The Evolution of Language

Article *in* Annals of the New York Academy of Sciences · April 2009

Impact Factor: 4.38 · DOI: 10.1111/j.1749-6632.2009.04423.x · Source: PubMed

Michael C Corballis
University of Auckland

The Evolution of Language

Michael C. Corballis

Department of Psychology, University of Auckland, Auckland, New Zealand

Language, whether spoken or signed, can be viewed as a gestural system, evolving from the so-called mirror system in the primate brain. In nonhuman primates the gestural system is well developed for the productions and perception of manual action, especially transitive acts involving the grasping of objects. The emergence of bipedalism in the hominins freed the hands for the adaptation of the mirror system for intransitive acts for communication, initially through the miming of events. With the emergence of the genus *Homo* from some 2 million years ago, pressures for more complex communication and increased vocabulary size led to the conventionalization of gestures, the loss of iconic representation, and a gradual shift to vocal gestures replacing manual ones—although signed languages are still composed of manual and facial gestures. In parallel with the conventionalization of symbols, languages gained grammatical complexity, perhaps driven by the evolution of episodic memory and mental time travel, which involve combinations of familiar elements—Who did what to whom, when, where, and why? Language is thus adapted to allow us to share episodic structures, whether past, planned, or fictional, and so increase survival fitness.

Key words: **language; evolution; gesture; *FOXP2*; mirror neurons**

Introduction

Is language evolution “the hardest problem in science,” as recently suggested (Christiansen & Kirby 2003, p. 1)? Certainly, it has had an extraordinarily difficult and contentious history. In early times, at least, part of the difficulty has been religious, but that in turn might relate to the seemingly miraculous flexibility and open-endedness of language. In Christian thought it was proclaimed that language must be a gift from God: “In the beginning,” says St. John in the Bible, “was the Word, and the Word was with God, and the Word was God.” This view constrained philosophical thinking well into the second millennium AD. In the 17th century, Réné Descartes (1985/1647) considered that language allowed such freedom of expression that it could not be reduced to mechanical principles, confirming for him, at least, that it must indeed be God-given. Nonhuman animals, on the other hand, were mere automata, and therefore incapable of human-like language. Descartes has been called the founder of modern philosophy, but his views on language clearly left little scope for an evolutionary account.

The publication of Darwin’s *Origin of Species* in 1859 should have encouraged a more evolutionary approach but was quickly followed in 1866 by the famous ban on all discussion of the origin of language by the Linguistic Society of Paris. The ban was reiterated in 1872 by the Philological Society of London. These suppressive moves may well have been influenced by the religious opposition to Darwin’s theory of natural selection but may have also been motivated by the apparent uniqueness of human language, the absence of material evidence as to how it evolved, and the speculative nature of any discussion.

In more recent times, discussion of language evolution was further stifled by the views of

---

Address for correspondence: Michael C. Corballis, Department of Psychology, University of Auckland, Auckland, New Zealand. m.corballis@auckland.ac.nz

The Year in Cognitive Neuroscience 2009: Ann. N.Y. Acad. Sci. 1156: 19–43 (2009).
doi: 10.1111/j.1749-6632.2009.04423.x

Noam Chomsky, the most influential linguist of the late 20th century. In 1975, he wrote:

> We know very little about what happens when 10^{10} neurons are crammed into something the size of a basketball, with further conditions imposed by the specific manner in which this system developed over time. It would be a serious error to suppose that all properties, or the interesting structures that evolved, can be 'explained' in terms of natural selection. (p. 59)

As an avowed Cartesian, Chomsky's views were based on the firm opinion that language is based on "an entirely different principle" (1966, p. 77) from animal communication. Unlike Descartes, though, he did not appeal to divine intervention.

The tide turned, though, with the publication of an influential paper by Pinker and Bloom in 1990. They took special issue with Chomsky's view that language could not have evolved through natural selection. Their argument rested primarily on the simple fact that language is complex, designed for the communication of propositional structures, and that the only rational explanation for complex structures lies in natural selection. Even Chomsky appears to have been somewhat influenced, since he was coauthor of an article accepting some degree of continuity between human and animal communication, but still insisting on a critical component unique to humans (Hauser et al. 2002). Pinker and Bloom's article spurred a flurry of books, articles and conferences on the evolution of language, offering such a diversity of opinion that one might begin to wonder whether the ban should be restored. In presenting my own views, then, it may be helpful to establish some preliminary assumptions.

First, pace Chomsky, I assume that language did indeed evolve through natural selection. This makes it at least reasonable to seek precursors to language in our primate forebears. Second, I assume that language did not appear abruptly in hominin evolution, as suggested by authors such as Bickerton (1995), who wrote that "true language, via the emergence of syntax, was a catastrophic event, occurring within the first few generations of *Homo sapiens sapiens*" (p. 69). The idea that grammatical language evolved gradually is more consistent with the notion of *grammaticalization* (e.g., Heine & Kuteva 2007; Hopper & Traugott 1993), whereby grammar emerges in an incremental manner, than with Chomsky's (1975) view that all humans are equipped with an innate *universal grammar*. In this respect, this article is in accord with growing skepticism about universal grammar as a useful or viable concept (e.g., Christiansen & Chater 2008; Everett 2005, 2007; Tomasello 2003, 2008; Wray 2002). Finally, I assume that language is fundamentally a gestural system, and evolved from manual gestures rather than animal calls, and indeed it persists in this form in signed languages. This assumption is in accord with the notion that language is an embodied system (e.g., Barsalou 2008), rather than a system based on amodal abstract symbols.

These assumptions will not find universal favor. They are chosen to support what I see as the most plausible account of how language might have evolved. But if we heed Dobzhansky's (1973, p. 125) famous statement that "[n]othing in biology makes sense except in the light of evolution," then a coherent evolutionary account should itself support the assumptions underlying it.

It is convenient, then, to begin with the controversial theory that language evolved from manual gestures.

The Gestural Theory

History

Despite religious objection to evolutionary accounts, the idea that language may have originated in manual gestures did surface in early writings. In 1644, Bulwer wrote on "the natural language of the hand," and Cordemoy (1668/1972) called gestures "the first of all languages," noting that they were universal and understood everywhere. These observations were based partly on the experiences of European traders, who discovered that they could communicate in foreign lands by using only bodily gestures.

Some authors found ways to avoid religious opprobrium through gestural theory. The Italian philosopher Giambattista Vico (1953/1744) accepted the Biblical story of the divine origin of speech but proposed that after the Flood humans reverted to savagery, and language emerged afresh. He suggested that one of the post-Deluge languages made use of gestures and physical objects, but later became vocal through onomatopoeia and interjections. Two years later the Abbé de Condillac (1971/1746) also proposed that language evolved from manual gestures but, as an ordained priest, he, too, was constrained by religious doctrine and presented his theory in the form of a fable about two children wandering about in the desert after the Flood. They began by communicating in manual and bodily gestures, which were eventually supplanted by vocalizations. Condillac's compatriot Jean-Jacques Rousseau endorsed the gestural theory more openly in an essay published in 1782.

In the following century, Charles Darwin made reference to the role of gestures in his book *The Descent of Man*: "I cannot doubt that language owes its origins to the imitation and modification of various natural sounds, and man's own distinctive cries, *aided by signs and gestures*" (1871/1896, p. 86, emphasis added). Shortly afterwards, the philosopher Friedrich Nietzsche chimed in with the following extract from his 1878 book *Human, All too Human*:

> Imitation of gesture is older than language, and goes on involuntarily even now, when the language of gesture is universally suppressed, and the educated are taught to control their muscles. The imitation of gesture is so strong that we cannot watch a face in movement without the innervation of our own face (one can observe that feigned yawning will evoke natural yawning in the man who observes it). The imitated gesture led the imitator back to the sensation expressed by the gesture in the body or face of the one being imitated. This is how we learned to understand one another; this is how the child still learns to understand its mother. In general, painful sensations were probably also expressed by a gesture that in its turn caused pain (for example, tearing the hair, beating the breast, violent distortion and tensing of the facial muscles). Conversely, gestures of pleasure were themselves pleasurable and were therefore easily suited to the communication of understanding (laughing as a sign of being tickled, which is pleasurable, then served to express other pleasurable sensations).
>
> As soon as men understood each other in gesture, a symbolism of gesture could evolve. I mean, one could agree on a language of tonal signs, in such a way that at first both tone and gesture (which were joined by tone symbolically) were produced, and later only the tone. (1986/1878, p. 129)

This extract also anticipates the later discovery of the mirror system. In 1900, Wilhelm Wundt, the founder of the first laboratory of experimental psychology at Leipzig in 1879, wrote a two-volume work on speech and argued that a universal sign language was the origin of all languages. He wrote, though, under the misapprehension that all deaf communities use the same system of signing, and that signed languages are useful only for basic communication, and are incapable of communicating abstract ideas. We now know that signed languages vary widely from community to community and can have all of the communicative sophistication of speech (e.g., Emmorey 2002; Neidle et al. 2000).

Modern Developments

In 1973, the anthropologist Gordon W. Hewes presented the gestural theory in more modern dress. He too drew on evidence from signed languages but made no claim for a universal signed language. He also referred to contemporary work showing that great apes were unable to learn to speak but could use manual gestures in a language-like way, with at least moderate success. Gestural theory languished, though, until the 1990s. The influential article by Pinker and Bloom (1990) had made no mention of Hewes' work but was followed by an increasing number of publications that picked up the gestural theme (e.g., Arbib 2005; Armstrong 1999; Armstrong et al. 1995; Armstrong & Wilcox 2007; Corballis 1991; 1992, 2002, 2003; Givón 1995; Place 2000; Pollick & de Waal 2007; Rizzolatti & Arbib 1998; Rizzolatti & Sinigaglia 2008; Ruben 2005; Skoyles 2000; Tomasello 2008).

A major impetus has been the growing evidence that the signed languages of the deaf have all of the grammatical and semantic sophistication of spoken languages, as exemplified by the fact that university-level instruction at Gallaudet University in Washington, DC, is conducted entirely in American Sign Language (ASL). Children exposed to signed languages from infancy learn them as easily and naturally as those exposed to speech learn to speak, even going through a stage of manual "babbling" (Petitto & Marentette 1991). Although it has been argued that manual and vocal babbling are evidence for language as a fundamentally amodal system (Petitto et al. 2004), others have used evidence from sign language to argue that language, whether manual or vocal, is fundamentally gestural rather than amodal, with manual gestures taking precedence in human evolution (Armstrong et al. 1995; Armstrong & Wilcox 2007). At least in principle, language could have existed in a form consisting entirely of manual and facial gestures, comparable to present-day signed languages. It is perhaps more likely, though, that a vocal component was blended in gradually, and even today speech is normally accompanied by manual gestures (Goldin-Meadow & McNeill 1999, McNeill 1985, 1992).

