Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

R. Wes Smith April 16th, 2012

Gestural Origins of Language as an Evolutionary Explanation for the Digital and Analogic Properties of Human Communication

by

R. Wes Smith

Dr. Bradd Shore
Adviser

The Program in Linguistics

Dr. Bradd Shore

Adviser

Dr. Dietrich Stout

Committee Member

Mr. Timothy McDonough

Committee Member

2012

Gestural Origins of Language as an Evolutionary Explanation for the Digital and Analogic Properties of Human Communication

By

R. Wes Smith

Dr. Bradd Shore

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

The Program in Linguistics

2012

Abstract

Gestural Origins of Language as an Evolutionary Explanation for the Digital and Analogic Properties of Human Communication

R. Wes Smith

This thesis explores several lines of evidence supporting the hypothesis that the evolution of manual object manipulation laid the foundations for the development of syntactic grammar. Once biological evolution had yielded a human brain and body capable of supporting language, the capacity for hierarchically organized action and thought involved in complex behaviors like stone tool construction was co-opted to support the feature of compositionality in language. These two abilities – hierarchical action and compositional syntax – were mutually selective in the cultural evolution of the modern mind and co-evolved in an expanding spiral from an initially analogic form of communication into a fully digital and productive linguistic system. Dexterous hands originally adapted for object manipulation came to serve the dual functions of analogic gesture and a digital system of initially manual arbitrary signs, although the truly compositional and productive qualities of language ultimately became restricted to the vocal modality. This process can be corroborated by the evolution of lithic technologies and other behavioral correlates in the archaeological record.

Gestural Origins of Language as an Evolutionary Explanation for the Digital and Analogic Properties of Human Communication

By

R. Wes Smith

Dr. Bradd Shore

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

The Program in Linguistics

2012

Acknowledgements

I would like to thank my committee members, Dietrich Stout and Tim McDonough, for their thorough reading of and constructive criticism on earlier drafts of this thesis. Special thanks are due to my adviser Bradd Shore, whose extensive feedback on each chapter, openness to conversation, and endless support and guidance throughout every stage of this project has been critical to its success.

Table of Contents

Tables and Figures

1. Introduction

Nearly a century and a half ago, the Linguistic Society of Paris established a ban on any discussion of the origins of language in response to widespread speculation on the subject inspired by Darwin's recently published theory of evolution (Hockett, 1960). As a result, most scholars avoided the topic completely for the better part of the twentieth century, dismissing the study of language evolution as futile or disreputable. However, significant advances in the fields of anthropology, theoretical linguistics, and evolutionary biology and psychology have once again made the study of language origins a worthy endeavor, and the last few decades have seen a proliferation of theories (e.g., Chomsky, 1965; Donald, 1991; Pinker, 1994; Dunbar, 1996; Christiansen & Chater, 2008; Arbib, 2005, 2008) and the birth of several new disciplines referred to collectively as the cognitive sciences. This renewed academic vigor has produced many fascinating revelations about the nature and beginnings of language, yet the subject remains as hotly contested as ever before.

The present paper aims to explore the notion of language as an evolutionary extension of a general hierarchical capacity for the complex manipulation and recombination of physical objects. This capacity has its roots in the gradual specialization of the primate hand for efficient object manipulation and necessarily predates modern language. As such, the argument being made here is not about the evolution of language per se, but rather the evolution of one very important "design feature" of language. This modular framework views language as a confluence of specific capabilities rather than a unitary skill and is adopted from Hockett (1960), who astutely pointed out that human language, like any communication system in any species, is composed of many different design features which can be isolated and compared with other features in other systems. For example, one design feature of communication in humans is that it relies primarily on the vocal-auditory channel, as opposed to communication in bees, which relies on dancing. Hockett (1960) identifies thirteen design features of human language, including semanticity, arbitrariness, and productivity. Through comparison with other species and with a little evolutionary reasoning, we find that most features of language are shared with other communication systems, especially those of our closest primate relatives, and were probably present in the communication systems of our common ancestors. According to Hockett (1960), the only features unique to human communication are displacement (being able to communicate about something remote from the communicator in space or time), traditional transmission (explicit teaching and learning), duality of patterning (the existence of words like "tack," "cat," and "act," which are completely distinct in meaning yet consist of different combinations of just three meaningless sounds [p. 92]), and productivity (the ability to produce novel utterances from the composite parts of previous utterances). It follows, then, that these four features developed at some point in the hominid line, and that the emergence of the full suite corresponds to the emergence of what we call modern human language.

This thesis is primarily concerned with just one of these four design features: productivity. Productivity in language is a result of the Principle of Compositionality, which basically states that the meaning of a complex utterance is determined by the meaning of its constituent parts. This is true on multiple levels: individual speech sounds are manipulated and recombined into different morphemes, morphemes into words, words into clauses, and so on all the way up to the level of discourse. In other words, language (or, more specifically, syntactic grammar) is hierarchically organized and subject to the manipulation and recombination of units (of sound and of meaning), yielding endless generativity from a finite set of arbitrary sounds and symbols. These properties are also characteristic of non-communicative, goal-directed human action: dexterous manual behaviors are composed of sub-actions and sub-goals that are manipulated and recombined into ever-larger constituents. For example, stone tool construction involves various sub-actions and sub-goals that are nested within several levels of hierarchical complexity (Stout, 2011). The same is true of something as simple as using a spoon, just with fewer levels of complexity (Greenfield, 1991), or as complicated as preparing a meal. Implicit in this compositional ability is the capacity to reduce the cognitive complexity of a process by categorization and hierarchical clustering or chunking, which facilitates the manipulation of these units in a sequence.

The fundamental properties of conventional language and goal-directed action described above may be attributed to the general digital characteristics of these systems. The term *digital* describes systems that hierarchically combine discrete units in a compositional fashion; in the context of communication, these units are arbitrary linguistic symbols. Digital systems contrast with analogic ones, which are not compositional and represent concepts or intentions iconically or indexically. Human language in the broadest sense lies at the intersection of digital and analogic channels of communication, relying heavily on both conventional linguistic forms – words – and non-conventional forms such as the spontaneous, idiosyncratic gestures people tend to produce as they speak. As analogic communication is common among non-human primates but digital communication is not, a fundamental question in the study of language evolution is how and why analogic communication came to be supplemented by the digital properties of non-communicative action.

This paper considers various lines of evidence for the following hypothesis: the compositional properties of a digital system originally evolved in support of goal-directed manual action became co-opted in the evolution of the hominin line to complement an initially analogic communication system and allow for open-ended symbolic expression. This new system of digital communication initially focused on the hands as a representational modality, but various selective pressures eventually confined this system to the vocal domain and left the hands as a medium of strictly analogic communication (with the exception of sign language). Biological evolution yielded a brain and body equipped with the cognitive capacity for digital and analogic systems and the anatomical structures to support them, and modern humans developed language as we know it today through an extended process of cultural evolution in Africa and subsequently dispersed throughout the world, bringing dynamic linguistic systems with them wherever they went. In supporting this hypothesis, I will draw widely on evidence from the cognitive sciences, theoretical and diachronic linguistics, structural descriptions of human behavior, developmental psychology, physical anthropology, and archaeology, as well as specific theoretical scenarios for the evolution of language.

In the following chapter, I present the phenomenon of embodied cognition as proof that language and thought are grounded in sensorimotor experience and shaped to a large extent by the form of the human body. This establishes the connection between language and action and sets the stage for further discussion of their evolutionary relationship. Chapter 3 describes the communicative function of manual action in modern language – gesture – and explains the distinction between analogic and digital systems in detail. Chapter 4 examines the digital systems of conventional language and goal-directed action and the hierarchical, compositional structure that they have in common, and argues that these systems are different manifestations of the same underlying cognitive capacity.

In Chapter 5, I shift from the description of digital and analogic systems to an investigation of their development in human ontogeny. This chapter demonstrates that language, gesture, and action follow parallel developmental patterns, which suggests interesting evolutionary implications. Chapter 6 presents a useful paradigm for understanding language evolution, which serves as a general framework for the subsequent discussion of specific theories of language origins in Chapter 7. This chapter considers whether digital communication originated in the vocal or the manual modality, and ultimately advocates a specific theory that supports the hypothesis outlined above. The final chapter of this paper examines primarily archaeological lines of evidence that support the view of language evolution developed in previous chapters, and attempts to place temporal and geographic limits on the origin of language based on anatomical evidence as well as the evolution of lithic technologies and correlates of behavior found in the fossil record.

The goal of this paper is to synthesize the findings of these various literatures into a cohesive and possibly novel account of the role of digital and analogic systems and the cognitive capacities that underlie them in the evolution of modern language, and to outline a productive paradigm for future research. This paper does not present any original empirical research or propose specific experimental methods, nor does it fully answer the important question of exactly how and why the analogic communication system of our primate ancestors was augmented with digital properties originally evolved for object manipulation. Those worthy goals will be left to future investigation.

2. Embodiment

This paper places a heavy emphasis on the relationship between language, thought, and action, and uses this connection as a foundation for further reasoning about the evolution of each. If these important capacities can be shown to overlap or rely on each other in significant ways, then it becomes reasonable to assume that they share a closely intertwined evolutionary history and enables researchers to identify appropriate lines of evidence for the study of language evolution. The notion of a deep relationship between body and mind is contrary to the intuitions of Cartesian dualism (i.e., that the mind and body are distinct entities; Hart, 1996) and cognitivism (i.e., that thought is a strictly computational process distinct from behavior; Lilienfeld et al., 2010); however, an understanding of language and other aspects of cognition as being fundamentally shaped by aspects of the body has been established by a large and relatively recent body of work across many disciplines. This chapter examines this theoretical approach, described as embodiment, and uses it to set the stage for a subsequent discussion of the evolution of modern language and cognition.

Over the past few decades, the phenomenon of embodiment has been established as a central aspect of human cognition. Also referred to as grounded cognition (Barsalou, 2010), this body of thought emerged in its modern form during the early 1980's with roots in philosophy, linguistics, psychology, and cognitive science. Over the next thirty years, the theory of the embodied mind continued developing in these areas and spread to such varied disciplines as artificial intelligence and neuroscience, among many others, and a growing amount of research on all of these fronts has led to its increasing acceptance as a fundamental property of human cognition.

2.1. Definition and early descriptions of the phenomenon

At its most basic level, embodiment means that we understand things (consciously or otherwise), including language, in relation to our bodies. That is to say, all aspects of cognition, such as high level mental constructs like concepts and categories and abilities like reasoning and judgment (and hierarchical organization), are grounded in aspects of the body: the motor system, the perceptual system, and our accumulated experiences of interacting with the environment (Lakoff & Johnson, 1999). The way in which the human mind works is based upon and even dictated by the design and function of the human body.

The modern conception of embodied cognition finds its precursors in the work of Heinz Werner & Bernard Kaplan (1963) and Jean Piaget (1983), whose hypotheses about the early development of cognition and symbol use anticipate later descriptions of how language and thought become grounded in sensorimotor experience. Werner & Kaplan (1963) discuss "physiognomized" symbols – linguistic forms that are "midway between onomatopoetic forms, which seem to mirror their referents, and conventional forms which are externally, though not internally, remote from their referents," (207) – and their inherently "organismic" quality – their "embeddedness in, and emergence from, postural-affective-motor states," (208). Words such as these allegedly possess a quasi-substantial, thing-like, pictorial-dynamic quality, and are characterized by "the primordial *lack of distance* between the organismic activity (postural-affective) involved in the bodily schematization of meaning and the meaningful symbol itself," (211; emphasis is original). According to these authors, we derive the meaning of a symbol through physiognomic apprehension of the concept it refers to.[1] While this is a very free and

---

[1] Werner & Kaplan (1963) provide several examples of physiognomized words, as described by German subjects of psychological experiments. For example:

unrestrained process in the development of nonconventional symbols (e.g., communicative gestures, which are largely iconic or analogic), the physiognomization of conventional linguistic forms (which bear an arbitrary or digital relation to their referents) must occur within the more or less pre-determined phonetic constraints imposed by the established language. Werner & Kaplan (1963) were precocious in their description of language as embodied within physical experience, which has since been corroborated by much more recent neurological studies (e.g., Lacey et al., 2012).

Piaget (1983) also attributes an important role to the perceptual and motor systems in supporting cognition. In his comprehensive theory of cognitive development, Piaget (1983) proposes a "sensorimotor stage," lasting from birth through the first two years of human life, in which knowledge is derived from physical experience and interaction with the world. This stage may be divided into several sequential sub-stages, ranging from the initial development of motor reflexes such as sucking and grasping through the emergence of patterned motor habits and hand-eye coordination, a fascination with objects and the ways in which they can be manipulated, playful experimentation with new behaviors and the discovery of new means to meet goals, and ultimately the ability to use basic symbols and form lasting mental representations. The development of object permanence – the understanding that objects continue to exist even when they cannot be perceived directly – is completed during this stage. Although Piaget (1983) is much more concerned with the ontogenesis of cognition in humans

---

braunrot (brownred): "Glancing over it, it's like touching the disagreeable sandy quality of rusty iron." (209)
faul (decayed): "One dips into the word without feeling resistance, like into rotten fruit." (209)
hart (hard): " At the sight of the word I immediately experience a definite 'steel-like' structuring of my body with the center in the back and the neck, particularly strong around the uppermost vertebrae. The structure coalesces fully with the visual structure of the word. (211-212).

(which will be addressed in some detail in Chapter 5) rather than its embodied nature, his theory provides insight into the way in which thought and language are scaffolded by sensorimotor knowledge and experience.

2.2. Current approaches to embodiment

The early descriptions of the embodied mind outlined above may be generally described as *metaphoric* and *sensorimotor*, respectively, although each incorporates physiognomic and developmental aspects as well. They are important precursors of modern approaches to embodied cognition, and their legacy can be observed in a wide range of much more recent work on the topic. The following sections examine several such approaches, all of which may be described as essentially metaphoric or sensorimotor in orientation but variously incorporate added dimensions such as the role of emotion or simulation in cognition. Each approach presented in this chapter provides a useful framework for understanding the relationship between language, thought, and the body. The metaphoric perspective is more concerned with meaning and semiotics, however, which are somewhat beyond the scope of the present discussion; for the purposes of this paper, it is more useful to conceive of cognition and linguistic structure as being grounded in the body and sensorimotor experience.

2.2.1. Embodiment in cognitive linguistics

Embodiment was first established as a linguistic phenomenon in 1980 with Lakoff and Johnson's description of conceptual metaphor. As is evident from the term, this is a metaphoric approach to embodied cognition that is closely allied with the physiognomic apprehension of the symbol described by Werner & Kaplan (1963). They argued that a great deal of everyday speech is metaphorical, in the sense that we consistently and systematically use ways of thinking about one conceptual domain to understand other conceptual domains. This framing or mapping across various abstract and concrete domains is called conceptual metaphor, and is informed by the nature of our bodies, how we interact with the environment and experience the world, and our social and cultural dispositions. For example, certain words that have literal meanings in one domain (i.e., words about vertical direction such as *up*, *down*, *rise*, *fall*, etc.) take on a more abstract, figurative meaning when systematically mapped to another domain such as quantity (2003 Afterword, p. 247-248). Thus, the conceptual metaphor MORE IS UP allows us to understand the abstract notion of increasing prices as "rising" by framing it in terms of the concrete physical domain of verticality. Importantly, this metaphorical relationship is determined by our bodily experience with the environment that informs us that when something physically rises, it is higher than it was before; this knowledge about physical height is applied to the separate domain of amounts when we describe prices as rising. Another example is the conceptual metaphor LOVE IS A JOURNEY, which allows us to understand the development of a relationship in terms of travelling along a physical path (e.g., "We're at a *crossroads*," "We'll just have to *go our separate ways*," "It's been a *long, bumpy road*," etc. [p. 44-45; emphasis is original]). In this way, the majority of language and abstract thought involves mapping across conceptual domains and is embodied, or grounded in bodily experience.

2.2.2. A broader view: basic sensorimotor grounding of experience

Lakoff and Johnson's contributions are no doubt fascinating and have wide ranging implications for linguistics and cognitive science and many other fields, but conceptual metaphor is a highly semantic phenomenon and not entirely relevant to the interests of the present paper. It is linked to an overall understanding of embodied cognition and does serve as an excellent illustration of this phenomenon, but it is very specifically concerned with the interaction of meaning across conceptual domains. The focus here is more on the general fact that language relies on the cognitive circuitry of bodily action and perception (e.g., the motor system and the perceptual system), rather than the complex semantic web that is built on top of this circuitry. For example, the evolution of speech organs was not always driven by the need for communication, but rather bootstrapped onto body parts originally adapted for other purposes (e.g., eating and breathing). Similarly, the evolution of certain aspects of language (the rapid manipulation and recombination of units in particular) may have co-opted cognitive structures underlying our ability to physically manipulate objects and perceive the actions of others as compositional and hierarchically organized.

There are many examples of how language and thought are directly grounded in the motor system. In an experiment investigating the role of embodiment in emotion and language comprehension, Glenberg et al. (2005) demonstrated that forcing participants to assume facial expressions associated with specific emotional states facilitated comprehension of sentences corresponding to those emotions. Specifically, participants forced to engage the facial muscles of a smile by holding a pen between their teeth were quicker to understand pleasant sentences than unpleasant ones, while those forced to engage the muscles of a frown by holding a pen between their lips understood unpleasant sentences more readily than pleasant ones. The same two conditions also affected participants' perception of emotionally charged stimuli: smiling participants rated cartoons as funnier than did their frowning counterparts. In a similar experiment, Havas et al. (2010) used botox injections to temporarily paralyze an important facial muscle engaged in frowning before presenting participants with sentences designed to evoke a negative emotional response that would normally elicit a frown. The experimenters found that participants took significantly longer to comprehend angry and sad sentences when their frowning muscles were paralyzed than they did before receiving the botox injections, and that the botox injections had no effect on the comprehension of happy sentences. Experiments such as these are reminiscent of a theory of emotion developed independently by William James and Carl Lange in the late 1800's, which posits that emotion is the mind's perception of physiological changes (e.g., rising heart rate, sweaty palms, muscular tensions, etc.) that result from some stimulus (Ellsworth, 1994); thus, according to James, we do not see a bear, fear it, and then run, but rather see the bear, run from it, and fear it as a consequence of that response. This theory and the experiments described above demonstrate the remarkable degree to which emotion – an important aspect of cognition – and language comprehension are grounded in the motor system. Human cognition and language are determined, at least in part, by innately programmed automatic motor responses that correspond to emotional states[2].

It is somewhat difficult and perhaps misguided to pull apart the motor and perceptual systems when discussing their role in grounded cognition. The motor system and perception are in fact deeply intertwined, and explaining embodiment in such a way creates an unnecessarily sharp distinction between the two; it is more productive to explain cognition and language as being grounded in sensorimotor experience. Indeed, Pulvermüller and Fadiga (2010) conclude from extensive neuroimaging data that perception and comprehension of environmental stimuli rely on neuronal circuits for motor action, and that language processing is based on these interdependent sensorimotor circuits. In the specific case of language processing, these circuits are actually formed during the learning process, as the production of speech sounds through motor circuits results in an auditory stimulus that is subsequently processed by the auditory cortex and thus fed back into the action-perception system, reinforcing and increasing the

---

[2] There is an evolutionary explanation for this: emotion helps drive adaptive behavior. For example, strong feelings of compassionate love that cause males and females to maintain a prolonged monogamous relationship will be beneficial to the survival of their offspring, thus conferring a selective advantage (Kaschak et al., 2009).

salience of the connection between motor output and perceived stimulus. Sensorimotor circuitry is engaged not only at the phonological level of language processing but the semantic level as well, as evidenced by lesion experiments in which the impairment of specific action-perception circuits resulted in impaired recognition of specific semantic word categories. Furthermore – and most relevant to the interests of this paper – Pulvermüller and Fadiga (2010) assert that syntactic processing of grammatical sentences is grounded in the sensorimotor circuits of the perisylvian cortex, which is also implicated in processing the more general "syntax" of sequential, hierarchically organized, intention-driven action. The fact that language processing at all levels – phonological, semantic, and syntactic – as well more general programs of intentional combinatorial action rely on congruent action-perception circuits constitutes overwhelming support for the argument that cognition and language are embodied and situated in sensorimotor experience.

2.3. The role of simulation

The fundamental role of simulation is implicit in the notion of an embodied mind. We are able to comprehend visual and auditory stimuli because they trigger activation in sensorimotor circuits conditioned by prior experience with similar stimuli; essentially, we process and understand external stimuli by simulating those stimuli in our own minds, grounding them in our own sensorimotor experience. Barsalou (1999) argues that individual perceptual experiences activate sensorimotor patterns that are stored by brain association areas in a bottom-up fashion, and as similar perceptual experiences accumulate, salient aspects of these experiences are recalled from association areas in a top-down manner to partially simulate the original sensorimotor activations; the aggregate of these partial simulations of similar sensorimotor experiences are called perceptual symbols. These symbols are constantly organized and re-organized into common frameworks, which constitute a conceptual system that can support categorization and categorical inferences (i.e., conceptual metaphor) and represent individual types through sensorimotor simulation.

According to the motor theory of speech perception (Liberman & Whalen, 2000; Galantucci, Fowler, & Turvey, 2006), sensorimotor simulation of perceptual stimuli enables humans to meet the parity requirement necessary for successful communication between speaker and listener: the perception of speech sounds produced by another results in a motor simulation of the articulatory movements previously associated with that particular phonetic representation, grounding the stimulus in bodily experience and making possible its comprehension. Although this theory's initial claim that speech processing is carried out by a specialized module no longer seems plausible, the idea of speech perception relying on the motor system has received strong support in recent years from the discovery of mirror neurons (Rizzolatti & Arbib, 1998) in the brains of monkeys that fire either when the monkey grasps or manipulates objects or when the monkey observes another monkey (or human experimenter) making similar actions. The mirror neurons in monkeys were found in the ventral premotor cortex (area F5), an area considered homologous to Broca's area in the human brain (an area long considered to be heavily involved in the processing and production of language, and situated within the perisylvian cortex mentioned by Pulvermüller and Fadiga [2010]); although it is generally not possible to study the behavior of individual neurons in the human brain, structural similarities between monkey and human brains suggest that neural mirroring systems are also present in humans. These neurons likely represent a system for matching observed actions to internal motor simulations, which would have huge implications for understanding the actions and intentions of others, imitation, and learning. (These and other important prerequisites for language will be addressed in a later section). In addition, further studies have discovered audiovisual mirror neurons in the same monkey homolog for Broca's area in humans that fire when monkeys perform an action, observe that same action in another, or hear a sound associated with that action (e.g., the distinctive sound made by tearing a piece of paper) (Kohler et al. 2002). The presence of a mirror neuron system mapping perceived audiovisual stimuli to the sensorimotor cortex goes a long way in explaining the phenomenon of grounded cognition, and has obvious significance in the relationship between action and language. Indeed, many scholars (e.g., Arbib, 2005, 2008, 2011; Corballis, 2010) have used the mirror neuron system as a foundation for entire theories on the evolution of language out of action.

