

To appear in: *The Oxford Handbook of Language Evolution*,
M. Tallerman & K. Gibson (eds). Oxford: Oxford University Press.

Primate social cognition as a precursor to language

By

Robert M. Seyfarth and Dorothy L. Cheney
Departments of Psychology and Biology
University of Pennsylvania
Philadelphia, PA 19104 USA

Abstract. The many homologous traits shared by human and nonhuman primates allow us to identify some of the precursors to language in our prelinguistic ancestors. Comparative data suggest that the prelinguistic ancestor had a small repertoire of relatively fixed calls which nonetheless yielded an open-ended, highly modifiable, and cognitively rich set of meanings, particularly in the domain of social relations. Baboons' knowledge of their companions, for example, is based on discrete-valued traits (identity, rank, kinship) that are combined to create a representation of social relations that is hierarchically structured, open-ended, and rule-governed. The prior evolution of social cognition created individuals who were preadapted to develop language.

Introduction

The goal of phylogenetic reconstruction is to group similar animals together. One method is based on measures of distance, and arranges species into a phylogeny such that each is grouped with those with which it shares the greatest number of characters. Other methods rely on parsimony, generating the phylogeny that requires the fewest evolutionary changes in character states (Ridley 1993). Among primates, both distance and parsimony yield a branching ‘tree structure’ in which humans are grouped most closely with apes, less closely with Old World monkeys, and progressively less closely with New World monkeys, Prosimians, and non-primate mammals. This phylogeny is consistent with both distance and parsimony: for example, morphological and genetic evidence indicate both that there is less evolutionary distance between humans and chimpanzees than between humans and any other primate and that a phylogeny that groups humans and chimpanzees together is more parsimonious (requires fewer evolutionary steps) than a phylogeny that does not (Boyd & Silk 2003).

As Ridley (1993) points out, the parsimony principle is reasonable because evolutionary change is improbable. Suppose we know, for example, that two modern species have the same character state. Parsimony suggests that all the intermediate, ancestral states in the continuous lineages between each species and their common ancestor possessed the same character state, and that the two species are alike in this character because they descend from a common ancestor. Parsimony, in other words, suggests that the two species are homologous in this character. Of course, an infinitely large number of changes could have occurred between ancestor and descendant; however, a change followed by a reversal of that change is unlikely.

“For the characters shared between humans and chimpanzees, the argument is particularly powerful. Chimpanzees and humans share whole complex organ systems like hearts and lungs, eyes, brains, and spinal cords. The initial evolution of each of these characters required improbable mutations, and natural selection operating over millions of generations. It is evolutionarily improbable to the point of near impossibility that the same changes would have evolved independently in the two lineages after their common ancestor. (Ridley 1993:450).

Parsimony is important for those interested in the evolution of language, for two reasons. First, it argues against the hypothesis that language evolved from gesture. The gestural origins hypothesis states that the emergence of speech ‘owes little to nonhuman vocalizations’ and concludes that ‘evolution did not proceed from monkey-like primate vocalizations to speech but rather proceeded from vocalization to manual gesture and back to vocalization again’ (Arbib 2005). Clearly, the gestural origins hypothesis requires many steps. To accept it, we must reject the more parsimonious view that many traits – in perception, cognition, behavior, and neurophysiology – common to nonhuman primate vocalizations and human language (see below) were present in the common ancestor of humans, apes and Old World monkeys and are therefore homologous: identical by virtue of descent from a common ancestor. Instead, the gestural origins view requires that these similarities either arose independently or were lost and regained in the human

lineage. The latter implies that natural selection first favored these traits, then reversed itself, and then reversed itself again. This seems highly unlikely.

Second, the principle of parsimony allows us to identify homologous traits in humans and other primates that were likely to have been present in the ancestral system of vocal communication from which language evolved. This makes it possible to specify those traits that may have served as precursors to language and to identify features that were most likely to have evolved only in the hominid lineage, after divergence from the common ancestor of humans and chimpanzees.

