls at high rates, others call less often, and still others remain silent (Cheney and Seyfarth 1990). In more controlled laboratory settings, the timing, duration, and rate of calling by monkeys can be brought under operant control (Pierce 1985; Egnor et al. 2007). Clearly, primates can control whether they vocalize or not depending upon variations in both the ecological, social, and acoustic environments.

Within a given context, nonhuman primates can also make subtle modifications in the acoustic structure of their calls (reviewed by Hammerschmidt and Fischer 2008; Seyfarth and Cheney in press). To cite just one example, wild chimpanzees in Uganda give long, elaborate pant-hoots either alone or in "choruses" with others. When two individuals have called together several times, the acoustic features of their pant-hoots begin to converge (Mitani and Brandt 1994; Mitani and Gros-Louis 1998). Apparently, individuals modify the acoustic structure of their calls depending upon auditory experience (see also Crockford et al. 2004).

Many other studies suggest that, whereas the basic acoustic structure of nonhuman primate calls is innately determined, the fine spectrotemporal features of vocalizations can be modified (for reviews see Seyfarth and Cheney 1997a, in press; Egnor and Hauser 2004; Hammerschmidt and Fischer 2008). Some features are more easily modified than others. As Janik and Slater (1997) first pointed out, temporal features like call duration and amplitude are more easily modified than frequency parameters. The distinction between relatively innate and more modifiable components of phonation is important, because it has significant implications for future research on the neurobiology of primate vocal production (see Egnor and Hauser 2004; Hammerschmidt and Fischer 2008 for further discussion).

In their relatively fixed vocal production, nonhuman primates are typical of most mammals and even the great majority of birds. In their 1997 review, Janik and Slater found evidence for learned, modifiable vocal production in only three orders of birds, cetaceans, harbor seals, and humans. Although we may yet be surprised by novel evidence of vocal imitation (e.g., Poole et al. 2005) or creative call combinations (Zuberbühler 2002; Crockford and Boesch 2003; Arnold and Zuberbühler 2006), for the moment it appears that the ability to modify vocal production depending upon experience is comparatively modest.

13.3 Call Perception

Whenever there is a predictable relation between a particular call type and a specific social context, a vocalization has the potential to inform nearby listeners about objects or events. The underlying mechanisms are irrelevant. A tone that informs a rat about the imminence of a shock, an alarm call that informs a vervet about the presence of an eagle, or a scream that informs a baboon that her offspring is involved in a fight all have the potential to provide a listener with precise information if they are predictably associated (Rescorla 1988) with a narrow range of events. The widely different mechanisms that lead to this association have no effect on the signal's potential to inform (Seyfarth and Cheney 2003).

But while they have this potential, do animal vocalizations really provide listeners with information about who is doing what to whom? The study of vocal perception in animal communication is fraught with both practical and conceptual difficulties because we cannot interview our subjects. As a result, the only way to

determine the "meaning" of a signal is to examine the recipient's response — a very crude measure of whether the recipient has or has not acquired information.

In some cases, animals respond in qualitatively different ways to different vocalizations, and the characteristics of their responses suggest that each call type conveys specific information. Consider, for example, the responses of Diana monkeys (*Cercopithecus diana*) to a playback of different alarm calls. When Diana monkeys hear the growl of a leopard, they respond with their own "leopard alarm", when they hear the shriek of an eagle, they give their own "eagle alarm" (Zuberbuhler et al. 1999).

In many other cases, the lack of qualitatively different responses complicates the interpretation of call meaning. Two methods help to circumvent this problem, at least partially. First, if observations suggest that some generic response, like looking toward the speaker, will be the only reaction elicited by a playback experiment, scientists can design a matched pair of trials, alike in all but one respect, and then compare subjects' responses under two different conditions. A consistent difference in the duration of response may permit inferences about the different sorts of information conveyed by different calls, or by the same call under different conditions. Alternatively, subjects may show no immediate response to playback of a call but their subsequent behavior may nonetheless be affected. Having heard a particular call from individual X, for example, a subject may be more likely to approach X in the next 30 min than if no call, or the call of a different individual, had been played. Such longer-term changes in behavior also allow one to make inferences, albeit indirect, about the meaning of a specific vocalization (Seyfarth and Cheney 1997b, for examples see Cheney and Seyfarth 2007).