2.4. The importance of embodiment

Why is embodiment important to the present argument? The fact that cognition and language are grounded in and shaped by sensorimotor circuits of the brain that are established and reinforced through repeated interactions with the environment – in other words, bodily action and experience – demonstrates a concrete relationship between action, thought, and language. This empirically established connection may now serve as a foundation for further arguments about the evolutionary relationship between these three central aspects of humanity. Given the combinatorial, hierarchical properties shared by action and language and the fact they both rely on congruent circuits in the brain, might not the two be separate manifestations of the same underlying cognitive ability? Sequential, intention-driven action is undoubtedly older than the human capacity for language: complex action is shared to a large degree with the other great apes, although no species possesses a communication system that even begins to approach the complexity and productivity of human language. Is it not a reasonable assumption that language is an evolutionarily recent extension of a more ancient and fundamental capacity for digital and hierarchical cognition? Chapter 4 explores the comparative structure of language and action in detail, so that a more informed assessment of their evolutionary relationship can be made later on. However, it will be useful to present first another important aspect of language: the predominantly analogic medium of gesture. The following chapter describes the nonconventional, intuitively expressive ways in which we use our hands to communicate, and presents one side of the fundamental digital/analogic distinction within language.

3. Hands and Language

We humans do some very complex things with our hands: we create and use tools, type on keyboards, play musical instruments, and even perform neurosurgery. In addition to those precise technical tasks, we also communicate by gesticulating as we speak, using conventional symbols like the thumbs up or the middle finger, or even foregoing speech altogether through the use of manual sign language. While the uses to which we put our hands may vary greatly, none of them would be possible without a high level of manual dexterity. This useful trait is a hallmark of the primate line, and has been developed most robustly in the hominin branch of that family. Manual dexterity is a fundamental prerequisite to both praxis – that is, skillful intentional action on objects – and communicative gesture, both of which share essential properties with spoken language. As such, the capacity for complex manual articulation and action is likely to have played an important role in language evolution, manifesting itself in the design features of human communication.

I will argue here that the human hand has in fact affected the evolution of language in two very different ways: first, the general ability to dynamically move and position the hands provided the physical capacity to support the later emergence of communicative gesture; second, the goal-directed actions (i.e., praxis) for which fine motor ability and precise manual articulation originally evolved have a hierarchical organization that was co-opted for linguistic structure. It is important to note that these two different aspects of language started out as closely related phenomena – that is, they are both founded in action, albeit different aspects of action – but diverged significantly as language evolved as a system. Crucially, gesture (and other paralinguistic phenomena such as facial expression, whole-body posturing, and all of the other indexes of communicative intent that are collectively referred to as “body language”) is an analogic form of communication not unlike the forms of communication observed among other primates. The major development in the evolution of human communication was the invention of a digital system of representing information; it is this aspect of language that makes it so distinct from other forms of animal communication. Language as a whole combines both analogic and digital aspects that are fundamentally different in form yet highly complementary, allowing its users to express their thoughts more fully than would be possible with either system on its own. What is interesting is that, in the grand evolutionary scheme of things, both analogic and digital communication (initially) found their expression in the hands rather than the vocal modality.

I will present the communicative function of manual action – the analogic medium of gesture – in some detail here, and return to it later in considering possible scenarios for language origins. I will discuss the hierarchical structure and digital characteristics of language and action in the following chapter.

3.1. The communicative function of manual action: gesture

The dexterous motor abilities underlying praxis are put to a communicative purpose in gesture. Although it is often ignored in formal linguistic analyses of language structure, everyone knows from experience that our hands inevitably become a part of the conversation: without thinking, we automatically wave our hands about as we talk, stabbing in the air to emphasize a point, pointing to direct someone's attention to something, and often mimicking in some intuitive representational way the events being discussed. Hand movements of this sort are very closely synchronized with speech, conveying relevant information to the topic at hand that often cannot be easily expressed in words. Kendon (1988) refers to this special medium of communication as "gesticulation," McNeill (1992, 2005) prefers "gesture," and Goldin-Meadow

(2003) uses the term “co-speech”. In face-to-face communication, gesture and speech form an integrated and complementary system, each encoding unique semantic content into the overall message.

In an experimental study, Skipper et al. (2009) used functional magnetic resonance imaging to demonstrate the integrated nature of the gesture-language system and the degree to which co-speech gestures add complementary and often disambiguating semantic information to the intended message. Citing the general fact that being able to see the hand movements of a speaker facilitates language comprehension, these authors looked at brain activity when spoken language was accompanied by a variety of manual actions, some communicative and some not. They found that language accompanied by co-speech gestures, but not other types of hand movements, produced strong activity in and between brain areas thought to mediate semantic aspects of language as well as areas involved in the planning and execution of motor activities. These results suggest that the motor system and neural circuits for language comprehension work together to interpret the meaning of co-speech gestures, and that the brain flexibly and opportunistically uses contextual non-verbal information to facilitate language comprehension in face-to-face communication. These findings are also strong evidence pointing to the embodied nature of language, discussed in the previous chapter.

Gestures of the sort described in Skipper et al.’s (2009) study contrast with other types of hand movements called *adaptors* and *emblems* (Goldin-Meadow, 2003). Adaptors are habitual, previously learned motions that are also produced unwittingly but serve no communicative function, such as playing with one’s hair or stroking one’s chin[3]. Emblems, on the other hand,

---

[3] While adaptors are not usually intentionally communicative, they still have communicative significance in a discourse setting. An absent-minded use of an adaptor serves an indexical function for the interpreter, letting him or her know whether the conversation partner is engaged

are culturally conditioned hand movements like the “thumbs up” or the “okay” sign that are consciously and intentionally used to communicate with others. Gestures lie on a continuum between these two extremes, and are interesting because they are at once intentionally communicative and yet unrestrained by a codified system, being created at the moment of speaking and free to take highly creative forms that can often express what conventionalized language cannot (Goldin-Meadow, 2003).

These distinct types of manual articulation are further manifestations of human dexterity, and exist along what McNeill (1992:37) has called Kendon’s Continuum:

Gesticulation → *Language-like Gestures* → *Pantomime* → *Emblems* → *Sign Languages*

This continuum, adapted from Kendon’s (1988) classification of manual forms of communication, ranges from spontaneous, idiosyncratic gesticulation on the left extreme (i.e., McNeill’s “gesture” and Goldin-Meadow’s “co-speech”) to formal sign languages on the right, with intermediate forms such as emblems and pantomime falling somewhere in the middle. As one moves from left to right, manual communication becomes increasingly independent of speech, taking on language-like, socially regulated forms and eventually supplanting the spoken word altogether.

3.2. Types of gesture

McNeill (1992) further divides the left-most end of Kendon’s continuum into four types of gesture, ranging from relatively transparent representational motions to more opaque abstract ones. *Iconic* gestures bear an obvious relationship to the semantic content of the speech they are coordinated with. For example, while uttering the phrase, “and he bends it way back,” (1992:12) a speaker simultaneously produced an iconic gesture of gripping an imaginary tree branch and

---

in the discussion, uncomfortable, etc. This information becomes very important in the pragmatic interpretation of an utterance (Bradd Shore, personal communication).

pulling it back in space. Another example of an iconic gesture is a child making a twisting motion in the air while saying, "I can't open this jar," (Goldin-Meadow 2003:6). These types of gesture are fairly transparent depictions of the semantic message conveyed by speech, although their meaning would not be apparent if executed without the utterance. Importantly, iconic gestures and speech present the same message in overlapping but not entirely congruent ways; they refer to the same event, but describe different aspects of or perspectives on it. For example, McNeill (1992:13) describes a speaker who said, "she chases him out again," – conveying the ideas of pursuit and recurrence – while iconically gesturing as if he was swinging an object through the air – depicting the umbrella used as a weapon in the act of chasing. Speech and gesture work together in a complementary way to give a more complete understanding of the speaker's communicative intention.

When describing an abstract idea rather than a concrete event or object, speakers often produce *metaphoric* gestures. These gestures are similar to iconic gestures in that they are pictorial in content, but they correspond to a concept that cannot be represented visually in a clear way. For example, when describing a particular cartoon – "it was a Sylvester and Tweety cartoon" – the speaker's hands raised up to offer the listener an "object" (McNeill, 1992:14). This metaphorical object image corresponds to the abstract notion of a specific genre of cartoon, presenting it as a concrete object that exists in space and can be offered to the listener; this type of gesture is analogous in many ways to the conceptual metaphors in spoken language described by Lakoff and Johnson (1980) and further demonstrates the embodied nature of language. Goldin-Meadow (2003:7) provides another example, mentioning that, when describing algebra word problems, adults typically produce gestures that are either smooth and continuous or disjointed and consist of several discrete movements. The type of gesture produced indicates whether the speaker conceptualizes the problem as one of gradual continuous change or composed of several discrete steps.

Two other types of gestures, *beats* and *deictics*, are not pictorial but still complement the act of speech with nuanced semantic information (McNeill, 1992). Beat gestures are rhythmically synchronized with speech – short, quick movements that punctuate the discourse and index the words they co-occur with as significant. Beats tend to have the same form regardless of semantic content, but still add meaning to speech by stressing particular words as important. Deictic gestures are instances of pointing to indicate objects, people, and locations in the real world, whether or not they are physically present at the time of speaking. Speakers may point to a physical entity (e.g., a chair or another person) while they are referring to it in speech, but deictic gestures are used just as often in a metaphoric way to refer to an entity that is removed in space and time or even an abstract concept that does not exist in the physical world.

3.3. Gesture conveys information in a way fundamentally different from speech

While gestures of the types just described are highly efficient vehicles for conveying meaning, their way of doing so is fundamentally different from the spoken language with which they are coordinated. Perhaps the most essential characteristic of all linguistic systems is hierarchical structure: language segments and linearizes meaning into a compositional string of units. As Saussure ([1916] 1959) noted, language operates in this way because it is restricted by the dimension of time, and is forced to communicate complex ideas by compartmentalizing meaning into discrete units that can be combined and strung together into a temporal sequence. McNeill (1992) stresses that gesture is completely different, describing it as “global” and “synthetic,” and cites a number of ways in which the gestural modality of communication contrasts with the spoken one. First and foremost, gesture is not compositional. A single gesture cannot be broken down into meaningful constituent parts as a sentence can be broken down into words or a word into morphemes, nor can multiple gestures produced together combine to form a more complex gesture with compositional meaning. Even when multiple gestures are produced within the same sentence or clause, each is a flat whole that represents or emphasizes in a unique way multiple semantic dimensions of what is being said verbally and cannot be analyzed at a lower level (i.e., gestures are not “synthetic”). This is not to say that gestures cannot be described hierarchically – they can, on the motor level (e.g., multiple articulations and movements of the hand are often combined in a single gesture) – but this level of hierarchy is analogous to the phonemic composition of morphemes rather than the morphological composition of words or syntactic composition of utterances. It is the gesture as a whole that gives meaning to its individual parts, and not the other way around (i.e., gestures are “global”).

Gesture also contrasts with language in other ways. Unlike language, gestures are not subject to standards of form. There are no conventionally established rules for forming gestures that must be followed in order for them to have communicative effect; gestures of the sort discussed here are by nature idiosyncratic, spontaneous expressions of what the speaker considers relevant or salient for the discussion. This definition is resonant with Werner & Kaplan’s (1963) description of the physiognomization of nonconventional symbols: because the symbolizer is not limited by standards of form, he or she is free to choose whatever aspects of the referent seem most salient as a model from which to create the symbolic vehicle. The gestures of people speaking the same language will be just as distinct from one another as the gestures of people speaking different languages (McNeill, 1992:22). Gesture also lacks the duality of patterning that is inherent to linguistic systems (cf. Hockett, 1960). While there is an arbitrary correspondence of form and meaning in words, the meaning of a spontaneous gesture (but not an emblem, which may be iconic to a degree but is akin to conventionalized onomatopoeia in spoken language) is necessarily a consequence of its form – the two cannot be separated (McNeill, 1992).

It is perhaps because gesture and language convey meaning in such diametrically opposed ways that they complement each other so well and form an integrated system: language provides the necessary structure and conventions to communicate in an unambiguous way, and gesture allows for a degree of unrestrained creativity that enables the speaker to express nuances of meaning that cannot easily be put into words. The two modes of communication clearly share a long and complex evolutionary relationship, the exact nature of which will be examined in the final chapters of this paper.

3.4. Analogic versus digital channels of communication

Gesture and speech convey information so differently because they differ in their formal character: gesture is an analogic system of communication, while spoken language (and sign language) is largely digital. Essentially, digital communication relies on conventional, externally arbitrary symbols that lend themselves to combination. "In analogic communication, however, real magnitudes are used, and they correspond to real magnitudes in the subject of discourse," (Bateson, 1972:373). Analogic communication is non-arbitrary, bearing direct relation in some iconic or indexical way to the topic of discussion, and grounded in sensorimotor perception and bodily experience of the point being conveyed. To use Werner & Kaplan's (1963) terms, analogic communication is a process of unconstrained physiognomic symbolization in which the communicator draws directly from perceptible qualities of the referent to describe it in a transparently motivated way.

This definition suits gesture as described by McNeill (1992) perfectly. Iconic gestures bear an obvious pictorial resemblance to the semantic content of the utterances they co-occur with, but other forms of gesture are equally analogic: beats visibly punctuate the discourse and mark certain words as important, and deictics indicate a clear and transparent point of reference. Metaphorics also demonstrate in a non-arbitrary way the most important aspects of the verbal message (e.g., in describing a “global” phenomenon or a “broad” category, one is likely to make correspondingly big gestures, with arms outstretched and hands held wide apart). Importantly, gesture is only one of many possible channels for analogic communication: facial expression, whole-body posturing, and the intonation, pitch, and volume of one’s voice are equally direct and non-arbitrary means of expressing communicative intent. As anyone who has ever attempted to communicate with someone who speaks a different language can report, analogic systems are more basic and universal than digital ones. When a common linguistic (i.e., digital) system is lacking, humans revert to more basic means of expression such as gesture, pantomime, and the other paralinguistic (i.e., analogic) channels just mentioned above.

Analogic communication is not unique to humans, but rather a vestigial form of the communication systems of all primates and indeed all land mammals. In any such communication system, all messages are fundamentally expressions of the relationship between the entities in conversation (Bateson, 1972). When a mother dog weans her pup by biting its neck, she is communicating about who is dominant in the relationship. Similarly, when one chimpanzee grooms another, it is reinforcing the social bond of trust and reciprocity between them. This communication is analogic, because it demonstrates in a very concrete, physical way the nature of the communicators’ relationship. In human communication, paralinguistic messages are also used to indicate characteristics of the relationship between the communicators (e.g., the familiarity of the speakers, their status relative to one another, their respective emotional states, etc.). This seems to be a basic feature of mammalian communication, and one can assume that it was present in the communication of human ancestors at the origin of language. The fundamental difference between human communication and that of other animals[4], then, is that humans employ a digital system as well as an analogic one; this raises important evolutionary questions that I will return to in Chapter 7.

The most important characteristic of digital systems is their combinatorial quality, which allows for the productive manipulation and combination of units to form compositional structures. Analogic systems lack this quality, as analogic expressions are complete unto themselves and cannot be broken down into meaningful constituent parts or combined with others to express more complex messages. (For example, the act of grooming between two chimpanzees mentioned above is a simple indicator of trust and friendship, and cannot be combined with another analogic expression, such as pointing to a third chimpanzee, to communicate a more complex message such as, "You should go and groom that fellow over there.")[5] In contrast, digital linguistic systems that utilize conventionalized arbitrary symbols are able to endlessly produce meaningful new messages through the combination and manipulation

---

[4] The communication of dolphins and whales is a potential exception, as adaptation to life in the sea has stripped their bodies of the expressive organs used by other mammals as means of analogic communication. Instead, these animals communicate largely through vocalizations, the meanings of which are entirely opaque to humans. This may prove that aquatic mammals have also developed (at least partially) digital systems of communication; given the phylogenetic distance between these species and humans, however, this would necessarily be a case of convergent evolution rather than a derived feature (Bateson, 1972).

[5] While humans may be able to gesturally convey more complex, "compositional" messages in the form of pantomime (e.g., presenting an object to someone while nodding in someone else's direction, with the intended message that they should present that object to the third person), this is a more stylized form of communication that is somewhere between analogic and digital (Tim McDonough, personal communication). Indeed, this intersection of analogic and digital channels uniquely characterizes human communication, and pantomime may have played an important role in the evolution of modern language (see Chapter 7).

of discrete units. While arbitrariness and convention are necessary characteristics of linguistic systems, however, this is a unique requirement of their communicative function that is not imposed on other digital systems. An example of a non-communicative digital system is praxis – goal-directed manual action on objects (e.g., the chimpanzee practice of fishing for termites with a stick, or the manufacture of stone tools by humans). Praxis involves the sequential combination of discrete units of action to meet specific goals, which are then combined on a higher level to meet more general goals and ultimately achieve some desired purpose (such as creating a stone tool). These units of action are by no means subject to conventional standards of form, nor are they arbitrary symbols, yet they are the building blocks of productive sequences of action.

Although digital praxis differs from analogic gesture in that it lacks a communicative function, they both rely on the use of the hands. However, each system uses the hands in different ways: gesture involves the dynamic movement and positioning of the whole hand and perhaps the pointing or vague wiggling of the fingers; praxis, on the other hand, requires the precise control and manipulation of individual digits (hence the term "digital")[6]. While gesture does require a certain degree of dexterity, it operates holistically and the exact articulation of each finger is not crucial to the overall communicative effect. Praxis, on the other hand, depends very much on fine digital manipulation for the success of a given action. This is also true of the one form of digital communication that relies on the hands instead of the vocal modality: sign language. While the gestures of hearing people convey information holistically (analogically) through fairly coarse movements of the hands, sign languages of the deaf are capable of complex, nuanced (digital) communication through the fine movement of individual fingers.

---

[6] I owe this insight about the different roles played by the hands in gesture and in praxis to Bradd Shore, who pointed out this distinction in personal communication.

While this does not necessarily mean that the earliest forms of digital communication in humans resembled modern sign languages, it does suggest that the manipulative abilities of the hands played a crucial role in the evolution of modern human language, which added a digital system to the analogical communicative repertoire shared with other animals. This point is not lost on Bateson (1972:371), either, who, in discussing the digital properties of human language, makes the following statement: "I do not think that any animal without hands would be stupid enough to arrive at so outlandish a mode of communication."

A crucial question in the study of language evolution, then, is how humans transitioned from purely analogic communication to a larger system that incorporates digital properties as well. The vital role of the hands in both analogic communication and digital praxis – both of which are shared in some form with other primates – points to their important evolutionary role in the development of a digital system of communication. The exact nature of this role remains unclear, however, and will be returned to in the final chapters of this paper. At present, it will be beneficial to delve a little deeper into both the communicative and non-communicative digital systems mentioned above to better understand the one property that is fundamental to both: manipulative combinatoriality and hierarchical organization. The following chapter examines this general hierarchical capacity in greater detail, describing its manifestation in language and praxis as well as the common brain structures shared by both.

4. The Parallel Structure of Language and Complex Action

In the previous chapter I described communicative gesture, one of many important functions of the human hand, in some detail, arguing that this medium conveys information differently than spoken or even signed language. Gesture is a highly idiosyncratic, holistic form of communication that bears a direct and fairly transparent relation to the semantic content of a given message – in other words, it is fundamentally analogic in nature. In contrast, formal language is subject to conventional standards of form and relies on shared symbols that denote concepts arbitrarily, which can be combined in specific ways to yield complex compositional meaning – in this way, formal language is fundamentally digital. This chapter focuses on that digital property and its expression both within and outside the linguistic realm.

Digital systems lend themselves to hierarchical organization, for which humans have a remarkable capacity. This ability is demonstrated by the syntax of language as well as by our penchant for complex, goal-directed action. Both of these skills are characterized by compositionality, involving the combination of units (of either language or action) on one level to form constituents that may be treated as composite units to be combined on a higher level of hierarchical complexity. These basic units can be recursively combined and manipulated in a variety of ways, yielding infinite possibilities from a discrete set of building blocks. I argue that these two distinct abilities – language and complex manual action – share these structural similarities because they are in fact different manifestations of a single underlying cognitive capacity.

4.1. A note about hierarchical behavior in general

Before delving into the parallel structures of language and action, it is necessary to qualify the following discussion of hierarchically organized behavior. Many aspects of human society and behavior are hierarchical in a general sense; in fact, virtually any sequential act can be described in this way, depending on the level of analysis. For example, the physical act of walking can be understood as a hierarchical process on the motor level (i.e., bipedal locomotion is the combined result of the coordinated contraction and expansion of several muscles, which are in turn produced at a more basic level of patterned activations of motor circuits in the brain), and the complex movement of large crowds can likewise be explained in terms of a hierarchy of crowds, groups, and individuals (Musse & Thalmann, 2001). In this paper, however, I am concerned with an intermediate level of hierarchical behavior lying somewhere between these extremes of micro- and macro-level organization.

This chapter draws a specific analogy between language and goal-directed manual action, which both possess similar ranges of hierarchical organization. At the lowest levels of these systems are the basic motor processes that are responsible for the phonological production process and for the fine dexterous manipulation involved in articulating the hands and fingers to act on objects; the simple grasps and hand motions of manual action are essentially similar to the phonemes of language. These basic units are combined at a higher level to form meaningful compounds of sound and achieve low-level goals within an action hierarchy; individual words and morphemes are comparable to simple sub-actions within a more complex behavior (e.g., grasping a knife as part of the larger activity of slicing an apple). Complete syntactic utterances appear at the next level, as do more complete actions like slicing the apple. The highest level of hierarchy in language is probably that of discourse, which can be compared with more temporally extended activities that combine several distinct actions (e.g., preparing a meal). While many behaviors can be described hierarchically at some level of analysis, language and action are remarkably similar in organization and seem to be governed by the same or at least significantly overlapping cognitive processes.

4.2. The structure of language

In the late 1950s, Noam Chomsky began to formulate a generative theory of language (1957, 1965) that initiated a major shift within the field of linguistics and established the study of the underlying structure of language as a central focus of the discipline. This approach operates on the principle that there is a finite set of rules that govern the ways in which different types of words can be combined to form grammatical expressions (i.e., expressions that could be uttered and understood by native speakers of a given language). Chomsky (1965) uses the term *linguistic competence* to refer to the full range of possible constructions in a language, all of which can theoretically be generated by such a system of underlying rules. According to him, the primary goal of linguistics is to deduce the rules of linguistic competence from the data of actual performance.

A central and enduring aspect of Chomskyan frameworks of syntax is the distinction between two different representations of a possible linguistic construction: deep structure (D-structure) and surface structure (S-structure). The D-structure of a given sentence is an internal representation in the mind of a speaker-hearer generated by syntactic rules, while the S-structure is the construction as it is uttered in actual performance; the S-structure may or may not match the D-structure exactly, and several different S-structures may correspond to a single D-structure. For example, consider the following D-structure and some of its possible S-structures:

(1) D-structure: John will go to the store tomorrow.

(2) S-structure: John will go to the store tomorrow.

(3) S-structure: Tomorrow, John will go to the store.

(4) S-structure: Will John go to the store tomorrow?

Each of these S-structures contains the same essential constituent parts, but some of them have been manipulated and reorganized to achieve slightly different communicative functions. These various possible S-structures are connected to the D-structure by specific transformation rules that form part of the underlying grammatical structure. This general approach to generative grammar is known as transformational grammar, and Chomsky has revised it many times (e.g., 1981, 1995) since its initial formulation.

