13

Communication

Brandon C. Wheeler, William A. Searcy,
Morten H. Christiansen, Michael C. Corballis, Julia Fischer,
Christoph Grüter, Daniel Margoliash, Michael J. Owren,
Tabitha Price, Robert Seyfarth, and Markus Wild

Abstract

This chapter reviews what has been learned about animal thinking from the study of animal communication, and considers what we might hope to learn in the future. It begins with a discussion on the importance of informational versus non-informational interpretations of animal communication and then considers what inferences can be drawn about the cognitive requirements of communication from the communicative abilities of simple organisms. Next, it discusses the importance of context to the meaning of animal signals and the possibility of asymmetries in the neural processes underlying production versus reception. Current theories on the evolution of human language are reviewed and how the study of animal communication informs these theories.

Information in Animal Communication

Accounts of animal communication have traditionally relied heavily on the concept of information. Bradbury and Vehrencamp (1998:2), for example, defined communication as the "provision of information from a sender to a receiver," and Otte (1974:385) defined communication signals as "behavioral, physiological, or morphological characteristics fashioned or maintained by natural selection because they convey information to other organisms." Running counter to these accounts of information transmission, however, has been another tradition, one that opposes interpreting animal communication in terms of information (Dawkins and Krebs 1978; Owings and Morton 1997; Owren and Rendall 2001; Rendall et al. 2009; Fischer, this volume). The anti-informational tradition has argued that, rather than informing receivers, signals should be viewed as managing (Owings and Morton 1997) or influencing (Rendall et al. 2009) receivers in the interests of the signalers. Thus receivers

are viewed as responding to signals in a way that benefits the signaler rather than themselves.

Part of the criticism of the use of information transmission in interpreting animal communication has been that the term "information" has been vaguely or inadequately defined (Rendall et al. 2009). Here we use information to mean a reduction of uncertainty on the part of the receiver about the state of the environment, including the signaler as part of that environment. More formally (following Dretske 1981), information is a change in the conditional probability that the environment is in a certain state given the signal (r) and what the receiver already knows (k), relative to the probability of that environmental state given k alone. Information in our usage, then, can be considered to be "perceiver information," in the sense that it depends on how the receiver's own internal assessment of the environment changes due to reception of the signal. Once this definition is specified, it seems fair to use "information transmission" as shorthand for cases in which signals reduce receiver uncertainty regarding the state of the environment.

The anti-informational stance on animal communication begs the question of why receivers would respond in a way that benefits signalers rather than themselves. A variety of answers has been given. Perhaps the best-known hypothesis is sensory bias (Ryan et al. 1990) or sensory drive (Endler and Basolo 1998). This hypothesis proposes that the sensory and neural systems of receivers have biases that may have been favored in contexts other than signaling, such as foraging or predator detection, or which may be nonselected consequences of the ways that sensory and neural systems are put together. Signalers then evolve to exploit these receiver biases; this side of the interaction is termed "sensory exploitation." A second hypothesis is that certain signals, such as screams given by subordinate monkeys in conflict situations, have a directly aversive effect on the receiver solely because of the signals' unpleasant acoustic properties (Owren and Rendall 2001).

A concrete illustration of a non-informational interpretation of a signaling system is provided by the well-known case of the mating call of the túngara frog (*Physalaemus pustulosus*) (Ryan et al. 1990; Ryan and Rand 1993, 2003). The basic mating call of túngara frogs is a whine to which a male can add one to several chucks. Female túngara frogs prefer calls with chucks to calls lacking chucks. Ryan and colleagues proposed that this preference was due to the auditory tuning of female túngara frogs: the whine stimulates mainly the ear's amphibian papilla whereas the higher frequencies of the chucks stimulate mainly the basilar papilla, and the greater overall auditory stimulation by the whine plus chuck(s) produces the female's preference. The sensory exploitation hypothesis predicts that the female preference (for calls with chucks) should have preceded the evolution of the preferred male trait (the chucks themselves), and testing of female preferences in related species combined with phylogenetic analysis seemed to confirm that prediction (Ryan and Rand 1993). Recently, Ron (2008) measured female preferences for chucks in additional species in

the túngara frog complex and, based on these data and a more extensive phylogenetic analysis, has concluded that the preference did not evolve before the chuck, thus undermining the sensory bias interpretation. A number of other cases of sensory bias have also been proposed, such as female preferences for swords in swordtail fish (Basolo 1990, 1996) and female responses to male vibratory signals in water mites (Proctor 1991, 1992).

The non-informational view of animal communication proposes that signalers manipulate receivers to behave in ways that benefit the signaler rather than the receiver. The informational view proposes that signalers evolve signals that change the behavior of receivers in ways that benefit the signalers, but adds the assumption that receiver response behavior evolves to benefit receiver fitness. Receivers will thus only respond to signals if it is, on average, advantageous to do so (see Fischer, this volume). Further, the only mechanism from which receivers can benefit by responding entails signals which vary consistently with some feature of the environment, including the sender's quality and/or subsequent behavior. In this sense, one may state that such signals provide information because they have the potential to reduce the recipient's uncertainty. The informational interpretation has been termed an "equilibrium" view of signaling (Hurd and Enquist 2005), because both signalers and receivers have evolved to a state in which their behaviors benefit them more than would any alternative, so selection does not favor any further change. Such an equilibrium is an essential feature of game theory models of animal communication, including models of mate attraction (Grafen 1990; Kokko 1997), aggressive signaling (Enquist 1985; Számadó 2008), and begging (Godfray 1991; Johnstone and Grafen 1992). A concept that only takes into account the sender's interest might represent a "non-equilibrium" view of signaling, because it is agnostic on whether receivers have evolved to a state in which their present behavior benefits them more than alternatives would. At equilibrium, signals are expected to be informative, but this is not necessarily an expectation for non-equilibrium signals (Hurd and Enquist 2005).

