7 Contextual Flexibility in Infant Vocal Development and the Earliest Steps in the Evolution of Language

D. Kimbrough Oller and Ulrike Griebel

Overview on Vocal Flexibility in Development and Evolution

Prelude

The natural laboratory of human development offers indirect evidence about the underlying capabilities that must have been evolved in the hominin line in order for language to have emerged. In this chapter we argue that in all likelihood the first foundational steps of hominin communicative evolution away from the primate background had much in common with stages that are observable in human infant development (Oller, 2000; Oller and Griebel, 2005). These stages pertain to vocal production in particular, since it is vocal production (see review in Snowdon, this volume), rather than perception and auditory processing, that appears to differentiate humans so strongly from nonhuman primates.

The parallelism we advocate between development and evolution is different from the concrete form of the idea attributed to Haeckel (1866) that ontogeny recapitulates phylogeny. Our view conforms to the abstract idea that ontogeny and phylogeny often abide by the same "natural logic," a logic that should be expected to have extremely broad application because it characterizes sequences of transparently increasing complexity or of sequences required by development. For example, in embryology single-cell structures must precede multicellular ones, and also in evolution the existence of single-cell creatures had to precede multicellular ones. The term "natural logic" can be applied in a variety of realms of science. For example, in chemistry, it is obvious that the building of organic molecules involves naturally logical steps combining first atoms, then simple molecules, and then more complex molecules.

Communicative complexity in many cases can be interpreted in terms of the number of capabilities that have to be coordinated in order for some type of communicative act to be performed. For example, the capability to control one word at a time is simpler than the capability to put two words together, because the latter requires

both the former plus something more. Similarly, increasing complexity of capabilities can be viewed as requiring increasingly complex neural subsystems to manage the actions in question. For example, the capability to produce one word at a time draws on neural subsystems that must be expanded to make syntactic combination of words possible.

Empirical study of early vocal development, to be reviewed below, has determined the existence of stages that similarly appear to occur in a naturally logical order. Vocal contextual flexibility provides foundations for the power of human language to communicate an indefinitely large variety of functions and ideas precisely and rapidly. Without vocal contextual flexibility, well-recognized features of language would be impossible, including learning new speech sounds by imitation, recombination of sounds to create words, acquisition of the lexicon of any language, generation of sentences composed over the lexicon, and so on.

Definitions of Signal Flexibility and Functional Flexibility

Our chapter focuses on the appearance of vocal contextual flexibility in the first months of life, including features of both signal flexibility and functional flexibility as defined in Griebel and Oller (this volume). The definitions are extended here.

Signal and function have been recognized as conceptually distinct in language since the founder of modern linguistics, de Saussure, drew the distinction between *signifiant* (that which signifies) and *signifié* (that which is signified) in language (de Saussure, 1968; first published in 1916, three years after his death). The distinction can be expanded to encompass all types of communication. The study of signal characteristics independent of the functions they may transmit has been termed "infraphonology," and the study of the functions that can be served by signals has been termed "infrasemiotics" (Oller, 2000). As we research the development and evolution of language, it proves crucial to maintain the distinction between these two aspects of communicative events.

We use the term "signal" in this chapter to refer both to actions that transmit communications systematically and to actions that have the *potential* to transmit systematic communications. The inclusion of potential signals in the definition is important because the young of a species may systematically produce sounds or other potential signals that lay foundations for later communication.

The term "signal flexibility" focuses on the (infraphonological) ability to produce systematic physical variations in actions that can be signals, but the term is not intended to encompass all such variations. Consider, for example, the hissing of the house cat. The sound itself is a signal of aggression, and systematic variations in intensity of the signal can be produced to denote systematic variations in intensity of its communicated function (aggression). However, the hissing of the house cat is stereotyped in form, that is, limited in acoustic variability. The signal has

been naturally selected to serve no other purpose than that of expressing aggression, and if it were not limited in variability it could fail to serve its purpose at points of urgency.

The term "signal flexibility" in our usage invokes in particular the ability to vary signal parameters systematically *beyond* the limited types of variations of signals that are naturally selected to serve a particular function, as in the case of hissing in the house cat. When signals vary beyond the limits of naturally selected stereotypy, a many-to-one mapping from signal to function can occur. In notable cases of this sort (especially in very young animals), variable vocal signals occur where the *social* function appears to be to provide a fitness indicator for conspecifics, especially caretakers. A nonsocial function, vocal practice, appears also to be involved. When human infants engage in vocal play (Stark, 1980) or developing songbirds produce subsong (Nooteboom, 1999), they vary the parameters of potential signals systematically, often in the absence of immediate communicative purpose or effect other than possible fitness signaling, and thus provide examples of "signal flexibility" as we intend the term. In accord with our proposal, learning new vocal signals (either through imitation or through vocal exploration) involves this kind of signal flexibility as a prerequisite foundation.