Chomsky's current model, known as the Minimalist Program (1995), emphasizes the highly economical nature of syntax in natural languages. This principle of economy minimizes computational demands on the brain during language processing and production. Accordingly, the Minimalist Program proposes just two fundamental transformation rules in the generation of grammatical constructions: Merge and Move. Merge places constraints on the ways in which syntactic units (either individual lexical items or higher-order constituents) can be combined to form new syntactic units. Merge operates at multiple levels: within a single lexical item, as in the case of inflection (e.g., the morpheme "dog" is combined with the inflectional morpheme "-s" through Merge in order to denote plurality); at the phrase level, as in the combination of individual lexical items into larger syntactic units (e.g., the noun "dogs" can be combined with the adjective "stray" and the determiner "the" to form the noun phrase "the stray dogs"; likewise, the verb "chased" can be combined with the noun phrase "the alley cats" to form the verb phrase "chased the alley cats"); and at the level of a full sentence, which involves the combination of at least a noun phrase and a verb phrase (e.g., the phrases mentioned above can be combined through Merge to produce the full sentence, "The stray dogs chased the alley cats"). The Merge operation is responsible for generating the D-structure of a sentence.

Move, on the other hand, refers to the set of possible operations to derive an actual, uttered S-structure from an underlying, internal D-structure. Essentially, the Move operation allows for any constituent (i.e., a word or group of words that functions as a single syntactic unit within the hierarchical structure of the sentence) to be relocated to a different position in a sentence. It is in this way that we can derive the S-structures (3) and (4) mentioned above from the single D-structure (1). For example, (3) is produced by moving the constituent "tomorrow" from its original position in the D-structure to the front of the sentence. In certain cases such as question formation, Move can also operate on a specific part of a constituent known as the "head" of the phrase (e.g., the movement of "will," which is not itself a full constituent, in (4)). This ability to manipulate and reorganize hierarchical structures is a hallmark of human language as well as more general types of structured behavior.

4.2.1. X-bar theory

In order to more fully comprehend the hierarchical nature of sentences produced by Merge and manipulated by Move, it is helpful to implement a system of notation for representing these structures visually. There is actually a variety of different notation systems used to represent syntactic structure, each with proponents who argue for its superiority for one reason or another; despite their differences, however, all such systems (at least within the tradition of transformational grammar) are fundamentally similar in that they are constituency-based (i.e., compositional), and any one will suffice here as a demonstration of that hierarchical structure. One of the most widely used frameworks is called X-bar theory[7], proposed by Chomsky (1970)

---

[7] In this theory, the symbol "X" is used to represent an arbitrary lexical category (i.e., noun, verb, adjective, or preposition; in analyzing a specific construction, the X is replaced with an N, V, A, or P). The lowest level of this schema is referred to as the "head", an individual word of category X that determines the category of the entire phrase. At the next hierarchical level, the head is combined with an optional complement (which is a constituent with its own subordinate

and further developed by Jackendoff (1977), which represents utterances in the form of branching tree structures according to a general schema for constructing phrases.

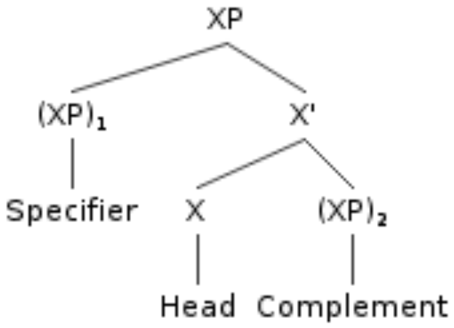


Figure 1. X-bar schema

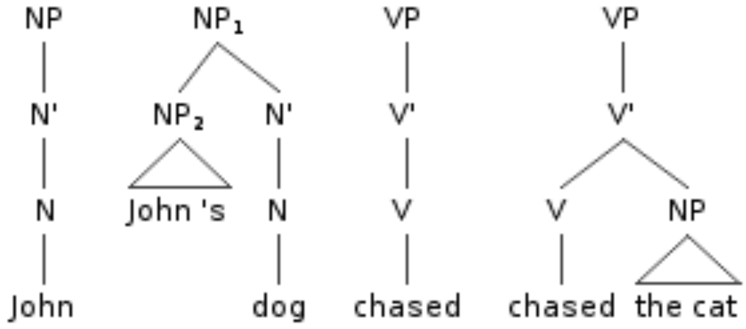


Figure 2. Tree structures for XP's with and without the optional Specifier and Complement

The X-bar schema can be used to combine phrases all the way up to the level of a complete sentence, although the further terminology and conventions required for implementing this theory are complex and beyond the scope of the present discussion. It suffices to say that X-

---

structure) under an intermediate syntactic unit notated as X' (read as "X-bar"). In turn, the X' and its subordinate structure is combined with an optional specifier (another constituent similar to the complement but different in grammatical function) to form a complete phrase, XP. Any constituent (i.e., a group of words that functions as a single unit and can be integrated into larger structures or moved to a different position within the sentence) consists of a complete XP structure. This means that the optional complement and specifier are actually complete XP's of their own which are incorporated into a superordinate phrase. This construction schema is represented visually in the Figures 1 and 2.

bar theory is a useful tool for representing the hierarchical structure of language, and can theoretically be used to analyze any grammatical sentence in any language.

4.2.2. Constituency tests

How can we be sure that this hierarchical constituent structure is in fact a real property of language and not an illusory design arbitrarily imposed on flat strings of words by overeager theoretical linguists? Syntacticians (e.g., Carnie, 2006) rely on a variety of handy diagnostics called constituency tests that, while not always foolproof, are useful ways of demonstrating the hierarchical dependencies within sentences. There are a variety of these tests, all of which manipulate some portion of a sentence to produce a result that native speakers either perceive as grammatical (indicating constituency) or ungrammatical (indicating that the string of words in question is not a constituent). One such test, called Topicalization or "fronting", involves moving a purported constituent to the front of the sentence. For example:

(5) John will go to the store tomorrow [to buy groceries].

(6) [To buy groceries], John will go to the store tomorrow.

(7) * [Buy groceries], John will go to the store tomorrow to.[8]

In these sentences, the words "to buy groceries" are proven to act as a single unit within the sentence because they can be grammatically moved to a different part of the structure. In contrast, "buy groceries" does not constitute a complete constituent because the Topicalization test does not yield a grammatical result. Another useful test is Pro-form Substitution, whereby a suspected constituent is replaced with an appropriate pronoun:

(8) John will go [to the store] tomorrow to buy groceries.

(9) John will go *there* tomorrow to buy groceries.

---

[8] In this example and those that follow, an asterisk marks the sentence that follows it as ungrammatical (i.e., it could not be uttered naturally by a native speaker).

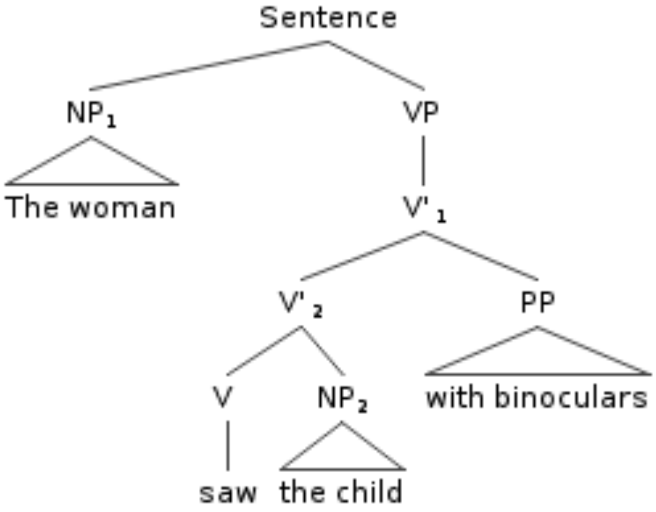
"To the store" is therefore a complete constituent, although "store" by itself is not:

(10) * John will go to the [store] tomorrow to buy groceries.

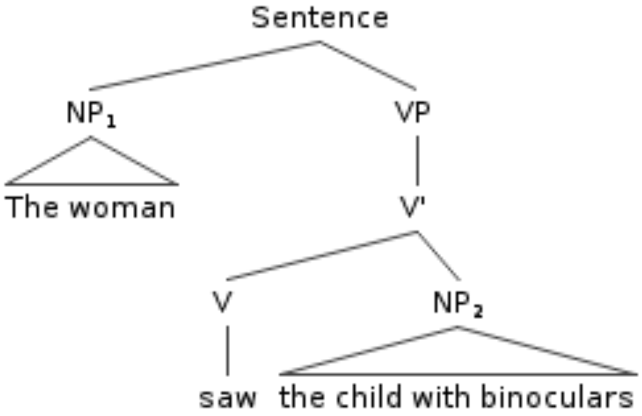
(11) * John will go to the *there* tomorrow to buy groceries.

A final piece of evidence pointing to the internal structure of sentences is the existence of ambiguous utterances. It is entirely possible to imagine a string of words that can be interpreted in multiple ways to yield different meanings:

(12) The woman [saw the child] [with binoculars].



(13) The woman [saw [the child with binoculars]].



The fact that this single string of words can be understood to mean two different things (i.e., "The woman saw the child through her pair of binoculars," or, "The woman saw a child who had a pair of binoculars") implies that these two distinct semantic interpretations each possess their own distinct hierarchical structure in the minds of language users. Without that structure, ambiguous sentences would not be possible.

4.2.3. Recursion in language

One very important aspect of language structure, alluded to previously but not explicitly mentioned, is the phenomenon of recursion. Fitch (2010) describes recursion in great detail, but put simply, it allows a constituent phrase to be embedded within another constituent of the same category. In terms of X-bar theory, this means that an entire XP can function as a single node within a larger XP tree structure (i.e., in the complement or specifier position). For example, consider the following phrases:

(14) $[_{NP}$ the sea]

(15) $[_{PP}$ beside $[_{NP}$ the sea]]

(16) $[_{NP}$ the house $[_{PP}$ beside $[_{NP}$ the sea]]]

In these examples, a complete noun phrase ("the sea") becomes the complement of a larger prepositional phrase, which in turn becomes the complement of an even larger noun phrase. This process can be applied to constituents of any size, even complete clauses, which can be incorporated into the larger syntactic structure either at the end of the sequence (as in the examples above) or in the middle. Corballis (2007:698) refers to these two types of recursion as "tail recursion" and "center-embedded recursion," which are demonstrated below in the modified examples from his original description:

(17) This is the rat [that ate the malt].

(18) This is the rat [that ate the malt [that lay in the house]].

(19) This is the rat [that ate the malt [that lay in the house [that Jack built]]].

(20) The malt [that the rat ate] lay in the house.

(21) The malt [that the rat [that the cat killed] ate] lay in the house.

In this way, recursion can technically be applied *ad infinitum* to produce sentences of infinite length, although computational limits on language processing prevent this from occurring in reality; as is evident in the examples above, especially in the case of center-embedded recursion, increasing levels of embedding become increasingly difficult to parse.

Recursion is of vital importance in language, being a crucial property of the Merge operation and allowing for the endless generativity of human communication. Indeed, Hauser, Chomsky, and Fitch (2002) identify recursion in the linguistic sense outlined above as the fundamental characteristic distinguishing human language from other forms of animal communication. While many other authors (i.e., Pinker and Jackendoff, 2005) have vigorously contested this claim that recursion is the only unique aspect of linguistic communication, few deny that it plays a very significant role in the human capacity to use language.

The past several pages have tried to demonstrate in some detail the structural nature of human language. This structure is inherently hierarchical and largely dependent on the property of recursion, which accounts for the generative, open-ended qualities of linguistic systems. In the pages that follow, I will argue that this structure is in fact not unique to language but is shared to a large extent by other domains, particularly goal-directed manual action.

4.3. The structure of human action

The importance of the overarching goal in a complex action must not be overlooked, as it distinguishes hierarchically structured manual behavior from gesture, a highly interesting and unique type of communicative activity that has communicative purpose but lacks compositionality. A complex manual action is composed of constituent sub-actions that are organized hierarchically, with the overarching goal as the highest level of that hierarchy; the goal of an action is in many ways (although not entirely, as we shall see) analogous to the highest XP in a syntactic tree. Through a behavioral study of imitation in children and adults, Wohlschläger et al. (2003) confirmed that an observed movement is not imitated as a whole, but rather broken down into separate aspects that are hierarchically organized into a main goal and subordinate goals.[9] It is from this perspective that I will refer to action throughout this section.

A useful term here is *praxis*, defined in this context by Arbib (2011:259) as "practical actions on objects with or without the use of tools." This description highlights the importance of objects and tools within a goal-directed activity as the subject and means, respectively, of hierarchical organization and manipulation. The goal of an action (as it is referred to here) is inevitably to manipulate some object in such a way as to produce a desired result, employing whatever devices are available, be they one's own hands or some other object (i.e., a tool), in order to effect that outcome. An analogy can be drawn here with language, if we consider content words as objects that are manipulated by function words and grammatical requirements (i.e., tools) in order to achieve a certain communicative goal (Arbib, 2011:258). Let us return to the structure of action, though, by considering some specific examples of praxis and breaking them down into their constituent parts.

---

[9] Note that the attentive observation and execution of an action involved in imitation are two sides of the same coin, analogous to the comprehension and production of language and derivative of the embodied relationship between perception and sensorimotor experience.

4.3.1. Grammars of action

Greenfield (1991:532) demonstrated that children employ three strategies for organizing motor action to achieve a goal of stacking nested cups: “pairing,” the “pot” strategy,” and the “subassembly” strategy.

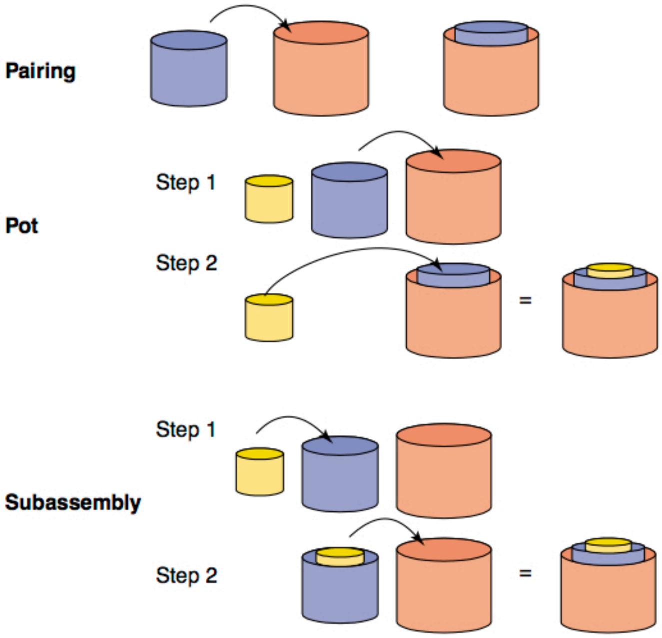


Figure 3. Cup stacking strategies used by human children (Conway & Christiansen, 2001:545)

The pairing and pot strategies involve one or more cups being placed one-by-one into a single static cup to produce the final stack in a chain-like combination. These strategies are to be distinguished from the subassembly strategy, a more complex, two-level hierarchy in which multiple cups are combined into a single constituent-like unit (the “subassembly), which is in turn combined with a single cup or another subassembly to produce the final structure. Greenfield termed these strategies “grammars of action,” describing them with a tree notation not unlike that of X-bar theory and likening the hierarchical organization of motor activity to that of language.

Greenfield's (1991) subassembly strategy is a very simple example of praxis in which objects (i.e., the cups) are acted on without the use of a tool, although the incorporation of a tool into the action does not necessarily change its hierarchical structure. Eating with a spoon, a very common tool in Western society, is another example of the subassembly strategy: spoon and food are combined into a single unit, which then acts on a final entity (the mouth) to achieve the goal of feeding oneself. Of course, this highly specific motor action is typically only a small constituent and sub-goal of a much larger action sequence that could be represented visually as a tree diagram. The highest level of this hierarchical structure would represent the ultimate goal of, for example, having dinner, which is very general and abstract and occurs over an extended period of time; intermediate levels would represent more particular sub-goals such as preparing ingredients, cooking food, setting the table, eating, and doing the dishes; and the lowest levels would represent specific motor actions such as the mechanics of spoon use mentioned above. As Stout (2011:1051) points out, high-level, general goals can be flexibly adapted to specific contexts through the appropriate selection of lower-level sub-goals and motor actions, while simultaneously being informed and directed by the outcome of these same subordinate actions (e.g., the general goal of turning on a light can be variably achieved by flipping a switch, turning a knob, or pulling a cord).

4.3.2. Stone tool production

Structural analyses such as this could be applied to any sort of goal-directed behavior, but the construction of stone tools has received special attention due to its powerful potential as a means for making inferences about the evolution of the hominin line. Moore (2010) uses Greenfield's (1991) tree notation to describe the hierarchical structure of stone toolmaking, and Stout (2011) expands on that model to further describe the organization of the major toolmaking strategies in hominin evolution, as inferred from modern experiments and the analysis of evidence from the archaeological record.

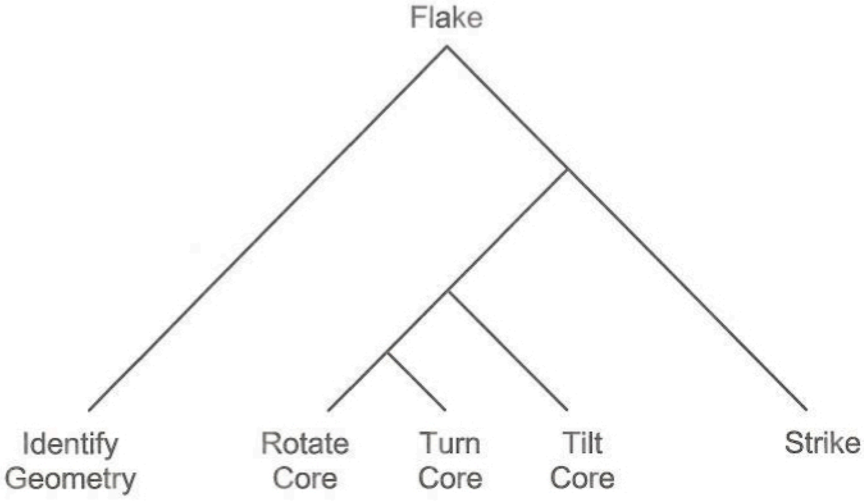


Figure 4. Tree structure for a basic flake unit (Moore, 2010:20)

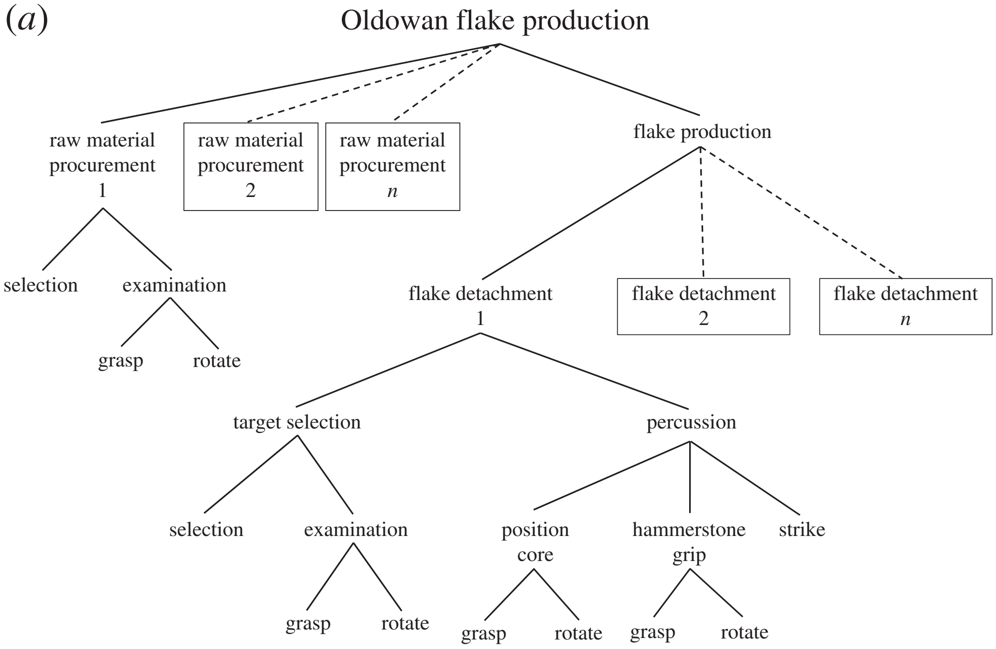


Figure 5. Tree structure partially depicting the action hierarchies involved in Oldowan stone tool production (Stout, 2011:1052).

Although the level of hierarchical complexity involved has increased throughout the history of human evolution, stone tool production can generally be described as directed by the ultimate goal of shaping raw stone materials into a preconceived tool category. This overarching superordinate goal combines the two temporally separated sub-goals of (1) procuring the raw materials to work with (which must be of the appropriate size, shape, and composition) and (2) the actual process of flaking a stone core with a hammerstone to produce the desired tool; both of these sub-goals are in turn composed of highly complex, hierarchically organized sequences of action. A closer examination of the flaking process reveals the further organization of subordinate goals and actions such as examining the stone core, determining where to strike it with the hammerstone, rotating or supporting the core accordingly, selecting an appropriate grip for the hammerstone, and accurately striking the core to detach a flake. All of this planning and motor coordination goes into the detachment of a single flake, and is 'chunked' into a single action constituent that can be repeated indefinitely; these individual detachments are directed by and connected to each other under a superordinate goal.

4.3.3. Describing grammars of action with a formal, syntax-like schema

The structure of language has been described in great detail with many different formal models (e.g., X-bar theory), and although grammars of action have been described systematically with tree structures (Greenfield 1991) and applied with great detail to specific contexts such as stone tool production (e.g., Moore, 2010; Stout, 2011), a formal model for the structure of action remains elusive. Pastra & Aloimonos (2012), however, have taken the formal conventions of Chomsky's (1995) Minimalist Program for describing linguistic structure as a reference model for a formal generative grammar of action. In this model, an action takes as its complements a tool and an object, and is orchestrated by a goal in the same way that a sentence is inflected according to a certain verb tense. The tool complement of an action is whatever device effects that action, be it a body part, a combination of body parts, or a graspable object that functions as an extension of a body part; the object complement is whatever object is acted on by the tool complement.

For example, consider the act of grasping a knife: the “syntactic head” is the act of grasping, which is effected by a tool (the hand) on an object (the knife). The act of grasping could be carried out differently with alternative tool and object complements (e.g., using tongs to grasp a pencil), but the general action involved is essentially the same. Importantly, the pairing of action and tool acts as a constituent, which is combined through the Merge operation with its object complement at a higher level in the action hierarchy. This larger action constituent, “grasp knife with hand,” can in turn become part of an even larger action structure such as “slice apple with knife,” in which the essential action “slice” is combined with the tool complement “knife” to affect an object “apple”. These embedded action constituents are sequential and connected to each other in time through a “temporal conjunction” to form the entire action sequence, which is orchestrated and in a sense “inflected” by the final goal of slicing the apple. A different final goal (e.g., “stab the apple”) would affect the realization of each of the subordinate parts of that action, resulting in a different grip on the knife, a different way of approaching the apple, etc.