The distinction between informational and non-informational interpretations is relevant to the cognitive demands of communication. Cognition has been defined as "mechanisms by which animals acquire, process, store, and act on information from the environment" (Shettleworth 1998:5). Thus if communication does not involve information, it cannot depend on cognition, at least on the part of the receiver. This conclusion matches with intuition: if female túngara frogs prefer males emitting whines plus chucks over males emitting only whines, and the preference is due solely to differential stimulation of the female's peripheral auditory system, then it would seem wrong to invoke cognition to explain the female's response.

Although cases of non-informational signaling may exist, as in the instances of sensory bias discussed above, the interpretation of many animal signaling systems has been couched in terms of the concept of information. Examples include the waggle dance of the honeybee (see Grüter, this volume), alarm calls

in vervet monkeys (Seyfarth et al. 1980), suricates (Manser 2001) and many other vertebrates, aggressive signaling in some songbirds (Searcy et al. 2006; Ballentine et al. 2008), and food calls in chickens and primates (Marler et al. 1986; Di Bitetti 2003). In these systems in which receivers are seen to process and act on information contained in signals, cognitive mechanisms may be involved in both producers and receivers.

If non-informational as well as informational signals exist, then definitions of communication or of signals that require signals to be informational are not sufficiently broad. One definition of signals that drops the information requirement is provided by Maynard Smith and Harper (2003:3): a signal is "any act or structure which alters the behavior of other organisms, which evolved because of that effect, and which is effective because the receiver's response has also evolved." The last criterion would seem to exclude most cases of non-informational signaling, as these systems involve signalers exploiting preexisting behavioral mechanisms of receivers that have not evolved to be affected by the signal. Therefore, we prefer a definition that omits the last criterion: signals are acts or structures that affect the behavior of other organisms and have evolved because of those effects.

A specific category of communication with possible implications for cognition is deception. In human communication, deception is said to occur when a signaler produces a signal that it knows to be false with the intention of creating a false belief in one or more receivers. Such a definition requires that the signaler be able both to form intentions and to attribute mental states to others. Biological definitions of deception, however, omit the stipulation concerning the intention to create a false belief and instead stipulate that the signaler derives some fitness benefit from conveying false information (Hauser 1996; Searcy and Nowicki 2005). Under this biological definition, we can classify as deception instances in which we are unable to determine whether signalers intend to deceive, including for example those cases in which animals give "false alarms" in contexts where they benefit from eliciting escape behavior from receivers in the absence of a predator (Wheeler 2009; Bro-Jorgensen and Pangle 2010). Moreover, we can even include cases in which all thought on the part of the signaler is precluded, as for example when orchids mimic the appearance and odor of the females of an insect species and thereby attract male insects to aid in pollination (Jersakova et al. 2006).

Communication from Single Cells to Complex Systems

As the orchid example illustrates, communication can be performed by organisms entirely lacking nervous systems, though in the orchid example it is only the signaler, and not the receiver, that fits this description. A case in which both signalers and receivers lack nervous systems is provided by quorum sensing in

bacteria (Miller and Bassler 2001). Here, bacteria secrete signaling molecules termed "autoinducers," which interact with receptors in other bacterial cells to affect expression of particular genes. Bacteria are able to assess population density via the concentration of the autoinducer, so that genes for certain traits are turned on only when some threshold density is reached. The traits controlled by quorum sensing are typically ones that are effective at high population densities but ineffective at lower ones: examples include bioluminescence in *Vibro fischeri* and biofilm production in *Pseudomonas aeruginosa*. In these and other cases, autoinducers appear to have evolved to affect receiving cells, so these chemicals meet our definition of signals (Diggle et al. 2007). The fact that these signaling systems operate in organisms entirely lacking any nervous system demonstrates that simple forms of communication can be accomplished without any cognitive ability at all.

Organisms that possess nervous systems are often capable of more complex communication than seen in orchids and bacteria; nevertheless, it cannot be claimed that any tight association exists overall between the degree of development of an animal's nervous system and the complexity of its communication. To the contrary, arguably the most complex communication systems found in nonhuman animals are seen in social insects (Grüter, this volume), whose nervous systems are relatively simple compared to those of birds and mammals. Within the social insects, the waggle dance of the honeybee (von Frisch 1967; Dyer 2002) provides the most extreme known example of communication complexity.

Honeybees use the waggle dance both when deciding on a new nest site and during foraging. In the foraging context, the dance is performed by a worker that has returned from a foraging trip during which she has successfully found a food source of high quality, and the dance functions to increase and direct the foraging activity of additional workers. The dance has several communicative components (von Frisch 1967; Seeley 1997; Grüter, this volume): the angle of the dance's waggle run relative to vertical conveys the angle of the food source relative to the sun; the duration of the waggle run conveys the distance to the food; and the number of waggle runs performed per dance communicates the quality of the food source to the workers as whole. Recruits that attend to a dance often fly off in the direction and for the distance indicated (von Frisch 1967). The precision of the waggle dance is impressive; as one example, recruits following the dance are able to orient within ± 7.5° (standard deviation) of a food source sited 700 meters from the hive (Towne and Gould 1988).