Playback experiments on a wide variety of species have now demonstrated that many nonhuman primate calls function referentially, providing listeners with information about what is happening and who is involved. This is true both of calls that are acoustically very different and calls that are acoustically graded. Baboons, for example, respond differently to *alarm* and *contest* wahoos, *alarm* and *contact* barks, and *move* and *infant* grunts (Rendall et al. 1999; Fischer et al. 2001b; Kitchen et al. 2003). Signaler identity and context also play crucial roles in informing listeners about what is occurring. Whatever the mechanisms underlying call production, therefore, listeners extract meaning from the calls they hear.

The call-meaning relationship in the listener's mind is interesting in several respects. First, it constitutes an arbitrary association between a sound and the thing for which it stands. There is nothing about the sound of a Diana monkey's *leopard* alarm call, for example, that sounds like a leopard, and nothing about the sound of a monkey's *eagle* alarm that would obviously link it to an eagle. In much the same way, there is nothing in the acoustic details of baboons' *alarm* and *contest* wahoos that would help a listener learn that one is given in response to predators while another is given during male-male aggression.

Second, the meaning of each call is defined not just by its relation to an object in the world but also by its relation to other calls in the monkey's repertoire. A male Diana monkey's *leopard* alarm is similar in meaning to a leopard's growl and a female's *leopard* alarm, but different in meaning from all of three eagle-associated

calls. In the monkey's mind, there exists a kind of semantic space in which the three leopard-associated sounds are closely linked in one cluster, whereas the three eagle-associated sounds are closely linked in another.

This leads to a third conclusion, that primate calls are acoustic units linked to particular concepts. When a Diana monkey hears a male's *leopard* alarm, she appears to form a mental representation of the call's meaning. Then, when she hears a leopard's growl coming from the same location shortly thereafter, she forms a second representation and compares the two calls on the basis of their meaning. Her lack of response to the now redundant growl is based on this assessment. The female, in other words, has a concept — a kind of mental image — of a leopard. The concept can be activated by any one of three quite different sounds that are linked together based on their shared meaning (e.g., Cohen et al. 2006). The concept is also amodal or multimodal, involving a combination of visual and auditory information (Gili-da-Costa et al. 2004; Ghazanfar et al. 2005).

As another example, consider the phenomenon of individual recognition by voice, which has been amply demonstrated in many species and underlies many of our playback experiments (e.g., Snowdon 1986; Rendall et al. 1996; Owen et al. 1997). Baboons clearly recognize other group members from their voices alone, regardless of whether they are giving a grunt, a *contact* bark, or a threat-grunt, and regardless of whether they are vocalizing in a calm or in an agitated manner. Despite wide variation in the acoustic cues that mark a call as a particular individual's, and the fact that the calls of one individual may grade acoustically into the calls of another, listeners still link each call with a unique individual in a discrete, categorical fashion. Individual recognition occurs in so many contexts, with so many vocalizations, that it is hard to escape the impression that listeners have a mental representation, or concept, of each group member as an individual. If monkeys were human, we would call this a concept of *person*.

In sum, whereas call production in primates is relatively fixed, the cognitive mechanisms that underlie call perception are considerably more complex. Underlying primates' assessment of call meaning is a rich conceptual structure, in which calls are linked both to objects and relations in the world and to other calls in the species' repertoire. When responding to calls, monkeys act as if they recognize individuals and have concepts like *leopard*, *eagle*, *close associate*, and so on. The contrast between impoverished production and rich, conceptually based perception argues strongly against the view that a concept cannot be acquired unless it is instantiated in one's language (reviewed by Gleitman and Papaflegou 2005). Monkeys and apes have many concepts for which they have no words.

13.4 Syntax

There is little evidence for rule-governed syntax in the calls of nonhuman primates. Recent work by Zuberbuhler and colleagues on the alarm calls of forest monkeys provide intriguing evidence that the presence of one call type can modify the

meaning of another (Zuberbuhler 2002; Arnold and Zuberbuhler 2006; Clarke et al. 2006), and a study by Crockett and Boesch (2003) suggests that a call combination in chimpanzees may carry new meaning that goes beyond the meaning of the individual calls themselves, but these rare exceptions meet few of the definitions of human syntax.