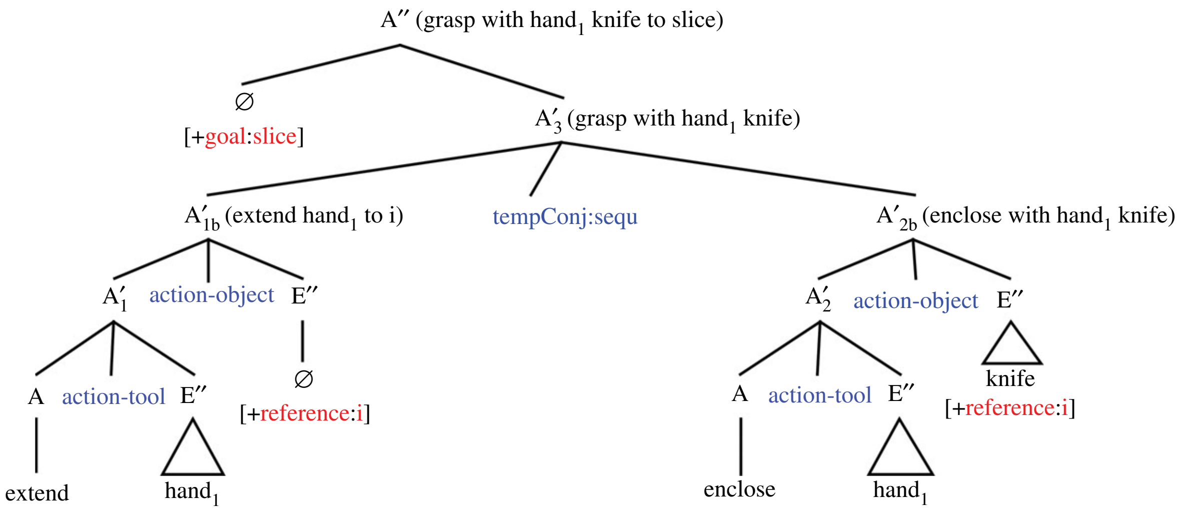


Figure 6. Tree structure for the action, "grasp knife with hand," with the goal of slicing something (Pastra & Aloimonos, 2012:108).

4.3.4. Recursion in praxis

If action and language really are parallel in structure, then surely the fundamental role of recursion in language will be mirrored in action. Pastra and Aloimonos (2012:112) account for this, claiming that both tail recursion and center-embedded recursion are important features of grammars of action. Tail recursion can be seen in the example used above, in which each completed action constituent becomes a tool used in the next (i.e., [extend hand, [grasp knife with extended hand, [slice apple with grasped knife]]]). Center-embedded recursion occurs when one action is started, interrupted by another action, and then completed (i.e., [grasp knife, [grasp apple, [position apple on cutting board]], [slice apple with grasped knife]]). Recursion is also at play in the simple merging of a combined action-tool constituent with its object complement to create a larger constituent, as mentioned above. Stout (2011) also notes that formal recursion is present in stone tool construction, as when a particular flake detachment is made possible by several preparatory recursive flake detachments (which may be understood as embedded within the larger process of detaching that particular flake). This recursive property accounts for the extreme productivity of complex action, which can be implemented in endlessly variable contexts to achieve any goal imaginable.

Just like language, complex goal-directed actions are characterized by the hierarchical organization of sequential constituent sub-actions. Linguistic utterances and goal-directed action sequences can both be described by tree notation systems that are governed by similar "grammatical" construction rules. Finally, both language and action share the property of recursion, which makes each system infinitely generative and productive in an endless number of specific contexts or situations. In these ways, language and action can be said to have parallel hierarchical structures. These striking similarities imply that the two systems are in fact analogous, relying on the same (or at least largely overlapping) structures within the brain.

4.4. Homologous brain structures for language and action

So far I have outlined the many ways in which the structures of language and action are parallel, but I have yet to provide any direct evidence suggesting that the two systems rely on the same cognitive hardware. There is in fact a great deal of experimental evidence, both behavioral and neurological, that points to this conclusion. Although it is beyond the scope of this paper and the technical competence of this author to delve into the neurological evidence in any great detail, it is important here for the evolutionary scenario that I am arguing to present some of the major findings pointing to homologous brain structures and circuits for language and action.

In one study, Allen et al. (2010) presented participants with various sentences describing action sequences that were structured in three distinct ways, and found that participants were quicker to comprehend sentences that were preceded by sentences describing the same type of action sequence. In a second experiment, the same authors paired sentences describing the same action sequences with short video clips of wooden blocks being assembled in a way that was either similar or dissimilar to the action structures described in the sentences; as expected, participants were quicker to comprehend sentences that were preceded by video clips of analogous block structures. This priming effect demonstrates that the structure of a given action sequence or sentence is abstractly represented in the mind and can be at least partially dissociated from the particular details involved, and furthermore suggests that brain processes involved in the perception of both language and action are the same or overlapping.

Greenfield (1991:531) comes to a very similar conclusion in her previously mentioned study, claiming that, at least during the first two years of life, "a common neural substrate (roughly Broca's area) underlies the hierarchical organization of elements in the development of speech as well as the capacity to combine objects manually." Broca's area has long been associated with the faculty of language in general, and although it was originally described as a discrete cognitive module ventrally located in the left frontal lobe of the cerebral cortex, it is now recognized that language relies heavily on the entire interior frontal gyrus as well as on neural circuits distributed throughout the brain (Stout & Chaminade, 2012). This area of language-relevant cortex is now understood to be responsible for a wide range of linguistic functions, including the production and comprehension of phonetic, syntactic, and semantic aspects of language. Furthermore, this brain region is known to be involved in the processing of a variety of hierarchically organized activities other than language, including object manipulation, visual search, music, and mathematics. Given these diverse functions, Koechlin & Jubault (2006) speculate that Broca's area has an underlying computational role in the processing of hierarchically structured information.

The large body of work on embodiment, which demonstrates the ways in which language and thought are grounded in sensorimotor circuits of action and perception, also supports the description of language and action as superficial manifestations of the same underlying ability. Pulvermüller & Fadiga (2010) point out that Broca's area is situated within the larger perisylvian cortex, which is implicated in the "syntactic" processing of goal-direction actions. Various lesion studies have shown that impairments anywhere within the perisylvian cortex can result in agrammatism, which refers to a wide range of deficiencies in linguistic processing traditionally associated with Broca's area, just as impairments within Broca's area can result in difficulties with more general cases of hierarchical processing (Pulermüller & Fadiga, 2010; Greenfield, 1991).

Of special relevance to the present discussion, brain regions implicated in tool use and goal-directed activities such as stone tool production have been demonstrated to overlap significantly with the neural correlates of language. In a functional imaging study Stout et al. (2008) show that brain activity during Oldowan and Late Acheulean toolmaking (as performed by modern experimenters) engages the ventral premotor and inferior parietal cortices in both hemispheres, as well as the right hemisphere homologue of Broca's area. This activation pattern overlaps significantly with previously established language circuits, and strongly suggests that language and toolmaking share an evolutionary basis in a more general capacity for hierarchically organized, goal-directed action (i.e., digital processing).

This section has attempted very briefly to review behavioral and brain evidence explaining the structural parallels between language and goal-directed action outlined in previous sections. In summary, present knowledge suggests that language is not represented in the brain as a discrete, individual module, but rather is dependent upon a widely distributed neural network that is responsible for the general processing of hierarchically organized information in digital systems. In the following section, I will discuss the development of this general capacity in humans as well as any clues that this ontogeny may hold about the evolutionary past.

5. Action and Language in Human Ontogeny

The previous chapter demonstrated in some detail the degree to which language and goal-directed manual action (i.e., praxis) are digital systems that share a parallel hierarchical structure and are grounded in the same or significantly overlapping neural circuits. It concluded that the two modalities are likely different manifestations of a single underlying capacity for digital processing. Given that praxis is present to a degree in other primates, and that it is implicit in the construction of stone tools that appear in the archaeological record long before the hominin line is thought to have had true linguistic abilities, it seems likely that the capacity for hierarchically organized behavior originally evolved to support practical actions on objects and was subsequently expanded and put to use for communication (Arbib, 2011; Stout & Chaminade, 2012; Stout, 2011).

This chapter will examine evidence from developmental psychology that supports the same conclusion, particularly the fact that the digital systems of manual action and conventional language as well as analogic communication systems such as gesture follow parallel patterns of development early in human life. These modalities are initially very similar in their basic structure and are localized in the same part of the brain; over the course of development, however, they become more specialized and distinct in structure, governed by increasingly differentiated neural networks and ultimately taking forms unique unto themselves (Greenfield, 1991:548). It is true that using developmental evidence from modern humans as a means of making inferences about the evolutionary past is dependent upon certain assumptions, and the validity of this line of reasoning will be assessed briefly at the end of the chapter. First, however, I will address some very insightful observations from several decades ago about the development of language and action in children that are too often overlooked in the study of language evolution despite their relevance to the subject.

5.1. The importance of play

In his investigation of the adaptive nature and function of immaturity (i.e., the extended period of vulnerable development in humans), Jerome Bruner (1972) focuses on the importance of play in the emergence of species-typical adult behavior. As Bruner uses it, “play” refers to the curiosity-driven practice and exploration through which immature individuals develop proficiency in and eventual mastery of culturally fixed patterns of behavior. Play is a means of violating fixity through which young individuals explore the boundaries of behavioral patterns, manipulating and varying their component aspects and acquiring competency in them through sheer repetition. Bruner adopts an evolutionary approach to understanding human development, comparing human infancy and childhood with that of other primates to assess its distinctiveness or generality. He notes that behavioral repertoires are much more limited among prosimians and monkeys compared with the apes, and that play in these species is correspondingly less varied, occurring over a shorter period of time and more or less disappearing in adulthood. Indeed, in another paper, Bruner (1975:8) speculates that, “the more productive the programs of action of the adult of the species, the greater the likelihood of mastery play early on in development in that species.” Higher primates and especially humans possess highly productive programs of action (e.g., elaborate patterns of tool use, complex social structure, language) and thus engage in a greater amount and degree of play during development (e.g., playing with building blocks, ‘playing house’, playing with linguistic forms).

Bruner (1972) is not the only scholar to recognize the importance of play in human ontogeny. Perhaps the best known author to tackle this issue was Piaget (1962), who placed emphasis on the roles of imitation and play in the child's transition through successive stages of cognitive development. According to his observations, imitation and manipulative play become increasingly complex throughout the sensorimotor stage (cf. Piaget, 1983), beginning with the systematic imitation of sounds and movements already made and seen by the child and progressing to the imitation of novel models of phonation and action. After about the first two years of life, imitation has become sufficiently complex and systematic to support conceptual schemas, marking the transition from the sensorimotor stage to cognitive representation. This new phase of development sees the elevation of play from the level of sensorimotor manipulation to that of cognitive and symbolic manipulation, with the emergence of games with rules and the imitation and manipulation of social roles. Play and imitation play a crucial role in the development of human cognition, growing more complex throughout early development.

Bruner's (1972) explanation of play – the practice, variation, and perfection by young individuals of the behavioral patterns of adults through endless manipulation and recombination of individual features of those patterns – is an implicit recognition of what is described in this paper as the fundamental human capacity for hierarchical organization. Bruner focuses specifically on the development of tool use through play, arguing that tool use and play were mutually selective in the evolution of the great apes, and draws an analogy between the playful ontogenesis of tool use in higher primates and the imitation and acquisition of linguistic forms in human children. He acknowledges the structural similarities between language and skilled action, and concludes from the developmental evidence that both rely on the same underlying cognitive ability: "the simultaneous appearance in man of language and tool using suggests that the two may derive from some common programming capacities of the enlarging hominid nervous system," (696). Along with several other authors (e.g., Huizinga, 1955; Neumann, 1974; Garvey, 1990; Lowenfeld, 1990), Bruner (1972) recognizes the evolutionary implications of these parallels: “The initial use of language is probably in support of and closely linked to action. The initial structure of language and, indeed, the universal structure of its syntax are extensions of the structure of action,” (700).

While these two digital systems are clearly linked and are dependent on the same underlying cognitive capacity, there is a major difference between them: language is used for communication and therefore must be subjected to the constraints of a common symbolic code (as noted by Werner & Kaplan [1963] in their discussion of the physiognomic process of symbolization). These constraints are responsible for the form of language: “there is also a communicative function of language; and it is this function, in all probability, that determines many of its design features (cf. Hockett, 1960),” (Bruner, 1972: 699). Because language relies on a digital, symbolic code (in addition to analogic channels of communication), these symbols become subject to the human tendency for manipulation. It is through symbolic play that children manipulate linguistic symbols and their underlying concepts, and in the process gain fluency in the rules and conventions of their society (Ervin-Tripp, 1991). Once manipulative play has been elevated to this conceptual level of language, it quickly becomes dissociated from the actions it initially supports, taking on new forms and structures unique unto themselves.

Bruner’s (1972, 1973) early penetrating insights into the close developmental and evolutionary relationship between language and action remain highly relevant to the study of language evolution, although they are often overlooked in that field. A wealth of more recent findings in developmental psychology and neuroscience has corroborated Bruner’s claims, detailing the parallel development of language and action through time as well as within the brain. I will turn to that literature now.

5.2.1. The parallel development of language and action in human children

Greenfield (1991) reports that the first two years of human life are characterized by the development of increasingly complex levels of hierarchical organization in both action and language. That this developmental program is true of action has been confirmed by a variety of different object combination tasks involving, for example, nesting cups, nuts and bolts, construction straws, blocks, and two-dimensional pictures. Greenfield's (1991) study focuses specifically on children's performance in a cup-stacking activity as an example of systematic development towards more complex levels of hierarchical organization. As mentioned in the previous chapter, children were observed to employ three strategies for stacking nested cups: "pairing", in which a single active cup acts on a single stationary cup, the "pot" strategy, in which multiple active cups act on a single stationary cup (i.e., successively placing several cups into a single final cup), and the "subassembly" strategy, in which two cups are combined into a pair, which is then manipulated as a single unit in the next action. Importantly, the first two strategies only involve one level of combination, while the third requires an added level of hierarchical complexity. These three strategies develop in this sequential order starting at 11 months of age, with the more hierarchically complex subassembly stage appearing last.

This development towards increasingly hierarchical programs of action is mirrored structurally and temporally in grammatical development. The child starts with one-word utterances (typically vocabulary items relating to eating, drinking, greeting, and other basic activities) between about 11 and 13 months of age (Bates & Dick, 2001), which are combined with other single words in the next developmental step between 18 and 20 months of age, as in the combination of an adjective and a noun to form a superordinate noun phrase (Greenfield, 1991). The period from 24 to 30 months typically sees an explosion of grammar, with two-word combinations giving way to more complex structures complete with inflections and function words (Bates & Dick, 2001); for example, complex noun phrases such as “more grapejuice” start to be combined with verbs at this stage, yielding more complete constructions such as “want more grapejuice,” (Greenfield, 1991).

These parallel developmental patterns in language and action are not coincidental, and appear to be governed by common underlying constraints on hierarchical complexity that are progressively expanded over the first several years of life. Even before the appearance of recognizable one-word utterances or the capacity to effectively manipulate objects with the hands, the constituent parts of those abilities are synchronized in development. Canonical babbling – the production of repeated consonant-vowel segments such as [ba], [didi], or [yaya] – corresponds to the onset of rhythmic hand banging or clapping around 6 to 8 months of age (Bates & Dick, 2001). These activities exemplify Bruner’s (1972) description of play, in which the child endlessly manipulates and recombines segments (of language or of action) until they have been mastered and can be combined into larger sequences; in this case, the child is playing with sound segments to master the phonemic inventory of the language it is acquiring while simultaneously playing with its hands to develop the fine motor skills necessary to produce praxic actions on objects.

Constraints on hierarchical complexity are successively lifted throughout development and are responsible for the similarities between the stages of increasing ability with language and with action. Each of the successively more complex combinatorial strategies for stacking cups described by Greenfield (1991) corresponds to a specific stage in linguistic development. For example, children only capable of the “pairing” strategy mentioned above are restricted to one-word utterances, and are no more capable of producing a more complex two-word combination than they are of using the “pot” strategy to stack nested cups. Just as the pot strategy is analogous to two-word constructions, the subassembly strategy corresponds to the ability to combine noun phrases with verbs to form simple sentences with two levels of hierarchical structure (e.g., “want more grapejuice”). Furthermore, according to Bates & Dick (2001), as grammatical ability starts to rapidly expand between 24 and 30 months of age, so too does children’s “ability to remember and imitate arbitrary sequences of manual actions (in scene construction tasks with novel objects),” (296). These parallels between the increasingly complex hierarchical organization of language and action in the first few years of life suggest that the two domains are simply different manifestations of a single developing ability.

5.2.2. How gesture fits into the developmental picture

Given the important complementary role gesture plays in the production and comprehension of language (discussed in the third chapter), it should come as no surprise that this nonverbal means of communication also fits into the same developmental framework as language and action. The manipulative manual abilities that start to come together through play in support of digital praxis around 6 to 8 months of age are also used to support language in the form of analogic co-speech gestures; in fact, children communicate through the gestural medium even before uttering their first words. Language and gesture may be considered as two sides – digital and analogic – of the same communicative coin, developing together and in parallel to non-communicative manual action. Bates & Dick (2001) provide a detailed account of correlated milestones in the early development of gesture and speech, summarized in the table below.

Table 1. Language and gesture milestones, by age[10]

Age in Months	Language Milestones	Gestural Correlates
6–8	Canonical babbling	Rhythmic hand movements
8–10	Word comprehension	Deictic gestures, gestural routines, first tool use
11–13	Word production (naming)	Recognitory gestures
18–20	Word combinations	Gesture–word & gesture–gesture combinations
24–30	Grammaticization	Sequences of 3–5 arbitrarily ordered gestures

As mentioned earlier, canonical babbling and rhythmic hand movements emerge at roughly the same time, around 6-8 months of age. Both milestones occur outside of a communicative framework, but are linked to each other and to later changes in word production and gestural communication, and as such constitute a significant event in the development of digital linguistic abilities (i.e., they are the "digits" that will eventually be productively combined in speech and in praxis). This particular speech-gesture link is even present among infants with language-related disorders such as Williams syndrome (albeit somewhat delayed in this population), indicating that gesture and speech do indeed form a closely integrated system (Bates & Dick, 2001).

The 8 to 10 month age range sees the emergence of the first truly communicative abilities in both domains: systematic word comprehension and deictic gestures (e.g., giving, showing,

---

[10] This is an abridged version of the original table in Bates & Dick (2001:294), which includes a fourth column of associated references for each developmental milestone.

pointing). Interestingly, this stage sees developments in both the digital realm (systematic comprehension of arbitrary symbols) and the analogic realm (deictic gestures are indexical rather than arbitrary), indicating that these systems are diverging or already distinct at this early point in development. These developments are also correlated with the emergence of other relevant abilities outside of communication, such as early tool use, feature-based categorization, and imitation of novel acts. This developmental stage is marked by significant changes in cognition that prove essential for subsequent communication and praxis.

The onset of first words, or “naming”, around the 1-year mark is accompanied or slightly preceded by the onset of “gestural naming”, recognitory gestures that are associated with and symbolically represent specific objects. The earliest gestural names are usually produced with the associated object in hand, although they are eventually produced even in the absence of the physical referent. These first words and gestures are remarkably similar in semantic content, usually referring to a shared vocabulary about salient events in the child’s life such as eating, drinking, bathing, greeting, dressing, household objects and activities, etc., and are typically brief and stylized in form. For example, a child may put a cup to his or her lip to refer to the act of drinking, but the child distinguishes between this act of gestural naming and real drinking, and may produce it with fake cups or upon seeing someone else drinking (Bates & Dick, 2001:296). Despite the similarity between vocalized and gestured names in semantic content, however, the two forms are already distinct in formal character: first words are arbitrary symbols and units of a digital system, whereas gestural names are iconic, stylized movements of the whole hand and instances of analogic communication. The distinction between digital and analogic communication seems to be firmly established even at the onset of intentional communication and symbol use.

Importantly, the phenomenon of gestural naming is only a transient phase between about 12 and 18 months of age in typically developing children (Bates & Dick, 2001). It seems that, in the initial stages of symbol formation, children focus equally on gesture and speech as a means of communication, but the focus on gesture quickly fades away as soon as they become sufficiently adept in the much richer and more productive linguistic domain. This fact echoes Bruner's (1972) observation that language quickly becomes dissociated from the actions it initially supports, and marks the beginning of a developmental trend in which language and action become increasingly differentiated in structure and function. This does not mean that digital communication becomes completely unhinged from its analogic foundations, however – as Werner & Kaplan (1963) explain, externally arbitrary linguistic (digital) forms retain a sense of internal psychological reality, grounded in sensorimotor experience and possessing some sort of intuitive analogical relation to the concepts they denote. This suggests that the relation between digital language and the action it initially supports simply becomes masked as the child expands his or her repertoire of arbitrary linguistic forms.

It is telling of this point that communicative gesture in adults is never more complex or compositional than the gestural repertoires of one-year-old children; it seems that the gestural medium develops to a point sufficient for concrete demonstration and indication (e.g., iconic and deictic gestures) – analogic functions that remain useful throughout life and support language as co-speech gestures – until the mastery of true linguistic systems (spoken or signed) allows for digital communication about a much wider range of more nuanced and abstract topics. Again, this is not to say that analogic gesture is replaced by digital language as the sole medium of communication, but rather that it is fully developed at an earlier point and remains present in a supporting role (although that role is sometimes hidden, as in the case of written language).

The first instances of two-word combinations around 18 to 20 months are accompanied or slightly preceded by gesture-word combinations, such as simultaneously pointing at and naming an object (Bates & Dick, 2001). These developments seem to be the result of a newly found hierarchical capacity for two-unit combinations, also seen in the "pot" strategy for object combination (Greenfield, 1991) observed in children around the same time, and are indicative of further expansion of digital capacities. At this point in development, language and gesture can be combined to convey complementary communicative messages, eliminating the need to produce independently meaningful gestures and consequently resulting in the gradual disappearance of the gestural naming phenomenon. By 24 to 30 months of age, further hierarchical constraints seem to be lifted to allow for more complex combinations of units and the explosion in grammar and more complex sequences of manual action mentioned earlier, officially relegating gesture to a secondary, supportive role in communication. Examples of these increasingly complex hierarchical structures include "more ornate sentences with inflections and free standing function words," in the linguistic medium and "the ability to remember and imitate arbitrary sequences of manual actions (in scene construction tasks with novel scene construction tasks with novel objects)," (Bates & Dick, 2001:296) in the realm of praxis.

5.2.3. Language and action in the developing brain

The initially similar but eventually differentiated forms assumed by language, gesture, and action are reflected in the neural circuitry of the brain. Broca's area has long been recognized to play an important role in processing language and a host of other similarly structured tasks and abilities, although the exact limits of this region are hard to define (Greenfield, 1991). Originally described as a discrete cognitive module ventrally located in the left frontal lobe of the cerebral cortex and acting in isolation, the various functions of Broca's area are now understood to be carried out by neural circuits that extend beyond the region's traditional boundaries and that are shared with other basic cognitive abilities such as perception and imitation (Greenfield, 1991; Bates & Dick, 2011). Regardless of precise definitional difficulties, Broca's area and the surrounding portion of the posterior prefrontal cortex have been demonstrated to contain a system of executive processes that control the hierarchical organization of action and thought in relation to internal goals (Koechlin & Jubault, 2006), and as such play a central role in processing language (that is, speech and communicative gesture) as well as praxis.