"Functional flexibility" is the (infrasemiotic) ability to produce a given signal type in such a way as to serve different communicative purposes on different occasions. The notion "function" generally refers to coherent classes of social communication. One coherent social communicative function might be called "aggression" or "threat" (as transmitted by the hissing house cat), another "warning" or "announcement of alarm," another "advertisement of fitness," yet another "greeting," and so on. Because there is coherency to these different purposes, it is possible for a particular signal to be naturally selected within a particular species to serve any one of them. When an individual signal is selected to serve a particular purpose and no other, it becomes a "fixed signal" in the terminology of classical ethology (Lorenz, 1951; Tinbergen, 1951). The hissing of the house cat is just such a signal. A fixed signal serves the same kind of purpose on *each* occasion of use—a fixed signal of threat, such as hissing in the house cat, cannot be used as a warning or greeting on some other occasion. The acoustic characteristics of fixed signals are stereotyped because the signals are naturally selected to serve the same purpose on each occasion, and the purpose cannot be consistently served unless the signal itself is unambiguous.

However, not all signals are fixed in these ways. Words or sentences in human languages are signals that are free to transmit different purposes (warning, naming, ridiculing, etc.) on different occasions. Consider the word "vampire." It can be used to warn a disbeliever, to name a flying mammal, to invoke a mythical concept, or to ridicule an ex-spouse, among many other possibilities. Functional flexibility is this ability to use the same signal to serve different functions on different occasions.

Furthermore, each root word in a human language has to be learned, and language learning requires a mechanism flexible enough to adapt to different sound systems in different languages, to different conventions for the relation of each word with its meaning, to different conventions about how words can be modified and combined in sentences, and so on. Both signal flexibility and functional flexibility are thus required in order for the process of learning to proceed, and until a species acquires at least rudimentary flexibility in these domains, evolution in the direction of the communicative power of language is not possible.

The Proposal in Short

We propose, then, that the earliest requirement of hominin communicative evolution beyond the vocal abilities seen in nonhuman primates, abilities that were presumably also present in our distant primate ancestors, was the establishment of vocal contextual flexibility, both signal flexibility and functional flexibility. This proposal takes stock of the vocal limitations of nonhuman primates both in signal flexibility and in functional flexibility (see Hammerschmidt and Fischer, Owren and Goldstein, and Snowdon, this volume, for reviews of empirical data and prevailing interpretations). Owren and Goldstein (this volume) provide a compatible proposal about the importance of scaffolding of stages of learning in the emergence of vocal flexibility and voluntarity in the human infant in contrast with the relative inflexibility that is observed in nonhuman primates.

We reason that in hominin evolution the capabilities for signal flexibility and functional flexibility must have been built in steps, applying first in simple, primitive communications resembling (at least in part) those of human infancy, and applying only much later in the context of the enormously more complex structures of mature human language. Human infancy, then, provides a model, at least in broad outlines, of how foundational steps might be taken to break free from the limitations of fixed signaling. The empirical evidence on early human development to be reviewed here indicates that systematic foundational steps toward vocal contextual flexibility are taken by humans in the first months of life. By the middle of the first year of human life, humans surpass nonhuman primates in vocal flexibility at any age (Oller, 2000).

Naturally Logical Interpretation of the First Steps of Infant Vocal Development

Natural Logic and the Ordering of Stages in Development and Evolution

Our proposal turns on the idea that the ordering of steps of vocal contextual flexibility in the first months of life occurs *by necessity*: The steps abide by a "natural logic," the broad outlines of which can be discerned through evaluation of the patterns of

development. Whenever developmental or evolutionary sequences can be explained as occurring by natural necessity, we invoke the term "natural logic" and try to specify the principles of logical dependency of one capability with respect to another that lie behind the natural necessity for sequencing. In some cases, it may be easy to correctly posit a sequence abiding by natural logic from our armchairs—for example, we need no research (at least not any more) to know that independent single cells precede multicellular clusters in both development and evolution. The relative complexity of single- and multicellular structures is unambiguous. However, there are many cases of development and evolution where empirical research illustrates systematic sequencing prior to scientific recognition of the natural logic that may predispose a sequence of development to occur.

   Vocal development in human infants provides clear examples that armchair speculation about natural sequences is often of little use. The naturally logical character of the sequences of vocal development that have been revealed were not at all obvious when systematic research on human infant vocalization began in the 1970s (Oller, 1978; Stark et al., 1975; Zlatin-Laufer and Horii, 1977). Furthermore, only after considerable interpretive progress in empirical developmental research did the importance of the posited developmental natural logic come into focus as a basis for evolutionary speculations. We now turn to a summary of stages of human vocal development and the natural logic that appears to predispose their orderly occurrence.