Evidence has also come from the study of communication in nonhuman primates, and in particular our closest nonhuman relatives, chimpanzees and bonobos. Such evidence is of course indirect, since present-day chimpanzees and bonobos have themselves evolved since the last common human–chimpanzee ancestor some 6 or 7 million years ago. Given the general consensus that chimpanzees and bonobos are incapable of true language (or have yet to demonstrate it), there is the added problem that language must have appeared *de novo* at some point in hominin evolution, and intermediate forms are not available. Nevertheless, the chimpanzee and bonobo probably provide the best available proxies for the communicative abilities of the earliest hominins and a starting point for the construction of evolutionary scenarios.

Consideration of communication capabilities in primates indeed supports the idea that language evolved from manual gestures. For one thing, nonhuman primates have little if any cortical control over vocalization, but excellent cortical control over the hands and arms (Ploog 2002). This is illustrated by the fact that attempts over the past half-century to teach our closest nonhuman relatives, the great apes, to speak have been strikingly unsuccessful, but relatively good progress has been made toward teaching them to communicate by a form of manual sign language, as in the case of the chimpanzee Washoe (Gardner & Gardner 1969), or by pointing to visual symbols on a keyboard, as in the bonobo Kanzi (Savage-Rumbaugh et al. 1998). These visual forms of communication scarcely match the grammatical sophistication of modern humans, but they are a considerable advance over the restricted nature of the speech sounds that these animals make. The human equivalents of primate vocalizations are probably emotionally-based sounds like laughing, crying, grunting, or shrieking, rather than words.

Although vocal communication is widespread in the animal kingdom, surprisingly few species are capable of vocal learning, which is of course critical to speech. These species include elephants, seals, killer whales, and some birds (Jarvis 2006). Among the primates, according to Jarvis, only humans are vocal learners. Studies nevertheless suggest that primate calls do show limited modifiability, but its basis remains unclear, and it is apparent in subtle changes within call types rather than the generation of new call types (Egnor & Hauser 2004). Where language is flexible and conveys propositional information about variable features of the world, animal calls are typically stereotyped and convey information in circumscribed contexts. For example, a study by Pollick and de Waal (2007) showed that both chimpanzees and bonobos communicate more often and more flexibly using bodily gestures than vocalizations. They recorded the incidence of 31 manual gestures and 18 facial/vocal gestures over six different contexts (affiliative, agonistic, food, grooming, play, and sex) and found that the facial/vocal gestures were much more tied to specific contexts than the manual gestures. Other studies have shown that the communicative bodily gestures of gorillas (Pika et al. 2003), chimpanzees (Liebal et al. 2004), and bonobos (Pika et al. 2005) are subject to social learning and are sensitive to the attentional state of the recipient—both prerequisites for language.

Tomasello (2008) nevertheless notes that sensitivity to the attentional state of the recipient is not sufficient for true language, at least in the Gricean sense of language as a cooperative enterprise directed at a joint goal (Grice 1975). The chimpanzee gestural repertoire seems largely confined to making requests that are essentially self-serving; for example, a chimp may point to a desirable object that is just out of reach, with the aim of getting a watching human to help. According to Tomasello, true language requires the further step of shared attention, so that the communicator knows not only what the recipient knows or is attending to, but knows also that the recipient knows that the communicator knows this! This kind of recursive mind-reading enables communication beyond the making of simple requests to the sharing of knowledge, which is one of the distinctive properties of language.

Tomasello (2008) goes on to summarize his and his colleagues' work on pointing behavior in human infants, suggesting that shared attention and cooperative interchange begins to emerge from about one year of age. For example, 1-year-old infants not only point to objects in order to request them, but sometimes point to things that an adult is already looking at, indicating the understanding that attention to the object is shared. During the second year of life, infants seem already to understand sharing and shared communication before language itself, at least as manifest as speech, has appeared. In this, they have already moved beyond the communicative capacity of the chimpanzee.

Chimpanzees, along with bonobos, are our closest living nonhuman relatives and provide the best estimate of what communication was like prior to the emergence of true language. Tomasello's work therefore provides strong evidence that language evolved from manual gestures. Intermediate forms can be seen in gestural requests made by chimpanzees and bonobos, and in the progression to more collaborative forms of gesture in human infants.

The Mirror System

The gestural theory was boosted in the 1990s by the discovery of mirror neurons, and later of a more general mirror system, in the primate brain (Arbib 2005; Rizzolatti & Arbib 1998; Rizzolatti & Sinigaglia 2008). Mirror neurons, first recorded in area F5 in the ventral premotor cortex of the monkey, are activated both when the animal makes grasping movements and when it observes another individual making the same movements (Rizzolatti et al. 1996). Although these neurons responded to manual actions, area F5 is considered the homologue of Broca's area in humans (Rizzolatti & Arbib 1998), an area long associated with the production of speech (Broca 1861). More precisely, Broca's area in humans can be divided into Brodman areas 44 and 45, with area 44 considered the true analogue of area F5. In humans, it is now evident that area 44 is involved not only in speech, but also in motor functions unrelated to speech, including complex hand movements, and sensorimotor learning and integration (Binkofski & Buccino 2004). Indeed it has been proposed that "Broca's area" should now be regarded as a collective term, involving many different functions, and no clearly demarcated subdivisions (Lindenberg et al. 2007).

It has also become apparent that mirror neurons are part of a more general "mirror system" involving other regions of the brain as well as F5. In monkeys, the superior temporal sulcus (STS) also contains cells that respond to observed biological actions, including grasping actions (Perrett et al. 1989), although few if any respond when the animal itself performs an action. F5 and STS are connected to area PF in the inferior parietal lobule, where there are also neurons that respond to both the execution and perception of actions. These neurons are now known as "PF mirror neurons" (Rizzolatti et al. 2001). Other areas, such as the amygdala and orbito-frontal cortex, may also be part of the mirror system. Moreover, the extended mirror system in monkeys largely overlaps with the homologues of the cortical circuits in humans that are involved in language, leading to the notion that language is just part of the mirror system itself (Fogassi & Ferrari 2007; but see Grodzinsky 2006, for some caveats).

From Hand to Mouth

The primate mirror system has to do mainly with manual gestures, and the signed languages of the deaf are also predominantly manual, although facial movements also play a prominent role. But of course the dominant mode of present-day language is speech, although movements of the hand and face play a secondary role in normal conversation (Goldin-Meadow & McNeill 1999; McNeill 1985, 1992). If language evolved from manual gestures, then, there must have been a switch from hand to mouth during the course of hominin evolution. Some have regarded this as a critical weakness of the gestural theory. The linguist Robbins Burling, for example, wrote that "the gestural theory has one nearly fatal flaw. Its sticking point has always been the switch that would have been needed to move from a visual language to an audible one" (2005, p. 123).

Part of the answer to this lies in the growing realization that speech itself is a gestural system, so that the switch is not so much from vision to audition as from one kind of gesture to another. The notion of speech as gesture underlies the so-called motor theory of speech perception, which holds that speech sounds are perceived in terms of how they are produced, rather than on the basis of acoustic analysis (Liberman et al. 1967). The issue leading to the motor theory was not the lack of information in the acoustic signal, but rather the fact that individual phonemes are perceived as invariant despite extreme variability in the acoustic signal. Liberman and colleagues proposed that invariance lay instead in the articulatory gestures; as Galantucci et al. (2006) put it in a recent review, "perceiving speech is perceiving gestures" (p. 361). If this theory is correct, the perception of speech might therefore be considered a natural function of the mirror system.

The idea of speech as a gestural system has led to the concept of articulatory phonology (Browman & Goldstein 1995; Goldstein et al. 2006). Speech gestures comprise movements of six articulatory organs, the lips, the velum, the larynx, and the blade, body, and root of the tongue. Each is controlled separately, so that individual speech units are comprised of different combinations of movements. The distribution of action over these articulators means that the elements overlap in time, which makes possible the high rates of production and perception. As support for the motor theory of speech perception, the speech signal cannot be decoded by means of visual representations of the sound patterns, such as that provided in a sound spectrograph, but can be discerned in mechanical representations of the gestures themselves, through X-rays, magnetic resonance imaging, and palatography (Studdert-Kennedy 2005). The transition from manual gesture to speech, then, can be regarded as one occurring within the gestural domain, with manual gestures gradually replaced by gestures of the articulatory organs, but with likely overlap at all stages.

Although the mirror system in nonhuman primates seems not to incorporate vocalization (Rizzolatti & Sinigaglia 2008), it is receptive to acoustic as well as visual input. Kohler et al. (2002) recorded neurons in area F5 of the monkey that responded to the sounds of manual actions, such as tearing paper or breaking peanuts. Significantly, there was no response to monkey calls. This is consistent with evidence that vocalizations in nonhuman primates are controlled by the limbic system, rather than by neocortex (Ploog 2002), and are therefore not part of the mirror system. Even in the chimpanzee, voluntary control of vocalization appears to be limited, at best (Goodall 1986). The incorporation of vocalization into the mirror system was therefore a critical development in the evolution of speech—which is not to say that it was critical to the development of language itself.

Other properties of the primate mirror system indicate the close links between hand and mouth. Some neurons in area F5 in the monkey fire when the animal makes movements to grasp an object with either the hand or the mouth (Rizzolatti et al. 1988). Petrides et al. (2005) have identified an area in the monkey brain just rostral to premotor area 6 that is involved in control of the orofacial musculature. This area is also considered a homologue of part of Broca's area. The close neural associations between hand and mouth may be related to eating rather than communication, but later exapted for gestural and finally vocal language. The connection between hand and mouth can also be demonstrated behaviorally in humans. In one study, people were instructed to open their mouths while grasping objects, and the size of the mouth opening increased with the size of the grasped object; conversely, when they opened their hands while grasping objects with their mouths, the size of the hand opening also increased with the size of the object (Gentilucci et al. 2001).

Grasping with the hand also affect the kinematics of speech itself. Grasping larger objects induces selective increases in parameters of lip kinematics and voice spectra of syllables pronounced simultaneously with action execution (Gentilucci et al. 2004). Even observing another individual grasping or bringing to the mouth larger objects affects the lip kinematics and the voice spectra of syllables simultaneously pronounced by the viewer (Gentilucci 2003). In the course of evolution, this mechanism of joint control of hand and mouth could have been instrumental in the transfer of a communication system, based on the mirror system, from movements of the hand to movements of the mouth (Gentilucci & Corballis 2006).

The relationship between representations of actions and spoken language is further supported by neuroimaging studies, which show activation of Broca's area when people make meaningful arm gestures (Buccino et al. 2001; Decety et al. 1997; Gallagher & Frith 2004; Grèzes et al. 1998), or even imagine them (Gerardin et al. 2000; Grafton et al. 1996; Hanakawa et al. 2003; Kuhtz-Buschbeck et al. 2003; Parsons et al. 1995).