While the general processing of hierarchically organized information seems to be focused in and around this particular region, brain damage and lesion studies have shown that damage to this region of the brain appears to impair certain linguistic and praxic abilities while leaving others intact. Specifically, the existence of different types of aphasia (a particular inability to produce or comprehend grammatical language) and apraxia (comparable impairments of sensorimotor skills) in adults imply that Broca's area is not unitary but rather subdivided into functionally distinct areas or circuits[11], at least within the fully developed human brain (Bates & Dick, 2001; Greenfield, 1991).

Given that language, action, and gesture initially display strong structural similarities, each reaching analogous milestones at roughly the same times throughout early development only to become increasingly differentiated in form and function after the first two years of life (Bates & Dick, 2001), Greenfield (1991) claims that "Broca's area might start out, early in development, as an undifferentiated neural region, programming both manual action and

---

[11] Unfortunately, current neuroimaging techniques lack sufficient resolution to determine this empirically.

language production," (537). To account for the eventual differentiation and specialization of these neural circuits in adults, she cites the fact that neuronal connections between cortical areas are established diffusely early on in development only to be refined later on by a process of selective synaptic pruning and growth. This process has been studied in great detail and further confirmed by more recent neurological work (e.g., Chechik et al., 1998; Iglesias et al., 2005; Craik & Bialystok, 2006; Low & Cheng, 2006). Essentially, the infant brain, equipped with a undifferentiated system for the digital processing of hierarchically organized information in general, eagerly establishes synaptic connections to facilitate the processing of new stimuli; as the child accumulates experience, these synaptic connections are either reinforced by recurring patterns (e.g., linguistic input, hierarchical action routines) and flourish into highly developed and functionally distinct circuits, or fall into disuse and disappear to better allocate cognitive resources. This process provides convincing evidence for the developmental basis of embodied cognition.

This is a compelling theory that elegantly accounts for the fundamental similarities of language, action, and other hierarchical capacities as well as the gradual differentiation of each of these abilities throughout development into highly specialized and structurally distinct systems that are closely related yet functionally distinct within the adult brain. Synaptic pruning links conventional linguistic systems and praxis as related but neurally differentiated digital systems in the adult brain, which complement analogic systems to yield complete language.

5.3. Evolutionary implications

Greenfield (1991) uses her theory of neural differentiation within Broca's area over the course of development to draw evolutionary conclusions about the relationship between language and object manipulation. She cites comparative behavioral evidence from primates indicating that language-trained chimpanzees such as Kanzi also possess analogous levels of hierarchical ability with tool use and symbolic communication, although these abilities are limited to the simple "subassembly" level of complexity and no more (i.e., Kanzi never produced an utterance more complex than, for example, "want more grapejuice," nor did he or other chimps ever demonstrate tool use that involved more than two levels of combination) and develop much more slowly (compare Kanzi's 3 ½ years at this stage to the typical child's 1 year). Nonetheless, this hierarchical constraint in chimps and the fact that it is controlled by a primate homologue of Broca's area in humans is consistent with the view that language and tool use abilities evolved together. Assuming a homologous origin of the hierarchical capacities shared by these two species, Greenfield (1991) argues that a common ancestor of humans and chimpanzees possessed the cognitive abilities to support the ontogeny of manual object combination, but not protolanguage: "In this scenario, neural organization of combinatorial manual activity serves as a preadaptation (or exaptation) for the combinatorial aspect of language," (549). Basically, manual object combination (i.e., praxis) provided the cognitive scaffolding on which language was built, and once this initial protolanguage emerged, it entered into a mutually selective relationship with praxis, each increasing the adaptive power of the other.

This scenario is very much in line with the argument presented in this paper. However, it does not account for the role of gesture, which develops in parallel with both action and language and is thus an equally integral part of the evolutionary puzzle. Goodwyn & Acredolo (1993) find that, although the same milestones are reached in gesture and language nearly simultaneously, "the onset of symbol use in the gestural modality very, very reliably preceded the onset of symbolic use of words," (698), although only by a very small margin (less than two months).

Although Bates & Dick (2001) do not mention the consistently precocious development of gestural communication relative to speech, they do concede that, for each of their developmental milestones, the vocal modality is accompanied or slightly preceded by gesture. On the one hand, this gestural advantage could simply be due in part to the fact that the gestural modality is more visible and easily imitated than vocal production, and that it grows out of well-rehearsed sensorimotor behaviors and tends to be more concretely descriptive than the vocal modality. However, it may also indicate that gesture is in fact primary to and constitutive of language – if ever so slightly – which could have implications for the evolutionary history of the two complementary systems.

Iverson & Goldin-Meadow (2005) also find that communicative gesture precedes language and that changes in gesture even predict changes in language within a single child. They ascribe this phenomenon to the fact just mentioned – that symbolic gesture is typically more transparent and easily accessible than the arbitrary mapping between words and their referents – and state that “gesture may thus serve as a transitional device in early lexical development,” (369). Again, the primacy of gesture to language and the transitional role it seems to play in linguistic development may echo its role in the evolution of language.

Kelly et al. (2002) take an explicitly evolutionary approach, speculating that gesture played an important role in language evolution based on its important roles in language development as well as the production and comprehension of language in real time. The authors identify three basic functions of communication: indication, demonstration, and description. Importantly, nonverbal behaviors (both manual and “orofacial” gestures) are especially well suited to indication and demonstration – analogic functions – though less so to description. They suggest that indication and then demonstration were probably the first communicative functions to evolve (judging by their simplicity, the degree to which they rely on gesture, and their immediate utility for the speaker), laying the groundwork for the more abstract descriptive function. The first two functions, which may have been executed nonverbally, eventually provided the scaffolding for the third, which is more abstract and probably required vocalization: "vocalizations that accompanied the physical actions of indicating and demonstrating could have become more conventional and codified with repetition, and could have eventually evolved into words and sentences," (326). Importantly, analogic gesture is perfectly adapted to the basic functions of indexing relations among things (indication) and iconically representing thoughts (demonstration), but the exigencies of description require the more nuanced and productive possibilities of a digital linguistic system and came to rely more heavily on speech than on gesture. Furthermore, digital language makes possible "redescription" – or paraphrase – and commentary on description, as well as the description of counterfactual impossibilities that cannot be analogically indicated or demonstrated; this capacity for description is the primary advantage of digital language over analogic communication.

Taking into consideration the supportive and transitional role of gesture in linguistic development (particularly the transient phenomenon of gestural naming; Bates & Dick, 2001), I propose to modify Greenfield's (1991) evolutionary scenario – that complex action scaffolded the subsequent emergence of language – with an intermediate step. In this scenario, complex object manipulation produced the dexterous manual abilities necessary for indicative and demonstrative gestural communication, which in turn provided the symbolic framework for formal language to fulfill the more precise and demanding function of description. This proposed sequence of evolutionary developments accounts for the closely interconnected nature of action, gesture, and language, and is reflected in human ontogeny.

5.4. Using developmental evidence as a line of evolutionary reasoning

Much of the previous section relies on a very important assumption about the relation between a given species' particular pattern of development and its evolutionary past, the validity of which has significant implications for a major line of evidence in the field of language evolution. The familiar catchphrase "ontogeny recapitulates phylogeny" was coined by Haeckel (1866, cited in Mayr, 1994:225) shortly after the publication of Darwin's theory of evolution, although the notion of recapitulation – that the developmental stages that a species passes through in ontogeny correspond to the most important mature forms of its evolutionary ancestors – dates back as far as the German natural philosophers of the 1790's (Mayr, 1994). In light of modern evolutionary theory, a strict belief in recapitulation of this sort is crude and something to be avoided; however, the generally conservative nature of evolution renders the notion not entirely outlandish, and in any case it makes available a very rich line of evidence for making inferences about the evolution of a species. Phylogeny is, after all, a history of ontogenies and is indubitably shaped in many ways by the developmental patterns of successive generations.

Kelly et al. (2002) appeal to this general rule of evolutionary conservatism – that evolution tends to build on what is already there. Greenfield (1991) points out a more basic reason for relying on developmental evidence to reconstruct an evolutionary scenario, however: "Homologous origins of capacities across species imply homologous ontogenetic histories," (547). It is established fact that the frontal lobe develops in a back-to-front fashion in all mammals, and that the most prefrontal portions of the cortex (those associated with hierarchical information processing in both humans and chimpanzees) develop last and most fully in humans. The increasingly differentiated patterns in the development of hierarchical behavior are similar across humans and chimpanzees because they are located in functionally and structurally related (and therefore presumably homologous) parts of the brain that are similarly immature at birth and differentiate over the course of ontogeny. Although the behavioral similarities in the development of hierarchical capacities in these species are compelling on their own, anatomical similarities in the underlying neural structures are hard evidence of common phylogenetic origins and justify the use of developmental patterns as a means of making inferences about the evolutionary past.

In discussing the ontogeny of the hierarchical capacity for action and language, this chapter has touched on a number of other cognitive skills that are implicitly required by this developmental process. The emergence of language and praxis in human children is intricately related to and dependent upon a wide range of other abilities including but not limited to working memory and attention span, a theory of mind, and imitation. The following chapter explores the biological evolution of these various skills that make up the language-ready brain of modern humans, which set the stage for the subsequent cultural evolution of language.

6. The Respective Roles of Biological and Cultural Processes in Language Evolution

The previous chapter explored developmental evidence suggesting that conventional language and praxis are intricately related phenomena, each a manifestation of an underlying domain-general ability for the hierarchical organization and digital processing of information. This capacity for digital processing is combined with an analogic system of communication in full language, and both digital and analogic systems develop in parallel in human ontogeny. This chapter aims to place these fundamental abilities in an evolutionary context, demonstrating that they rely on a number of more basic skills and in combination are but one (albeit the most important) of several pre-adaptations for language. This view makes the connectionist assumption that "'language is a new machine built out of old parts' (Bates & Goodman, 1997), emerging from a nexus of skills in attention, perception, imitation, and symbolic processing that transcend the boundaries of 'language proper'," (Bates & Dick, 2001), and is opposed to the Chomskyan notion of an innate capacity for language that evolved as a distinct circuit in the brain adapted for communication (e.g., Hauser et al., 2002). Rather, each of these individual capacities emerged at various points in the evolutionary history of the hominin line (and some even earlier) as adaptations to serve a variety of functions, ultimately forming a cognitive suite with the combinatorial, manipulative, and symbolic qualities necessary to support the cultural evolution of language. For example, the hierarchical capacity discussed at length in this paper initially evolved in support of praxis and tool use, but had to be subsequently combined with essential other skills in order to support the complexities of language.

The position argued for in this chapter is largely inspired by the work of Morten Christiansen and his colleagues (e.g., Christiansen & Chater, 2008; Christiansen & Kirby, 2003), which posits that language may be understood as a complex organism that evolved to fit the brain and is thus shaped by constraints imposed by its various preexisting features. This framework for understanding language evolution has great explanatory power, accounting for the striking structural and developmental similarities as well as the neurological overlap of language and non-communicative praxis discussed in the earlier chapters of this paper.

6.1.1. Basic pre-adaptations for language

Certain abilities are of somewhat obvious importance in the larger capacity to use and understand language, and must necessarily have existed prior to the emergence of such a skill. A very basic requirement for language is that working memory and attention span be sufficient for retaining a complex sequence of symbols in one's mind long enough to interpret its compositional meaning. This ability is also of primary importance outside of communicative contexts, such as in keeping track of social networks, the execution of a planned action in praxis (e.g., tool production), or the imitation of an observed action that enables learning. Working memory places strict constraints on the length and complexity of sequences (of language or of action) that can be learned.

Another basic requirement for language is self-awareness and the ability to understand other individuals as intentional beings with minds of their own, which can be influenced by the actions of others. If this were not the case, it would be pointless to communicate with them through language because the messages contained therein would fall on uncomprehending and unresponsive ears. This ability, explained by some as a theory of mind (Premack & Woodruff, 1978) and by others as simulation theory (Gallese & Goldman, 1998), is a product of social living and arises from the need to interact with other members of the group in a variety of different ways (e.g., indicating, demonstrating, bonding and forming alliances, and even deceiving); indeed, the social life of our primate ancestors has undoubtedly shaped the nature of human communication (Dunbar, 1996). Christiansen & Kirby rightfully point out that working memory and attention span, theory of mind, and certain other pre-adaptations to language are "shared with other species, in particular other primates, and that differences in these skills may be more quantitative in nature rather than qualitative," (2003:302). There is little doubt that these skills are present in some form in other primates, but they are developed in humans to a much higher degree.

These basic skills are extremely important prerequisites to language, and a great deal more can be and has been said about them (citations?). However, it is precisely because they are shared with other primates that I will not go into them here any further; this chapter is concerned with the uniquely human pre-adaptations to language that have evolved through natural selection since the divergence of the hominin line from chimpanzees 4 to 6 million years ago, which together form a "language-ready brain" (Arbib, 2009) that enabled modern *homo sapiens* to collectively develop language through an extended process of cultural evolution.

6.1.2. Imitation

If one adopts the view just articulated – that language emerged through cultural evolution on top of a biologically evolved, language-ready neural substrate – then imitation must play a crucial role in that process. If language is not an innately programmed instinct that requires only minimal exposure to a particular grammar for successful acquisition, as Chomsky (1965) and others (e.g., Pinker, 1994) have argued, its development must have been fundamentally shaped by cultural transmission and the important learning mechanism of imitation. Arbib (e.g., 2008, 2009, 2011) has focused on imitation as a crucial factor in language evolution, making important distinctions between varying degrees of imitative complexity.

While grasping and manual praxic actions are present in monkeys, these primates have at best a very limited capacity for imitation enabled by a mirror neuron system for grasping (Arbib, 2008). Importantly, however, this mirror system in and of itself does not yield true imitation: while monkeys may have neurons that fire when executing or observing a given action, they do not necessarily repeat the action when it is observed in another, nor do they use observation to incorporate a novel action into their repertoire of manual praxis. These basic skills are far overshadowed by what Arbib (2008) calls *simple imitation*, as exhibited by apes. Unlike monkeys, apes are able to acquire completely novel actions through observation of repeated examples, usually over a long period of time. For example, chimpanzees required at least 12 trials to successfully imitate a novel behavior in a laboratory setting, and young gorillas in the wild take months to acquire the complex feeding strategies observed in adults. In these cases of simple imitation, apes tend to focus more on reproducing the observed outcome of an action rather than faithfully adhering to each of the individual steps that comprise the whole behavior. The successful imitation of novel behaviors in apes typically takes such a long time because these primates require a great many exposures to the action in order to recognize the relevant subgoals common to each performance (Arbib, 2008:5).

The simple imitative capacities of apes contrast sharply with those of humans, who are typically able to make sense of and reproduce quite complex behaviors and their constituent actions and subgoals after only a few exposures. Arbib (2008) calls this ability *complex imitation*, arguing that it is unique to the hominin line and was not present in the last common ancestor of humans and chimpanzees. Complex imitation enables the quick and effective transfer of novel skills within a community, and implicitly involves the ability to recognize patterns and parse hierarchically organized sequences of information. Indeed, complex imitation as described by Arbib (2008), coupled with the manipulative play described by Bruner (1972), is analogous with the general hierarchical capacity underlying praxis and language discussed at length throughout this paper. The emergence of complex imitation in the hominin line was likely in support of praxis, and probably developed in parallel with increasingly complex programs of action. For example, as hominins developed increasingly complicated strategies for tool production, involving more and more levels of hierarchical complexity, and passed them on to successive generations, a greater ability to identify and reproduce the relevant subroutines of these tool production techniques would have become increasingly important; in this way, complex imitation and tool production were mutually selective skills in hominin evolution. Importantly, the uniquely human ability to immediately recognize the constituent parts of a hierarchically organized sequence is precisely what enables us to comprehend and produce completely novel utterances in language. The biologically evolved ability for complex imitation was an essential prerequisite for the subsequent cultural evolution of language.

6.2. Language as shaped by the brain: constraints on the cultural evolution of language

With the biologically evolved capacities mentioned above in place – a large working memory, a theory of mind and understanding of intentionality, and complex imitation and the hierarchical information processing involved in it – *homo sapiens* was able to use this skillset (which was initially developed in support of various praxic actions) to support a much more complex and productive form of communication than was present in our last common ancestor with other primates. Christiansen & Chater (2008:489) liken this communicative system to a complex organism that evolved within the environment of the human brain, shaped by selective pressures from human learning and processing mechanisms. This approach suggests that the structure of language is a result of domain-general cognitive abilities and the constraints imposed by thought, the perceptual and motor systems, cognition, and pragmatics.

The first of these constraints, that of thought, rests on the assumption that thought is prior to and independent from language. This is an unsubstantiated and controversial assumption, as the exact relationship between thought and language is a largely philosophical question and the two have not been convincingly and empirically dissociated from one another. If one accepts this view, however, then it is possible to conclude that, "fundamental properties of language, such as compositionality, function-argument structure, quantification, aspect, and modality, may arise from the structure of the thoughts language is required to express," (Christiansen & Chater, 2008:501). While the relationship between language and thought is a potentially rich line of evidence explaining the structure and evolution of language, its entirely speculative nature renders it insufficient as a sole justification for the argument that language was shaped by the brain.

A much more concrete and convincing reason that language may have been shaped by the brain is that the perceptual and motor systems do in fact impose very real constraints on all human behavior, including language. The nature of human perception and motor activity prevents communication from taking certain forms (e.g., telepathy is not a viable means of communication because we do not have the perceptual and motor hardware to support such an endeavor), and limits such a system to certain channels. The expressive capabilities of the human body and the complex nature of the type of information conveyed through language forces us to produce messages that consist of sequences of discrete units varying along the single dimension of time (cf. de Saussure, 1916/1959). The form of these units – spoken or signed symbols – is determined by the ability to articulate existing bodily structures such as the vocal tract or the hands; in other words, linguistic structure is bootstrapped onto body parts and brain mechanisms that were initially adapted for other functions and force the productive communication of complex ideas to assume the form of a digital system. Christiansen & Chater's (2008) description of perceptual and motor constraints on language evolution is essentially a reference to the phenomenon of embodiment in thought and language (discussed in Chapter 2). The demonstrated role of embodiment in language is strong evidence that language was shaped through processes of cultural transmission and iterated learning to fit the already established structure of the human body and brain.

Christiansen & Chater's (2008) third constraint on the cultural evolution of language derives from general cognitive limitations on learning and processing. These authors focus particularly on the connection between general sequential learning (e.g., planning, praxis) and language, calling attention to the fact that "both involve the extraction and further processing of discrete elements occurring in complex temporal sequences" (2008:502). These similarities have long been recognized by psychologists (e.g., Bruner, 1972; Greenfield, 1991) and are explained by Arbib (2008) as stemming from the ability for complex imitation and attributed by me in this paper to a general underlying capacity for processing hierarchically organized information. Christiansen & Chater (2008) suggest that theoretical constraints in sequential learning and processing can explain universal tendencies in language. For example, even though transformational grammars and Chomskyan notions of linguistic competence theoretically allow for infinite embedding of constituents, general processing abilities tend to limit the number of recursive center-embedded clauses to at most two or three in actual linguistic performance, with further levels of complexity being perceived as ungrammatical. In a similar vein, computer simulations taking into account these theoretical processing limitations have indicated that "basic word order patterns may thus derive from memory constraints related to sequential learning and processing of linguistic material" (Christiansen & Chater, 2008:502). The notion of language being culturally shaped by general constraints on human cognition is a compelling one, and eliminates the need to explain language universals as a product of a biologically evolved instinct for Universal Grammar.

Christiansen & Chater's (2008) final theoretical constraint on the cultural evolution of language is imposed by pragmatics. In actual conversation, the intended meaning of any given utterance is intricately related to and determined by the wider context of the discourse (Austin, 1962), and evidence from diachronic linguistics shows that syntactic structures emerge over time from the reduction of discourse structures (Christiansen & Chater, 2008:503). There is a gradual diachronic shift from general pragmatic constraint to rigid syntactic rule, which is mirrored by the process of grammaticalization in which open-class lexical items gradually lose semantic meaning and become closed-class functional items through phonological reduction and increasingly strict dependencies with other words.[12] This brings up an important implication of the view that language is not a biologically evolved capacity, but rather a complex system that has emerged through a very prolonged process of cultural transmission across hundreds or even thousands of generations of learners: the well-studied topic of language change may have a direct bearing on language origins, as historical linguistics has documented the tail-end of a continuous process of cultural evolution.

It is important to note that language as a cultural product was shaped by each of the aforementioned constraints in combination. The combined effect of these constraints is

---

[12] Take for example English "to go," which in certain (but not all) usages has lost the original connotation of movement and serves as a functional marker of the future tense, as in "I am going to [do something]," which has been phonologically reduced to "I'm gonna" or even "I'ma" in some dialects.

significant, because none of them in isolation is qualitatively different from abilities possessed by apes. Rather, each of the mechanisms involved in Christiansen & Chater's (2008) constraint-based scenario for language evolution are quantitative refinements of older systems still present in other primates; it is the unprecedented degree of complexity that these mechanisms have reached in humans and their interaction in the brain that gives us the unique ability for language. This suite of cognitive functions constitutes a biologically evolved pre-adaptation for language – a "language-ready" brain – on top of which language evolved through general learning and processing mechanisms and cultural transmission across countless generations. Christiansen & Chater (2008:503) concede that, "initial changes, if functional, could have been subject to further amplification through the Baldwin effect, perhaps resulting in multiple cognitive shifts in human evolution," which would account for young children's uncanny predisposition to acquiring language. The fact that language is unique to humans does not require the postulation of a biologically evolved instinct, and language evolution is better explained as a process of cultural transmission constrained by a confluence of primate systems that are developed in humans to an unprecedented degree.

This chapter has outlined a compelling framework for understanding language evolution, in which language emerged through a process of cultural transmission across hundreds or thousands of generations. This development was shaped by constraints imposed by essential pre-adaptations, namely sufficient memory capacity to process linguistic input, a theory of mind that enabled an understanding of intentionality in other individuals as well as various other aspects of cognition arising from the social life of primates, and a capacity for complex imitation and sequential learning that enables humans to efficiently process hierarchically structured information and readily identify the constituent parts involved. While this framework outlines a highly plausible and probable scenario for the emergence of language, it does not specify the means by which simple and holistic communication in primates evolved into the highly complex and compositional linguistic system possessed by humans. The following chapter examines some more explicit possibilities for the origin of language, which consider the vital role played by hands in communication and praxis and account for the striking structural and developmental parallels between spoken language, gestural communication, and skilled manual action.

7. Plausible Scenarios for the Cultural Evolution of Language

The previous chapter presented a compelling account of the way language may have developed through cultural processes atop a biologically evolved suite of necessary pre-adaptations. Such a scenario describes linguistic abilities in humans as an extension of more general types of intelligence, made possible by learning and processing mechanisms basic to the human brain. This view nicely explains the remarkable structural and developmental similarities between language, gesture, and praxis discussed in Chapters 4 and 5. In this way, language is "a new machine built out of old parts," (Bates & Dick, 2001), the result of a quantitative expansion of systems shared by other primates. Crucially, language lies at the intersection of a digital channel and an analogic one, both of which can be observed in simpler forms in other primates (e.g., the basic grammars of action involved in chimpanzee tool use, the limited gestural repertoires and other "paralinguistic" media such as facial expression and innate vocalizations utilized in ape communication), but have been greatly expanded over the course of human evolution. This is an attractive alternative to the nativist views of Chomsky (e.g., 1965) and others, whose claim that language is a qualitative break from systems preceding it seems evolutionarily implausible and does not adequately explain certain similarities across modalities.