Reconstructing the pre-linguistic ancestor

1. Vocal Production

With a few exceptions, all modern mammals have a relatively small repertoire of call types (a variety of grunts, threatening vocalizations, alarm calls, screams, and so on), each of which has adult-like acoustic features when first produced and shows little modification during development. Mammals can control whether they vocalize or remain silent and make some limited modification in the acoustic features of their calls, but for the most part vocal production is highly constrained (for review see Hammerschmidt & Fischer in press). The rare exceptions are some marine mammals (cetaceans, harbour seals) and humans (Janik & Slater 1997). If we assume, following the principle of parsimony, that flexible vocal production, once evolved, would be highly adaptive and unlikely to revert to a system of fixed, unmodifiable calls, it follows that constrained vocal production was the ancestral mammalian condition and that flexible phonation is a derived characteristic that, among primates, evolved relatively recently in humans. Because vocal production in the Great Apes (chimpanzees, bonobos, gorillas, and orangutans) appears to be typical of that found in other mammals, flexible phonation in humans probably evolved during the last six million years, sometime after the divergence of human ancestors from the common ancestors of humans, chimpanzees, and bonobos (Enard et al. 2002).

2. Vocal usage

Nonhuman primates use acoustically different vocalizations in different social contexts, suggesting that the mechanisms underlying call usage have a strong genetic component, although perhaps not as strong as the mechanisms underlying call production.

For example, vervet monkeys (*Chlorocebus aethiops*) give acoustically different alarm calls to leopards, eagles, and snakes. Each call type elicits a different, adaptive response. Individuals on the ground run into trees when they hear a leopard alarm, look up in the air when they hear an eagle alarm, and peer into the grass around them when they hear a snake alarm (Cheney & Seyfarth 1990). Confronted with a wide variety of potential predators, adult vervets are highly selective, giving 'leopard alarms' to mammalian carnivores, 'eagle alarms' primarily to martial (*Polemaetus bellicosus*) and crowned (*Stephanoetus coronatus*) eagles, and 'snake alarms' to pythons (*Python sebae*). Infants and juveniles, by contrast, make many more mistakes, by giving alarm calls to species like warthogs or pigeons that pose no danger to them. Their mistakes, however, are not entirely random. They give leopard alarms almost exclusively to terrestrial mammals, eagle alarms to birds, and snake alarms to reptiles and snake-like objects. Vervet infants behave as if they are predisposed from birth to divide other species into

broadly different classes: predator versus non-predator, and, within the former class, to distinguish among terrestrial carnivores, eagles, and snakes. With time and experience they sharpen the relation between each alarm call type and the stimulus that elicits it (Cheney & Seyfarth 1990).

Although rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques both produce coo and gruff calls, individuals in the two species use them in different ways. In play, for example, Japanese macaques give coos whereas rhesus macaques give gruffs. In a three-year cross-fostering experiment, two infant rhesus and two infant Japanese macaques were raised in a group of the other species. Despite their complete social integration, the cross-fostered infants showed little or no modification in call usage. They behaved as if the link between call and context was difficult to modify (Owren et al. 1993).

The mixture of innate mechanisms and experience found in these two studies is typical of vocal usage in virtually all mammals and birds (Seyfarth & Cheney 2009). Clearly, however, it differs from vocal usage in humans, who can learn to use any word in any context. If we assume that completely flexible vocal usage, once evolved, would be highly adaptive and unlikely to revert to a system with more fixed, innate links between call and context, it follows that the highly flexible call usage found in humans evolved relatively recently, after the divergence of the human lineage from the common ancestor of humans and Great Apes, and that the ancestral, pre-linguistic condition was one in which vocal usage was partially innate and partially modifiable by experience.