Wenner and colleagues argue that the behavior of the recruits can be explained by a simpler "olfactory hypothesis," which proposes that, rather than following the dance parameters given above, recruits follow odors picked up by the dancer at the food source (Wenner and Johnson 1967; Wenner et al. 1969; Wenner 2002). The dance hypothesis, however, is firmly supported by an array of experimental results: when dancers are manipulated to dance in an incorrect direction or report an incorrect distance, recruits follow the dance

to the predicted incorrect site (Gould 1975; Esch et al. 2001); and recruits displaced to a new starting point when leaving the hive fly the direction and distance indicated by the dance to a site similarly displaced from the original goal (Riley et al. 2005).

The honeybee waggle dance has the property of "functional reference." Functional reference means that a signal functions to refer to things external to the signaler (Macedonia and Evans 1993), without implying that the signal causes receivers to call up a representation of those things. The signal passes both criteria established by Macedonia and Evans (1993) for functional reference: dances exhibit production specificity, in that there is a close relationship between the signal that is produced and properties of the external object (its direction, distance, and quality), and they satisfy the perception criterion, in that the signal alone is sufficient to allow receivers to choose the appropriate response. Although referential in this sense, the waggle dance does not fulfill a number of criteria for language: it does not involve a true symbol system (with largely arbitrary relationships between signals and referents), let alone complex recursive structure (see Menzel, this volume). As functionally referential signals, the waggle dance can best be compared to vertebrate calls that also have this property, such as the alarm calls of certain mammals (Seyfarth et al. 1980; Zuberbühler et al. 1999; Manser 2001; Fischer and Hammerschmidt 2001) and birds (Gyger et al. 1987; Templeton et al. 2005). Alarm calls of some species have been shown to vary with both predator type and "response urgency." For example, suricates (*Suricata suricatta*) simultaneously vary their alarm calls based on both predator type and the proximity of the predator (Manser 2001), whereas Barbary macaque alarm calls vary with predator type and appearance of the predator (Fischer et al. 1995). Nevertheless, it is clear that for the systems of vertebrate alarm calls thus far studied, all fall short of the honeybee waggle dance in terms of communicative complexity and precision.

Honeybees have some advanced cognitive abilities. They are able, for example, to learn the concepts of sameness and difference, and to transfer the concept from one modality (e.g., olfaction) to another (e.g., vision) (Giurfa et al. 2001). Honeybees also have numerical abilities akin to counting (Dacke and Srinivasan 2008) and sophisticated spatial memory (Menzel, this volume). Their brains, though tiny in absolute terms, are large for their body size, though not necessarily large relative to other social insects (Mares et al. 2005). Although the navigational skills and learning abilities of honeybees are impressive, further claims for honeybee cognition are limited; they are not claimed, for example, to have episodic memory or theory of mind. In other words, a highly developed communicative system does not necessarily entail highly developed cognitive machinery.

Pragmatics and Contingency in Animal Communication

In human language, the interpretation of a given utterance depends not only on the linguistic knowledge of speaker and listener but also on the context in

which an utterance is made. Linguistic pragmatics studies how context can influence the way in which an utterance is understood. "Context" in language can include ongoing events; memory of past events; the status, age, or sex of those involved; and the inferred intent of the signaler.

As in language, responses of recipients to animal signals can be influenced by external context or prior knowledge to varying degrees. For example, worker honeybees which have recently observed a food source in location *B* and then view a waggle dance that indicates a food source in location *A* will often not use the vector of the dance but instead fly to location *B*, particularly if the food source at *B* is of high quality. The bees' "private" information (Grüter et al. 2008) apparently overrides the more "public" information they acquire from observing the dance (see also Grüter and Farina 2009).

Territorial songbirds offer another example. Akcay et al. (2009) demonstrated that the response of male song sparrows (*Melospiza melodia*) to the songs of neighbors was contingent on the past behavior of those neighbors. The song of a neighbor was played from a loudspeaker set in the center of a subject's territory, simulating a territorial intrusion. After a lapse of 45 minutes, a second playback was staged from the subject's boundary, using either the song of the first "bad neighbor" or the song of another unoffending neighbor as a control. Subjects responded more aggressively to the song of the bad neighbor. In a subsequent experiment, Akcay et al. (2010) used song playback to simulate an intrusion by a bad neighbor, not on the subject's own territory, but on the territory of another neighbor—the "victim." In response to subsequent playback from the subject's boundary, subjects were more aggressive toward songs of bad neighbors than toward songs of victims. Response to song is thus contingent not only on an individual's own experience with the singer but on what is inferred from the experience of others.

In baboons (*Papio hamadryas ursinus*), individuals appear to use social context when inferring the intent of the signaler. In one set of experiments, Engh et al. (2006) waited until one adult female, *D*, had directed aggression against a lower-ranking female, *E*. After the two had separated, the experimenters played female *D*'s threat-grunt to female *E*. On another day, the experiment was repeated after *D* and *E* had groomed. After prior aggression, *E* responded strongly to *D*'s threat-grunt: she acted as if the call was directed at her. By contrast, after prior grooming *E* showed little response to the threat-grunt: she acted as if the call was directed at someone else. *E*'s responses, moreover, were specific to particular individuals: prior aggression or grooming with *D* did not affect *E*'s responses to the threat-grunts of other high-ranking animals. Female baboons, therefore, used their memory of prior interactions with particular individuals to infer a speaker's intent and to decide how to respond to a vocalization (for other examples, see Cheney and Seyfarth 2007).