While Christiansen & Chater (2008) provide a powerful and productive framework for thinking about language evolution, their argument is very broad and does not take an explicit stance on what form language initially took – that is, whether a fully productive and compositional digital system of communication originated in the vocal modality or the gestural modality. This is a major debate within the field of language evolution, and this chapter will consider arguments for both sides. Some (e.g., Lieberman, 2003; Dunbar, 1996, 2003) argue that language emerged in the vocal modality, with speech gradually developing out of basic primate vocalizations; others (e.g., Corballis, 2003; Donald, 1991; Arbib, 2008) claim that spoken language was preceded by a productive system of conventionalized representational gestures. I argue for the latter view, as an initially manual system of digital communication would account nicely for similarities between digital language and praxis and analogic gesture, and proposes a smoother evolutionary transition than does the alternative.

Such a theory is not without its difficulties, however. Although this initial system of conventionalized representational gestures would have necessarily developed out of non-arbitrary, analogic forms of manual communication (i.e., iconic gestures, pantomime), it must have at some point co-opted the endlessly generative properties of previously evolved digital systems (i.e., non-communicative praxis) in order to yield the full expressive power of modern language. A fundamental question for the study of language evolution, then, is how and why analogic communication became supplemented by a digital system, and what that early digital system may have looked like. I argue that the answer lies in the human hand, as this is the most visible and easily manipulated organ of the body. Presented with no better candidate, each of these divergent systems focused on the hands as a representational modality (albeit in different ways), despite their contrasting formal characteristics. This convergence of digital and analogic systems on the same representational modality set the stage for the double evolution of manual movement: the gradual evolution of arbitrary and compositional protosigns on the one hand, and, on the other, the evolution of a parallel but different system of analogic gestures that accompany speech but never became compositional.

7.1. The modality debate: speech vs. gesture

There have traditionally been two contrasting positions explaining the evolution of spoken language out of an earlier, essentially primate communication system: one is that speech evolved directly out of a primate vocalization system (e.g., Dunbar, 2003; Lieberman, 2003), and the other is that it evolved via an intermediate stage of gestural communication out of an earlier primate capacity for praxis (e.g., Arbib et al., 2008; Corballis, 2003; Donald, 1991).

Lieberman (2003) argues for the former scenario on the grounds of his considerable anatomical study of the human vocal apparatus, specifically the position of the larynx and shape of the human vocal tract and tongue. These structures have been modified significantly in humans relative to other primates, allowing us to produce a wide range of vowel and consonant sounds impossible in our closest living primate relatives and even in Neanderthals, with whom we shared a much more recent common ancestor. This structural change in humans creates many possible problems for the basic functions of eating and breathing, however, and its only selective advantage is that it yields the ability to produce a wide range of discrete speech sounds. Lieberman (2003) claims that this evidence suggests that the neural capacity for speech production must have been in place before evolution of anatomically modern humans, presumably in the form of systems for earlier primate vocalizations.

Dunbar (2003) focuses more on vehemently rejecting the possibility of gestural origins than on actually articulating a plausible scenario in which human speech may have evolved out of primate vocalizations similar to the ones observed in monkeys and apes today. His main criticisms are that a gestural language would have required line-of-sight contact between communicators, making it useless in the dark and rendering its “speakers” unable to use their hands for other purposes (e.g., foraging) when engaged in communication. These are valid observations, but they do not necessarily eliminate the possibility of initial language taking a gestural form; the difficulties pointed out by Dunbar (1996, 2003) may well have placed constraints on an early gestural language and favored the selection of the less limited vocal apparatus. However, Dunbar (2003) goes to lengths to dismiss the possibility of gestural origins entirely, and instead argues that spoken language evolved in stages directly out of a system of primate vocalizations. He points to the fact that some modern monkeys possess a repertoire of distinct calls that correspond to specific semantic meanings (e.g., distinct vocalizations referring to important events such as food discovery or the presence of specific predators) and suggests that semantically laden calls could have at some point been combined to convey more complex messages. He also claims that these call systems, which at most serve to keep track of other group members in modern primates, may have developed into a richer form of maintaining contact at a distance to facilitate social bonding in groups of increasing size, eventually employing grammatical structures to convey social information with symbolic, arbitrary vocalizations.

Arguments that speech evolved directly out of primate vocalizations are weak at best, as they ignore significant limitations of primate calls (i.e., they are innate and holistic and do not lend themselves to productive combination to yield new meaning) and do not account for the important role still played by the hands in language today. Furthermore, the most valid criticisms made by the gestural theory's opponents do not demonstrate that gestural origins of language are implausible, but rather highlight the most important limitations of such a system that inevitably led to the selection of the vocal modality as a more flexible and efficient medium for linguistic communication. An early manual system of digital communication was a necessary intermediate step between the complex praxic abilities of apes and the uniquely compositional form of communication used by modern humans, and would have emerged alongside of and been closely coordinated with analogic systems such as gesture as we know it today and other paralinguistic phenomena.

Corballis (2003) and Arbib et al. (2008) present evidence on primate vocalizations that severely diminishes their likelihood of having been directly modified to yield human speech. First and foremost, primate calls are holistic and non-compositional: they can neither be broken down into smaller, interchangeable units of meaning nor meaningfully combined into larger sequences. While these calls are paired with distinct semantic meanings, they seem to be for the most part innate and involuntary and cannot be dissociated from the situation that triggers them. Chimpanzees are unable to suppress the distinctive call signaling the discovery of food even when they evidently wish to keep their find to themselves, nor can they produce such a call in the absence of the appropriate emotional state (Corballis, 2003:202). The involuntary nature of primate calls may actually be adaptive, as it prevents individuals from deceiving each other and is precisely what makes distinct calls reliable analogical indexes of the situations they signal. This property of primate vocalizations and of paralinguistic phenomena in general (e.g., coarse manual movements that do not rely on the individual manipulation of the fingers) makes them ill suited to exaptation for intentional communication. Analogic indicators such as these are limited in the quality of information they can convey on their own because they are holistic and lack compositionality.

On the other hand, complex manual actions can be and are imitated by apes, making them excellent candidates for the initial medium of intentional digital communication. Imitation of complex forms of praxis in the hominin line may have been repurposed to serve a communicative function in the form of pantomime or *mimesis*, as suggested by Donald (1991; see below). Furthermore, the parity requirement of language – whereby a speaker's intended message is analogous with the listener's interpretation of it – may have been made possible by a putative mirror system for manual action in human ancestors, as advocated by Arbib (2005).

7.2.1. Major transitions in the evolution of cognition, culture, and language

Merlin Donald's (1991) theory of cognitive evolution is very much concerned with the development of culture and language in humans, and suggests that the modern manifestation of these two traits was dependent upon earlier transitions in the biological evolution of the hominin line. This theory is closely aligned with that of Christiansen & Chater (2008), but provides a finer level of detail in describing the evolutionary stages preceding the emergence of the modern mind. Donald (1991) specifically proposes three cognitive transitions in the evolution of modern humans since the last common ancestor shared with other apes, of which the first two are biological and the third is cultural.

The first transition is described as a movement from the "episodic" level of culture of apes to the "mimetic" level of culture of *Homo erectus*. Ape culture is described as episodic because "their lives are lived entirely in the present, as a series of concrete episodes, and the highest element in their system of memory seems to be at the level of event representation," (Donald, 1991:149). The mental representations of apes are bound to the concrete situation or episode at hand, whereas humans are capable of abstract, symbolic, semantic representations that can be stored in memory and recalled in the future and as such are not subject to situational limitations. Apes are highly skilled at event perception, capable of perceiving complex patterns of behavior as cohesive, goal-directed actions; indeed, we consider apes to be so intelligent because they are able to read aspects of behavior and recognize patterns of much greater complexity than can other animals. Arbib (2008) describes this capacity as simple imitation, the ability to extract and reproduce the constituent sub-goals and actions of a given behavior after observing it many times. The imitative abilities of apes are simple – that is, they are dependent upon many repeated exposures to a behavior – precisely because apes are limited to an episodic culture. Whereas humans can quickly and easily abstract generalities of an observed behavior and compare them with similar symbolically represented experiences across different contexts and situations (i.e., complex imitation), apes are limited to the episode at hand and struggle to dissociate the general, abstract behavior from its immediate context.

Donald (1991) claims that *Homo erectus* was the first in the hominin line to break free of the constraints imposed by ape-like episodic culture. The transition was made possible by the advent of abstract, symbolic mental representation unbounded by situational context. This shift yielded complex imitative skills and a "mimetic" culture characterized by "the ability to produce conscious, self-initiated, representational acts that are intentional but not linguistic," (168). Mimesis is distinct from complex imitation in that it serves a communicative purpose, adding a symbolic and representational dimension to hierarchically complex action sequences. It makes use of many actions and modalities, such as "tones of voice, facial expressions, eye movements, manual signs and gestures, postural attitudes, patterned whole-body movements of various sorts, and long sequences of these elements [to] express many aspects of the perceived world," (169). Mimesis is the ability for pantomime, a symbolic representation of bodily action that can be produced and understood outside the context in which it usually appears. Importantly, pantomime possesses both digital and analogic properties: it is still mimetic and represents information in a non-arbitrary, embodied, analogic way, yet it is able to combine these analogic symbols fairly productively in a digital fashion. The mimetic culture of *Homo erectus* was an important precursor to full language – it marks the first use of digital capacities for intentional communication – and was a necessary intermediate stage in the biological evolution from our ape-like ancestors into modern humans. Although the human mind has since undergone additional evolutionary transitions, vestigial forms of the mimetic stage are still present in modern language and culture (e.g., in the form of analogic channels of communication).

Donald's (1991) second transition in cognitive evolution was from the mimetic culture of *Homo erectus* to the modern culture of *Homo sapiens*. This stage apparently saw the completion of the biological evolution of modern humans, including the speech apparatus that facilitates modern spoken language. Donald (1991) does not propose that human ancestors moved from simple mimetic capacities to a fully arbitrary system of vocally articulated symbols in one fell swoop; there had to be some intermediary step in which mimetic skills were expanded to not only represent an action or emotion in an abstract and symbolic way, but to facilitate the invention of a wholly new, arbitrary class of symbols. He identifies the gestural modality as the most likely candidate for original arbitrary symbols, as mimetic performance (i.e., the digital combination of analogic representations) became conventionalized and ultimately dissociated from its original context of meaning.[13] The mimetic mind may indeed have produced fairly elaborate systems of gestural symbolism that preceded the speech adaptation during the cognitive transition from *erectus* to *sapiens*. The development of these symbolic systems laid the foundations for a new type of mental model, entirely distinct from the purely perceptual models that underlay episodic and mimetic representation. A system of arbitrary symbols does not directly model its referents – it only signifies them. This is a wholly new type of representational system, in which knowledge of the physical world and the symbols that refer to that knowledge have become dissociated. At this point, the system of arbitrary symbols becomes complete unto itself, and the invention of new symbols becomes much simpler in the sense that new meanings

---

[13] This is an important point that needs further clarification. Donald (1991) focuses on the analogic system of mimesis as the transition point leading to modern language and cognition, but these abilities are dependent on a compositional system that is more closely related to digital praxis. This point is elaborated on below in the discussion of Arbib's (2005, 2008) theory.

can be attributed to old symbols and old symbols can be modified to accommodate new meanings; furthermore, because these symbols have been dissociated from their referents, it becomes easier to manipulate their configuration and recombine them in new ways to produce new meanings. However, a purely gestural repertoire of arbitrary symbols is subject to important constraints (e.g., it requires a direct line of sight and cannot be used effectively at a distance or in the dark; a "speaker" is unable to use his or her hands for other purposes when engaged in communication; Dunbar, 1996, 2003), and thus there is a selective advantage for an alternate means of using arbitrary symbols: the speech apparatus.

Importantly, the cognitive and evolutionary transitions proposed by Donald (1991) up to this point can effectively support full-blown language; the cognitive hardware is there, it is simply missing an expedient vehicle (i.e., speech) to put this underlying linguistic ability to use. This cognitive hardware essentially constitutes the biologically evolved language-ready brain argued for by Christiansen & Chater (2008): the digital capacity to represent the physical world symbolically and the ability to combine those symbols into hierarchically organized, compositional structures (just as non-communicative constituent actions are combined in praxis) are there, they just need to be integrated into a cohesive system alongside pre-existing forms of analogic communication through collective cultural processes. However, cultural processes alone could not have bridged the gap from a system of arbitrary symbolic gestures to the modern system of symbolic communication mediated by a "high-speed phonological device," (Donald, 1991). There must necessarily have been a subsequent evolutionary process in which fine voco-motor skills and the production of discrete units of sound were selected for as a biological adaptation for symbolic communication. This fact is not incompatible with Christiansen & Chater's (2008) argument for a language-ready brain, however; their main point is that the capacity for language was not a functional adaptation for communicating information, but rather a serendipitous result of the confluence of several more general cognitive functions. This capacity did emerge serendipitously out of the mimetic culture of *Homo erectus*. The fact that digital linguistic abilities were initially realized in the manual modality and that the vocal modality was subsequently enlisted and functionally adapted to better serve those abilities is beside the point. Regardless of modality, the essential development was a hierarchically organized symbolic system for digital communication (which emerged alongside of increasingly elaborate analogic forms of communication and ultimately combined and conventionalized these representations through pantomime), the structure of which was constrained by general properties of cognition and shaped by cultural transmission across countless generations of human ancestors. This early digital system of manual communication would lie somewhere near the right extreme of Kendon's Continuum (McNeill, 1992), although it would not be as fully productive and conventionalized as modern sign language.

According to Donald (1991), this second cognitive transition yielded a fully productive system of spoken language; however, this transition could just as well be described as containing a number of individual transitions within it, such as the cultural invention of arbitrary symbolic systems and the subsequent biological adaptation for speech. Regardless, Donald (1991) attributes this ability to archaic *Homo sapiens*, and proposes a third and final transition that was relatively recent and entirely cultural. This was the invention of "visuographic" representation, including basic self-adornment (e.g., body-painting, ritual scarring of the body, elaborate costume), truly pictorial representation (e.g., cave paintings), and eventually writing systems. Donald (1991) refers to this transition as the advent of external symbolic storage, and compares this invention to "providing the CPU of a computer with an external storage device, or more accurately, with a link to a network," (17). The emergence of abilities such as reading and writing has certainly changed the way in which humans store and access information, and has resulted in rapid cultural change that would not be possible without it. While this is an interesting observation and a valid point, however, this third and final proposed transition in cognitive evolution is not entirely relevant to the discussion at hand. I am concerned with the origin of full, digital linguistic ability characterized by arbitrary reference, compositionality, and generativity – which Donald (1991) claims to have emerged during the transition from *Homo erectus* to archaic *Homo sapiens* – and the medium to which that original linguistic ability may have been confined.

7.2.2. The mirror system hypothesis for gestural origins of language

Michael Arbib (e.g., 2005, 2008) has argued at length for a gestural theory of language origins based on the discovery of a system of mirror neurons in both premotor area F5 in monkeys and Broca's area in humans, which are homologous brain regions. This mirror neuron system is activated for both the execution and observation of manual actions, and its presence in both humans and monkeys implies that a mirror system was also present in some form in the common ancestor of modern humans, apes, and monkeys. Arbib (2005) claims this mirror neuron system served as a neural substrate that underlay the evolution of manual praxis in monkeys and apes as well as the emergence of more complex abilities in the hominin line, culminating in a language-ready brain in *Homo sapiens*. The subsequent development of full-blown language as observed in modern humans was a historical process fueled by cultural transmission and shaped by constraints imposed by the already-established structure of the brain, involving a minimal degree if any of further biological adaptation. This hypothesis resonates deeply with the theory proposed by Christiansen & Chater (2008), although Arbib (2005) argues specifically for a gesture-first scenario and provides a much greater level of detail than these authors or Donald (1991), explaining precisely how the necessary pre-adaptations of language came together. Arbib's (2005) mirror system hypothesis outlines seven distinct stages in the evolution of language out of the abilities of our distant nonhuman ancestors.

The necessary starting point of this evolutionary progression was the establishment of cortical control over the hands, endowing the ancestor of the great apes and monkeys with the ability for manual praxis. The next stage was the evolution of a mirror neuron system for grasping, still shared with the common ancestor of humans and monkeys. At some point before the following stage, the ancestral line of the great apes diverged from that of monkeys. The third stage saw the development of a simple imitation system for manual actions and was unique to the great apes. As mentioned in the previous chapter, a mirror system alone is not sufficient for imitation; monkeys may have neurons that fire when observing a manual action, but it does not use that neural activation to incorporate novel actions into its own praxic repertoire. Apes, on the other hand, are able to imitate novel actions through repeated observation, and Arbib et al. (2008) point out that this capacity is sufficient to support a rudimentary system of novel communicative gestures, albeit a very limited one.

The fourth stage represents the divergence of the hominin line from our last common ancestor with chimpanzees, seeing the emergence of a complex system of imitation for manual actions. Arbib (2005) describes this stage as "the ability to recognize another's performance as a set of familiar actions and then repeat them, or to recognize that such a performance contains novel actions which can be approximated by variants of actions already in the repertoire," (107). Unlike the simple imitation system of other apes, hominins equipped with the capacity for complex imitation could recognize the component parts of a novel action and successfully imitate them after only a few exposures to the behavior. Complex imitation may have been the highest achievement of the episodic hominin culture described by Donald (1991), making possible his proposed cognitive transition to the mimetic culture of *Homo erectus*. Arbib's (2005) fourth stage marks the first appearance of the uniquely human ability for complex hierarchical and combinatorial manipulation underlying the eventual development of linguistic syntax.

Each of the stages of Arbib's (2005) mirror system hypothesis described up to this point are solely involved with manual praxis, and not with intentional communication. Although Arbib et al. (2008) claim that simple and complex imitation may have supported a rudimentary system of novel gestures – possibly acquired through a process of ontogenetic ritualization[14] – although such a system would have been severely limited and essentially closed to innovation. Hominins at this evolutionary stage would in all likelihood have also possessed a fixed repertoire of involuntary, innately specified vocalizations (although the occasions for using such vocalizations may change with experience in modern apes; Arbib, 2008); essentially, the communication system of human ancestors was at this point entirely analogic, possessing the full range of holistic paralinguistic phenomena observed in modern analogic communicative systems. The fifth stage of the mirror system hypothesis, *protosign*, broke through the fixed repertoire of primate vocalizations and rudimentary system of novel gestures to yield a manual-based digital communication system with an open repertoire; importantly, this new digital system did not replace the already existing analogic system, but rather diverged from it and continued along a parallel and complementary yet distinct evolutionary path. This evolutionary step is analogous

---

[14] Ontogenetic ritualization is a process in which an "increasingly abbreviated and conventionalized form of an action may come to stand in for that action, an example being a beckoning gesture recognized by the child as standing for the parent's action of reaching out to grasp the child and pull it closer," (Arbib, 2008:5).

with Donald’s (1991) notion of a mimetic hominin culture, in which individuals used pantomime to create novel communicative utterances that could be understood upon first observation. Arbib (2008) argues that pantomime would have first been used to demonstrate grasping and manual praxic actions, and later to signal objects and events that did not involve manual action at all (e.g., flapping one’s hands to mime a flying bird). Although pantomime would have marked a watershed in the evolution of communicative ability – it marks the first use of digital capacities for intentional communication[15] – inventing an entirely novel pantomime for every act of communication would be a taxing and inefficient endeavor. For this reason, Arbib (2008) claims that elaborate pantomimes were soon conventionalized into a stylistic, socially learned and shared set of protosigns, possibly through ontogenetic ritualization within hominin groups. This development would essentially represent the digitization of the originally analogic symbols combined by pantomime.

The sixth stage of Arbib’s (2005) hypothesis is the development of *protospeech*, which combined with protosign to create *protolanguage*. Once the use of protosign had been established in the brain and in hominin culture, neural mechanisms evolved for the open-ended creation of arbitrary manual signs could have been extended to control the vocal apparatus with increasingly flexibility, eventually allowing for the incorporation of non-innate vocalizations (i.e., “protowords”) into the communicative repertoire. In this way, protosign provided the

---

[15] This intermediate stage combining both digital and analogic systems may have been mirrored in the non-communicative domain in the form of finger counting and the emergent capacity for numeracy. Basic systems for counting on the hand and fingers – which are diverse and have been documented in many cultures throughout recorded history (Ifrah, 1981) – are analogic in the sense that they use real entities (i.e., the hand, fingers, joints) to directly and non-arbitrarily represent quantities in the actual world, yet they are digital in the sense that they are combinatorial and productive (e.g., one system documented in China allows one to count up to 100,000 on a single hand; Ifrah, 1981:67). This dual digital-analogic character is similar to that of pantomime, and may have emerged around the same time in the hominin lineage (Bradd Shore, personal communication).

essential scaffolding for protospeech, creating an evolutionary niche and selective pressure for Donald's (1991) high-speed phonological device. Arbib (2008) characterizes protolanguage as "an expanding spiral of conventionalized manual, facial, and vocal communicative gestures," (5). Protolanguage included more than just the digital systems of protosign and later protospeech, however, incorporating the more ancient analogic system of communication as well. These two systems were highly complementary, mutually reinforcing each other throughout Arbib's (2008) expanding spiral and eventually developing into more complex and nuanced systems (i.e., complete linguistic systems and fully modern co-speech gestures), all while retaining their contrasting formal characteristics.

By stage six of the mirror system hypothesis, the hominin line had completed the evolution of the language-ready brain. Stage seven was the cultural evolution of modern language, with human ancestors recruiting novel manual and vocal gestures to communicate an expanding set of meanings. It is unclear whether this emergent lexicon was initially word-like, resembling the familiar verbs and nouns of modern language, or holophrastic constructions that could be best approximated by several words in English (Arbib, 2008). In the first case, syntax would have emerged as a means of employing the digital capacity for hierarchical and combinatorial manipulation to put protowords together in such a way as to yield a compositional meaning; in the latter case, syntax would be the result of that same digital capacity being used to decompose holophrases into meaningful units and figuring out how to recombine those units into new and meaningful configurations. Either way, syntax was a product of the same capacity for complex imitation and hierarchical processing that emerged in the hominin line after we diverged from our common ancestor with chimpanzees, albeit a highly refined version of that capacity. The development of all of the world's modern languages may be understood as a historical process of cultural transmission across countless generations, fueled by the same mechanisms of language change studied by diachronic linguists today (e.g., grammaticalization).

Arbib's (2005, 2008) account of language evolution is a much more specific and detailed version of the theory espoused by Christiansen & Chater (2008). By proposing a gestural origin of language (i.e., a manual digital system), Arbib's (2008) scenario is able to account for the remarkable similarities that exist between praxis and language. The way in which protosign and protospeech and a wider system of paralinguistic phenomena evolved together in an expanding spiral is an especially compelling argument for the co-evolution of co-speech gesture and conventional language as a single hybrid system of communication; although the vocal modality ultimately overtook the manual modality as the primary medium for linguistic communication, analogic gestures are vestigial reminders of the way in which language must have first been expressed.