3. Comprehension

In contrast to the data on production and usage, data on perception and comprehension reveal several similarities between human and nonhuman primates. Like human speech, primate vocalizations comprise a series of acoustically intergraded signals that are perceived, roughly speaking, as discretely different calls. In baboons, different call types are distinguished according to the placement of vowel-like formants (see Seyfarth 2005 for references). Nonhuman primate call perception also exhibits parallels with human speech in its underlying neural mechanisms. In macaques, for example, the left hemisphere is specialized for processing species-specific vocalizations but not other auditory stimuli (Poremba et al. 2004). This and other results (Ghazanfar & Hauser 2001) suggest that many of the neural mechanisms that underlie human speech processing are general primate characteristics, shared among humans, apes, and Old World monkeys. Once again following the principle of parsimony, the simplest explanation is that these traits are homologous, and were present in the common ancestor of all Old World primates.

Further parallels are evident in the development of call comprehension, which -- in contrast to production and usage -- is flexible, open-ended, and can be modified by experience. Infant vervet monkeys, for example, respond to playback of leopard, eagle, and snake alarm calls by running to their mothers or showing some other, often inappropriate, reaction. They require several months' experience before they respond to the different alarm calls in an appropriate, adult-like manner (Cheney & Seyfarth 1990). In their natural habitats, many primates learn to recognize the alarm calls of other species even though these calls are acoustically different from their own. In the cross-fostering experiments described earlier, cross-fostered subjects learned to recognize their foster

mothers' calls – and the foster mothers learned to recognize theirs – even in contexts in which the two species used acoustically different vocalizations (Seyfarth & Cheney 1997). And throughout their lives monkeys and apes must continually learn to identify the voices of individuals who join their group or are born into it. For group living animals, it is clearly adaptive to learn these associations and to retain an open-ended ability to identify new sound-meaning pairs throughout one's life. Because the ability to form learned associations is widespread among birds and mammals, we assume that it has a long evolutionary history and was present in the pre-linguistic ancestor of modern humans.

4. Summary: the prelinguistic ancestor had limited vocal production but open-ended comprehension

Modern monkeys, apes, and other mammals share an oddly asymmetric system of communication in which a small repertoire of relatively fixed calls, each closely linked to a particular context, nonetheless gives rise to an open-ended, highly modifiable, and cognitively rich set of meanings (Seyfarth & Cheney 2009). Because these traits are widespread across so many taxonomic groups, it seems highly unlikely that they evolved independently in each case. Instead, we assume that relatively fixed production and open-ended comprehension are homologous traits, and were present in the common ancestor Old World monkeys, apes, and humans. To illustrate the implications for theories of language evolution, consider the communication and cognition of modern baboons.

Social knowledge in baboons

Baboons (*Papio hamadryas*) are Old World monkeys that shared a common ancestor with humans roughly 30 million years ago (Steiper et al. 2004). They live throughout the savannah woodlands of Africa in groups of 50 to 150 individuals. Although most males emigrate to other groups as young adults, females remain in their natal groups throughout their lives, maintaining close social bonds with their matrilineal kin (Silk et al. 2006a, b). Females can be ranked in a stable, linear dominance hierarchy that determines priority of access to resources. Daughters acquire ranks similar to those of their mothers. The stable core of a baboon group is therefore a hierarchy of matrilines, in which all members of one matriline (for example, matriline B) outrank or are outranked by all members of another (for example, matrilines C and A, respectively). Ranks within matrilines are as stable as those between matrilines: for example, A1>A2>A3>B1>B2>C1, where letters are used to denote matrilineal kin groups and numbers denote the different individuals within them (Cheney & Seyfarth 2007).

Baboon vocalizations are individually distinctive (e.g. Owren et al. 1997), and playback experiments have shown that listeners recognize the voices of others as the calls of specific individuals (reviewed in Cheney & Seyfarth 2007). The baboon vocal repertoire contains a number of acoustically graded signals, each of which is given in predictable contexts. Because calls are individually distinctive and each call type is predictably linked to a particular social context, baboon listeners can potentially acquire quite specific information from the calls that they hear.