Research on the role of contextual cues carries important implications for the study of communication and cognition in animals. Context, after all, is ubiquitous in nature. It is extremely difficult to imagine an animal signal whose

meaning does not have the potential to be influenced by the context in which it is given. The pervasiveness of contextual cues suggests that natural selection has acted strongly to favor receivers who can integrate relevant signal properties with cues acquired from memory or from the circumstances in which the signal is given.

Moreover, while virtually all animals have a relatively small repertoire of signals, these limited signal types can generate an enormous variety of responses (Smith 1977). The richness of animal communication can, in many cases, be traced to the cognitive operations by which receivers integrate signal and context to create meaning. This integration of signal and context may be the step where cognitive abilities are taxed, and where the performance of higher vertebrates such as birds and primates exceeds that of honeybees and other social insects. This hypothesis deserves further testing, by additional experiments on the ability of higher vertebrates to modify response based on complex contextual variables, and especially by parallel experiments on effects of context on communication in social insects such as honeybees. Testing a broad range of species would be particularly valuable for understanding at which point (or points) such abilities evolved (see Bshary et al., this volume).

Some contextual cues play an important role in communication whereas others, apparently, do not. Signaler identity, for instance, affects communication in many species (Tibbetts and Dale 2007); by contrast, it remains unclear whether any animal species includes knowledge about the signaler's thoughts or beliefs as part of context. A baboon may attribute to others an intent to communicate *to her*, but the attribution of other mental states remains controversial (Penn, this volume).

The importance of context provides a possible tie between animal communication systems and human language.

Animal Communication and Human Language

Asymmetries in Signal Production and Perception

An important observation for animal communication is that signaling is not an inherently symmetrical process—at any level of analysis. While symmetry can and does exist in some aspects, the best starting assumption is that the development, mechanisms, adaptive function, and phylogenetic history of communication can be significantly different for signalers and perceivers. In ontogeny, for example, full-fledged signal production may emerge before, simultaneously with, or after functional responding to the very same communicative events. An illustration from nonhuman primates is that infant vervet monkeys show semantic-like, predator-specific alarm call production well before any understanding of the significance of those calls is evident on the receiver side (Seyfarth and Cheney 1986). Mechanistically, that outcome may be traceable

to an asymmetry in the strength of direct cortical control of vocal production versus response that is evident in nonhuman primates, and which has implications for the evolution of human speech (Seyfarth and Cheney 2010).

The vocal pathway in terrestrial mammals (and many other taxa) involves different subsystems, contributing to different degrees in the initiation and structural properties of vocalizations. The first pathway runs from the anterior cingulate cortex via the midbrain periaqueductal gray (PAG) into the reticular formation of the brainstem, and from there to the phonatory motoneurons. The second pathway runs from the motor cortex via the reticular formation to the phonatory motoneurons. This pathway has been shown to include two feedback loops: one involving the basal ganglia and the other involving the cerebellum (Jürgens 2009). Both pathways are linked to the different motoneurons that innervate the respective muscles for vocal fold, lip, jaw, and tongue movements via the reticular formation. The comparison of vocalization pathways among terrestrial mammal species has revealed that only humans exhibit strong direct connections from the motor cortex to the motoneurons controlling the laryngeal muscles, which can be understood as a third pathway. While this appears to be a derived trait in humans, connections between the limbic cortex and the motoneurons constitute an ancestral trait found in many nonhuman species (for reviews, see Jürgens 2002, 2009; Hammerschmidt and Fischer 2008).

The degree to which these pathways exhibit ancestral or derived characteristics needs to be evaluated cautiously. Because long-distance tract tracing in postmortem human brains is challenging, the putative direct pathway is supported only by electrophysiological studies. Neuroanatomy remains poorly explored in cetaceans, the other comprehensive example of mammalian vocal learning. In birds, direct forebrain projections to syringeal and laryngeal motoneurons—an equivalent "third" pathway—is well-established in songbirds and parrots (e.g., Wild 1993; Striedter 1994) and may also be represented in other groups. There is compelling evidence and broad (but not universal) consensus of homology between avian and mammalian forebrain (Reiner et al. 2004; Wang et al. 2010), so descending pathways might also be preserved. The evidence that basic pattern-generating circuitry in the brainstem has been conserved for perhaps 450 million years since early fishes also indicate that descending forebrain vocomotor projections target ancestral networks (Bass et al. 2008).

The first pathway described above, involving the anterior cingulate cortex and the PAG, seems to be responsible for the initiation of some classes of vocalizations. The PAG apparently controls the production of involuntary sounds, such as a cry of pain given in response to a painful stimulus (Jürgens 2009). The anterior cingulate cortex controls the voluntary production of such sounds. Macaques, in which this area is intact, can learn to increase their vocalization rate for a food reward, whereas individuals with lesions in this area are unable to master this task (Sutton et al. 1974). The second pathway, running from the motor cortex through the reticular formation, is responsible for

the patterning of vocalizations (Jürgens 2009). The third pathway, in which the motor cortex connects directly to phonatory motoneurons, is presumably the one that allows humans the ability to perform vocal learning (i.e., the learning of vocal production through imitation). Vocal learning is defined in distinction to auditory learning, which is the ability to learn the meaning of sounds produced by others. Vocal learning is known to be present in three groups of birds (songbirds, parrots, and hummingbirds) and in four clades of mammals (bats, cetaceans, elephants and humans) (Jarvis 2004; Jarvis et al. 2000; Poole et al. 2005). Evidence for vocal learning is scarce for other mammals and in particular for nonhuman primates (Egnor and Hauser 2004; Snowdon 2008). Auditory learning, by contrast, seems to be universal among the higher vertebrates.