In evolutionary terms, a gradual shift from hand to mouth for purposes of intentional communication might well have begun with the increasing involvement of the hands in manufacture, and perhaps in transporting belongings or booty from one location to another. Manufactured stone tools, often considered to be a conceptual advance beyond the opportunistic use of sticks or rocks as tools, appear in the fossil record from some 2.5 million years ago, perhaps in *Homo rudolfensis*, a likely precursor to *Homo erectus* (Semaw et al. 1997). From some 1.8 million years, *erectus* began to migrate out of Africa into Asia and later into Europe (Tattersall 2003), and the Acheulian industry emerged, with large bifacial tools and handaxes that seemed to mark a significant advance over the simple flaked tools of the earlier Oldowan industry (Gowlett 1992).

With the hands increasingly involved in such activities, the burden of communication may have shifted to the face, which provides sufficient diversity of movement and expression to act as a signaling device. It also naturally conveys emotion and can serve to direct attention. Signed languages involve communicative movements of the face as well as of the hands (Sutton-Spence & Boyes-Braem 2001), and Muir and Richardson (2005) found that native signers watching discourse in British Sign Language focused mostly on the face and mouth, and relatively little on the hands or upper body. Facial expressions and head movements can turn an affirmative sentence into a negation, or a question. Facial gestures serve to disambiguate hand gestures and provide the visual equivalent of prosody in speech (Emmorey 2002). The face may play a much more prominent role in signed languages than has been hitherto recognized and may have been critical in the transition from manual gesture to speech.

The face also plays a role in the perception of normal speech. Although we can understand the radio announcer or the voice on the cell phone, there is abundant evidence that watching people speak can aid understanding of what they are saying. It can even distort it, as in the McGurk effect, in which dubbing sounds onto a mouth that is saying something different alters what the hearer actually hears (McGurk & MacDonald 1976). Evidence from an fMRI study shows that the mirror system is activated when people watch mouth actions, such as biting, lip-smacking, oral movements involved in vocalization, when these are performed by people, but not when they are performed by a monkey or a dog (Buccino et al. 2004). Actions belonging to the observer's own motor repertoire are mapped onto the observer's motor system, while those that do not belong are not—instead, they are perceived in terms of their visual properties. Watching speech movements, and even stills of a mouth making a speech sound, also activates the mirror system, including Broca's area (Calvert & Campbell 2003).

These observations are consistent with the idea that speech may have evolved from visual displays that included movements of the face.

Adding Sound

Despite the close association between hand and mouth in primates, the one missing ingredient is vocalization. The incorporation of vocalization into the mirror system may have naturally followed involvement of the face. Many facial gestures, especially those of the tongue, are internal to the mouth and largely hidden from sight. With the addition of sound through vibrations of the vocal folds, these hidden gestures are potentially recoverable through the mirror system, in much the same way as the primate mirror system responds to the sounds of actions such as tearing paper or cracking nuts (Kohler et al. 2002). Speech, then, is facial gesture half swallowed, and rendered partly invisible. Once sound was introduced, though, these gestures became accessible through audition, not vision. The problem, though, is that vocalization in nonhuman primates is controlled subcortically and appears not to be part of the mirror system, so its incorporation must have occurred at some point in hominin evolution.

One clue as when this might have occurred comes from genetics. A mutation of the forkhead box transcription factor, *FOXP2*, in some members of an English family known as the KE family has resulted in a severe deficit in vocal articulation (Watkins et al. 2002). Moreover, the members of the family affected by the mutation, unlike their unaffected relatives, show no activation in Broca's area while covertly generating verbs (Liégeois et al. 2003). This might be taken to mean that the *FOXP2* gene in humans is involved in the cooption of vocal control by Broca's area (Corballis 2004a). In songbirds, knockdown of the *FOXP2* gene impairs the imitation of song (Haesler et al. 2007), and insertion of the *FOXP2* point mutation found in the KE family into the mouse critically impairs synaptic plasticity and motor learning (Groszer et al. 2008).

Although highly conserved in mammals, the *FOXP2* gene underwent two mutations since the split between hominin and chimpanzee lines. According to one theoretical estimate, the more recent of these occurred "not less than" 100,000 years ago (Enard et al. 2002), although the error associated with this estimate makes it not unreasonable to suppose that it coincided with the emergence of *Homo sapiens* around 170,000 years ago. Contrary evidence, though, comes from a report that the mutation is also present in the DNA of a 45,000-year-old Neandertal fossil (Krause et al. 2007), suggesting that it goes back as much as 700,000 years ago to the common ancestor of humans and Neandertals (Noonan et al. 2006). But this is challenged in turn by Coop et al. (2008), who used phylogenetic dating of the haplotype to reestimate the time of the most recent common ancestor carrying the *FOXP2* mutation. Their answer was 42,000 years ago, with an estimated 95% confidence interval from 38,000 to 45,500 years ago. Even allowing for distortions in their assumptions, this is much more consistent with the estimate reported by Enard et al. (2002) than with the estimate implied by Krause et al. (2007).

Coop et al. argue that the presence of the mutation in Neandertal was more likely due to contamination of the Neandertal DNA, or to low rates of gene flow between human and Neandertal, on the assumption that the allele was globally beneficial. Recent evidence suggests that microcephalin, a gene involved in regulating brain size, may have entered the human gene pool through interbreeding with Neandertals (Evans et al. 2006), so the reverse possibility of *FOXP2* entering the late Neandertal gene pool from *Homo sapiens* is not completely ruled out. Our forebears might have been slightly friendlier toward the Neandertals than is generally thought.

Of course, mutation of the *FOXP2* gene need not have been the only factor in the switch from manual gesture to speech. Fossil evidence suggests that the anatomical requirements for fully articulate speech were probably not complete until the emergence of *Homo sapiens*. For example, the hypoglossal canal is much larger in humans than in great apes, suggesting that the hypoglossal nerve, which innervates the tongue, is also much larger in humans, perhaps reflecting the importance of tongued gestures in speech. The evidence suggests that the size of the hypoglossal canal in early australopithecines, and perhaps in *Homo habilis*, was within the range of that in modern great apes, while that of the Neandertal and early *H. sapiens* skulls was contained well within the modern human range (Kay et al. 1998), although this has been disputed (DeGusta et al. 1999). A further clue comes from the finding that the thoracic region of the spinal cord is relatively larger in humans than in nonhuman primates, probably because breathing during speech involves extra muscles of the thorax and abdomen. Fossil evidence indicates that this enlargement was not present in the early hominins or even in *Homo ergaster*, dating from about 1.6 million years ago but was present in several Neandertal fossils (MacLarnon & Hewitt 2004).

The Neandertals would have been incapable of fully articulate speech, based on reconstructions of their vocal tract (D.E. Lieberman 1998; P. Lieberman et al. 1972). Robert McCarthy of Florida Atlantic University has recently simulated how the Neanderthal would have sounded when articulating the syllable /i/ (or *ee*), based on the shape of the vocal tract.[a] One observer described it as sounding more like a sheep or a goat than a human. P. Lieberman's claims have nevertheless been controversial (e.g., Gibson & Jessee 1999), but there is other evidence that the cranial structure underwent changes subsequent to the split between anatomically modern and earlier "archaic" *Homo*, such as the Neandertals, *Homo heidelbergensis*, and *Homo rhodesiensis*, and that these changes bear on the ability to speak.

One such change is the shortening of the sphenoid, the central bone of the cranial base from which the face grows forward, resulting in a flattened face (P. Lieberman 1998). D. E. Lieberman et al. (2002) speculate that this is an adaptation for speech, contributing to the unique proportions of the human vocal tract, in which the horizontal and vertical components are roughly equal in length. This configuration, they argue, improves the ability to produce acoustically distinct speech sounds, such as the vowel [i]. It is not seen

---

[a] It can be found on http://anthropology.net/2008/04/16/reconstructing-Neanderthal-vocalizations/.

in Neandertal skeletal structure (see also Vleck 1970). Another adaptation unique to *H. sapiens* is neurocranial globularity, defined as the roundness of the cranial vault in the sagittal, coronal, and transverse planes, which is likely to have increased the relative size of the temporal and/or frontal lobes relative to other parts of the brain (D.E. Lieberman et al. 2002). These changes may reflect more refined control of articulation and also, perhaps, more accurate perceptual discrimination of articulated sounds.

Indirect support for the late emergence of speech comes from African click languages, which may be residues of prevocal language. Click sounds are made entirely in the mouth, and some languages have as many as 48 click sounds (Crystal 1997). Clicks could therefore have provided sufficient variety to carry a form of language prior to the incorporation of vocalization. Two of the many groups that make extensive use of click sounds are the Hadzabe and San, who are separated geographically by some 2000 kilometers, and genetic evidence suggests that the most recent common ancestor of these groups goes back to the root of present-day mitochondrial DNA lineages, perhaps as early as 100,000 years ago (Knight et al. 2003), prior to the migration of anatomically modern humans out of Africa. The date of this migration is still uncertain. Mellars (2006) suggests that modern humans may have reached Malaysia and the Andaman Islands as early as 60,000 to 65,000 years ago, with migration to Europe and the Near East occurring from western or southern Asia, rather than from Africa as previously thought. Those who migrated may have already developed vocal speech, leaving behind African speakers who retained click sounds. The only known non-African click language is Damin, an extinct Australian aboriginal language. *Homo sapiens* may have arrived in Australia as early as 60,000 years ago (Thorne et al. 1999), not long after they migrated out of Africa. This is not to say that the early Australians and Africans did not have full vocal control; rather, click languages may be simply a vestige of earlier languages in which sound production through the vocal folds had not yet been incorporated.

The most persistent advocate of the late emergence of speech is P. Lieberman, and as recently as 2007 he summarized a review of the evidence as follows:

> fully human speech anatomy first appears in the fossil record in the Upper Paleolithic (about 50,000 years ago) and is absent in both Neanderthals and earlier humans (p. 39).

This statement is not inconsistent with the dating of the most recent *FOXP2* mutation, discussed earlier and is later than the estimate of the origin of click languages, discussed in the preceding paragraph. It suggests that articulate speech arose even later than the emergence of our species, *Homo sapiens*, thought to have originated in East Africa not later than around 120,000 years ago (e.g., Ray et al. 2005). The question of precisely when articulate speech emerged is by no means resolved, and P. and D.E. Lieberman's provocative claims heighten the challenge to discover unequivocal data.

Why the Switch?

Given the fairly extensive anatomical and neurophysiological changes required, along with the heightened risk choking due to the lowering of the larynx, evolutionary pressure for the switch must have been strong. Since signed languages are as sophisticated linguistically as spoken ones (Armstrong et al. 1995; Emmorey 2002, Neidle et al. 2000), the pressure was almost certainly based on practical rather than linguistic considerations. One advantage of speech is that it can be carried on at night, or when the line of sight between sender and receiver is blocked. Communication at night may have been critical to survival in a hunter-gatherer society. The San, a modern hunter-gatherer society, are known to talk late at night, sometimes all through the night, to resolve conflict and share knowledge (Konner 1982).