Throughout the day, baboons hear individuals giving vocalizations to each other. Some interactions involve aggressive competition; for example, when a higher-ranking animal gives a series of threat-grunts to a lower-ranking animal and the latter screams.

Threat-grunts are aggressive vocalizations given by higher-ranking to lower-ranking individuals, whereas screams are submissive signals, given primarily by lower- to higher-ranking individuals. A threat-grunt-scream sequence, therefore, potentially provides information not only about the identities of the opponents involved but also about who is threatening whom. Baboons are sensitive to both types of information. In playback experiments, listeners respond with apparent surprise to sequences of calls that appear to violate the existing dominance hierarchy. Whereas they show little response upon hearing the sequence “B2 threat-grunts and C3 screams”, they respond strongly – by looking toward the source of the call – when they hear “C3 threat-grunts and B2 screams” (Cheney & Seyfarth 2007). Between-family rank reversals (C3 threat-grunts and B2 screams) elicit a stronger violation of expectation response than do within-family rank reversals (C3 threat grunts and C1 screams) (Bergman et al. 2003).

A baboon who ignores the sequence “B2 threat-grunts and C3 screams” but responds strongly when she hears “C3 threat-grunts and B2 screams” reveals, by her responses, that she recognizes the identities of both participants, their relative ranks, and their family membership. Baboons who react more strongly to call sequences that mimic a between-family rank reversal than to those that mimic a within-family rank reversal act as if they classify individuals simultaneously according to both rank and kinship. In all of these cases, listeners act as if they assume that the threat-grunt and scream have occurred together not by chance but because one vocalization caused the other to occur. Without this assumption of causality there would be no violation of expectation when B2’s scream and C3’s threat-grunt occurred together.

Baboons’ ability to deduce a social narrative from a sequence of sounds reveals a rich cognitive system in which listeners extract a large number of complex, nuanced messages from a relatively small, finite number of signals. A baboon who understands that “B2 threat-grunts and C3 screams” is different from “C3 threat-grunts and B2 screams” can make the same judgment for all possible pairs of group members as well as any new individuals who may join. This open-ended system of classification is, in at least one respect, abstract, because the categories of rank and matrilineal kinship persist despite changes in the individuals who comprise them (Cheney & Seyfarth 2007).

In addition to making judgments based on social causation, rank, and kinship, baboons appear to recognize other individuals’ intentions and motives. Baboon groups are noisy, tumultuous societies, and an individual would not be able to feed, rest, or engage in social interactions if she responded to every call as if it were directed at her. In fact, baboons seem to use a variety of behavioral cues, including gaze direction, learned contingencies, and the memory of recent interactions with specific individuals when making inferences about the target of a vocalization. For example, when a female hears a recent opponent’s threat-grunts soon after fighting with her, she responds as if she assumes that the threat-grunt is directed at her, and she avoids the signaler. However, when she hears the same female’s threat-grunts soon after grooming with her, she acts as if the calls are directed at someone else and ignores the calls (Engh et al. 2006).

The attribution of motives is perhaps most evidence in the case of ‘reconciliatory’ vocalizations. Like many other group-living animals, baboons incur both costs and benefits from joining a group. In an apparent attempt to minimize the disruptive effects of within-group competition, many primates ‘reconcile’ with one another, by coming together, touching, hugging, or grooming after aggression. In baboons, reconciliation

among females occurs after roughly 10% of all fights, and typically occurs when the dominant animal grunts to the subordinate. Playback experiments have shown that, even in the absence of other behavior, grunts alone function to restore former opponents' behavior to baseline levels (Cheney & Seyfarth 2007).