Nonetheless, monkeys and apes *hear* different calls in combination all the time. These calls are usually given by different individuals, allowing listeners to construct narratives about what is happening and who is involved. In assessing call sequences, listeners attend simultaneously to the signalers' identities, call type, the rank and kinship of the signalers, and the temporal and spatial juxtaposition of different individuals' calls. Baboons, for example, respond much more strongly to a call sequence that suggests a reversal in the female dominance hierarchy than one that is consistent with it (Cheney et al. 1995a; Bergman et al. 2003). When played a threat-grunt, scream sequence in which a high-ranking female – say, Sylvia – threat-grunts and a lower-ranking female – say, Hannah – screams, they show little if any response. Their responses are much stronger, however, if they hear a sequence that appears to violate their knowledge of the female dominance hierarchy: for example, Hannah threat-grunts and Sylvia screams. Although the callers' identities and the call types are the same, subtle changes in the elements of the sequence cause its meaning to change fundamentally. Listeners also seem to assume a causal relationship between calls that are closely juxtaposed in time: Hannah's threat-grunts *caused* Sylvia to scream. These assessments are done instantaneously and probably large unconsciously. They are based on discrete properties, such as the signalers' identities, kinship, and rank, that are combined in a combinatorial system and that encode propositional information: for example, A is threatening B; A is mating with B; and so on. All of these features are also present in language.

13.5 Attributing Intentions to Signalers

During conversation, humans routinely make inferences about the motives and beliefs of their intended recipients. Baboons, too, seem to recognize the intended recipient of someone else's calls.

Baboon groups are noisy, tumultuous societies, and a baboon could not manage her social interactions if she interpreted every vocalization she heard as directed at her. Inferences about the “directedness” of vocalizations are probably often mediated by gaze direction and relatively simple contingencies. Even in the absence of visual signals, however, baboons are able to make inferences about the intended recipient of a call based on their knowledge of a signaler's identity and the nature of recent interactions. For example, when females were played the “reconciliatory grunt” of a recent aggressor within minutes after being threatened, they were more likely to approach their former opponent and to tolerate their opponent's approaches than after hearing either no grunt or the grunt of another

dominant female unrelated to their opponent (Cheney and Seyfarth 1997). Call type was also important, because females avoided their recent opponent if they heard her threat-grunt rather than her reconciliatory grunt (Engh et al. 2006). Thus, baboons use their memory of recent interactions to make inferences about whether a call is being directed at themselves or at some other individual.

In some cases, these inferences are complex and indirect, and call upon baboons' knowledge of the kinship relationships of other group members. For example, when female baboons were played the threat-grunts of their aggressor's relative soon after being threatened, they avoided members of their aggressor's matriline. In contrast, when they heard the same threat-grunts in the absence of aggression, they ignored the call and acted as if they assumed that the call was directed at someone else (Witig et al. 2007a). Similarly, when subjects heard the “reconciliatory” grunt of their aggressor's relative after a fight, they were more likely to approach both their aggressor and the relative whose grunt they had heard (Witig et al. 2007b). They did not do so, however, if they had heard the “reconciliatory” grunt of another, unrelated female, nor did they approach their aggressor's other relatives. Here again, subjects behaved as if they believed that a grunt from their aggressor's relative must be directed at *them*, as a consequence of the fight. What is especially interesting in these experiments is that subjects inferred that they were the target of the vocalization even though they had not recently interacted with the signaler, but with her relative. They could only have done so if they recognized the close bond that existed between the two females.

In primates, faces and voices are the primary means of transmitting social signals, and monkeys recognize the correspondence between facial and vocal expressions (Ghazanfar and Logothetis 2003). Presumably, visual and auditory signals are somehow combined to form a unified, multimodal percept in the mind of a monkey. In a study using positron emission tomography (PET), Gil-da-Costa et al. (2004) showed that when rhesus macaques hear one of their own species' vocalizations, they exhibit neural activity not only in areas associated with auditory processing but also in higher-order visual areas, including STS. Auditory and visual areas also exhibit significant anatomical connections (Poremba et al. 2003).