7.3. From gesture to speech

But how exactly did predominantly manual systems of digital communication give way to speech? Arbib's (2005) seven-stage account of language evolution proposes this transition matter-of-factly and without detailed explanation. It is important to note here that language is an inherently multi-modal form of communication, employing not only the vocal modality but also manual gestures and facial expressions and even broader forms of "body language" to convey nuanced semantic and pragmatic content. This would have been true of earlier stages in the evolution of language as well, as hominins made use of whatever medium of communication was available to them; as Arbib (2005) suggests, protolanguage necessarily evolved as an expanding spiral of manual, facial, and vocal communicative gestures. A reasonable theory of language origins should not claim that initial language was entirely restricted to one modality or the other

– early hominins would have been equipped with the same manual skills and involuntary vocalizations as their ape-like ancestors. Arbib's gestural theory simply argues that those highly advanced manual skills and the capacity for complex imitation were more amenable to conveying the first communicative symbols. That does not mean, however, that those early symbolic gestures were not accompanied by the same noisy chatter observed in other apes today.

Arbib (2008) provides comparative neurological evidence to further explain how protosign paved the way for protospeech. As mentioned at the beginning of this section, the putative mirror system in the human Broca's area corresponds to area F5 in the monkey premotor cortex. F5 in monkeys is not concerned with vocalization but rather with grasping and other manual actions. This suggests that Broca's area in humans initially evolved on top of a neural substrate for manual praxis rather than vocalizations, and only later expanded to vocal and oral areas of the motor cortex to allow more flexible control of vocalizations in coordination with manual gestures. In this way, manual gesture and pantomime were able to ground protospeech via the earlier adaptation for protosign.

This chapter has considered more specific scenarios for the evolution of language within the language-ready brain paradigm proposed by Christiansen & Chater (2008). It seems implausible that spoken language evolved directly out of a system of primate vocalizations, considering the innate and involuntary nature of these calls. A much more likely explanation is that modern language emerged via an intermediate stage of digital manual communication, yielding an open system of arbitrary symbols that paved the way for a system of non-innate vocalizations. Arbib (2005, 2008) outlines a very explicit series of evolutionary progressions that culminated in a language-ready brain, which could then yield modern language through cultural evolution. An exciting implication of this fine level of evolutionary detail is that it may be possible to identify when and where these successive stages in the development of language occurred. Although brains do not fossilize, it is still possible to find behavioral correlates of cognitive abilities (e.g., stone tools and cultural artifacts) in the archaeological record. The following chapter will explore archaeological lines of evidence in an attempt to ascertain approximately when and where language evolved.

8. Archaeological Lines of Evidence for the Study of Language Evolution

In the previous chapter, I discussed specific scenarios for the evolution of language compatible with the language-ready brain theory of Christiansen & Chater (2008), ultimately advocating a theory of gestural origins. Arbib (2005, 2008) argues persuasively that open-ended symbolic communication was probably initially supported by the hands rather than vocalizations, and clearly outlines several progressive stages of hominin evolution whereby older primate systems were modified and integrated to yield modern language. A great advantage of this hypothesis is that it makes it possible to trace the evolution of language through a deductive process of piecing together the evolutionary histories of the various anatomical features and cognitive abilities that make language possible. While language itself does not fossilize, some of the body parts involved do, and the skills that it requires are also necessary for the production of other material cultural artifacts that can be found in the archaeological record.

This chapter examines two different lines of archaeological evidence for language evolution: anatomical changes in the hominin line since its divergence from other apes and material items manufactured by human ancestors, the complexity of which can serve as an index of the cognitive abilities of their creators. Based on their position in the archaeological record, evidence of these two types as well as certain types of genetic data can help place upper and lower limits on when language may have emerged in our evolutionary past, as well as where that may have happened. Keeping in mind Christiansen & Chater's (2008) theory, it is worth mentioning that the first type of archaeological evidence discussed here is concerned with the biological evolution of a modern brain that could support language, while the second type of evidence is more indicative of the processes of cultural evolution that yielded language as we now know it. Therefore, it will only really be possible to roughly identify the time and place at which the language-ready brain evolved; the cultural evolution of modern language would have necessarily occurred in many different places and possibly on different time frames among different human groups.

8.1. Anatomical change and the biological evolution of modern humans

The most direct means of making inferences about the evolution of language is by examining hominin fossils and determining what physical capacity for using language these creatures may have possessed. Certain anatomical features are particularly useful in making such an assessment, namely: the hands (of obvious importance for a theory of gestural origins), the brain (or more specifically, the size and shape of the cranium), and the vocal tract (because modern language is, after all, a predominately spoken phenomenon). Even without focusing in on the specific body parts important for language and their respective histories, however, it is possible to establish a broad temporal window within which language must have emerged through a relatively simple process of evolutionary reasoning.

8.1.1. Temporal limits for the biological evolution of the language-ready brain

Considering that humans are unique in their capacity for fully expressive language, this skill must have evolved in the hominin line since its divergence from other apes after the last common ancestor with chimpanzees approximately 4-6 million years ago (mya). Language is unlikely to have appeared until much later, however, as general consensus is that the earlier members of the hominin line, the *Australopithecines*, were much more apelike in their cognitive capacities than human-like (e.g., Stanyon et al., 1993; Ruff et al., 2007; Holloway, 1993); even the earliest hypothesized origins of language do not attribute this capacity to any species prior to the *Homo* lineage, which began approximately 2.3-2.4 mya with *Homo habilis* (or *H. ergaster* – the taxonomic classification of early *Homo* fossils is a matter of contention in physical anthropology; Hoffecker, 2007). Interestingly, the emergence of *Homo* roughly coincides with evidence for the earliest stone tool use around 2.5 mya (Stout, 2011), which as discussed in Chapter 4 bears striking structural similarities to hierarchical systems of linguistic communication, and may indicate the beginnings of linguistic abilities in human ancestors (this possibility is discussed in greater detail in section 8.2). Taken as a whole, these evolutionary facts provide a reasonable uppermost limit of roughly 2.5 mya for the earliest possible emergence of the modern capacity for language.

Establishing a lower limit for language evolution – the point at which the biological evolution of a language-ready brain absolutely must have been complete – is a considerably more difficult task. Physical anthropology has documented the staggeringly complex evolutionary trajectory from a common ancestor with chimpanzees in Africa to anatomically modern humans, drawing an intricate family tree of species branching out of, coexisting with, and superseding one another through successive waves of dispersal out of Africa and throughout the world. Despite the wide diversity of hominin species that has roamed the earth over the past several million years, however, only one remains today: *Homo sapiens sapiens*. The fact that every human on the planet is a member of the same species means that we have all descended from a single ancestral group of hominins. This hypothesis, espoused by Darwin in his 1871 publication *The Descent of Man*, remained largely speculative until it was corroborated by molecular genetic evidence in the 1980's.

With each successive generation, the genetic makeup of two parents is mixed together to yield unique offspring; this is the biological mechanism on which natural selection acts to yield genetic diversity and drive the evolutionary process. While most of the human genome is admixed with that of the other parent, however, the mother's mitochondrial DNA (mtDNA) is always inherited intact. Based on a comparison of the mtDNA of a large group of modern humans from all over the world, Cann et al. (1987) hypothesized that all living humans inherited their mtDNA from a single woman living in Africa between 150,00 and 200,000 years ago, and that modern humans have therefore descended from a small ancestral gene pool from that place and time. Similar analyses of much larger subject pools have since confirmed this hypothesis (e.g., Vigilant et al., 1991; Liu et al., 2006), all implicating a "Mitochondrial Eve" that lived in East Africa about 160,000 years ago.

These and other studies of mtDNA also show that genetic diversity is greatest within Africa and decreases steadily with geographic distance from this purported origin point, implying that Africa has a longer history of population by modern humans than any other region. Collectively, genetic data suggest that modern humans descending from a common ancestor such as Mitochondrial Eve 160,000 years ago may have first spread throughout Africa and eventually outwards into the Levant and Eurasia and the rest of the world. The first ventures out of the African continent may have occurred as early as 125,000 years ago, as indicated by archaeological evidence from the Arabian peninsula, but these early attempts at migration were apparently unsuccessful; current genetic data ties all modern human groups outside of Africa to a small founding group of approximately 1000 effective individuals which probably left the continent about 60,000 years ago (Liu et al., 2006). These modern humans rapidly established new groups with high population-growth rates and were the first to succeed in dispersing throughout the globe, with both the initial migratory wave as well as subsequent migrations out of Africa reaching all corners of the major continents and displacing any earlier hominin species (e.g., Neanderthals, populations of *H. erectus* that had left Africa at an earlier date) that may have already been living there.

The genetic theory of a recent common origin of *H. sapiens sapiens* in Africa and its subsequent migration out of that continent receives further support from the archaeological record. The oldest fossils of anatomically modern humans were found in the Middle Awash region of Ethiopia and dated to 160,000 years ago (White et al., 2003). Virtually all fossils of fully modern humans outside of Africa have been dated to more recent times that fit with Liu et al.'s (2006) suggested initial successful migration 60,000 years ago (e.g., Brown, 1992), and the few exceptions to this pattern have been found in the Near East and are typically interpreted as early offshoots of the ancestral population that either did not survive or retreated back to Africa.

Genetic and archaeological evidence strongly suggest a lower limit of 160,000 years ago for the completed biological evolution of modern humans with the capacity for language. Furthermore, all evidence overwhelmingly points to African origins of *Homo sapiens sapiens* and, presumably, its linguistic abilities. The idea of an African origin of language is not without empirical support, either: in a recent global linguistic survey, Atkinson (2011) has found that African languages possess the greatest phonemic diversity, with phonemic inventories growing less and less varied with increasing geographical distance from that continent. Atkinson's (2011) claim that these findings support a theory of African language origins has been met with considerable controversy and criticism, and it is true that these findings on their own are not a strong endorsement of such a theory; when considered alongside other evidence from genetics and archaeology, however, the geographic distribution of modern language diversity becomes yet another piece in a very convincing case for the African origins of both modern humans and at least the beginnings of modern language approximately 160,000 years ago.

Based on various lines of evidence and sound evolutionary reasoning, it is reasonable to assume that the language-ready brain evolved in Africa sometime between 2.5 mya and 160,000 years ago. Prior to and within that temporal range, however, several important anatomical changes must have taken place to differentiate modern humans from earlier hominins and endow us with the physical capacity to use language.

8.1.2. Evolution of the hand in relation to language

Even without assuming gestural origins, the evolutionary history of the human hand is of vital importance to that of language. Language and goal-directed manual action share a parallel hierarchical structure, and the general capacity for one is undeniably related in some fashion to the general capacity for the other (Chapter 4). However, the form and function of the human hand is obviously much older than language, as demonstrated by comparative primate evidence, and was in all likelihood an important prerequisite for the development of linguistic abilities. This is not to say that the modern human hand has been inherited as is from some earlier primate species; human hands differ in certain ways from those of our ape-like ancestors, and important anatomical changes must have occurred in the evolutionary past to bridge that gap and make language possible.

All primate hands are well adapted for prehension – that is, grasping objects. While the hands and fingers may be put to other purposes (e.g., pushing, poking, communicative gesture), their evolutionary history and anatomical form have been shaped by the adaptive function of grasping objects, specifically tree branches (Connolly & Elliott, 1974). Prehension in primates can be classified in two ways: the power grip – formed by clamping the fingers and the palm around an object and applying added pressure with the thumb – and the precision grip – formed by pinching an object between the fingers and the thumb (Napier, 1960). The power grip provides greater force at the expense of the fine dexterity and mobility afforded by the independent use of digits in the precision grip. The power grip exists in similar forms throughout the primate world, well suited to life in the trees, but precision grips are less common and humans are capable of greater precision and dexterity than any other species. Importantly, a hand capable of only the power grip is sufficient for analogic communication systems, although digital systems require the dexterous control of individual fingers that comes with the precision grip; as such, one would not expect digital language to have evolved in the hominin line prior to the emergence of exceedingly precise gripping capabilities.

Young (2003) claims that, once hominins had left the trees and developed bipedal locomotion, the power and precision grips were modified and put to the unique uses of clubbing and throwing, respectively. These functions would have conferred a selective advantage on human ancestors and fueled anatomical changes in favor of more efficient clubbing and throwing, which in turn served as pre-adaptations for the later development of stone tool use and construction. Indeed, “a fully modern human hand seems to emerge only after 1.7 mya with *Homo erectus* and is broadly correlated with the handaxes of the Acheulean,” (Hoffecker, 2007:365); this time frame falls within the temporal limits established in the previous section. As suggested in Chapter 4 and argued later in this chapter, the development of lithic technologies then paved the cognitive path for the eventual emergence of language.

8.1.3. Evolution of the brain in relation to language

The adaptation of the hands for distinctly human behaviors culminating in the physical capacity for using and constructing stone tools was likely accompanied by significant changes in the hominin brain. Fossil remains indicate that the absolute size of the brains of the earliest members of the hominin line, *Australopithecus*, was not significantly greater than that of modern chimpanzees (which have cranial capacities of 275-500 cubic centimeters) and presumably that of our common ancestor with those species, implying that they were more similar cognitively and therefore behaviorally to modern apes than humans (who, in contrast, have cranial capacities ranging from 1,000-1,850 cc) (Stanyon et al., 1993). The first significant increase in absolute brain size as well as brain size relative to the size of the body occurred with the beginning of the *Homo* lineage, as indicated by fossil skulls of *H. habilis* dated to as early as 2.3 mya with cranial capacities greater than 600 cc, and the fossil record shows a steady increase in cranial capacity in subsequent hominin evolution. Interestingly, the onset of this trend of increasing cranial capacity fits closely with the upper limit for the emergence of the language-ready brain around 2.5 mya identified earlier.

Cranial capacity alone is not a sufficient measure of the intelligence of early hominins, however, as absolute brain size is not necessarily correlated with cognitive capacity (for example, elephants have larger brains than humans but are not considered to be more intelligent). A better indicator is relative brain size, specifically the encephalization quotient (EQ) of a species, defined as the ratio between actual brain mass and the brain mass expected of an animal of a given size. This measure is widely regarded as a more reliable indicator of animal intelligence, as it takes into consideration allometric effects of increasing body size. Modern humans have the highest EQ of any living species at 7.4-7.8, compared to the 2.2-2.5 of chimpanzee brains (Ruff et al., 2007). An analysis of estimated body mass and brain mass in the genus *Homo* paints a similar picture as data on cranial capacity, with a steady increase in EQ since the earliest members of *Homo* (Ruff et al., 2007).

While these two measures are useful tools for making inferences about the evolution of the brain and cognition, they are indirect. Cognitive function is determined more by the internal structure and organization of the brain than its overall size and mass; unfortunately, the brain is soft tissue and is not preserved in the archaeological record. In certain exceptional cases, however, the hollow cavity inside the skull of a human ancestor can become fossilized and preserved in the form of an endocast (this process can also be simulated artificially on well preserved crania). Endocasts do not reveal the inner structure of hominin brains, but they often preserve the outer shape of the brain in considerable detail and make it possible to gauge the size of brain areas situated close to the surface (e.g., Broca's area). Holloway (1983) examined endocasts of several different hominin species at various stages in human evolution and found that, "at 1.8-2.0 million years, there is clear fossil evidence for a *Homo* lineage showing a more modern and enlarged third frontal inferior convolution, expanded brain size (e.g., 750 ml [cc]), and strong cerebral asymmetries identical to those known for modern *Homo sapiens*," (105). The enlarged brain region he refers to contains Broca's area and other important circuits for language processing in modern humans, and cerebral asymmetry in *Homo sapiens* results in the unique human trait of handedness and is thought to be a result of restructuring due to the evolution of language systems in the brain. Once again, the collective archaeological evidence of brain evolution indicates that sufficient brain capacity and some of the important neural structures for language began to co-emerge with the beginning of the *Homo* lineage, fitting nicely with the proposed upper limit for language evolution of 2.5 mya.

8.1.4. Evolution of the vocal tract in relation to language

The two anatomical changes discussed so far – those involving the hands and the brain – would have contributed to the development of the language-ready brain and occurred before its evolution was complete; as such, these anatomical changes were necessarily the result of evolutionary processes originating long before that of language and reaching completion sometime prior to or within the temporal range outlined earlier. The final important change in hominin anatomy that was physically required to support language is the fully modern vocal tract, which, in the gestural theory of language origins advocated here, developed after the emergence of a language-ready brain as a biological adaptation to better utilize the open-ended communicative system of arbitrary symbols established by gesture. Lieberman's (2003) work on the human vocal tract sheds some light on when it may have evolved, even though he uses this evidence to argue for an opposing theory of language origins in which gesture played little to no role.

Like the brain, the vocal tract does not fossilize and its evolutionary history must be reconstructed indirectly through analysis of related parts of the skeleton. Some of these indexical features include the shape of the basicranium, length of the neck, size of the hypoglossal canal, and – in rare cases where it can be found – the size and position of the hyoid bone (Lieberman, 2003; Hoffecker, 2007). The larynx is positioned considerably lower in humans compared to other primates, enabling us to produce the wide range of consonant and vowel sounds used in languages around the world. The position of the hyoid bone in a Neanderthal skeleton dating to approximately 60,000 years ago (Arensburg et al., 1989) indicates that even this closely related hominin species lacked the descended larynx required for the production of modern language sounds. Humans and Neanderthals share a relatively recent common ancestor at 650,000 years ago and possibly later (Castro et al., 1997), and the fact that this close evolutionary cousin did not possess the physical capacity for speech as we know it confirms that this is a recent adaptation unique to *Homo sapiens*. There seems to be a consensus within the archaeological community that a modern vocal tract was present in our species by 100,000 years ago at the latest (Hoffecker, 2007); considering the lower limit of 160,000 years ago for the existence of anatomically modern humans suggested by genetic evidence, it is reasonable to believe that a modern vocal tract had developed by that earlier date.

A wide range of evidence from the archaeological record suggests that many features of a language-ready brain began to come together after 2.5 mya, and that the biological evolution of modern humans with the full physical capacity for modern language was complete by 160,000 years ago. Furthermore, all of these developments occurred in Africa. I now turn away from direct anatomical remains and genetic evidence to consider behavioral correlates of cognitive capacity in hope of further clarifying the evolutionary transition that took place within that temporal range.

8.2. Behavioral correlates of cognitive and cultural evolution

The earliest evidence of complex behavior indicative of language-like cognition in the hominin lineage is in the form of stone tools, which begin to appear consistently in the archeological record at about 2.5 mya (Stout, 2011) with the emergence of the genus *Homo*. These artifacts provide a fairly abundant and continuous record of technological change up until the present, "documenting the gradual expression of new behavioural capabilities," (Stout, 2010:1050). The evolution of lithic technologies progressed at a very slow rate initially, passing through two major phases or "industries," but has developed at an increasingly rapid rate over the last few hundred thousand years, culminating in an explosion of creativity in the production of tools and other cultural artifacts in the past 100,000 years (Arbib, 2011) and the emergence of "behavioral modernity" around 50,000 years ago (Hoffecker, 2007; Mellars, 2006). These developments are compelling evidence for the progression of cognitive and cultural evolution from the beginning of the *Homo* lineage, and Arbib (2011) has related the evolution of lithic technologies and more recent cultural innovations to specific stages of his mirror system hypothesis of a gestural origin of language.

8.2.1. The Oldowan Industry

The earliest stone tools are dated to 2.6 mya and assigned to the Oldowan Industry (Stout, 2011). These tools were produced by striking sharp stone flakes from cobble cores by direct percussion with another stone, and consisted of simple yet relatively consistent choppers and sharp flakes (used, for example, to break bones and scrape out nutritious marrow). The basic form of Oldowan tools remained the same for a very long time, from ca. 2.6-1.4 mya (Stout, 2011), and is associated with the earliest *Homo* species including *H. habilis* and early *H. erectus* (Arbib, 2011).

Oldowan tool production involves a number of important skills. The physical activity of banging two rocks together to produce a specific outcome minimally requires a high level of dexterity and demands expert and simultaneous use of both power and precision grips. It also requires the ability to smoothly and reliably execute single actions (i.e., deftly removing a flake from a core with the percussive strike of a hammerstone), which can only be acquired through extended practice (and manipulative play during ontogeny; Bruner, 1972) (Arbib, 2011). On a more abstract cognitive level, tool production is a hierarchically structured, goal-directed action sequence bearing general similarity to language (Chapter 4), and the Oldowan industry would have required of its users a basic capacity for identifying and reproducing the relevant sub-goals and sub-actions that comprise a given complex sequential behavior. Indeed, Arbib (2011) claims that the Oldowan corresponds to the capacity for simple imitation in early human ancestors (or the "episodic" stage of cognitive evolution described by Donald [1991]) and compares this type of tool production to that observed in modern chimpanzees. Hominins capable of producing Oldowan tools would presumably have also have been able to support a very limited, essentially closed system of rudimentary novel communicative gestures similar to those observed in present-day apes (Arbib et al., 2008).

8.2.2. The Acheulean Industry

Starting around 1.6 mya, several technological innovations begin to appear in the archaeological record that mark a qualitative break with the earlier Oldowan Industry (Stout, 2011). This new industrial complex involved a technical repertoire that yielded more standardized and specialized tool types such as axes, cleavers, and picks. The Acheulean Industry also lasted quite a long time and progressed very slowly, but a distinction can be made between Early and Late Acheulean tool production. The Early Acheulean was characterized by more elaborate flake production and the emergence of larger cutting tools, and lasted from about 1.6 mya to 900,000 years ago. The Late Acheulean lasted from about 700,000 to 250,000 years ago, and saw the production of “blank” cores – large flakes intentionally removed in preparation for subsequent flake detachment and further shaping, which led to an unprecedented degree of variety in the form and function of stone tools as well as smaller, thinner, and more finely crafted products that may have been combined with organic materials to form composite tools through hafting (Stout, 2011). The earliest Acheulean tools are associated with *H. erectus*, but this industry evolved with later hominin species all the way up to the immediate predecessor of modern humans (Arbib, 2011).

The Acheulean Industry required a greater degree of cognitive complexity than did the Oldowan Industry. Most importantly, the much more elaborate and intentional flake production involved in Acheulean tools requires much more hierarchically complex action sequences. Human ancestors were no longer picking up rocks more or less at random and banging them together based on the most suitable affordance for the next flake removal; the tool maker now had to be able to consciously choose appropriate materials to produce a desired standard tool type, and to carefully choose each flake removal in order to arrive at the predetermined result (Stout, 2011). Hoffecker (2007) claims that the Acheulean industry shows the earliest signs of recursion (i.e., nested sub-goals of flake removal; Hauser et al. [2002] identify recursion as the fundamental trait distinguishing human language from other forms of animal communication) and external representation (i.e., pre-existing mental templates of specific tool types) in the archaeological record. Arbib (2011) attributes this greater degree of hierarchical complexity and intentionality to the capacity for complex imitation – unique to the hominin line – and the development of protolanguage (Donald's [1991] "mimetic" culture). By the end of the Acheulean period roughly 250,000 years ago, biological evolution would have yielded early *Homo sapiens*, the first hominin species equipped with a language-ready brain. The rapid emergence of later, more sophisticated lithic technologies and complex cultural artifacts and social behaviors would have been directly linked to the emergence and development of full-blown modern language; however, Arbib (2011) cautions that it may have taken "more than 100,000 years for the developing power of protolanguage to yield to the first true languages with their consequent impact on the acceleration of human evolution," (267).