In some cases, the behavior of subordinates after aggression seems to involve more complex and indirect causal reasoning about both other animals' motives and their kinship bonds. For example, playback experiments have shown that baboons will accept the 'reconciliatory' grunt by a close relative of a recent opponent as a proxy for direct reconciliation by the opponent herself (Wittig et al. 2007). If individual D1 has been threatened by individual A1 and then hears a grunt from A2, in the hour that follows she is more likely to approach, and more likely to tolerate the approaches of, A1 and A2 than if she had heard no grunt or a grunt from another high-ranking individual unrelated to the A matriline. Intriguingly, D1's behavior toward other members of the A matriline does not change. Subjects in these experiments act as if they recognize that a grunt from a particular female is causally related to a previous fight, but only if the caller is a close relative of her former opponent (see Wittig in press for similar behavior in chimpanzees).

Social cognition as a precursor of language

Baboon vocalizations (and, by extension, those of other monkeys and apes) exhibit no properties that we would be tempted to call syntactic. Nevertheless, their social knowledge, assessment of call meaning, and parsing of call sequences display a number of features that we may think of as syntactic precursors.

First, knowledge is representational. When a monkey hears a vocalization she acquires information that is highly specific – about a particular sort of predator, or about a particular individual, her motivation to interact in specific ways with another, or the other animal's reaction.

Second, knowledge is based on properties that have discrete values (Worden 1988), such as individual identity, kinship, dominance rank, and call type.

Third, animals combine these discrete-valued traits to create a representation of social relations that is hierarchically structured. Baboons, for example, create a nested hierarchy in which others are placed in a linear rank order and simultaneously grouped according to matrilineal kinship in a manner that preserves ranks both within and across families.

Fourth, knowledge is rule-governed and open-ended. Baboons recognize that vocalizations follow certain rules of directionality that must, for instance, correspond to the current dominance hierarchy. Threat-grunts are given only by dominant animals to subordinates, fear barks are given only by subordinates to dominants, but infant and move grunts can be given in either direction. Knowledge is open-ended because new individuals can be added or eliminated without altering the underlying structure, and because the set of all possible interactions is very large (Worden 1998; Seyfarth and Cheney 2007). Taken together, these properties lead to a cognitive system in which animals comprehend a huge number of messages from a finite number of signals.

Fifth, knowledge involves the recognition of motives and causality and is therefore propositional. Baboons evaluate the meaning of call sequences in terms of other individuals' identities and motives and the causal relations that link one individual's behavior with another. That is, they represent in their minds (albeit in a limited way) the

individuated concepts of “Sylvia”, “Hannah”, “threat-grunt”, and “scream”, and combine these concepts to create a mental representation of one individual’s intentions toward another. In so doing, they interpret a stream of sounds as a dramatic narrative: “Sylvia is threatening Hannah and causing her to scream.” Once we accept the notion of causality – and there would be no violation of expectation without it -- call sequences resemble the words in a sentence in at least one respect: they are compositional. Individual calls preserve their meaning but the sequence as a whole conveys a meaning that is greater than the sum of its parts.

Sixth, knowledge is independent of sensory modality. While playback experiments allow us to explore the structure of primates’ social knowledge and demonstrate that such knowledge can be acquired through vocalizations alone, social knowledge is also obtained visually. Indeed, we now know that at the neurophysiological level visual and auditory information are integrated to form a multimodal representation of call meaning (Gil da Costa et al 2004; Ghazanfar et al. 2005).

These properties of nonhuman primates’ social knowledge, while by no means fully human, bear important resemblances to the meanings we express in language, which are built up by combining discrete-valued entities in a structured, hierarchical, rule-governed, and open-ended manner. This leads to the hypothesis that the internal representations of language meaning in the human brain initially built upon our prelinguistic ancestors’ knowledge of social relations (Cheney and Seyfarth 1998; Worden 1998). Indeed, as Worden (1998:156) argues, ‘no other candidate meaning structure has such a good fit to language meanings’.