Although some nonhuman primates have vocalizations that are functionally referential (Seyfarth et al. 1980; Manser 2001), the number of such vocalizations is consistently quite limited within any one species. The expansion of the repertoire of referential vocalizations necessary for the evolution of human speech presumably required the acquisition of vocal learning, and the evolution of vocal learning, in turn, presumably required the evolution of more direct cortical control of phonatory motoneurons, as seen in humans. The evolution of neural pathways allowing such cortical control of phonation can thus be considered one of the major steps in the evolution of human speech. Comparative work on the neural pathways controlling vocal production in cetaceans, bats, and elephants might aid in understanding the evolution of such pathways in the human lineage.

Syntax and Recursion

One controversial view of human language holds that the crucial cognitive ability that allows language, and which only humans possess, is the capacity for recursion (Hauser et al. 2002). To determine the importance of recursion for the evolution of language and its relation to nonhuman communication systems, it is imperative to evaluate the empirical data on recursive linguistic behavior. It is important to note that there are different kinds of recursive structures in language. Simple kinds of recursion involve left- and right-branching structure (also known as tail recursion), as for example, when using multiple adjective phrases in "the big, fat, gray cat." This kind of recursion can be accommodated within a finite-state grammar, in which only the transition from the current state to the next state is represented. More complex recursion can be found in the form of center-embedding and cross-serial dependencies; this is the kind of recursion that is typically at the center of discussions about recursion in language. In English, complex recursion is employed when phrases are embedded within phrases such as "the cat that the dog chased ran away." A further level of center-embedding would be "the cat that the dog that Alex owned chased ran away." The existence of complex recursion has been said to require sophisticated grammar machinery beyond so-called context-free grammars

(Chomsky 1957). Importantly, though, computational machinery of this sort is only needed if infinite depths of recursion have to be processed.

Gentner et al. (2006) challenged the assertion that recursion is unique to human language by testing the ability of starlings to distinguish sequences of rattles (R) and warbles (W) that either had a complex recursive structure, $RnWn$, or a tail recursive structure, $(RW)n$. The two categories of song elements, R and W, each contained eight different exemplars, so that the actual combinations of rattles and warbles were randomly chosen. An example of a complex recursive series with three levels of recursion would be $R7R1R5R3W6W7W2W5$. The corresponding tail recursive version of this series would be $R7W6R1W7R5W2R3W5$. Strictly speaking, the complex recursive series represent counting recursion rather than center-embedding, as the latter requires that embedded elements exhibit dependencies between the two categories of elements (such as noun-verb agreement in number) that are not actually present here. After extensive training, the starlings were eventually able to discriminate both the complex and tail recursive sequences from those not obeying these conditions. Since the actual examples of rattles and warbles were varied randomly from trial to trial, the birds could not have been learning specific sequences, but must have somehow grasped something about the different underlying structures.

It has been suggested that the starlings in the Gentner et al. (2006) study could have discriminated between the test sequences they were presented with using simpler heuristics (Corballis 2007; Hilliard and White 2009; ten Cate et al. 2010; but see Gentner et al. 2010). Many of the alternative mechanisms for discrimination, such as learning that complex recursive patterns always start with two rattles and tail recursive patterns do not, are eliminated by the starlings' responses to additional agrammatical probe stimuli presented by Gentner et al. (2006). Corballis (2007) has suggested a further alternative: that the starlings might have determined, by counting or subitizing, the number of successive Rs and then the number of successive Ws, and accepted the sequence as $RnWn$ if the numbers matched. Whether this alternative is more parsimonious than mastering complex recursion is arguable.

Although definitive evidence is not available that any nonhuman animal can master true center-embedded recursion, it should be realized that human performance on complex recursive constructions is in turn rather limited. For example, in corpus analyses of seven European languages, Karlsson (2007) found that doubly center-embedded sentences practically never occur. Psycholinguistic data show that people are unable to understand such sentences (Blaubergs and Braine 1974; Hakes et al. 1976; Hamilton and Deese 1971; Wang 1970) and receive little benefit from explicit training on them (Blaubergs and Braine 1974; Stolz 1967). Moreover, children appear to acquire their (limited) ability for recursive sentence processing gradually in a piecemeal fashion (Dickinson 1987)—construction by construction—indicating that recursion is not a fundamental part of the grammar that is initially limited

by memory or other developmental constraints (Christiansen and MacDonald 2009). Importantly, there is also considerable variation across languages with regard to the amount of recursion used (Evans and Levinson 2009). Finally, the same type of recessive recursive construction can vary dramatically across languages in how easy it is to process (Hawkins 1994; Hoover 1992).

Further work is clearly needed on the ability of both humans and nonhumans to utilize recursion and, in particular, to understand complex forms of recursive constructions such as center-embedding. In undertaking such work, it is important to approach human abilities on recursion tasks with the same skeptical stance that has been applied to work on recursion in nonhuman animals. It is also important to keep in mind that the ability to process recursive material, like any other biological trait, almost certainly evolved gradually (Margoliash and Nusbaum 2009), so that precursors at some level are very likely to be present in nonhuman animals.