Stages in Emergence of Vocal Contextual Flexibility in Human Infancy

On the infraphonological side of vocal development, human infants manifest emergence of a capability for signal flexibility through the following stages:

1. Spontaneous production of phonation in comfort, usually from the first month of life.

2. Rapid elaboration of spontaneous vocalizations in nonstereotyped displays of vocal raw material, including primitive supraglottal articulations, usually within the first two months.

3. Creation by the infant of gross but quite identifiable categories of (precanonical) vocalization from the elaborated vocal raw material (of stage 2) by the third or fourth month.

4. Further elaboration of vocal categories to include well-formed or "canonical" syllables, which incorporate signal flexibility both in terms of laryngeal (vocal) control and in terms of supralaryngeal, articulatory control by the middle of the first year of life or shortly thereafter.

   On the infrasemiotic side of development, infants show systematic growth of functional flexibility as follows:

1. By establishing a foundation for functional flexibility during the same time period as stage 1 of signal flexibility, with spontaneous production of flexible vocalizations that have no fixed social communicative function (unlike, e.g., crying).

2. By utilizing the nonstereotyped vocal raw material of stage 2 both in playful vocalization (play can be viewed *as* a function) when the infant is alone as well as in interactive face-to-face vocalization.

3. By showing free expressivity, where any of the newly developed categories indicated in stages 3 and 4 are used in multiple and often opposite affective contexts (exulting, complaining, or simply engaging in vocal play). Free expressivity is seen by the fourth or fifth month and continues to be elaborated throughout language development.

These patterns of emerging signal and functional flexibility, having been described in longitudinal studies to be cited below, are now the focus of even more intense research in our laboratories (e.g., Buder et al., in press; Oller et al., 2007). It is expected that the methods being developed to quantify the patterns will be applicable to description of the extent of contextual flexibility not only in human infants but in a wide variety of species whose vocal repertoires (both during development and in mature form) provide a backdrop for the understanding of human development and evolution.

The Stages of Vocal Development and Their Naturally Logical Basis

Infraphonological Stage 1 (The Phonation Stage): Spontaneous Production of Phonatory Acts

Spontaneous vocalizations occur from the first month of life. These "quasivowel" sounds, produced in comfort, begin as vocal acts of modal or normal phonation with no articulation and no within-utterance interruption. Quasivowels are produced with no systematic posturing of the supraglottal tract (which can be said to be "at rest"), and consequently these sounds can be viewed as pure phonatory events. Prior research has documented both acoustic and functional characteristics of quasivowels (Boliek et al., 1996; Koopmans-van Beinum et al., 2001; Koopmans-van Beinum and van der Stelt, 1986; Oller, 1980; Stark et al., 1993; Stark, 1978). Quasivowels are not fixed signals, because they are spontaneously produced, with no apparent stimulus that elicits them—they occur when the infant is alone *and* when the infant is attended. Quasivowels do not appear to be driven by strong emotional content, for they tend to occur when the infant is awake but otherwise quiescent. Quasivowels appear to function sometimes as state and fitness indicators to parents, who often seem pleased and pacified to hear the infant vocalizing in this way (instead of crying, the principal other vocalization of the first month). However, the sounds are inherently

ambiguous as to function and are not always interpreted as indicators of comfort—they can also be treated as fussing on some occasions.

"Close calls" appear to provide the most appropriate analogy in nonhuman primates to the acoustic characteristics and functions of quasivowels (Snowdon, 2004), but close calls have been relatively little studied in nonhuman primates, partly because they tend to occur at low intensity (they do not have to be loud to be heard by their targeted audiences) and are thus difficult to record (Becker et al., 2003). It has been speculated that close calls are affiliation expressions that may include several (perhaps many) subtypes in some species (see Snowdon, this volume) and may even include substantial contextual flexibility of production. Developing further quantitative description of human quasivowels and the methodology to characterize the relation of the sounds with contexts is important in order to form a foundation for more fruitful cross-species comparison. No one has yet quantified the degree of contextual flexibility occurring in close calls and quasivowels in order to compare nonhuman primates and human infants—a stable basis for scaling that comparison remains to be developed.

We interpret the ability to produce any sound spontaneously as a critical foundation for that sound's further flexible production, either in playful exploration, category formation, or variable communication usage. While quasivowels are primarily socially undirected in the first month of life, they are incorporated into face-to-face interactions with caregivers in stage 2. Later in the first year of life, quasivowels are often utilized by infants to communicate assent or acknowledgment (McCune et al., 1996). The capability for spontaneous production lays a naturally logical foundation for these later developments.