We are not suggesting that all of the syntactic properties found in language are present in primate social knowledge. Such a claim would be entirely unjustified, given the many features of language – like case, tense, subject-verb agreement, or recursion -- that have no counterpart in the communication of any nonhuman primate and that almost certainly evolved long after the divergence of the hominid line from the common ancestors of humans and chimpanzees. Instead, focusing on the early, prelinguistic stages of language evolution, we suggest that the precursor of the hominoid mind evolved in an environment characterized by social challenges and that such competition created selective pressures favoring structured, hierarchical, rule-governed intelligence. Because this social intelligence shares several features with language, many of the rules and computations found in human language may have first appeared as an elaboration of the rules and computations underlying social cognition.

Talmy (2007) argues that during the course of evolution a crucial bottleneck was overcome when our ancestors’ vocal communication changed from analog to digital, and he poses the question: Where did language get its digitalness from? The answer, we believe, lies in perception and social cognition. Long before they could engage in the computations that underlie modern grammar, our ancestors perceived calls as discrete signals, recognized individuals, and performed the computations needed to understand their societies. As a result, the discrete, compositional structure we find in spoken language did not first appear there. It arose, instead, because understanding social life and predicting others’ behavior requires discrete, compositional thinking.

Similarly, Hurford (1990) asks whether propositional structures (among other features) are “elements of the structure of languages” or whether they “somehow existed before language” in another domain. Here again, data on primate social cognition provide

an answer: The propositions that are expressed in language did not originate with language. They arose, instead, because to succeed in a social group of primates one must understand an elementary form of propositional relations.

Beyond precursors

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 (Enard 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 gave rise to these 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 (defined as the ability of both speakers and listeners to make attributions about each other's beliefs, knowledge, and other mental states when communicating with each other -- Grice 1957). How might these traits 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 came first, 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 & Seyfarth 2007). We make this argument on both empirical and theoretical grounds.

Empirically, 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. Recent work on the alarm calls of forest monkeys suggests that the presence of one call type can 'modify' the meaning of another (Zuberbuhler 2002; Arnold & Zuberbuhler 2006), and a study by Crockford 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.

By contrast, there is growing evidence that both Old World monkeys and apes possess rudimentary abilities to attribute motives or knowledge to others and engage in simple forms of shared attention and social referencing (see Cheney & Seyfarth 2007 for review). 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. More speculatively, we suggest that the prior appearance of a theory of mind acted as a prime mover in the evolution of language because, while it is easy to imagine a scenario in which a rudimentary theory of mind preceded and provided the impetus for the evolution of large vocabularies and syntax, any alternative sequence of events seems less likely.

Consider, for example, word learning in 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 (Tomasello 2003). One-year old children seem to understand 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 (see Fisher & Gleitman 2002 for review). 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. As Fisher and Gleitman (2002) argue, these ‘conceptual primitives’ provide children with a kind ‘conceptual vocabulary onto which the basic linguistic elements (words and structures) are mapped.’ Moreover, in contrast to monkeys, apes, and other animals, one year-old children are highly motivated to share what they know with others (Tomasello & Carpenter 2007). While animals are concerned with their own goals and knowledge, young children are motivated to make their thoughts and knowledge publically available.

The acquisition of a theory of mind thus creates a cognitive environment that drives the acquisition of new words and 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 (Fisher & Gleitman 2002; Tomasello 2003).

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 non-human 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 (as already noted, baboons clearly do this). 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, sea lions, and African gray parrots, can be taught to understand and in some cases even produce verbs, modifiers, and prepositions (see Cheney & Seyfarth 2007 for review). Even in their natural behavior, nonhuman primates and other animals certainly seem capable of thinking in simple sentences, but 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 recognizing when thoughts should be made explicit.