Ghazanfar et al. (2005) explored the neural basis of sensory integration using the *coos* and *grunts* of rhesus macaques as stimuli. They found clear evidence that cells in certain areas of the auditory cortex are more responsive to bi-modal (visual and auditory) presentation of species-specific calls than to unimodal presentation. Although significant integration of visual and auditory information occurred in trials with both vocalizations, the effect of cross-modal presentation was greater with *grunts* than with *coos*. The authors speculate that this may occur because *grunts* are usually directed toward a specific individual in dyadic interactions, whereas *coos* tend to be broadcasted generally to the group at large. The greater cross-modal integration in the processing of grunts may therefore have arisen because, in contrast to listeners who hear a *coo*, listeners who hear a *grunt* must immediately determine whether or not the call is directed at them.

In sum, when deciding “Who, me?” upon hearing a vocalization, baboons must take into account the identity of the signaler (who is it?), the type of call given

(friendly or aggressive?), the nature of their prior interactions with the signaler (were they aggressive, friendly, or neutral?), and the correlation between past interactions and future ones (does a recent grooming interaction lower or increase the likelihood of aggression?). Learned contingencies, doubtless, play a role in these assessments. But because listeners' responses depend on simultaneous consideration of all of these factors, this learning is likely to be both complex and subtle.

Although baboons and other monkeys may be able to recognize other individual's intentions when inferring whether or not they are the target of another individual's call, there is no evidence that they take into account their audience's knowledge or beliefs when producing or assessing calls. Both monkeys and apes give alarm calls, for example, without any apparent recognition of whether listeners are ignorant or already informed about the presence of a predator (reviewed by Cheney and Seyfarth 2007). Similarly, although the "food calls" of capuchin monkeys (Gros-Louis 2004) and the pant hoots of chimpanzees (Clark and Wrangham 1994) attract others to food, signalers show no evidence of recognizing whether their audience is already aware of the presence of food. Baboons often give *contact* barks when separated from others. When several individuals are calling simultaneously, it often appears that they are answering each other's calls in order to inform others of the group's location. Playback experiments suggest, however, that baboons call primarily with respect to their own separation from the group, not their audience's. They "answer" others when they themselves are separated, and they often fail to respond to the calls of even their offspring when they themselves are in close proximity to other group members (Cheney et al. 1996; Rendall et al. 2000). In this respect, the vocalizations of monkeys and apes are very different from human speech, where we routinely take into account our audience's beliefs and knowledge during conversation (Grice 1957).

13.6 Primate Communication and the Evolution of Language

The vocal communication of nonhuman primates is very different from human language, especially in the domain of call production. At the same time, however, comparisons between primate communication and human language have tended to focus on the differences, ignoring some of the intriguing continuities in perception and cognition.

As already noted, the striking difference between production and comprehension in animal communication is puzzling because producers are also perceivers: why should an individual who can deduce an almost limitless number of meanings from the calls of others be able to produce only a limited number of calls of her own? The difference may arise because call production depends on mechanisms of phonation, which are largely innate, whereas comprehension depends on mechanisms of learning (classical and operant conditioning), which are considerably more malleable. But this explanation offers no answer to a crucial question: Why has natural

selection so rarely acted to favor flexible vocal production? Here, we offer some speculations as they apply to human and nonhuman primates.

At some point in our evolutionary history – probably after the divergence of the evolutionary lines leading to chimpanzees and bonobos on the one hand and humans on the other hand (Enhard et al. 2002) – our ancestors developed much greater control over the physiology of vocal production. As a result, vocal output became both more flexible and considerably more dependent on auditory experience and imitation (Lieberman 1991; Fitch 2007). What selective pressures might have given rise to these physiological changes?

Vocal communication in nonhuman primates lacks three features that are abundantly present in human language: the ability to generate new words, lexical syntax, and a theory of mind. By the latter, we mean the ability of both speakers and listeners to make attributions about each others' beliefs, knowledge, and other mental states (Grice 1957). These are the simplest, most basic features that distinguish human and nonhuman primate vocal production, and it is with these traits that language evolution, we need not worry about the more complex properties of language that probably came later – properties like case, tense, subject–verb agreement, open- and closed-class items, recursion, long-distance dependency, subordinate clauses, and so on.

How might the ability to generate new words, lexical syntax, and a theory of mind have evolved: simultaneously, in response to the same selective pressures, or more serially, in some particular order? We propose that the evolution of a theory of mind preceded language, creating the selective pressures that gave rise to the ability to generate new words and lexical syntax, and to the flexibility in vocal production that these two traits would have required (Cheney and Seyfarth 2005, 2007).