Developmental Genetics and Language Evolution

Recent progress on the genetic foundations of speech may have implications for the evolution of language. One gene that appears to be involved in vocal production is *FOXP2*, which was identified in a British family with specific language impairments (Hurst et al. 1990). Molecular analysis has revealed that there is only a single amino-acid difference in the FoxP2 protein of chimpanzees and mice, but two additional amino-acid differences between chimpanzees and humans (Enard et al. 2002). These findings suggest that the substitutions in the human lineage underwent positive selection, perhaps due to effects on some aspects of speech and language (Fisher and Scharff 2009). The amino-acid changes that created the human version of FOXP2 are likely to have taken place before about 500,000 years ago, because analysis of the Neanderthal genome indicates that they had the human version of FOXP2 (Krause et al. 2007).

Importantly, *FOXP2* is not a language gene but is rather a gene for a transcription factor that affects the function of many genes, including ones involved in the development of the lungs, heart, and other organs (Fisher and Marcus 2006). Its precise effects in the phenotype affecting language development have been a matter of some debate. Affected individuals have problems with sequential speech production that can lead to major problems with intelligibility. They also have more general difficulties with language, made evident in their written language and in language comprehension (Bishop 2009). Whereas prior human studies of *FOXP2* have involved rare mutations, Mueller et al. (in preparation) found that a common polymorphism in the promotor region of this gene is associated with variation in language ability as well as the ability to learn visually presented sequential structure.

Studies of the effects of FoxP2 protein in animal models revealed that complete absence of the protein in mice leads to premature death, while conditional knock-out in birds impairs the accuracy with which birds learn to sing (Haesler

et al. 2007; Fischer and Hammerschmidt 2011). Mice carrying the "human variant" of the *Foxp2* gene show a higher density of medium spiny neurons in the striatum and slightly altered ultrasonic vocalizations (Enard et al. 2009). Intriguingly, the striatum plays a crucial role in the kind of sequential learning that Mueller et al. (in preparation) found to be associated with a common polymorphism in *FOXP2*. In sum, while there is some evidence that links *FOXP2* to vocal behavior, the precise mechanisms remain unclear. Continued research on the *FOXP2* gene will be critical to determine in what way it was involved in the emergence of human language.

Ancestral Stages in the Evolution of Human Speech

A major question in the evolution of language is the nature of the ancestral stages that preceded the evolution of spoken language. Here a natural theory is that spoken language evolved from the systems of vocal communication found in nonhuman primates (Seyfarth 1987; Cheney and Seyfarth 2005). An alternative theory holds that human language emerged initially from manual gestures, with human language only secondarily becoming spoken (Hewes 1973; Corballis 2002; Tomasello 2008; Arbib et al. 2008).

The ability to use symbols has been suggested as an important precursor to language (Christiansen and Kirby 2003; Jackendoff 1999); consequently, the extent to which other animals use vocal and gestural signals to refer to objects or events in the environment is relevant to the two theories. The vocal theory is supported by the finding that many species of primates produce "functionally referential" alarm calls, whereby the production of a specific call type is dependent on the appearance of a particular predator type, and reception of the call allows receivers to choose a response appropriate to that predator type. Functional reference in this sense has been supported for the alarms of vervet monkeys, guenons, lemurs, tamarins, and capuchin monkeys (Seyfarth et al. 1980; Zuberbühler et al. 1999; Fichtel and Kappeler 2002; Kirchhof and Hammerschmidt 2006; Wheeler 2010). Although the alarm calls of great apes have not been shown to have a similar level of functional reference, food calls of captive chimpanzees have been found to vary for different food types, and playback of the calls guided the search behavior of one test subject (Slocombe and Zuberbühler 2005). Other primate vocalizations provide listeners with detailed cues to events in their social environment. Among baboons, for example, certain calls are given only in highly predictable social circumstances: threat-grunts are solely given by higher-ranking to lower-ranking animals, and screams are given only by lower-ranking to higher-ranking animals. Playback experiments indicate that listeners monitor the vocalizations exchanged by others in their group, and in this way learn about changes in their social relationships. When listeners hear a higher-ranking animal give threat-grunts and a lower-ranking animal scream, they show little response, but when they hear a lower-ranking animal's threat-grunt followed by a higher-ranking animal's

scream they respond strongly (Cheney et al. 1995). Listeners' response to an apparent rank reversal is particularly strong if the interaction suggests that the member of a lower-ranking matriline has risen in rank above the member of a higher-ranking matriline (Bergman et al. 2003). These results suggest that highly specific, functionally referential vocalizations in nonhuman primates are not limited to alarm calls but instead can be found throughout the animals' repertoire of vocal signals.

By contrast, evidence for the referential use of manual gestures by nonhuman primates is scarce (Arbib et al. 2008) and mostly limited to imperative "pointing" gestures by captive apes during interactions with humans (Leavens et al. 1966; Call and Tomasello 1994; Miles 1990). A referential gesture, "the directed scratch," has been suggested for chimpanzees in the wild; here, an exaggerated scratching movement on a part of the body is used to elicit grooming of this area (Pika and Mitani 2006). The use of iconic gestures (gestures that bear a physical resemblance to an external referent) has been reported only in one bonobo and one gorilla (Tanner and Byrne 1996); therefore, the use of iconic gestures by great apes remains controversial (Tomasello and Call 2007a). Whether pointing gestures function referentially or whether receivers respond as a result of stimulus enhancement remains unclear. Gestures, especially in the wild, are often more difficult to record, categorize, and play back than are vocalizations, so the lack of evidence for functional reference in primate gestures may be due to a lack of appropriate research. More work on the use of primate gestures in natural communication with conspecifics is needed, especially work that employs experimental methods. The majority of what is known about gestural communication comes from studies of captive apes, compared to vocal research that is more often carried out on free-ranging monkeys. Studying vocal and gestural communication simultaneously within a species would also address this imbalance in knowledge and make direct comparisons of these two signaling modalities easier.