Speech is much less energy-consuming than manual gesture. Anecdotal evidence from courses in sign language suggests that the instructors require regular massages in order to meet the sheer physical demands of sign-language expression. In contrast, the physiological costs of speech are so low as to be nearly unmeasurable (Russell et al. 1998). In terms of expenditure of energy, speech adds little to the cost of breathing, which we must do anyway to sustain life.

But it was perhaps the freeing of the hands for other adaptive functions, such as carrying things, and the manufacture of tools, that was probably the most critical. Vocal language allows people to use tools and at the same time explain verbally what they are doing, leading perhaps to pedagogy (Corballis 2002). Indeed, this may explain the dramatic rise of more sophisticated tools, bodily ornamentation, art, and perhaps music, in our own species. These developments have been dubbed a "human revolution" (Mellars & Stringer 1989), dating from some 40,000 years ago in Europe, but more recent evidence suggests that the rise toward technological and behavioral modernity originated in Africa. Two phases of technological innovation in southern Africa have been dated at around 70,000 to 75,000 years ago and 60 to 65,000 years ago, respectively (Jacobs et al. 2008), just preceding the estimated date of human dispersal from Africa at around 55,000 to 60,000 years ago (Mellars 2006). These developments, associated with *Homo sapiens*, are often attributed to the emergence of language itself, but I have proposed elsewhere, and suggest here, that the critical innovation was not language, but speech (Corballis 2004b).

In summary, the view adopted here is that language is a gestural system, based on the primate mirror system. In the course of time, it shifted from a general bodily system to one occupying a small region of the body, namely the articulatory organs of the vocal tract, and to a system requiring minimal energy. This freed the rest of the body from obligatory involvement in communication, allowing the hands to be used for other manipulative activities. This account, though, does not explain the extraordinary complexity and flexibility of human language relative to other forms of animal communication.

Grammar

In parallel with the gradual switch from manual to vocal mode, language must have acquired the distinctive grammatical properties that allow us the unlimited freedom of expression. Following Chomsky (1975), this has commonly been attributed to universal grammar, considered to be an innate endowment unique to humans. More recent developments suggest, though, that grammar arises from more general cognitive capacities (Christiansen & Chater 2008, Tomasello 2003, 2008; Wray 2002).

Mental Time Travel

One such capacity may be the ability to travel mentally in time. The concept of mental time travel was based initially on a distinction, drawn by Tulving (1972), between two forms of memory. Semantic memory is our vast storehouse of facts about the world, the combined dictionary and encyclopedia of the mind, whereas episodic memory is the memory for events, the mind's personal diary. Both are regarded as forms of what has been called declarative memory—or memory that can be declared—which already suggests a connection with language. Episodic memory, unlike semantic memory, implies a mental journey into the past, as when one mentally relives or imagines some past episode. Tulving (1972, 1985) proposed that episodic memory is unique to humans.

To some extent, these two forms of memory must be related. Tulving (2002) has argued, for instance, that the storage of episodic memories must depend on semantic memories that are already in place, but are then related to the self in subjectively-sensed time. This allows the experience of an event to be stored separately from the semantic system. Yet there is also evidence that semantic and episodic memory can be doubly dissociated. In most cases of amnesia, episodic memory is lost while semantic memory remains largely intact (e.g., Tulving et al. 1988). Conversely, people with semantic dementia, a degenerative neurological disorder that afflicts some people in late adulthood, show severe decline in semantic memory, but their episodic memories remain remarkably and surprisingly intact (Hodges & Graham 2001).

The idea that episodic memory involves reenactment of past episodes can be extrapolated to the more general idea of mental time travel, the human capacity to travel mentally both forward and backward in time (Suddendorf & Corballis 1997, 2007). Evidence from brain imaging that remembering the past and imagining the future activate a common "core" network (e.g., Addis et al. 2007), and patients with amnesia typically have as much difficulty imagining future episodes as in recalling past ones (e.g., Hassabis et al. 2007). Mental time travel into the future does not of course imply precognition but rather refers to the ability to imagine possible future episodes, whether for the purpose of planning detailed activities or for comparing and evaluating different strategies. Episodic memory provides the vocabulary of scenarios that enable us to envisage particular future scenarios, its survival value must lie, not in the memory component per se, but rather in what it contributes to present and future survival.

Just as Tulving proposed episodic memory to be uniquely human, so it has been proposed that only humans have the capacity for mental time travel (Suddendorf & Corballis 1997, 2007). This has been challenged (e.g., Clayton et al. 2003, and see commentaries on Suddendorf & Corballis 2007), but whether or not genuine counter-examples can be found in nonhuman animals there seems little doubt that the human ability to conjure past and future episodes, and indeed purely imaginary ones, exceeds any capacity so far demonstrated in nonhuman animals. Even if nonhuman animals can be shown to travel mentally in time, the remarkable feature of mental time travel in humans may not be its existence but rather its flexibility and combinatorial structure. Our past and future imaginings are typically made up of different combinations of elements that are otherwise familiar, such as people, objects, actions, situations, and emotions.

Human language is exquisitely designed to transmit exactly this kind of information (Corballis & Suddendorf 2007). As Pinker (2003, p. 27) put it, language seems to have evolved to express "who did what to whom, when, where, and why," thus allowing personal experiences to be shared, with consequent benefits in social bonding and social and practical understanding. The sharing of past, future, and indeed imaginary experiences places a much greater burden on the communication channel than if events are experienced only in the present. Events in the present are shared by mutual experience, and it may take only a few signals to direct attention, or to convey the importance of some components rather than others. Animals will sometimes use simple signals to draw attention to events that their conspecifics may not be able to see, as when chimpanzees use pant hoot calls to signal the discovery of food (Goodall 1986), or vervet monkeys use different calls to warn of different predators (Cheney & Seyfarth 1990), but no syntax or combinatorial structure is necessary.

The time dimension vastly increases the mental canvas, since reference to different times generally involves different places, different actions, different actors, and so on. In order to represent or refer to episodic elements that are not available in the present, we need very large vocabularies of concepts, as well as of words to represent them. And we need rules to represent the way in which the elements of an event are combined, and corresponding rules to convey these combinations to others in the form of language. If there is adaptive advantage to be gained from mental time travel through one's own personal experiences, that advantage can be multiplied by adding the experiences of others. It provides information about possible scenarios, about how individuals behave in different circumstances, about how the social world works. This perhaps explains the human predilection for gossip, and also for fiction, as in stories, novels, plays, TV soaps, and the like. Language, then, was born of the value of sharing, and indeed inventing, scenarios involving the interactions of humans with each other and with the world.

It is likely that the distinctive features of both language and mental time travel evolved during the Pleistocene. This era is usually dated from about 1.8 million years to about 10,000 years ago (e.g., Janis 1993)—although it has been argued that it should be dated from as early as 2.58 million years ago (Suc et al. 1997), which corresponds more closely to the emergence of the genus *Homo*. With the global shift to cooler climate after 2.5 million years ago, much of southern and eastern Africa probably became more open and sparsely wooded (Foley 1987). This left the hominins not only more exposed to attack from dangerous predators, such as sabertooth cats, lions, and hyenas, but also obliged to compete with them as carnivores. The solution was not to compete on the same terms, but to establish what Tooby and DeVore (1987) called the "cognitive niche," relying on social cooperation and intelligent planning for survival. The problem is that the number of combinations of actions, actors, locations, time periods, implements, and so forth that define episodes becomes very large, and a system of holistic calls to describe those episodes rapidly taxes the perceptual and memory systems. Syntax may then have emerged as a series of rules whereby episodic elements could be combined.

Further evidence that mental time travel evolved during the Pleistocene comes from stone tools, which appear to have been transported for repeated use. As we have seen, the relatively primitive Oldowan industry goes back some 2.5 million years (Semaw et al. 1997), but more direct evidence that mental time travel was involved comes from the reconstruction of knapping routines suggesting that, by the Middle Pleistocene at least, our hominin ancestors produced tools at one site for use at another (Hallos 2005). Tool manufacture remained fairly static, though, for most of the Pleistocene until the dramatic surge in technological innovation associated with our own species from around 60,000 to 75,000 years ago, discussed earlier.

The increased memory demands due to both mental time travel and language may well have driven the dramatic increase in brain size associated with the Pleistocene. The brain size of the early hominins was about the same, relative to body size, as that of the present-day great apes, but from the emergence of the genus *Homo* some 2 to 2.5 million years ago it increased, and had doubled by about 1.2 million years ago. It reached a peak, not with *Homo sapiens*, but with the Neandertals, who shared a common ancestry with modern humans until about 700,000 years ago (Noonan et al. 2006). In some individual Neandertals, brain capacity seems to have been as high as 1800 cc, with an average of around 1450 cc. Brain size in our own species, *Homo sapiens*, is a little lower, with a present-day average of about 1350 cc (Wood & Collard 1999). This is about three times the size expected for a great ape of the same body size. Of course brain size depends on factors other than cognitive demands, such as body size. Fossil evidence suggests that the absolute size of the human brain has decreased from around 35,000 years ago, but this was paralleled by a decrease in body size (Ruff et al. 1997).

Let's consider in more detail, then, some of the ways in which mental time travel has shaped language.

Symbols and Mime

In order to communicate about events at points in time other than the present, we require ways of referring to them *in absentia*. Here, the use of manual gesture has something of an advantage over speech, especially as bipedalism emerged, freeing the hands for other purposes, including communication. A feature of the hands and arms, once freed from locomotory and postural duties, is that they can move in four dimensions (three of space and one of time), and so mimic real-world events. The early hominins, though, were facultative bipeds, retaining some adaptive features of arboreal life, and there is little evidence that they were different in cognitive or behavioral terms from the other great apes. Obligate bipedalism emerged later, with the genus *Homo*, and Donald (1991) suggested that this marked the beginning of what he called "mimetic culture," in which events were related through mimed movements of the body, with the hands assuming special importance. Communication of this sort persists in dance, ballet, and mime, and we all resort to mime when trying to communicate with people who speak a language different from our own.

The modern sign languages of the deaf are also partially dependent on mime, or on direct copying of real-world actions. It has been estimated, for example, that in Italian Sign Language some 50% of the hand signs and 67% of the bodily locations of signs stem from iconic representations, in which there is a degree of spatiotemporal mapping between the sign and its meaning (Pietrandea 2002). Emmorey (2002) notes that in ASL some signs are purely arbitrary, but many more are iconic. For example, the sign for "erase" resembles the action of erasing a blackboard, and the sign for "play piano" mimics the action of actually playing a piano. But signs can be iconic without being transparently so, and they often cannot be guessed by naïve observers (Pizzuto & Volterra 2000). They also tend to become less iconic and more arbitrary over time, in the interests of speed, efficiency, and the constraints of the communication medium. This process is known as *conventionalization* (Burling 1999).