There is no evidence in nonhuman primates for anything close to the large vocal repertoire we find even in very young children. Similarly, nonhuman primates provide few examples of lexical syntax. By contrast, there is growing evidence that both Old World monkeys (Flombaum and Santos 2005; Engh et al. 2006; Cheney and Seyfarth 2007) and apes (Hare et al. 2001; Tomasello et al. 2005; Buttelmann et al. 2007) may possess rudimentary abilities to attribute motives or knowledge to others, and engage in simple forms of shared attention and social referencing. Taken together, these data suggest that a rudimentary theory of mind appeared among primates long before flexible vocal production, the ability to generate new words, and lexical syntax.

A rudimentary theory of mind seems to be crucially important for word learning in young children. Beginning as early as 9–12 months, children exhibit a nascent understanding of other individuals' motives, beliefs, and desires, and this skill forms the basis of a shared attention system that is essential for early word learning (Gloom and Markson 1998; Tomasello 2003). One-year old children understand implicitly that words can be mapped onto objects and actions. Crucially, this understanding is accompanied by a kind of "social referencing" in which the child uses other people's direction of gaze, gestures, and emotions to assign labels to objects (Baldwin 1991, reviewed in Pinker 1994; Fisher and Gleitman 2002).

Gaze and attention also facilitate word learning in dogs and other animals. Children, however, rapidly surpass the simpler forms of shared attention and word learning demonstrated by animals. Long before they begin to speak in sentences, young children develop implicit notions of objects and events, actors, actions, and those that are acted upon. Moreover, in contrast to monkeys, apes, and other animals, 1-year-old children are highly motivated to share what they know with others (Tomassello and Carpenter 2007). While animals are concerned with their own goals and knowledge, young children are concerned with making their thoughts and knowledge publicly available. The acquisition of a theory of mind thus creates a cognitive environment that drives the acquisition of new words and new grammatical skills. Indeed, results suggest that children could not increase their vocabularies or learn grammar as rapidly as they do if they did not have some prior notion of other individuals' mental states (Pinker 1994; Fisher and Gleitman 2002; Tomassello 2003). In sum, data on children provide an excellent illustration of how a theory of mind can *drive* language development.

By contrast, it is much more difficult to imagine how our ancestors could have learned new words or grammatical rules if they were unable to attribute mental states to others. The lack of syntax in nonhuman primate vocalizations cannot be traced to an inability to recognize argument structure – to understand that an event can be described as a sequence in which an agent performs some action on an object. Baboons, for example, clearly distinguish between a sequence of calls indicating that Sylvia is threatening Hannah, as opposed to Hannah is threatening Sylvia. Nor does the lack of syntax arise because of an inability to mentally represent descriptive verbs, modifiers, or prepositions. In captivity, a variety of animals, including dolphins (Herman et al. 1993), sea lions (Schusterman and Krieger 1986), and African gray parrots (Pepperberg 1993), can be taught to understand and in some cases even to produce verbs, modifiers, and prepositions. Even in their natural behavior, nonhuman primates and other animals certainly seem capable of *thinking* in simple sentences. However, this ability does not motivate them to *speak* in sentences. Their knowledge remains largely private.

This limitation may arise because nonhuman primates and other animals cannot distinguish between what they know and others know and cannot recognize, for example, that an ignorant individual might need to have an event explained to them. As a result, although they may mentally tag events as argument structures, they fail to map these tags into a communicative system in any stable or predictable way. Because they cannot attribute mental states like ignorance to others, and are unaware of the causal relation between behavior and beliefs, monkeys and apes do not actively seek to explain or elaborate upon their thoughts. As a result, they are largely incapable of inventing new words or of recognizing when thoughts should be made explicit.

We suggest, then, that long before our ancestors spoke in sentences, they had a language of thought in which they represented the world – and the meaning of call sequences – in terms of actors, actions, and those who are acted upon. The linguistic revolution occurred when our ancestors began to express this tacit knowledge, and to use their cognitive skills in speaking as well as listening. The prime mover

behind this revolution was a theory of mind that had evolved to the point where its possessors did not just recognize other individuals' goals, intentions, and even their own intentions, beliefs, and knowledge do – but were also motivated to share motivated to make public thoughts and knowledge with others. It led to a mind that was private. The evolution of a theory of mind spurred the evolution of words and grammar. It also provided the selective pressure for the evolution of words and ical adaptations that enabled vocal modifiability. Whatever the selective pressures that prompted this change, the complex suite of skills that we call human speech built upon mental computations that had their origins and foundations in social interactions.