Although nonhuman primates typically possess rather small repertoires of vocal signals, the communicative power of these restricted repertoires can be substantially augmented by the receiver's ability to integrate vocalizations with contextual cues. We have already provided one example of how context can influence one primate's response to another's vocalization: baboons' response to the threat-grunt of another individual depends upon their recent interactions with that individual (Engh et al. 2006). In much the same way, the response of Diana monkeys to playback of a Diana monkey's leopard alarm call can be influenced by contextual cues. Normally, whenever a Diana monkey hears another Diana's leopard alarm call, she gives a leopard alarm call of her own. However, if the listener has recently heard (and responded to) the growl of a leopard coming from the same area, she no longer responds to the sound of an alarm call with her own alarm, presumably because she already knows that a leopard is in the area (Zuberbühler et al. 1999).

Research in functional linguistics (e.g., Clark 1996; Levinson 2000) suggests that pragmatic context also plays a crucial role in language processing and makes it possible to interpret the linguistic signal given the context. The importance of context in both human language and primate vocal communication provides an element of evolutionary continuity between the two, continuity that is less apparent in semantics and syntax. This interpretation does not preclude, of course, that humans may have evolved more sophisticated sociopragmatic skills than nonhuman primates, but the difference and impact on communication would be a matter of degree and not of kind.

One problem already discussed for the theory of language evolution from primate vocalizations is the weakness of direct cortical control of phonatory neurons in nonhuman primates—a weakness that largely precludes vocal learning and limits the ability to control when to produce or not produce vocalizations. Gestural theory does not suffer from this problem, as neocortical control of manual movements is well developed in primates (Gentilucci and Corballis 2006). This difference in cortical control of vocalizations versus gestures may explain why attempts to teach great apes to speak have failed, whereas at least moderate success has been attained from the use of simplified forms of sign language, or keyboards containing abstract symbols (Gardner and Gardner 1969; Savage-Rumbaugh et al. 1998).

Proponents of the gestural theory emphasize flexibility in the form and use of gestures, as an important step in the evolution of symbolic communication and language (Tomasello and Call 2007b). Studies of all great ape gestural repertoires have identified idiosyncratic gestures used by single individuals (Goodall 1986; Pika et al. 2005), suggesting that some gestures are invented through ontogenetic ritualization (Tomasello and Call 2007b). However, variability alone does not increase signal meaning. For gestures to be used communicatively, their variation must relate in some consistent way either to internal states or to external objects and events. For vocal signals, acoustic analysis and playback experiments present a reliable way to assess whether structural variation is meaningful to conspecifics (e.g., see Fischer 1998). Similarly, quantitative measures of variability and experimental validation of whether signal variation affects receiver response are needed for gestural signals, before such signals are accepted as communicative.

The observation that signed languages of the deaf have essentially all of the linguistic, semantic, and pragmatic properties of spoken language has also been used as an argument for the plausibility of the gestural theory (Armstrong et al. 1995; Stokoe 2001; Armstrong and Wilcox 2007). Others would argue, however, that these properties of signed languages have been found in fully evolved humans, rather than human ancestors, and thus may be a reflection of communication abilities evolved in another context, such as the vocal one. Finally, the dominance of the left hemisphere of the brain is apparent both in right-handedness and in the control of speech, suggesting a close tie between manual and vocal activity (Corballis 1989). There is, however, ample evidence

that nonhuman primates also reveal a left hemisphere dominance in the processing of vocalizations, indicating that hemispheric lateralization per se is not a good diagnostic (reviewed in Fitch 2010).

Gestural theory has recently received support from the discovery of mirror neurons in primates in areas of the cortex responsible for control of manual movements. These neurons fire both when a monkey makes intentional movements with its hands and when the monkey sees another individual making the same movements. These neurons also respond to the *sounds* of manual gestures, but they do not respond to vocalizations (Kohler et al. 2002). Mirror neurons were first discovered in monkeys in areas of the brain considered to be homologous with Broca's area. Mirror neurons are now understood to be part of a larger network, called the *mirror system*, which includes areas in addition to Broca's that are homologous to ones important to language. Altogether, the parieto-frontal mirror system in primates corresponds very closely to the language circuits in the left hemisphere of the human brain (Rizzolatti and Sinigaglia 2010). Evidence for mirror neurons in humans, however, is a subject of debate (e.g., Turella et al. 2009).

Rizzolatti and colleagues have proposed that the mirror system in monkeys is in essence a system for understanding action (Rizzolatti et al. 2001; Rizzolatti and Craighero 2004). That is, the monkey understands the actions of others in terms of how it would itself perform those actions. This is the basic idea underlying what has been called the *motor theory of speech perception*, which holds that we perceive speech, not in terms of the acoustic patterns it creates, but rather in terms of how we ourselves would articulate it (Liberman et al. 1967). The mirror system provides a natural substrate, though grounded in gesture rather than vocalization. Still, there is ample evidence that the perception of sounds in general and the acquisition of sound-referent relationships are widespread and independent of the ability to produce these sounds. An extreme example, perhaps, is, Rico: a border collie that learned the names of over two hundred toys and was able to retrieve them correctly on command (Kaminski et al. 2004). In contrast, his vocal repertoire was limited to some barks and growls, suggesting that a close perception-action link is not a prerequisite for the processing of acoustic stimuli (reviewed in Fischer 2010).