Once the principle of conventionalization is established, there is no need for communication to retain an iconic component, or even to depend on visual signals. We are quick to learn arbitrary labels, whether for objects, actions, emotions, or abstract concepts. Manual gesture may still be necessary to establish links in the first place—the child can scarcely learn the meaning of the word *dog* unless someone draws her attention to the animal itself—but there is otherwise no reason why the labels themselves need not be based on patterns of sound. Of course some concepts, such as the moo of a cow or miaow of a cat, depend on sound rather than sight, and it is not surprising that the words for these sounds tend to be onomatopoeic. Another example is *zanzara*, the evocative Italian word for mosquito, and Pinker (2007) notes a number of newly minted examples: *oink*, *tinkle*, *barf*, *conk*, *woofer*, *tweeter*. But most spoken words bear no physical relation to what they represent.

The Swiss linguist Ferdinand de Saussure (1977/1916) wrote of the "arbitrariness of the sign" as a defining property of language, and on this basis it is sometimes supposed that signed languages, with their strong foundation in iconic representations, are not true languages. The arbitrariness of words (or morphemes) is not so much a necessary property of language, though, as a matter of expedience, and of the constraints imposed by the language medium. Speech, for example, requires that the information be linearized, squeezed into a sequence of sounds that are necessarily limited in terms of how they can capture the physical nature of what they represent. The linguist Charles Hockett (1978) put it this way:

> when a representation of some four-dimensional hunk of life has to be compressed into the single dimension of speech, most iconicity is necessarily squeezed out. In one-dimensional projections, an elephant is indistinguishable from a woodshed. Speech perforce is largely arbitrary, if we speakers take pride in that, it is because in 50,000 years or so of talking we have learned to make a virtue of necessity (pp. 274–275).

The symbols of signed languages are not so constrained. The hands and arms can mimic the shapes of real-world objects and actions, and to some extent lexical information can be delivered in parallel instead of being forced into rigid temporal sequence. Even so, conventionalization allows signs to be simplified and speeded up, to the point that many of them lose most or all of their iconic aspect. For example, the ASL sign for *home* was once a combination of the sign for *eat*, which is a bunched hand touching the mouth, and the sign for *sleep*, which is a flat hand on the cheek. Now it consists of two quick touches on the cheek, both with a bunched handshape, so the original iconic components are effectively lost (Frishberg 1975).

The increasing complexity of human culture as it developed during the Pleistocene and beyond no doubt accelerated the drive toward conventionalization and increased efficiency. It has been estimated that the average literate person today knows some 50,000 concepts (Pinker 2007), and even with the degrees of freedom afforded by hand and arm movements it would be slow and cumbersome in the extreme to represent even the spatial concepts manually—although we often point or incline the head to indicate direction ("He went *that* way"). Conventionalization can be regarded simply as a device, dependent on associative learning, to streamline the representational system for maximum efficiency. Just as signs become more compact, so words tend to become shorter with more frequent use. This is captured by Zipf's law, which states that the length of a word is inversely proportional to its rank in frequency. The reason for this is evident from the title of Zipf's 1949 book, *Human Behavior and the Principle of Least-Effort*. Hence we have the progression from *television* to *telly* to *TV*, or in my own country from *universit*y to *varsity* to *uni*.

Marking Time

If it is to convey information about episodes, language must include some mechanism for conveying when in time they occurred, or will occur. In many languages this is accomplished by tense markers. In English, for example, verbs describing actions and states are endowed with tense to indicate different points in time, as well as distinctions between conditional and unconditional, continuous and non-continuous, and so on. Thus the words *walk*, *walked*, and *walking*, along with auxiliaries (e.g., *will walk, might have been walking*), refer to different times or timing conditions to do with a perambulatory event. Some languages have no tenses as such, but have other ways of indicating time. Chinese, for example, has no tenses, but the time of an event can be indicated by adverbs, such as *tomorrow*, and what are called aspectual markers, as in a sentence that might be roughly rendered as *He break his leg before* (Lin 2005).

The variety of ways in which time is marked in different languages suggests cultural influences rather than the operation of universal grammar. A revealing example comes from the language spoken by the Pirahã, a tribe of some 200 people in Brazil, which has only a very restricted way of talking about relative time. This takes the form of two tense-like morphemes which indicate simply whether an event is in the present or not. Pirahã also includes a few words serving as temporal markers, such as *nigh*t, *day*, *full moon*, and so on. The Pirahã are said to live largely in the present, with no creation myths, no art or drawing, no individual or collective memory for more than two generations past (Everett 2005).

One might suppose that the apparent paucity of mental time travel in Pirahã life results from the lack of ways to express time, as suggested by the Whorfian hypothesis. But the reverse is more likely true—that is, the language of the Pirahã adapted to their lack of concern for time (cf. Christiansen & Chater 2008). The Pirahã language is limited in other ways besides the dearth of time markers. It has no numbers or system of counting, no color terms, and may even be said to lack verbs, in the sense of a verb as a linguistic class, the Pirahã learn verbs one by one as individual entities. There is no recursive embedding of clauses (Everett 2005). One might be tempted to believe that the Pirahã suffer from some genetic defect, but this idea is rejected by Everett, who describes them as "some of the brightest, pleasantest, and most fun-loving people that I know" (p. 621).

Everett suggests that even these additional features derive fundamentally from their very limited sense of time, supporting the idea that the characteristics of language derive from mental time travel. He writes:

> [the] apparently disjointed facts about the Pirahã language—gaps that are very surprising from just about any grammarian's perspective—ultimately derive from a single cultural constraint in Pirahã, namely, *the restriction of communication to the immediate experience of the interlocutors.* (p. 622).

Everett's work on the Pirahã is understandably controversial (see the critique by Nevins et al. 2007 and the response by Everett 2007). Despite its seeming simplicity, though, Pirahã language is rich in morphology and prosody. Everett insists that it should not be regarded as in any way "primitive" and suggests that it is probably not especially unusual. Other languages of nonliterate peoples may have similar characteristics. For example, the Iatmul language of New Guinea is also said to have no recursion (Karlsson 2007). Tomasello (2003) suggests that theories of language have been unduly influenced by the characteristics of written language, and remarks that "there are very few if any specific grammatical constructions or markers that are universally present in all languages" (p. 5).

Generativity

The most distinctive property of language is that it is generative (Chomsky 1966). We can both construct and understand sentences that we have never used or heard before. A classic example comes from the British philosopher Alfred North Whitehead. In 1934 he had been seated at dinner next to B.F. Skinner, who was trying to explain how the principles of behaviorism would change the face of psychology. Obliged to challenge this view, Whitehead uttered the sentence "No black scorpion is falling upon this table," and asked Skinner to explain the behavioral principles that might have led him to say that. It was not until the publication of *Verbal Behavior* 23 years later that Skinner (1957) attempted an answer. In an appendix to that book, Skinner proposed that Whitehead was unconsciously expressing the fear that behaviorism might indeed take over, likening it to a black scorpion that he would not allow to tarnish his philosophy. Skinner's explanation is ironic, because it seems to owe more to psychoanalysis than to behaviorism, and Skinner was well known for anti-Freudian views.

We now know, largely through the efforts of Chomsky (1957 1959), that language cannot be explained in terms of learned sequences. Instead, it depends on rules. These rules combine words in precise ways to enable us to create and extract an essentially unlimited number of different meanings. As the German philosopher Gottlob Frege (1980/1914, p. 79) put it:

> The possibility of our understanding sentences that we have never heard before rests evidently on this, that we can construct the sense of a sentence out of parts that correspond to words.

The combinatorial structure of sentences, I suggest, derives in large part from the combinatorial structure of episodes, and words provide the access to the components of episodes. Most of the episodes we witness, remember, or construct in our minds are combinations of the familiar. Indeed it is generally the combinations that count, rather than the individual elements. In Whitehead's sentence, the notions of a black scorpion, falling, and a table are of themselves of less interest than the unusual combination of a scorpion in downward motion above the very table at which the two savants sat—and there may have been relief that this unusual event was not occurring. The manner in which the words describing such episodes are arranged depends on the conventions that make up grammar.

One such convention has to do with the order in which words are uttered or signed. The most basic episodes are those involving objects and actions, so the first "words" were probably nouns and verbs—an idea that goes back to the 19th-century English philologist John Horne Tooke (1857), who regarded nouns and verbs as "necessary words." The prototypical episode of someone doing something to someone or something else, then, requires one noun to be the subject, another to be a verb describing the action, and another noun to be the object of the action. How these are ordered is simply a matter of convention. In English, the convention is to place them in the order *subject verb object* (SVO). To use a well-chewed example, the sentence "dog bites man" means something very different from "man bites dog," the latter is news, the former simply a personal misfortune. Among the world's languages, the most common word order is SOV, with the verb at the end, but all possible combinations exist among the world's languages.

In speaking, we can only utter one word at a time, so word order can be critical. Some languages, though, mark the roles played by different words with changes to the words themselves. In Latin, for example, the subject and object of a sentence are signaled by different inflections—changes to the end of the word—and the words can be reordered without losing the meaning. So *canis virum mordet* means "dog bites man," while *canem vir mordet* means "man bites dog," although it would be normal still to place the subject first. The Australian aboriginal language Walpiri is a more extreme example of an inflected language in which word order makes essentially no difference. Such languages are sometimes called *scrambling languages*. Chinese, by contrast, is an example of an *isolating language*, in which words are not inflected and different meanings are created by adding words or altering word order. English is closer to being an isolating language than a scrambling one.

Unlike spoken languages, signed languages are less constrained to present words in sequence, since they make use of spatial information as well as sequential information, and different words can be signed simultaneously. For example, the hand can take the shape of an object while its movement indicates the action. Even so, the order in which signs are displayed can be critical. In ASL the basic order is SVO, while the newly emerged Al-Sayyid Bedouin Sign Language (ABSL) is an SOV language (Aronoff et al. 2007). In any event, episodes themselves are typically sequential, and it is natural for the sequence of events to be copied into the language that expresses them.

Grammaticalization

Grammar can be regarded as a device for making communication more efficient and streamlined. For example, many of the words we use do not refer to actual content, but serve functions that are purely grammatical. These are called function words, and include articles, such as *a* and *the*, prepositions such as *at*, *on*, or *about*, and auxiliaries such *will* in "They will come." Function words nevertheless almost certainly have their origins in content words, and the process by which content words are stripped of meaning to serve purely grammatical functions is known as *grammaticalization* (Heine & Kuteva 2007; Hopper & Traugott 1993). A classic example is the word *have*, which progressed from a verb meaning to "seize" or "grasp" (Latin *capere*), to one expressing possession (as in *I have a pet porcupine*, Latin *habere*), to a marker of the perfect tense ("I have gone") and a marker of obligation ("I have to go"). Similarly, the word *will* probably progressed from a verb (as in "Do what you will") to a marker of the future tense ("They will laugh").

Another example comes from the word *go*. It still carries the meaning of travel, or making a move from one location from to another, but in sentences like "We're going to have lunch" it has been bleached of content and simply indicates the future. The phrase *going to* has been compressed into the form *gonna*, as in "We're gonna have lunch," or even "I'm gonna go." In the US, people make an additional compression when they say "Let's go eat," where we less hungry Kiwis say "Let's go and eat." I'm waiting to hear someone to say "Let's go go-go."[b]

---

[b] I have since learned that "Let's go go-go" is the battle song of the Chicago White Sox.