8.2.3. Later innovations in lithic technology and "behavioral modernity"

The Acheulean and especially Oldowan Industries were remarkably stable, showing little or no change for hundreds of thousands of years at a time and representing periods of technological stasis. Such stasis in technological complexity implies a lack of cumulative cultural evolution in the Lower and Middle Paleolithic, which may be explained by underlying cognitive constraints (Stout, 2011). Arbib (2011) identifies these slowly expanding cognitive constraints as the capacities for simple and then complex imitation, which emerged as distinct stages in the biological evolution of the language-ready brain. The punctuated equilibrium (or at least extremely slow rate of cumulative cultural change) of these earliest lithic technologies contrasts sharply with the rapid technological change observed in the archaeological record since the emergence of early modern humans (who probably possessed fully modern vocal tracts; Hoffecker, 2007) around 200,000 years ago. This period was almost certainly characterized by a high degree of cumulative cultural change, with each successive generation readily acquiring the technological skills established by its predecessors and gradually expanding on the variety and hierarchical complexity of existing skillsets (Stout, 2011).

Stone tools found from 200,000 years ago and later involved production processes of greater hierarchical depth and represent unprecedented levels of recursion and external representation in human ancestors (Hoffecker, 2007). This is evidenced by the emergence of complicated composite tools (e.g., spears, handled axes) and highly refined varied forms of stone blades. A veritable explosion of cultural artifacts appears in the European archaeological record from 40,000 years ago, with examples of human art such as beads, tooth necklaces, cave paintings, stone carvings, figurines, and a variety of bone and antler tools; this period is referred to as the Upper Paleolithic and lasted until the Neolithic Revolution about 12,000 years ago, and is associated with behavioral modernity requiring fully modern language (Arbib, 2011). This period also shows evidence of other behaviors observed in modern hunter-gatherer populations such as musical compositions (inferred by the presence of instruments), ritual burial, and organization of domestic space (Hoffecker, 2007). The cultural transmission of these skills and behaviors was facilitated by the high fidelity of imitation (Stout, 2011) and even overimitation (whereby every aspect of a given behavior is interpreted as causally meaningful, even when that is not the case) characteristic of modern humans, which Arbib (2011) suggests may provide a glimpse into the "origins of the human propensity to do as others do within our culture," (261).

While the sudden appearance of behavioral modernity in Europe around 40,000 years ago was long interpreted as evidence of a "cultural revolution" caused by a dramatic restructuring of the brain accompanying the emergence of modern language, many components of this purported revolution are documented significantly earlier in Africa at sites widely separated in space and time (Arbib, 2011). The African record indicates that behavioral modernity and full language emerged more gradually in Africa and was subsequently exported to Eurasia and the rest of the world with the first successful migrations of anatomically modern humans about 60,000 years ago (Hoffecker, 2007; Mellars, 2006).

8.3. Tying it all together

So when and where did language evolve, exactly? Important components of the language-ready brain – namely sufficient brain capacity and organization coupled with more or less anatomically modern hands and the capacity for simple imitation – began to come together 2.5 mya in Africa with the emergence of the genus *Homo* and the Oldowan Industry of stone tool production. Cultural and technological change progressed very slowly with long periods of relative stability for the next two million years until the biological evolution of anatomically modern humans with brains and vocal tracts that could support the cognitive and physical demands of modern language was completed between 250,00 and 160,000 years ago. After this point, human ancestors very much like ourselves took about another 100,000 years to develop the productive system of symbolic manual gestures (protosigns) and non-innate vocalizations (protospeech) – made possible by complex imitative abilities supporting the development of the Acheulean Industry – into fully modern language and cultural behavior. This all happened among human groups in Africa, a small population of which experienced a demographic boom about 60,000 years ago and expanded throughout the African continent and the rest of the world, ultimately displacing all other members of the hominin lineage. The full linguistic diversity of the world today is the result of continuous processes of cultural transmission and historical change among various populations of modern humans.

9. Conclusion

In recent decades, a great deal of research in various disciplines has created a solid foundation to support productive reasoning about the evolution of language and the origins of modern human cognition. The embodied nature of language – first described by Piaget (e.g., 1983) and Werner & Kaplan (1963) – indicates that the form of language is largely determined by constraints imposed by the human bodies in which it evolved. This hypothesis was further developed by Lakoff & Johnson (1980) and has been corroborated on the neural level by many researchers (e.g., Rizzolatti & Arbib, 1998; Barsalou, 1999; Pulvermüller & Fadiga, 2010), demonstrating that language relies heavily on general cognitive abilities and parts of the brain that were initially developed for other purposes. The general notion of language and cognition as grounded in the body receives further support from the striking parallels in both structure (i.e., hierarchical organization) and ontological development between language and manual action. Crucially, spoken or signed language and goal-directed manual action – praxis – are digital processes by nature, as opposed to the analogic characteristics of communicative gesture. Human communication in its broadest sense is a combination of complementary digital and analogic systems; language emerged from the intersection of these two systems, which have both found expression in the hands.

The remarkable similarities between language, gesture, and praxis are best explained within the language-ready brain framework proposed by Christiansen & Chater (2008), in which biological evolution yielded a brain and a body capable of supporting language, and only after that process was complete did the world's diverse linguistic systems emerge through cultural transmission across hundreds or thousands of generations of otherwise anatomically and cognitively modern humans. Furthermore, the importance of the hands in both language and action implies a specifically gesture-first scenario (i.e., a digital system of manual symbols) for language origins (Arbib, 2005, 2008). Such an evolutionary picture is supported by the archaeological record and other lines of evidence.

9.1. Productive areas for future research

The language-ready brain theory outlined by Christiansen & Chater (2008) and the specifically gestural origins model proposed by Arbib (2005, 2008) provide a very productive paradigm for future research on the enduring problem of language evolution. Increased resolution and understanding of each stage of this complex evolutionary process can be achieved through further pursuit of many different lines of evidence, briefly discussed below.

The precise details of the evolutionary history of language can best be described by the archaeological record. As discussed in detail in the last chapter, this rich line of evidence can provide powerful insight into the development of both the anatomical features necessary for the physical production of language as well as the cognitive abilities required to support such an activity. Further discovery and analysis of the fossil remains of human ancestors as well as the predecessors of other living apes may help narrow the temporal window in which language is believed to have evolved. An understanding of exactly when the human cranium and vocal tract in particular assumed their modern morphologies would be of great use in placing a lower limit on language evolution. While skeletal remains of human ancestors are often incomplete and more representative of certain temporal and geographic ranges than others, cultural artifacts and particularly stone tools provide a relatively rich and continuous record of behavioral complexity over the past 2.5 million years. Continued discovery and description of these artifacts as well as an appreciation of the cognitive complexity required for their production will provide a valuable and increasingly nuanced understanding of the cultural and cognitive evolution of skills that also underlie the capacity to use language.

Structural description of the stone tool production process as well as other hierarchically organized behaviors is another useful endeavor, as it can help establish a formal means of comparing these goal-directed actions to language. Chapter 4 describes the parallel hierarchical structure of language and praxis in detail, and although numerous formal models have been developed to describe linguistic structure, praxis lacks comparable tools for structural analysis. Stone knapping has been represented visually with syntax-like tree structures (e.g., Moore, 2010; Stout, 2011) and at least one attempt has been made to develop a formal model of a generative grammar of goal-directed action (Pastra & Aloimonos, 2012), but a further standardization of these structural descriptions is needed to enable more scientifically rigorous and consistent comparisons of language and action. Work in this area could establish more reliable analogies between specific levels of hierarchical complexity in both of these domains, strengthening the argument that they are different manifestations of the same underlying cognitive function.

Likewise, developmental psychology and comparative primatology may also be able to play a role in identifying basic cognitive abilities underlying language and action. Examining the rate at which these two skills develop relative to each other in human ontogeny may be telling of the degree to which they rely on shared neural substrates. Comparing these abilities in humans with the praxic and communicative abilities of other apes can provide clues as to exactly what properties of language are unique to humans and merit special attention.

The claim that modern language is the product of cultural evolution on top of the biologically evolved substrate of a language-ready brain (and body) places a very high value on the study of cultural processes such as the transmission of knowledge and iterated learning.

Advocated by Christiansen & Kirby (2003), computational modeling is a powerful tool for exploring the origins of linguistic systems. Such models make it possible to simulate the ways in which systems may evolve through iterated individual learning, explore how theoretical constraints may affect such transmission, and evaluate the plausibility of purely theoretical explanations of language origins, possibly exposing hidden problems with such designs and/or suggesting creative new experiments that may shed light on those problems. The proposal that language emerged through an extended process of cultural transmission across hundreds or thousands of generations also makes the study of diachronic linguistics and established processes of language change highly relevant to the question of language origins. Grammaticalization – the historical process by which open-class words can be converted into functional grammatical forms – may have been of special importance in the development of syntactic structure (Heine & Kuteva, 2002). A deeper understanding of this phenomenon in particular may clarify the nature of the earliest protosigns and protowords (cf., Arbib 2005, 2008) and whether these symbols were compositional from the very start or initially holophrastic (for a review, see Arbib & Bickerton, 2010).

This paper has presented a fundamental distinction within human communication – that between digital and analogic systems of representation – and applied it to the study of language evolution. If kept in mind, this distinction provides a new lens and a useful framework for exploring each of the relevant lines of evidence mentioned above. Further investigation of analogic and digital systems in and of themselves will yield a greater understanding of the evolution of modern human communication and cognition. This should include a more precise definition of exactly what each term entails as well as what intermediate systems sharing properties of both, such as pantomime and finger counting, would have looked like. These last two systems are of particular interest, as they may have emerged simultaneously as a result of a crucial development in cognitive capacity allowing for the integration of digital and analogic channels to yield more productive and expressive abilities that paved the way for modern language and numeracy.

The question of language origins is a particularly thorny one, involving an intricate tangle of diverse lines of evidence from physical anthropology, theoretical linguistics, evolutionary biology, developmental psychology, comparative primatology, archaeology, and many other disciplines. As such, its resolution is dependent upon a multidisciplinary approach and will require a great deal of awareness and cooperation across academic fields. Although the study of language evolution is riddled with contentious issues that will not be agreed upon for many years to come, the field has come a very long way from the muddled confusion that led to the Linguistic Society of Paris' famous ban on discussion of the topic so many years ago. This paper has attempted to synthesize a wide range of work and has advocated a language-ready brain theory in which the analogic forms of communication observed in other apes became supplemented at some point in hominin evolution by a digital system of initially gestural symbols, yielding the productive compositionality characteristic of modern language. This evolutionary picture establishes a useful paradigm for further reasoning about the origins of language, and future research along the lines mentioned above should result in an enhanced understanding of how and why humans came to possess this remarkable and unique form of communication.

Works Cited

Allen, K., Ibara, S., Seymour, A., Cordova, N., & Botvinick, M. (2010). Abstract Structural Representations of Goal-Directed Behavior. *Psychological Science, 21*(10), 1518-1524.

Arbib, M. A. (2008). From grasp to language: Embodied concepts and the challenge of abstraction. *Journal of Physiology - Paris, 102*, 4-20.

Arbib, M. A. (2009). Evolving the language-ready brain and the social mechanisms that support language. *Journal of Communication Disorders, 42*, 263-271.

Arbib, M. A. (2011). From Mirror Neurons to Complex Imitation in the Evolution of Language and Tool Use. *Annual Review of Anthropology, 40*, 257-273.

Arbib, M. A., Liebal, K., & Pika, S. (2008). Primate Vocalization, Gesture, and the Evolution of Human Language. *Current Anthropology, 49*(6), 1053-1076.

Arbib, M. A., & Bickerton, D. (Eds.). (2010). *The Emergence of Protolanguage: Holophrasis vs Compositionality* (Vol. 24). Amsterdam: John Benjamins Publishing Co.

Arensburg, B., Tillier, A. M., VanDerMeersch, B., Duday, H., Schepartz, L. A., & Rak, Y. (1989). A Middle Palaeolithic human hyoid bone. *Nature, 338*, 758-760.

Atkinson, Q. D. (2011). Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science, 332*, 346-349.

Austin, J. L. (1962). *How to do Things with Words*. Oxford: Clarendon.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577-660.

Barsalou, L. W. (2010). Grounded Cognition: Past, Present, and Future. *Topics in Cognitive Science, 2*, 716-724.

Bates, E., & Dick, F. (2001). Language, Gesture, and the Developing Brain. *Developmental Psychobiology, 40*(3), 293-310.

Bates, E., & Goodman, J. C. (1997). On the Inseparability of Grammar and the Lexicon: Evidence from Acquisition, Aphasia and Real-time Processing. *Language and Cognitive Processes, 12*(5), 507-584.

Bateson, G. (1972). Problems in Cetacean and Other Mammalian Communication. *Steps to an Ecology of Mind* (pp. 364-378). Chicago: The University of Chicago Press.

Bruner, J. S. (1972). The nature and uses of immaturity. *American Psychologist*. 27: 687-708.

Bruner, J. S. (1973). Organization of early skilled action. *Child Development, 44*, 1-11.

Brown, P. (1992). Recent human evolution in East Asia and Australasia. *Phil. Trans. R. Soc. Lond. B, 337*, 235-242.

Cann, R. L., Stoneking, M., & Wilson, A. C. (1987). Mitochondrial DNA and human evolution. *Nature, 325*, 31-36.

Carnie, A. (2006). *Syntax: A Generative Introduction* (2 ed.). Oxford: Blackwell.

Castro, J. M. B. d., Arsuaga, J. T., Carbonell, E., Rosas, A., Martínez, I., & Mosquera, M. (1997). A Hominid from the Lower Pleistocene of Atapuerca, Spain: Possible Ancestor to Neandertals andModern Humans. *Science, 276*, 1392-1395.

Chechik, G., Meilijson, I., & Ruppin, E. (1998). Synaptic Pruning in Development: A Novel Account in Neural Terms. *Neural Computation, 10*(7), 1759-1777.

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

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, N. (1970). Remarks on nominalization. In R. Jacobs & P. Rosenbaum (Eds.), *Readings in English Transformational Grammar*. Waltham, MA: Blaisdell.

Chomsky, N. (1981). *Lectures on government and binding: the Pisa lectures*. Dordrecht: Foris.

Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.

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

Connolly, K., & Elliott, J. (1974). The Evolution and Ontogeny of Hand Function. In N. B. Jones (Ed.), *Ethological Studies of Child Behaviour*. Cambridge: Cambride University Press.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5*(12), 539-546.

Corballis, M. C. (2003). From Hand to Mouth: The Gestural Origins of Language. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution*. Oxford: Oxford University Press.

Corballis, M. C. (2007). Recursion, Language, and Starlings. *Cognitive Science, 23*(2), 157-205.

Corballis, M. C. (2010). Mirror neurons and the evolution of language. *Brain & Language, 112*, 25-35.

Craik, F. I. M., & Bialystok, E. (2006). Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences, 10*(3), 131-138.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge: Harvard University Press.

Dunbar, R. (1996). *Grooming, Gossip, and the Evolution of Language.* Cambridge, MA: Harvard University Press.

Dunbar, R. (2003). The Origin and Subsequent Evolution of Language. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution*. Oxford: Oxford University Press.

Ellsworth, P. C. (1994). William James and emotion: is a century of fame worth a century of misunderstanding? *Psychological Review, 101*(2), 222-229.

Ervin-Tripp, S. M. (1991). Play in language development. In B. Scales, M. Almy, A. Nicolopoulou & S. M. Ervin-Tripp (Eds.), *Play and the social context of development in early care and education* (pp. 84-98). New York: Columbia Teachers College.

Fitch, W. T. (2010). Three meanings of "recursion". In R. K. Larson, V. Deprez & H. Yamakido (Eds.), *The Evolution of Human Language: Biolinguistic Perspectives* (pp. 73-90). Cambridge, UK: Cambridge University Press.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361-377.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(2), 493-501.

Garvey, C. (1990). *Play*. Cambridge: Harvard University Press.

Glenberg, A. M., Havas, D., Becker, R., & Rinck, M. (2005). Grounding Language in Bodily States: The Case for Emotion. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking* (pp. 115-128). Cambridge: Cambridge University Press.

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences, 14*, 531-595.

Goldin-Meadow, S. (2003). *Hearing Gesture: How Our Hands Help Us Think*. Cambridge: Harvard University Press.

Goodwyn, S. W., & Acredolo, L. P. (1993). Symbolic Gesture versus Word: Is There a Modality Advantage for Onset of Symbol Use? *Child Development, 64*(3), 688-701.

Haeckel, E. (1866). *Generelle Morphologie der Organismen*. Berlin: Georg Reimer.

Hart, W. D. (1996). Dualism. In S. Guttenplan (Ed.), *A Companion to the Philosophy of Mind* (pp. 265-267). Oxford: Blackwell.

Hauser, M., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science, 298*(5598), 1569-1579.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic Use of Botulinum Toxin-A Affects Processing of Emotional Language. *Psychological Science, 21*(7), 895-900.

Heine, B., & Kuteva, T. (2002). On the evolution of grammatical forms. In A. Wray (Ed.), *Transitions to language* (pp. 376-397). Oxford: Oxford University Press.

Hockett, C. (1960). The Origin of Speech. *Scientific American* 203:88.

Holloway, R. L. (1983). Human paleontological evidence relevant to language behavior. *Human Neurobiology, 2*, 105-114.

Huizinga, J. (1955). *A study of the play element in culture*. London: Beacon Press.

Ifrah, G. (1981). *From One to Zero* (L. Bair, Trans.). New York: Viking.

Iglesias, J., Eriksson, J., Grize, F., Tomassini, M., & Villa, A. E. (2005). Dynamics of pruning in simulated large-scale spiking neural networks. *Biosystems, 79*(1-3), 11-20.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science, 16*(5), 365-371.

Jackendoff, R. (1977). *X-bar syntax: a study of phrase structure*. Cambridge, MA: MIT Press.

Kaschak, M. P., Maner, J. K., Miller, S., & Coyle, J. M. (2009). Embodied social cognition: Bodies, emotions, and blackberries. *European Journal of Social Psychology, 39*(7), 1255-1256.

Kelly, S. D., Iverson, J. M., Terranova, J., Niego, J., Hopkins, M., & Goldsmith, L. (2002). Putting Language Back in the Body: Speech and Gesture on Three Time Frames. *Developmental Neuropsychology, 22*(1), 323-349.

Kendon, A. (1988). How gestures can become like words. In F. Poyatos (Ed.), *Cross-cultural perspectives in nonverbal communication*. Cambridge: Cambridge University Press.

Koechlin, E., & Jubault, T. (2006). Broca's Area and the Hierarchical Organization of Human Behavior. *Neuron, 50*, 963-974.

Kohler, E., Keysers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science, 297*(5582), 846-848.

Lacey, S., Stilla, R., & Sathian, K. (2012). Metaphorically feeling: Comprehenging textural metaphors activates somatosensory cortex. *Brain and Language, 120*(3), 416-421.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: The University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5), 187-196.

Lieberman, P. (2003). Motor Control, Speech, and the Evolution of Human Language. In M. H. Christiansen & S. Kirby (Eds.), *Language Evolution*. Oxford: Oxford University Press.

Lilienfeld, S. O., Lynn, S. J., Namy, L. L., & Woolf, N. J. (2010). *Psychology: A Framework for Everyday Thinking (2nd edition)*. Boston: Pearson Education Incorporated.

Liu, H., Prugnolle, F., Manica, A., & Balloux, F. (2006). A Geographically Explicit Genetic Model of Worldwide Human-Settlement History. *Am. J. Hum. Genet., 79*(2), 230-237.

Low, L. K., & Cheng, H.-J. (2006). Axon pruning: an essential step underlying the developmental plasticity of neuronal connections. *Phil. Trans. R. Soc. Lond. B Biol. Sci., 361*(1473), 1531-1544.

Lowenfeld, M. (1990). *Play in childhood* (2 ed.). London: Mac Keith Press.

Mayr, E. (1994). Recapitulation Reinterpreted: The Somatic Program. *The Quarterly Review of Biology, 69*(2), 223-232.

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

McNeill, D. (2005). *Gesture and Thought*. Chicago: The University of Chicago Press.

Mellars, P. (2006). Why did modern human populations disperse from Africa *ca.* 60,000 years ago? A new model. *Proc Natl Acad Sci USA, 103*(25), 9381-9386.

Moore, M. W. (2010). 'Grammars of Action' and Stone Flaking Design Space. In A. Nowell & I. Davidson (Eds.), *Stone Tools and the Evolution of Human Cognition* (pp. 13-43). Boulder, CO: University Press of Colorado.

Musse, S. R., & Thalmann, D. (2001). Hierarchical model for real time simulation of virtual human crowds. *IEEE Transactions on Visualization and Computer Graphics, 7*(2), 152-164.

Napier, J. R. (1960). Studies of the hands of living primates. *Proc. Zool. Soc. London, 134*, 647-657.

Neumann, E. A. (1974). *The elements of play*. New York: Ardent Media.

Pastra, K., & Aloimonos, Y. (2012). The minimalist grammar of action. *Phil. Trans. R. Soc. B, 367*, 103-117.

Piaget, J. (1962). *Play, Dreams and Imitation in Childhood*. New York: Norton.

Piaget, J. (1983). Piaget's theory. In P. Mussen (Ed.), *Handbook of Child Psychology* (4 ed., Vol. 1). New York: Wiley.

Pinker, S. (1994). *The Language Instinct: How the Mind Creates Language*. New York: William Morrow and Company.

Premack, D. G., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515-526.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience, 11*, 351-360.

Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neurosciences, 21*(5), 188-194.

Ruff, C. B., Trinkaus, E., & Holliday, T. W. (1997). Body mass and encephalization in Pleistocene *Homo*. *Nature, 387*(8), 173-176.

Saussure, F. d. ([1916] 1959). *Course in General Linguistics* (R. Harris, Trans.). Chicago: Open Court.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2009). Gestures Orchestrate Brain Networks for Language Understanding. *Current Biology, 19*(8), 661-667.

Stanyon, R., Consigliere, S., & Morescalchi, M. A. (1993). Cranial Capacity in Hominid Evolution. *Human Evolution, 8*(3), 205-216.

Stout, D. (2011). Stone toolmaking and the evolution of human culture and cognition. *Phil. Trans. R. Soc. B, 366*, 1050-1059.

Stout, D., Toth, N., Schick, K., & Chaminade, T. (2008). Neural correlates of Early Stone Age toolmaking: technology, language and cognition in human evolution. *Phil. Trans. R. Soc. B, 363*, 1939-1949.

Stout, D., & Chaminade, T. (2012). Stone tools, language and the brain in human evolution. *Phil. Trans. R. Soc. B, 367*, 75-87.

Vigilant, L., Stoneking, M., Hapending, H., Hawkes, K., & Wilson, A. C. (1991). African Populations and the Evolution of Human Mitochondrial DNA. *Science, 253*, 1503-1507.

Werner, H., & Kaplan, B. (1963). *Symbol Formation: An Organismic-Developmental Approach to Language and the Expression of Thought*. New York: John Wiley & Sons, Inc.

White, T. D., Asfaw, B., DeGusta, D., Gilbert, H., Richards, G. D., Suwa, G., & Howell, F. C. (2003). Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature, 423*, 742-747.

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: an instance of the ideomotor principle. *Phil. Trans. R. Soc. Lond. B, 358*, 501-515.

Young, R. W. (2003). Evolution of the human hand: the role of throwing and clubbing. *Journal of Anatomy, 202*(1), 165-174.