We suggest, then, that while our pre-linguistic ancestors represented the world – and the meaning of call sequences – in terms of actors, actions, and those who are acted upon, a crucial step in the evolution of language occurred when these ancestors began to express their tacit knowledge and use their cognitive skills in speaking as well as listening. The prime mover behind this revolution was a theory of mind that spurred individuals not only to recognize other individuals’ goals, intentions, and even knowledge – as monkeys and apes already do – but also to share their own goals, intentions, and knowledge with others. Whatever selective pressures prompted this

change, it led to a mind that was motivated to make public the thoughts that had previously remained private. The evolution of a theory of mind thus spurred the evolution of words, grammar, and the vocal modifiability that these traits required.

Acknowledgments. We thank Thore Bergman, Mark Liberman, and Leonard Talmy for comments. Research supported by NSF, NIH, The Leakey Foundation, and the University of Pennsylvania.

- Arbib, M. (2005) From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behav. Brain Sci.* 28: 105-167.
- Arnold, K. and Zuberbuhler, K. (2006) Language evolution: Compositional semantics in primate calls. *Nature* 441: 303.
- Bergman, T.J., Beehner, J.C., Cheney, D.L., and Seyfarth, R.M. (2003) Hierarchical classification by rank and kinship in baboons. *Science* 302: 1234-1236.
- Boyd, R. & Silk, J.B. (2003) *How Humans Evolved*. New York: W.W. Norton.
- Cheney, D.L. and Seyfarth, R.M. (1990) *How Monkeys See the World*. Chicago: University of Chicago Press.
- Cheney, D.L. and Seyfarth, R.M. (1998) Why monkeys don't have language. In G. Petersen (ed.) *The Tanner Lectures on Human Values, Vol. 19*. Salt Lake City: Univ. of Utah Press, pp. 175-219.
- Cheney, D.L. and Seyfarth, R.M. (2007) *Baboon Metaphysics*. Chicago: University of Chicago Press.
- Crockford, C. and Boesch, C. (2003) Context-specific calls in wild chimpanzees: *Pan troglodytes verus*: Analysis of barks. *Anim. Behav.* 66: 115-125.
- Enard, W., Przeworski, M., Fisher, S.E., Lai, C., Wiebe, V., Kitano, T., Monaco, A., and Paabo, S. (2002) Molecular evolution of FOXP2: A gene involved in speech and language. *Nature* 418: 869-872.
- Eng, A.E., Hoffmeier, R., Cheney, D.L. and Seyfarth, R.M. (2006) Who, me? Can baboons infer the target of a vocalization? *Anim. Behav.* 71: 381-387.
- Fisher, C. & Gleitman, L.R. (2002) Language acquisition. In H.F. Pashler & C.R. Gallistel (eds.) *Stevens Handbook of Experimental Psychology, Vol. 3: Learning and Motivation*. Wiley: New York, pp. 445-496.
- Fitch, W.T. (2007) The evolution of language: A comparative perspective. In G. Gaskell (ed.) *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press.
- Ghazanfar, A. and Hauser, M. (2001) The auditory behavior of primates: A neuroethological perspective. *Curr. Op. in Neurobiol.* 11: 712-720.
- Ghazanfar, A., Maier, J.X., Hoffman, K.L. and Logothetis, N. (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, M., Carson, R. Herscovitch, P. and Martin, A. (2004) Toward an evolutionary perspective on conceptual representation: Species-specific calls activate visual and affective processing systems in the macaque. *Proc. Nat. Acad. Sci. USA* 101: 17516-17521.
- Grice, H.P. (1957) Meaning. *Phil. Review*, 66: 377-388.

Hammerschmidt, K. and Fischer, J. (In press). Constraints in primate vocal production. In: U. Griebel & K. Oller (Eds.), *The Evolution of Communicative Creativity: From Fixed Signals to Contextual Flexibility*. Cambridge, MA: MIT Press, pp. 00-00.

Hurford, J. (1990) Beyond the roadblock in linguistic evolution studies. *Behav. Brain Sci.* 13: 736-737.