There are other ways in which grammaticalization operates to make communication more streamlined. One has to do with the embedding and concatenization of phrases. For example, the statements "He pushed the door" and "The door opened" can be concatenated into "He pushed the door open." Statements like "My uncle is generous with money" and "My uncle helped my sister out" can be concatenated by embedding the first in the second: "My uncle, who is generous with money, helped my sister out." One can also alter the priority of the two statements by reversing the embedding: "My uncle, who helped my sister out, is generous with money."

Efficiency can also be improved by breaking down concepts into component parts, which can then be recombined to form new concepts. An interesting example comes from a signed language. In Nicaragua deaf people were isolated from one another until the Sandinista government assumed power in 1979 and created the first schools for the deaf. Since that time, the children in these schools invented their own sign language, which has blended into the system now called Lenguaje de Signos Nicaragüense (LSN). In the course of time, LSN has changed from a system of holistic signs to a more combinatorial format. For example, one generation of children were told a story of a cat that swallowed a bowling ball, and then rolled down a steep street in a "waving, wobbling manner." The children were then asked to sign the motion. Some indicated the motion holistically, moving the hand downward in a waving motion. Others, however, segmented the motion into two signs, one representing downward motion and the other representing the waving motion, and this version increased after the first cohort of children had moved through the school (Senghas et al. 2004).

One need not appeal to universal grammar to explain how this kind of segmentation occurs. Computer simulations have shown that cultural transmission can change a language that begins with holistic units into one in which sequences of forms are combined to produce meanings that were earlier expressed holistically (see Kirby & Hurford 2002, for a review).

Putting It Together

In this article, I have suggested two ways in which language evolved. First, I have argued that language evolved from intentional manual gestures, with vocal gestures gradually assuming dominance, perhaps with the emergence of *Homo sapiens* within the last 200,000 years. In this sense, language is an extension of the so-called mirror system, whereby primates understand the actions of others. Second, I have proposed that the evolution of mental time travel and the awareness of time led to a more complex form of language for the communication of episodes that take place at times other than the present. The demands of referring to episodes that are not immediately accessible to the senses led to the construction of grammar, or grammars, which probably took place over the past 2 million or so years from the emergence of the large-brained genus *Homo*.

Although mental time travel may have set the initial stage for language, the two must also have co-evolved. Thus Gärdenfors (2004) writes that, in his view, "there has been a co-evolution of cooperation about future goals and symbolic communication" (p. 243). Language itself adds to the capacity for mental time travel, since it provides a means by which people can create the equivalent of episodic memories in others, and therefore contributes to their episodic thinking. By telling you what happened to me, I can effectively create an imagined episode in your mind, and this added information might help you adapt more effectively to future conditions. And by telling you what I am about to do, you may form an image in your own mind, and work out a plan to help me, or perhaps thwart me.

Just when language became grammatical in relation to when speech became the dominant mode has been a matter of conjecture. Arbib

(2005) proposed that language evolved from manual gestures, but suggested that manual communication did not progress beyond what he called *protosign*. This is the signed equivalent of what Bickerton (1995) called *protolanguage*, which is effectively language without grammar, as exemplified by forms of communication acquired by captive great apes such as Washoe and Kanzi, or by the 2-year-old child. It is roughly equivalent to the mimetic stage proposed by Donald (1991). In Arbib's view, the transition was then from protosign to protospeech, and grammatical language evolved from there.

Since the signed languages of the deaf are fully grammatical, there seems no reason in principle why gestural language should not have achieved the status of full language prior to the emergence of speech as the dominant mode. It is perhaps more likely, though, that the switch from protolanguage to language was itself gradual and occurred in parallel with the switch from a primarily manual form of communication to a primarily vocal one, with varying degrees of admixture. The distinction between protolanguage and language is generally depicted as all-or-none, perhaps encouraged by the notion that grammatical language depends on the innate, uniquely human endowment known as universal grammar (e.g., Chomsky 1975). But if grammar emerges gradually, as suggested earlier, and is culturally rather than biologically tuned, then the evolution of grammar and the switch from manual to vocal modes may have been contemporaneous and largely independent.

Just as grammars vary considerably between cultures, so different cultures may vary in the extent to which speech dominates. At one extreme, of course, are the signed languages developed in deaf communities. Less extreme are signed languages developed by some native Australian tribes (Kendon 1988), and in the so-called Plains Sign Talk of Native North American tribes (Mithun 1999), in both cases, theses tribes also speak, but use signed language for special purposes. Signing may also have been more prominent among African tribes than among those of the industrialized West, as suggested by the following provocative quote from the 19th-century British explorer, Mary Kingsley (1965/1897):

> [African languages are not elaborate enough] to enable a native to state his exact thought. Some of them are very dependent upon gesture. When I was with the Fans they frequently said "We will go to the fire so that we can see what they say", when any question had to be decided after dark, and the inhabitants of Fernando Po, the Bubis, are quite unable to converse with each other unless they have sufficient light to see the accompanying gestures of the conversation (p. 504).

While this may seem condescending, it may well be the case that some cultures may make more use of manual gesture than others, through cultural rather than biological necessity. In suggesting that the African languages she observed were not elaborate, Kingsley also overlooked her own observation that elaboration was provided by manual gestures, not by spoken words.

I suggested above that grammar and speech evolved independently, but this may not be completely true. As suggested earlier, speech may have freed the hands for manufacture, vastly increasing the number of objects to be named. This may have increased the pressure for language to grammaticalize and become more efficient and sophisticated. We saw earlier that the languages of nonliterate societies may be simpler in terms of such features as the recursive embedding of clauses. The rise of technology may also have increased the demand for language to function as a pedagogical device. To that extent, then, Arbib (2005) may be correct in suggesting that grammaticalization accelerated after speech became the dominant mode.

Summary and Conclusions

I have argued that language evolved from the mirror system in primates, which provides a platform for both the production and perception of intentional bodily acts. This system was adapted for communication in hominins in part as a result of bipedalism, which freed the hands for more varied and elaborate gestures. But the climb to true language probably began with the emergence of the genus *Homo*, and the pressure to more cooperative behavior during the Pleistocene, when our forebears were forced from the forest canopy onto the more open and dangerous environment of the savanna. Memory systems evolved from simple learning and pattern recognition to the storage and retrieval of particular episodes, enabling more precise planning and prediction of the future. Mental time travel into past and future also gave rise to a sense of the self through time. One consequence may have been an understanding of mortality, leading to the emergence of religions promising a life after death.

Language was adapted to the sharing of episodic information, whether based on memory, future plans, or on fiction. Grammatical language, whether signed or spoken, seems uniquely adapted for the sharing of this information in an efficient, streamlined manner. Along with the development of grammar, vocal elements were increasingly introduced, so that with the emergence of our own species, *Homo sapiens*, speech became the dominant mode. This freed the hands for the development of more sophisticated manufacture and use of tools, as well as other artifacts such as bodily ornamentation, clothing, and perhaps musical instruments. This may have given further impetus to the development of sophisticated grammar and the use of language for more extensive and varied activities, such as pedagogy or argument.

A recurrent theme of this article has been the cultural shaping of language. The sheer variety of different language structures argues against Chomsky's (1975) notion of an innate universal grammar, or what Pinker (1994) called the "language instinct." This point has been elaborated in the recent article by Christiansen and Chater (2008). At some level, though, there must be an innate component, since grammatical language is universally and uniquely human. The critical question is whether the concept of universal grammar is useful in helping us understand the different forms that languages can take, from manual to vocal, from Pirahã to standard English. It may be more useful to view the constructive nature of language as the product of what Locke and Bogin (2006), after Marler (1991), called an "instinct for inventiveness" that goes beyond language per se. This instinct may well be uniquely human but is evident in many activities other than language, including mental time travel, manufacture, art, music, and other modes of storytelling, such as dance, drama, movies, and television. We are at once the most articulate and time-conscious of all species, and I dare any other species to contradict me.

Acknowledgments

I thank Thomas Suddendorf and an anonymous referee for their constructive comments on an earlier draft. Others who have contributed significantly to my thinking on language evolution include Nicola Clayton, Karen Emmorey, Tecumseh Fitch, Maurizio Gentilucci, Russell Gray, Nicholas Humphrey, Jim Hurford, Giacomo Rizzolatti, and Michael Studdert-Kennedy—but not all of them agree with me.

Conflicts of Interest

The author declares no conflicts of interest.

References

Addis, D. R., Wong, A. T., & Schacter, D. L. (2007). Remembering the past and imagining the future: Common and distinct neural substrates during event construction. *Neuropsychologia*, *45*, 1363–1377.

Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral & Brain Sciences*, *28*, 105–168.

Armstrong, D. F. (1999). *Original Signs: Gesture, Sign, and the Source of Language*. Washington: Gallaudet University Press.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the Nature of Language*. Cambridge: Cambridge University Press.

Armstrong, D. F., & Wilcox, S. E. (2007). *The Gestural Origin of Language*. Oxford: Oxford University Press.

Aronoff, M., Meir, I., Padden, C. A., et al. (2007). The roots of linguistic organization in a new language. *Interaction Studies*, *9*, 133–153.

Barsalou, L. W. (2008). Grounded cognition. *Ann. Rev. Psych.*, *59*, 617–645.

Bickerton, D. (1995). *Language and Human Behavior*. Seattle: University of Washington Press.

Binkofski, F., & Buccino, G. (2004). Motor functions of the Broca's region. *Brain & Language*, *89*, 362–389.

Broca, P. (1861). Remarques sur le siège de la faculté de la parole articulée, suivies d'une observation d'aphémie (perte de parole). *Bulletin de la Société d'Anatomie (Paris)*, *36*, 330–357.

Browman, C. P., & Goldstein, L. F. (1995). Dynamics and articulatory phonology. In T. van Gelder & R. F. Port (Eds.), *Mind as Motion* (pp. 175–193). Cambridge, MA: MIT Press.

Buccino, G., Binkofski, F., Fink, G. R., et al. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *Eur. J. Neurosci.*, *13*, 400–404.

Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., et al. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An fMRI study. *J. Cogn. Neurosci.*, *16*, 114–126.

Bulwer, J. (1644). *Chirologia: On the Natural Language of the Hand*. London.

Burling, R. (1999). Motivation, conventionalization, and arbitrariness in the origin of language. In B. J. King (Ed.), *The Origins of Language: What Nonhuman Primates Can Tell Us* (pp. 307–350). Santa Fe, NM: School of American Research Press.

Burling, R. (2005). *The Talking Ape*. New York: Oxford University Press.

Calvert, G. A., & Campbell, R. (2003). Reading speech from still and moving faces: The neural substrates of visible speech. *J. Cogn. Neurosci.*, *15*, 57–70.

Cheney, D. L., & Seyfarth, R. M. (1990). *How Monkeys See the World*. Chicago: University of Chicago Press.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1959). A review of B. F. Skinner's "Verbal behavior." *Language*, *35*, 26–58.