References

- Arnold K, Zuberbühler K (2006) Language evolution: compositional semantics in primate calls. *Nature* 441:303
- Baldwin DA (1991) Infants' contribution to the achievement of joint reference. *Child Dev* 62:875–890
- Bergman TJ, Beehner JC, Cheney DL, Seyfarth RM (2003) Hierarchical classification by rank and kinship in baboons. *Science* 302:1234–1236
- Bloom P, Markson L (1998) Capacities underlying word learning. *Trends Cogn Sci* 2:67–73
- Burichmann D, Carpenter M, Call J, Tomassello M (2007) Enculturated chimpanzees imitate rationally. *Dev Sci* 10:F31–F38
- Cheney DL, Seyfarth RM (1990) How monkeys see the world: inside the mind of another species. University of Chicago Press, Chicago
- Cheney DL, Seyfarth RM (1997) Reconciliatory grunts by dominant female baboons influence victims' behaviour. *Anim Behav* 54:409–418
- Cheney DL, Seyfarth RM (2005) Constraints and preadaptations in the earliest stages of language evolution. *Ling Rev* 22:135–159
- Cheney DL, Seyfarth RM (2007) Baboon metaphysics: the evolution of a social mind. University of Chicago Press, Chicago
- Cheney DL, Seyfarth RM, Silk JB (1995a) The responses of female baboons (*Papio cynocephalus ursinus*) to anomalous social interactions: evidence for causal reasoning? *J Comp Psychol* 109:134–141
- Cheney DL, Seyfarth RM, Silk JB (1995b) The role of grunts in reconciling opponents and facilitating interactions among adult female baboons. *Anim Behav* 50:249–257
- Cheney DL, Seyfarth RM, Palombi R (1996) The function and mechanisms underlying baboon 'contact' barks. *Anim Behav* 52:507–518
- Clark AP, Wrangham RW (1994) Chimpanzee arrival pant-hoots: do they signify food or status? *Int J Primatol* 15:185–205
- Clarke E, Reichard UH, Zuberbühler K (2006) The syntax and meaning of wild gibbon songs. *PLoS One* 1:e73. doi:10.1371/journal.pone.0000073
- Cohen YE, Hauser MD, Russ BE (2006) Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biol Lett* 2:261–265
- Crookford C, Boesch C (2003) Context-specific calls in wild chimpanzees (*Pan troglodytes verus*): analysis of barks. *Anim Behav* 66:115–125
- Crookford C, Hertringer L, Vigilant L, Boesch C (2004) Wild chimpanzees produce group-specific calls: a case for vocal learning? *Ethology* 110:221–243