Janik, V.W. and Slater, P.J.B. (1997). Vocal learning in mammals. *Adv. Stud. Behav.* 26: 59-99.

Lieberman, P. (1991) *Uniquely Human*. Cambridge, MA: Harvard University Press.

Owren, M.J., Dieter, J.A., Seyfarth, R.M. and Cheney, D.L. (1993). Vocalizations of rhesus and Japanese macaques cross-fostered between species show evidence of only limited modification. *Dev. Psychobiol.* 26: 389-406.

Poremba, A., Malloy, M., Saunders, R., Carson, R., Herskovitch, P. and Mishkin, M. (2004) Species-specific calls evoke asymmetric activity in the monkey's temporal poles. *Nature* 427: 448-451.

Ridley, M. (1993) *Evolution*. Cambridge, MA: Blackwell Scientific.

Seyfarth, R.M. (2005) Continuities in vocal communication argue against a gestural origin of language. *Behav. Brain Sci.* 28: 144-145.

Seyfarth, R.M. and Cheney, D.L. (1997). Some general features of vocal development in non-human primates. In: M. Husberger & C.T. Snowdon (eds.) *Social Influences on Vocal Development*. Cambridge: Cambridge University Press, pp. 249-273.

Seyfarth, R. M. and Cheney, D. L. (2003). Signalers and receivers in animal communication. *Ann. Rev. Psychol.* 54: 145-173.

Seyfarth, R.M. and Cheney, D.L. (2009) Production, usage, and comprehension in animal vocalizations. *Brain & Language*.

Silk, J.B., Altmann, J., and Alberts, S.C. (2006a) Social relationships among adult female baboons (*Papio cynocephalus*). I. Variation in the strength of social bonds. *Behav. Ecol. Sociobiol.* 61: 183-195.

Silk, J.B., Altmann, J., and Alberts, S.C. (2006b) Social relationships among adult female baboons (*Papio cynocephalus*). II: Variation in the quality and stability of social bonds. *Behav. Ecol. Sociobiol.* 61: 197-204.

Steiper, M.E., Young, N.M. and Sukarna, T.Y. (2004) Genomic data support the hominoid slowdown and an early Oligocene estimate for the hominoid-cercopithecoid divergence. *Proc. Nat. Acad. Sci. USA* 101: 17021-17026.

Talmy, L. (2007) Recombinance in the evolution of language. In J. E. Cihlar, David Kaiser, Irene Kimbara & Amy Franklin (eds.) *Proceedings of the 39th Annual Meeting of the Chicago Linguistic Society: The Panels*, vol. 39-2. Chicago: Chicago Linguistic Society.

Tomasello, M. (2003) *Constructing a Language: a Usage-based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. and Carpenter, M. (2007) Shared intentionality. *Dev. Sci.* 10: 121-125.

Wittig R., Crockford C., Wikberg E., Seyfarth R.M. and Cheney D.L. (2007) Kin-mediated reconciliation substitutes for direct reconciliation in baboons. *Proc. Roy. Soc. Lond. B* 274:1109-1115.

Wittig, R. (in press) Function and cognitive underpinnings of post-conflict affiliation in wild chimpanzees. In E. Lonsdorf, S. Ross and T. Matsuzawa (eds.) *The mind of the*

chimpanzee: Ecological and experimental perspectives. Chicago: University of Chicago Press., pp. 000-000.

Worden, R. (1998) The evolution of language from social intelligence. In J. Hurford, M. Sytuddert-Kennedy and C. Knight (eds.) *Approaches to the Evolution of Language*.

Cambridge: Cambridge University Press, pp. 148-1268.

Zuberbuhler, K. (2002). A syntactic rule in forest monkey communication. *Anim. Behav.* 63: 293–299.

Zuberbuhler, K., Cheney, D.L. and Seyfarth, R.M. (1999). Conceptual semantics in a nonhuman primate. *J. Comp. Psychol.* 113: 33-42.