Chomsky, N. (1966). *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper & Row.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral & Brain Sciences*, *31*, 489–558.

Christiansen, M. H., & Kirby, S. (2003). Language evolution: The hardest problem in science? In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution* (pp. 1–15). Oxford: Oxford University Press.

Clayton, N. S., Bussey, T. J., & Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.*, *4*, 685–691.

Condillac, E. B. de. (1971). *An Essay on the Origin of Human Knowledge*. T. Nugent (Tr.), Gainesville, FL: Scholars Facsimiles and Reprints. (Work originally published 1746).

Coop, G., Bullaughev, K., Luca, F., & Przeworski, M. (2008). The timing of selection of the human FOXP2 gene. *Mol. Biol. Evol.*, *25*, 1257–1259.

Corballis, M. C. (1991). *The Lopsided Ape*. New York: Oxford University Press.

Corballis, M. C. (1992). On the evolution of language and generativity. *Cognition*, *44*, 197–226.

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.

Corballis, M. C. (2003). From mouth to hand: Gesture, speech, and the evolution of right-handedness. *Behavioral & Brain Sciences*, *26*, 199–260.

Corballis, M. C. (2004a). FOXP2 and the mirror system. *Trends Cogn. Sci.*, *8*, 95–96.

Corballis, M. C. (2004b). The origins of modernity: Was autonomous speech the critical factor? *Psychol. Rev.*, *111*, 543–522.

Corballis, M. C., & Suddendorf, T. (2007). Memory, time, and language. In C. Pasternak (Ed.), *What Makes Us Human* (pp. 17–36). Oxford, UK: Oneworld Publications.

Cordemoy, G.de. (1972). *Discours physique de la parole* [A philosophical discourse concerning speech, conformable to the Cartesian principles]. Delmar, NY: Scholars' Facsimiles & Reprints, Inc (Work originally published 1688).

Crystal, D. (1997). *The Cambridge Encyclopedia of Language* (2nd ed.). Cambridge, UK: Cambridge University Press.

Darwin, C. (1859). *The Origin of Species*. London: John Murray.

Darwin, C. (1896). *The Descent of Man and Selection in Relation to Sex*. London: William Clowes. (Work originally published 1871).

Decety, J., Grezes, J., Costes, N., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain*, *120*, 1763–1777.

DeGusta, D., Gilbert, W. H., & Turner, S. P. (1999). Hypoglossal canal size and hominid speech. *Proc. Nat. Acad. Sci. USA*, *96*, 1800–1804.

Descartes, R. (1985). Discourse on method. In J. Cottingham, R. Stootfoff, & D. Murdock (Eds. & tr.), *The Philosophical Writings of Descartes*. Cambridge: Cambridge University Press. (Work originally published 1647).

Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Amer. Biol. Teacher*, *35*, 125–129.

Donald, M. (1991). *Origins of the Modern Mind*. Cambridge, MA: Harvard University Press.

Egnor, S. E. R., & Hauser, M. D. (2004). A paradox in the evolution of primate vocal learning. *Trends Neurosci.*, *27*, 649–654.

Emmorey, K. (2002). *Language, Cognition, and Brain: Insights from Sign Language Research*. Hillsdale, NJ: Erlbaum.

Enard, W., Przeworski, M., Fisher, S. E., et al. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*, 869–871.

Evans, P. D., Mekel-Bobrov, N., Yallender, E. J., et al. (2006). Evidence that the adaptive allele of the brain size gene *microcephalin* introgressed into *Homo sapiens* from an archaic *Homo* lineage. *Proc. Nat. Acad. Sci. USA*, *103*, 18178–18183.

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã. *Curr. Anthro.*, *46*, 621–646.

Everett, D. L. (2007). *Cultural Constraints on Grammar in Pirahã: A Reply to Nevins, Pesetsky, and Rodrigues (2007)*. Available online at http://ling.auf.net/lingBuzz/000427

Fogassi, L., & Ferrari, P. F. (2007). Mirror neurons and the evolution of embodied language. *Curr. Direct. Psych. Sci.*, *16*, 136–141.

Foley, R. (1987). *Another Unique Species: Patterns in Human Evolutionary Ecology*. Harlow: Longman Scientific and Technical.

Frege, G. (1980). Letter to Jourdain. In G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, & A. Veraart (Eds.), *Philosophical and Mathematical Correspondence* (pp. 78–80). Chicago: Chicago University Press (Work originally published 1914)

Frishberg, N. (1975). Arbitrariness and iconicity in American Sign Language. *Language*, *51*, 696–719.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonom. Bull. Rev.*, *13*, 361–377.

Gallagher, H. L., & Frith, C. D. (2004). Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia*, *42*, 1725–1736.

Gärdenfors, P. (2004). Cooperation and the evolution of symbolic communication. In D. K. Oller & U. Griebel (Eds.), *Evolution of Communication Systems* (pp. 237–256). Cambridge, MA: MIT Press.

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science*, *165*, 664–672.

Gentilucci, M. (2003). Grasp observation influences speech production. *Eur. J. Neurosci.*, *17*, 179–184.

Gentilucci, M., Benuzzi, F., Gangitano, M., et al. (2001). Grasp with hand and mouth: A kinematic study on healthy subjects. *J. Neurophysiol.*, *86*, 1685–1699.

Gentilucci, M., Santunione, P., Roy, A. C., et al. (2004). Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *Eur. J. Neurosci*, *19*, 190–202.

Gentilucci, M., & Corballis, M. C. (2006). From manual gesture to speech: A gradual transition. *Neurosci. Biobehav. Rev.*, *30*, 949–960.

Gerardin, E., Sirigu, A., Lehericy, S., et al. (2000). Partially overlapping neural networks for real and imagined hand movements. *Cerebral Cortex*, *10*, 1093–1104.

Gibson, K. R., & Jessee, S. (1999). Language evolution and expansions of multiple neurological processing areas. In B. J. King (Ed.), *The Origins of Language: What Nonhuman Primates Can Tell Us*. Santa Fe, NM: School of American Research Press.

Givòn, T. (1995). *Functionalism and Grammar*. Philadelphia: Benjamins..

Goldin-Meadow, S., & McNeill, D. (1999). The role of gesture and mimetic representation in making language the province of speech. In M. C. Corballis & S. E. G. Lea (Eds.), *The Descent of Mind* (pp. 155–172). Oxford, UK: Oxford University Press.

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to Language via the Mirror Neuron System* (pp. 215–249). Cambridge, UK: Cambridge University Press.

Goodall, J. (1986). *The Chimpanzees of Gombe: Patterns of Behavior*. Harvard University Press.

Gowlett, J. A. J. (1992). Early human mental abilities. In S. Jones, R. Martin, & D. Pilbeam (Eds.), *The Cambridge Encyclopedia of Human Evolution* (pp. 341–345). Cambridge: Cambridge University Press.

Grafton, S. T., Arbib, M. A., Fadiga, L., et al. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp. Brain Res.*, *112*, 103–111.

Grèzes, J., Costes, N., & Decety, J. (1998). Top-down effect of strategy on the perception of human biological motion: A PET investigation. *Cogn. Neuropsych.*, *15*, 553–582.

Grice, P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics, Vol. 3: Speech Acts* (pp. 43–58). New York: Academic Press.

Grodzinsky, Y. (2006). The language faculty, Broca's region, and the mirror system. *Cortex*, *42*, 464–468.

Groszer, M., Keays, D. A., Deacon, R. M. J., et al. (2008). Impaired synaptic plasticity and motor learning in mice with a point mutation implicated in human speech deficits. *Curr. Biol.*, *18*, 354–362.

Haesler, S., Rochefort, C., Georgi, B., et al. (2007). Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *PLoS Biol.*, *5*, 2885–2897.

Hallos, J. (2005). "15 minutes of fame:" Exploring the temporal dimensions of Middle Pleistocene lithic technology. *J. Hum. Evol.*, *29*, 155–179.

Hanakawa, T., Immisch, I., Toma, K., et al. (2003). Functional properties of brain areas associated with motor execution and imagery. *J. Neurophys.*, *89*, 989–1002.

Hassabis, D., Kumaran, D., Vann, S. D., et al. (2007). Patients with hippocampal amnesia cannot imagine new experiences. *Proc. Nat. Acad. Sci. USA*, *104*, 1726–1731.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*, 1569–1579.

Heine, B., & Kuteva, T. (2007). *The Genesis of Grammar*. Oxford: Oxford University Press.

Hewes, G. W. (1973). Primate communication and the gestural origins of language. *Curr. Anthro.*, *14*, 5–24.

Hockett, C. (1978). In search of love's brow. *American Speech*, *53*, 243–315.

Hodges, J. R. & Graham, K. S. (2001). Episodic memory: Insights from semantic dementia. *Phil. Trans. R. Soc. Lond. B: Biol. Sci.*, *356*, 1423–1434.

Hopper, P. J., & Traugott, E. C. (1993). *Grammaticalization*. Cambridge, UK: Cambridge University Press.

Horne Tooke, J. (1857). *Epea Pteroenta or the Diversions of Purley*. London.

Jacobs, Z., Roberts, R. G., Galbraith, R. F., et al. (2008). Ages for the Middle Stone Age of Southern Africa: Implications for human behavior and dispersal. *Science*, *322*, 733–735.

Janis, C. (1993). Victors by default: The mammalian succession. In S. J. Gould (Ed.), *The Book of Life* (pp. 169–217). New York: W.W. Norton.

Jarvis, E. D. (2006). Selection for and against vocal learning in birds and mammals. *Ornithol. Sci.*, *5*, 5–14.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*, 365–392.

Kay, R. F., Cartmill, M., & Barlow, M. (1998). The hypoglossal canal and the origin of human vocal behavior. *Proc. Nat. Acad. Sci. USA*, *95*, 5417–5419.

Kendon, A. (1988). *Sign Languages of Aboriginal Australia*. Melbourne: Cambridge University Press.

Kingsley, M. (1965). *Travels in West Africa, Congo Française, Corisco and Cameroons*. London: F. Cass (Work originally published 1897).

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the Evolution of Language* (pp. 121–148). London: Springer-Verlag.

Kohler, E., Keysers, C., Umilta, M. A., et al. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, *297*, 846–848.

Konner, M. (1982). *The Tangled Wing: Biological Constraints on the Human Spirit*. New York: Harper.

Knight, A., Underhill, P. A., Mortensen, H. M., et al. (2003). African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr. Biol.*, *13*, 464–473.

Krause, J., Lalueza-Fox, C., Orlando, L., et al. (2007). The derived FOXP2 variant of modern humans was shared with Neandertals. *Curr. Biol.*, *17*, 1908–1912.

Kuhtz-Buschbeck, J. P., Mahnkopf, C., Holzknecht, C., et al. (2003). Effector-independent representations of simple and complex imagined finger movements: A combined fMRI and TMS study. *Eur. J. Neurosci.*, *18*, 3375–3387.