- Darwin C [1871] (1981) The descent of man and selection in relation to sex. Princeton University Press, Princeton
- Egnor SER, Hauser MD (2004) A paradox in the evolution of primate vocal learning. *Trends Neurosci* 27:649–654
- Egnor SER, Wickelegren JG, Hauser MD (2007) Tracking silence: adjusting vocal production to avoid acoustic interference. *J Comp Physiol A* 193:477–483
- Elowson AM, Snowdon CT (1994) Pygmy marmosets, *Cebuella pygmaea*, modify vocal structure in response to changed social environment. *Anim Behav* 47:1267–1277
- Enard W, Przeworski M, Fisher SE, Lai CSL, Wiebe V, Kitano T, Monaco AP, Pääbo S (2002) Molecular evolution of FOXP2: a gene involved in speech and language. *Nature* 418:869–872
- Engel AL, Hofmeier RR, Cheney DL, Seyfarth RM (2006) Who me? Can baboons infer the target of vocalisations? *Anim Behav* 71:381–387
- Fischer J, Hammerschmidt K, Cheney DL, Seyfarth RM (2001a) Acoustic features of female chacma baboon barks. *Ethology* 107:33–54
- Fischer J, Metz M, Cheney DL, Seyfarth RM (2001b) Baboon responses to graded bark variants. *Anim Behav* 61:925–931
- Fischer J, Hammerschmidt K, Cheney DL, Seyfarth RM (2002) Acoustic features of male baboon loud calls: influences of context age and individuality. *J Acoust Soc Am* 111:1465–1474
- Fischer J, Kitchen DM, Seyfarth RM, Cheney DL (2004) Baboon loud calls advertise male quality: acoustic features and their relation to rank, age, and exhaustion. *Behav Ecol Sociobiol* 56:140–148
- Fisher C, Gleitman LR (2002) Language acquisition. In: Pashler HF, Gallistel CR (eds) *Stevens' handbook of experimental psychology*, vol 3. Learning and motivation. Wiley, New York, pp 445–496
- Fitch WT (2007) The evolution of language: a comparative perspective. In: Gaskell G (ed) *Oxford handbook of psycholinguistics*. Oxford University Press, Oxford, pp 787–804
- Flombaum JI, Santos LR (2005) Rhesus monkeys attribute perceptions to others. *Curr Biol* 15:447–452
- Ghazanfar AA, Logothetis NK (2003) Facial expressions linked to monkey calls. *Nature* 423:937–938
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK (2005) Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci* 25:5004–5012
- Gil-da-Costa R, Braun A, Lopes M, Hauser MD, Carson RE, Herksovitch P, Martin A (2004) Toward an evolutionary perspective on conceptual representation: species-specific calls activate visual and affective processing systems in the macaque. *Proc Natl Acad Sci USA* 101:17516–17521
- Gleitman LR, Papirerou A (2005) Language and thought. In: Holyoak K, Morrison R (eds) *Cambridge handbook of thinking and reasoning*. Cambridge University Press, Cambridge, pp 633–661
- Grice HP (1957) Meaning. *Philos Rev* 66:377–388
- Gros-Louis J (2004) The function of food-associated calls in white-faced capuchin monkeys, *Cebus capucinus*, from the perspective of the signaler. *Anim Behav* 67:431–440
- Hage SR, Jürgens U, Ehret G (2006) Audio-vocal interaction in the pontine brainstem during self-initiated vocalization in the squirrel monkey. *Eur J Neurosci* 23:3297–3308
- Hammerschmidt K, Fischer J (2008) Constraints in primate vocal production. In: Oller DK, Griebel U (eds) *The evolution of communicative creativity: complexity, creativity, and adaptability in human and animal communication*. MIT Press, Cambridge, MA, pp 93–119
- Hare B, Call J, Tomasello M (2001) Do chimpanzees know what conspecifics know? *Anim Behav* 61:139–151
- Hauser MD (1992) Articulatory and social factors influence the acoustic structure of rhesus monkey vocalizations: a learned mode of production? *J Acoust Soc Am* 91:2175–2179
- Herman LM, Pack AA, Morrel-Samuels P (1993) Representational and conceptual skills of dolphins. In: Roitblat HL, Herman LM, Nachtigall PE (eds) *Language and communication: comparative perspectives*. Comparative cognition and neuroscience. Lawrence Erlbaum Associates, Hillsdale, NJ, pp 403–442
- Janik VM, Slater PJB (1997) Vocal learning in mammals. *Adv Study Behav* 26:59–99
- Jürgens U (1995) Neuronal control of vocal production in human and nonhuman primates. In: Zimmernann E, Newman JD, Jürgens U (eds) *Current topics in primate vocal communication*. Plenum Press, New York, pp 199–206
- Kaminski J, Call J, Fischer J (2004) Word learning in a domestic dog: evidence for 'fast mapping'. *Science* 304:1682–1683
- Kitchen DM, Cheney DL, Seyfarth RM (2003) Female baboons' responses to male loud calls. *Ethology* 109:401–412
- Lieberman P (1991) Uniquely human: the evolution of speech, thought, and selfless behavior. Harvard University Press, Cambridge, MA
- Mitani JC, Brandt KL (1994) Social factors influence the acoustic variability in the long-distance calls of male chimpanzees. *Ethology* 96:233–252
- Mitani JC, Gros-Louis J (1998) Chorusing and call convergence in chimpanzees: tests of three hypotheses. *Behaviour* 135:1041–1064
- Owren MJ, Dieter JA, Seyfarth RM, Cheney DL (1993) Vocalizations of rhesus (*Macaca mulatta*) and Japanese (*M. fasciata*) macaques cross-fostered between species show evidence of only limited modification. *Dev Psychobiol* 26:389–406
- Owren MJ, Seyfarth RM, Cheney DL (1997) The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cynocephalus ursinus*): implications for production processes and functions. *J Acoust Soc Am* 101:2951–2963
- Pepperberg IM (1993) A review of the effects of social interaction on vocal learning in African grey parrots (*Psittacus erithacus*). *Neth J Zool* 43:104–124
- Pierce JD Jr (1985) A review of attempts to condition operantly alloprimate vocalizations. *Primates* 26:202–213
- Pinker S (1994) *The language instinct*. William Morrow and Sons, New York
- Poole JH, Tyack PL, Stoeger-Horwath AS, Watwood S (2005) Animal behaviour: elephants are capable of vocal learning. *Nature* 434:455–456
- Potvin A, Saunders RC, Crane AM, Cook M, Sokoloff L, Mishkin M (2003) Functional mapping of the primate auditory system. *Science* 299:568–572
- 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
- Rendall D, Rodman PS, Emmond RE (1996) Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim Behav* 51:1007–1015
- Rendall D, Seyfarth RM, Cheney DL, Owren MJ (1999) The meaning and function of grunt variants in baboons. *Anim Behav* 57:583–592
- Rendall D, Cheney DL, Seyfarth RM (2000) Proximate factors mediating 'contact' calls in adult female baboons (*Papio cynocephalus ursinus*) and their infants. *J Comp Psychol* 114:36–46
- Rescorla RA (1988) Pavlovian conditioning: it's not what you think it is. *Am Psychol* 43:151–160
- Schusterman RJ, Krieger K (1986) Artificial language comprehension and size transposition by a California sea lion (*Zalophus californianus*). *J Comp Psychol* 100:348–355
- Schusterman RJ, Reichmuth Kastak C, Kastak D (2002) The cognitive sea lions: meaning and memory in the lab and in nature. In: Bekoff M, Allen C, Burghardt G (eds) *The cognitive animal: empirical and theoretical perspectives on animal cognition*. MIT Press, Cambridge, MA, pp 217–228
- Seyfarth RM, Cheney DL (1997a) Some general features of vocal development in nonhuman primates. In: Husbberger M, Snowdon CT (eds) *Social influences on vocal development*. Cambridge University Press, Cambridge, pp 249–273
- Seyfarth RM, Cheney DL (1997b) Behavioural mechanisms underlying vocal communication in nonhuman primates. *Anim Learn Behav* 25:249–267
- Seyfarth RM, Cheney DL (2003) Signalers and receivers in animal communication. *Annu Rev Psychol* 54:145–173