Although the discovery of mirror neurons responsive to manual movements has widely been taken to support the gestural theory of the origin of language, it should be noted that mirror neurons have recently been shown to also exist in the song system of songbirds (Prather et al. 2008). Mirror neurons may eventually prove to be widely distributed in the brains of higher vertebrates, and thus not to be strong evidence in favor of any particular theory of language origin (see also Hurford 2004 for discussion). Moreover, it is also important to remember the many continuities that exist between human and nonhuman primates in the perception of conspecific vocalizations (Gil-da-Costa et al. 2004), lateralization in the perception of such calls (Poremba et al. 2003; Poremba et al. 2004), the integration of faces and voices (Ghazanfar et al. 2005), and

the recognition of individual speakers (Scott 2008). These continuities are just what we would expect to find if human language had emerged from an ancestral vocal system of communication.

The greatest problem for the primate vocal theory for the evolution of language, as stated above, is the absence of direct cortical control of vocal production in nonhuman primates and the presumed difficulty of evolving such control. The greatest problem for the gestural theory is that, even if our ancestors started with gestural language, the switch to spoken language still must have been made at some point (Burling 2007). That switch, whenever it occurred, would have required the evolution of cortical control of vocal production. A communication system dependent on manual gestures does not require cortical control of vocal production, and thus does nothing to pave the way for such an adaptation. Put another way, the vocal theory assumes the sequence primate vocal communication to human spoken language, whereas the gestural theory assumes the sequence primate gestural communication to gestural language to spoken language. Corballis (2010) has suggested that the latter transition might have occurred by face movements gradually becoming more important than hand movements in gestural communication, with the eventual addition of voicing and movements of the vocal tract to facial gestures. Nevertheless, if the intermediate step of gestural language does not simplify the second transition, then the vocal theory, with one transition, can be claimed to be more parsimonious than the gestural theory, with two.

An intermediate view might be that the vocal-facial and manual pathways coevolved. There is good evidence for a close link between specific facial expressions and specific vocalizations (Haesler et al. 2007). Further studies should examine the link between hand movements and vocalizations, including its neural basis, in more detail.

Coevolution of Language and Cognition

Corballis (this volume) argues for the importance of episodic memory and mental time travel in providing the selective impetus for the evolution of language. The ability to review past events and to plan for the future is enhanced by improvements in language skills; thus the fitness benefits of episodic memory and future planning might have provided a primary selective advantage for the evolution of language. Others have made a similar argument with respect to theory of mind. A full blown theory of mind requires language, for example, to derive explanations for the behavior of others that depend on inferences about their mental states (Malle 2002). If so, the selective benefits of having a theory of mind would also provide a selective advantage for the evolution of language.

The argument can be reversed; that is, it can be argued that cognitive skills, such as theory of mind and episodic memory, enhance language skills. Thus the ability to engage in joint attention, an aspect of theory of mind, seems to

be important to word acquisition in humans (Baldwin 1993). It is also sometimes claimed that the ability to infer the intention to communicate in others, another theory of mind skill, is important to language learning (Malle 2002). According to Grice (1989), ordinary conversation is dependent on knowing what is in the minds of others, to the point of requiring a specialized theory of mind module (Sperber and Wilson 2002), although Millikan (1984) and others have criticized this idea as making conversation more complicated than it actually is. Improvements in memory certainly must have been important as human ancestors expanded their vocabulary of signals beyond those found in other primates, and episodic memory may have played a role here.

The conclusion, then, is that language enhances cognitive skills and cognitive skills enhance language. The primary benefit of both sets of skills may have been in dealing with social complexity (Cheney and Seyfarth 1990b; Dunbar 1998b; see also chapters in the section on Knowledge, this volume). Cognition and language then would have coevolved, though not in the sense that species coevolve (e.g., as in host-parasite interactions) by putting reciprocal selection pressures on each other. Rather, cognition and language would coevolve in the sense that both attributes would evolve gradually over the same time period, with each enabling improvements in the other. The degree to which such coevolution would have resulted in specific biological adaptations for language is, however, unclear given the possibility that language itself might have evolved by way of cultural evolution (e.g., Chater et al. 2009; Christiansen and Chater 2008).

Conclusions

Although the bulk of animal communication involves the integration of signal and context, and thus potentially relies on cognition, the actual cognitive demands of many signaling systems seem rather modest. This conclusion is brought home by the observation that some forms of communication are managed by organisms, such as bacteria, which lack nervous systems altogether, as well as by the fact that the most complex forms of communication known in nonhuman organisms are accomplished by social insects with relatively modest nervous systems. Nevertheless, we see a role of advanced cognitive abilities in certain aspects of animal communication, especially in the integration of signals with context. Such integration occurs more obviously at the receiver end, when the response of receivers to signals is contingent on a combination of present circumstances and memories of past circumstances. Integration can also occur at the signaler end, when senders make decisions about whether or not to signal or on the form of the signal, which is again contingent on context. Moreover, the complex context-dependence of signal production and interpretation found in particular in the communication of nonhuman primates provides an important element of continuity between animal communication

and human language. Eventual understanding of the evolution of language will depend on an analysis of how the context-dependent, partially referential signaling systems of our ancestors were gradually elaborated through the co-evolution of language and the other cognitive abilities that both support and require language.