Liebal, K., Call, J., & Tomasello, M. (2004). Use of gesture sequences in chimpanzees. *Amer. J. Primat.*, *64*, 377–396.

Liberman A. M., Cooper F. S., Shankweiler, D. P., et al. (1967). Perception of the speech code. *Psychol. Rev.*, *74*, 431–461.

Lieberman, D. E. (1998). Sphenoid shortening and the evolution of modern cranial shape. *Nature*, *393*, 158–162.

Lieberman, D. E., McBratney, B. M., & Krovitz, G. (2002). The evolution and development of cranial form in *Homo sapiens*. *Proc. Nat. Acad. Sci. USA*, *99*, 1134–1139.

Lieberman, P. (1998). *Eve Spoke: Human Language and Human Evolution*. New York: W.W. Norton.

Lieberman, P. (2007). The evolution of human speech. *Curr. Anthro.*, *48*, 39–46.

Lieberman, P., Crelin, E. S., & Klatt, D. H. (1972). Phonetic ability and related anatomy of the newborn, adult human, Neanderthal man, and the chimpanzee. *Amer. Anthropol.*, *74*, 287–307.

Liégeois, F., Baldeweg, T., Connelly, A., et al. (2003). Language fMRI abnormalities associated with FOXP2 gene mutation. *Nat. Neurosci.*, *6*, 1230–1237.

Lin, J-W. (2005). Time in a language without tense: The case of Chinese. *J. Semantics*, *23*, 1–53.

Lindenberg, R., Fangerau, H., & Seitz, R. J. (2007). "Broca's area" as a collective term? *Brain & Language*, *102*, 22–29.

Locke, J. L., & Bogin, B. (2006). Language and life history: A new perspective on the development and evolution of human language. *Behavioral & Brain Sciences*, *29*, 259–325.

MacLarnon, A., & Hewitt, G. (2004). Increased breathing control: Another factor in the evolution of human language. *Evol. Anthro.*, *13*, 181–197.

Marler, P. (1991). The instinct to learn. In S. Carey & B. Gelman (Eds.), *The Epigenesis of Mind: Essays on Biology and Cognition*. Hillsdale, NJ: Erlbaum.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychol. Rev.*, *92*, 350–371.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago, IL: University of Chicago Press,

Mellars, P. (2006). Going east: New genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science*, *313*, 796–800.

Mellars, P. A., & Stringer, C. B. (Eds.). (1989). *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans*. Edinburgh: Edinburgh University Press.

Mithun, M. (1999). *The Languages of Native North America*. Cambridge: Cambridge University Press.

Muir, L. J., & Richardson, J. E. G. (2005). Perception of sign language and its application to visual communications for deaf people. *J. Deaf Studies & Deaf Edu.*, *10*, 390–401.

Neidle, C., Kegl, J., MacLaughlin, D., et al. (2000). *The Syntax of American Sign Language*. Cambridge, MA: The MIT Press.

Nevins, A., Pesetsky, D., & Rodrigues, C. (2007). *Pirahã Exceptionality: A Reassessment*. Available online at http://ling.auf.net/lingBuzz/000411

Nietzsche, F. W. (1986). *Human, All Too Human: A Book for Free Spirits* (Trans. R. J. Hollingdale). Cambridge: Cambridge University Press. (Work originally published 1878).

Noonan, J. P., Coop, G., Alessi, J., et al. (2006). Sequencing and analysis of Neanderthal genomic DNA. *Science*, *314*, 1113–1118.

Parsons, L. M., Fox, P. T., Downs, J. H., et al. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature*, *375*, 54–58.

Perrett, D. I., Harries, M. H., Bevan, R., et al. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *J. Exp. Biol.*, *146*, 87–113.

Petrides, M., Cadoret, G., & Mackey, S. (2005). Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature*, *435*, 1325–1328.

Petitto, L. A., Holowka, S., Sergio, L. E., et al. (2004). Baby hands that move to the rhythm of language: Hearing babies acquiring sign language babble silently on the hands. *Cognition*, *93*, 43–73.

Pettito, L. A., & Marentette, P. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science*, *251*, 1493–1496.

Pietrandrea, P. (2002). Iconicity and arbitrariness in Italian Sign Language. *Sign Language Studies*, *2*, 296–321.

Pika, S., Liebal, K., & Tomasello, M. (2003). Gestural communication in young gorillas (*Gorilla gorilla*): Gestural repertoire, and use. *Amer. J. Primat.*, *60*, 95–111.

Pika, S., Liebal, K., & Tomasello, M. (2005). Gestural communication in subadult bonobos (*Pan paniscus*): Repertoire and use. *Amer. J. Primat.*, *65*, 39–61.

Pinker, S. (1994). *The Language Instinct*. New York: Morrow.

Pinker, S. (2003). Language as an adaptation to the cognitive niche. In M. H. Christiansen, & S. Kirby (Eds.), *Language evolution* (pp. 16–37). Oxford: Oxford University Press.

Pinker, S. (2007). *The Stuff of Thought*. London: Penguin Books.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral & Brain Sciences*, *13*, 707–784.

Pizutto, E., & Volterra, V. (2000). Iconicity and transparency in sign languages: A cross-linguistic cross-cultural view. In K. Emmorey & H. Lane (Eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima* (pp. 261–286). Mahwah, NJ: Lawrence Erlbaum Associates.

Place, U. T. (2000). The role of the hand in the evolution of language. *Psycoloquy*, *11*(7).

Ploog, D. (2002). Is the neural basis of vocalisation different in non-human primates and *Homo sapiens*? In T. J. Crow (Ed.), *The Speciation of Modern* Homo sapiens (pp. 121–135). Oxford, UK: Oxford University Press.

Pollick, A. S., & de Waal, F. B. M. (2007). Apes gestures and language evolution. *Proc. Nat. Acad. Sci. USA, 104*, 8184–8189.

Ray, N., Currat, M., Berthier, P., & Excoffier, L. (2005). Recovering the geographic origin of early modern humans by realistic and spatially explicit simulations. *Genome Res.*, *15*, 1161–1167.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends Cogn. Sci.*, *21*, 188–194.

Rizzolatti, G., Camardi, R., Fogassi, L., et al. (1988). Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Exp. Brain Res.*, *71*, 491–507.

Rizzolatti, G., Fadiga, L., Gallese, V., et al. (1996). Premotor cortex and the recognition of motor actions. *Cogn. Brain Res.*, *3*, 131–141.

Rizzolatti, G., Fogassi, L., & Gallese V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev.*, *2*, 661–670.

Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain*. Oxford: Oxford University Press.

Rousseau, J.-J. (1969). *Essai sure l'origine des langues*. Paris: A. G. Nizet. (Work originally published 1782).

Ruben, R. J. (2005). Sign language: Its history and contribution to the understanding of the biological nature of language. *Acta Oto-Laryngolica*, *125*, 464–467.

Ruff, C. B., Trinkaus, E., & Holliday, T. W. (1997). Body mass and encephalization in Pleistocene *Homo. Science*, *387*, 173–176.

Russell, B. A., Cerny, F. J., & Stathopoulos, E. T. (1998). Effects of varied vocal intensity on ventilation and energy expenditure in women and men. *J. Speech Lang. Hear. Res.*, *41*, 239–248.

Saussure, F.de. (1916). *Cours de linguistique générale*, C. Bally and A. Sechehaye (Ed.), with the collaboration of A. Riedlinger. Lausanne and Paris: Payot; Translated as W. Baskin (1977), *Course in General Linguistics*. Glasgow: Fontana/Collins.

Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, Language, and the Human Mind*. New York: Oxford University Press.

Semaw, S., Renne, P., Harris, J. W. K., et al. (1997). 2.5-million-year-old stone tools from Gona, Ethiopia. *Nature*, *385*, 333–336.

Senghas, A., Kita, S., & Ōzyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, *305*, 1780–1782.

Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.

Skoyles, J. R. (2000). Gesture, language origins, and right handedness. *Psycoloquy*, *11*(24).

Studdert-Kennedy, M. (2005). How did language go discrete? In M. Tallerman (Ed.), *Language Origins: Perspectives on Evolution* (pp. 48–67). Oxford, UK: Oxford University Press.

Suc, J.-P., Bertini, A., Leroy, S. A. G., et al. (1997). Towards the lowering of the Pliocene/Pleistocene boundary to the Gauss-Matuyama reversal. *Quarter. Inter.*, *40*, 37–42.

Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Gen. Soc. Gen. Psych. Mono.*, *123*, 133–167.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral & Brain Sciences*, *30*, 299–351.

Sutton-Spence, R., & Boyes-Braem, P. (Eds.). (2001). *The Hands Are the Head of the Mouth: The Mouth as Articulator in Sign Language*. Hamburg, Germany: Signum-Verlag.

Tattersall, I. (2003). Once we were not alone. *Scientific American*, *13*(2), 20–27.

Thorne, A., Grün, R., Mortimer, G., et al. (1999). Australia's oldest human remains: Age of the Lake Mungo human skeleton. *J. Hum. Evol.*, *36*, 591–612.

Tomasello, M. (2003). Introduction: Some surprises for psychologists. In M. Tomasello (Ed.), *New Psychology of Language: Cognitive and Functional Approaches to Language structure* (pp. 1–14). Mahwah, NJ: Lawrence Erlbaum.

Tomasello, M. (2008). *The Origins of Human Communication*. Cambridge, MA: MIT Press.

Tooby, J., & DeVore, I. (1987). The reconstruction of hominid evolution through strategic modeling. In W. G. Kinzey (Ed.), *The Evolution of Human Behavior: Primate Models* (pp. 183–237). Albany, NY: SUNY Press.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). New York: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist*, *26*, 1–12.

Tulving, E. (2002) Episodic memory: From mind to brain. *Ann. Rev. Psych.*, *53*, 1–25.

Tulving, E., Schacter, D. L., McLachlan, D. R., et al. (1988). Priming of semantic autobiographical knowledge: A case study of retrograde amnesia. *Brain & Cognition*, *8*, 3–20.

Vico, G. B. (1953). *La Scienza Nova*. Bari: Laterza (Work originally published 1744).

Vleck, E. (1970). Etude comparative onto-phylogénétique de l'enfant du Pech-de-L'Azé par rapport à d'autres enfants néanderthaliens. In D. Feremback (Ed.), *L'enfant Pech-de-L'Azé* (pp. 149–186). Paris: Masson.

Watkins, K. E., Dronkers, N. F., & Vargha-Khadem F. (2002). Behavioural analysis of an inherited speech and language disorder: Comparison with acquired aphasia. *Brain*, *125*, 452–464.

Wood, B., & Collard, M. (1999). The human genus. *Science*, *284*, 65–71.

Wray, A. (Ed.) (2002). *The Transition to Language*. Oxford: Oxford University Press.

Wundt, W. (1900). *Die Sprache* (2 Vols.). Leipzig: Enghelman.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least-Effort*. New York: Addison-Wesley.