- Seyfarth RM, Cheney DL. Primate vocal communication. In: Platt M, Chazanfar AA (eds) *Primate neuroethology*. Oxford University Press, Oxford (in press)
- Snowdon CT (1986) Vocal communication. In: Mitchell G, Erwin J (eds) *Comparative primate biology*, vol 2A. Behavior, conservation, and ecology. Alan R Liss, New York, pp 495–530
- Tomassello M (2003) *Constructing a language: a usage-based theory of language acquisition*. Harvard University Press, Cambridge, MA
- Tomassello M, Carpenter M (2007) Shared intentionality. *Dev Sci* 10:121–125
- Tomassello M, Carpenter M, Call J, Behne T, Moll H (2005) Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci* 28:675–691
- Witig RM, Crockett C, Wikberg E, Seyfarth RM, Cheney DL (2007a) Kin-mediated reconciliation substitutes for direct reconciliation in female baboons. *Proc Roy Soc Lond B* 274:1109–1115
- Witig RM, Crockett C, Seyfarth RM, Cheney DL (2007b) Vocal alliances in chacma baboons. *Papio hamadryas ursinus*. *Behav Ecol Sociobiol* 61:899–909
- Zuberbühler K (2002) A syntactic rule in forest monkey communication. *Anim Behav* 63:293–299
- Zuberbühler K, Cheney DL, Seyfarth RM (1999) Conceptual semantics in a nonhuman primate. *J Comp Psychol* 113:33–42