Infraphonological Stage 2 (The Primitive Articulation Stage): Elaboration of Spontaneously Produced Vocalizations to Include Primitive Articulation

Infants usually show rapid elaboration of spontaneous vocalizations into nonstereotyped displays of vocal raw material by the second or third month of life. New sounds such as squeals and growls begin to appear. Quasivowels also show elaboration in the first weeks, becoming increasingly variable in acoustic character, exhibiting both long and short types, louder and softer ones, and glottal interrupts begin to be heard within quasivowels, that is, the breathing cycle is halted during vocalization. By the fifth or sixth week it is clear that infants adjust breathing by taking in extra air before beginning quasivowel-like vocalization (Boliek et al., 1996).

Quasivowels are initially produced as pure phonatory events with no systematic supraglottal posturing or movement—the supraglottal vocal tract tends to be at rest. But by stage 2, supraglottal articulation does begin in "gooing," which involves seemingly uncoordinated movements where the tongue dorsum is brought into contact with other structures erratically. The articulations produce a kind of "primitive

syllabification," which seems primitive precisely because it is unpredictable in time and extent (Zlatin, 1975), unlike the well-formed canonical syllables that occur systematically at a later point in development.

During the same period that gooing appears, we also note elaboration of vocalization at the larynx. High-pitched and low-pitched sounds (without articulation), that is squeals and growls, now occur along with sounds in the midrange of normal phonation for human infants. These sounds manifest emergence of elaborate vocal raw material.

The lack of stereotypy, or to put it another way, the tendency for infant vocalizations in stage 2 to include a tremendous range of sound qualities, is the subject of intense investigation currently in our laboratories in Memphis. Complexity of vocalization is the rule, not the exception, by this point in time, in vocalizations produced both when the infant is alone and in interaction (Buder et al., in press; Oller et al., 2007). Lack of stereotypy, along with the fact that nonstereotyped production occurs both when the infant is alone and when the infant is in social interaction, provides a further indication of signal flexibility.

In general, descriptions of nonhuman primate vocal patterns have emphasized stereotypy and ritualization of particular sounds. Further, the descriptions have emphasized unitary functions for each vocal type. While Snowdon subscribes to this point of view in the main, he has also described instances, especially in New World monkeys, that appear to provide challenges to the general view on nonhuman primates (Snowdon, 2004; Snowdon et al., 1997; Snowdon and Hodun, 1981; Snowdon, this volume). Hammerschmidt and Fischer (this volume) also review interesting cases of vocal flexibility in nonhuman primates. However, research has not yet made direct comparisons, utilizing a well-defined common methodology, between human infant and nonhuman primate vocalizations. At present the most comparable data available across species suggest that human infants by stage 2 produce quantitatively more elaborate vocalizations than nonhuman primates at any age.

Still, the most important point about stage 2 is one of natural logic, and that point does not depend on data from nonhuman primates: Stage 2 builds on stage 1, because spontaneously produced quasivowels are elaborated in stage 2 to include more variable raw material both in phonatory and supraglottal articulation.

Infraphonological Stage 3 (The Expansion Stage): Primitive Vocal Category Formation from Vocal Raw Material

In the months following the first appearance of elaborated displays, infants show creation of new categories of vocalization from the raw material of complex sounds developed during the prior period. In this "expansion stage," repetitive sequences of vocalizations having a particular property (e.g., high pitch) are alternated systematically with sequences having another property (midpitch, e.g., or harsh vocal quality).

These sound types and their repetitive occurrence in vocal play have been described spontaneously by parents through more than thirty years of longitudinal research in our laboratories and others' as "squeals," "vowel-like sounds," and "growls." Squeals and growls are often perceived as opposites, and vowel-like sounds are viewed as neutral in vocal quality and pitch with respect to the other two. All three categories tend to be produced with the supraglottal vocal tract open rather than in the at-rest position that is typical of quasivowels.

We recognize categories in stage 3 in several quantitative ways (Buder et al., 2003; Oller et al., 2003). In the first method, we merely ask listeners to judge utterances presented from real samples as having squeal, growl, or vowel-like quality, and we find "good" agreement among untrained judges as measured by Cohen's kappa. A second method we use is to acoustically analyze the same utterances presented in the interobserver agreement studies and to plot them in multidimensional acoustic space. Even with only two dimensions, the categories identified by observers auditorily segregate significantly (Kwon et al., 2006b), and in further work we are seeking to characterize the degree of segregation that can be obtained when additional acoustic dimensions are utilized. A third method invokes lag sequential analysis of utterances produced by infants in various circumstances (alone, in interaction, while in the same room with the parent but separated in space and not obviously interacting). The utterances are first categorized as fitting into one of the three categories, and thereafter it can be shown the categories do not appear at random but are ordered in repetition and/or alternation by the fourth or fifth month (Kwon et al., 2006a). All these kinds of quantification suggest that new vocal categories have been developed by human infants by stage 3. Current work in our laboratories with neural networks provides additional confirmation of acoustic patterns in infant vocalization that suggest category formation by stage 3.

New vocal category formation within an individual at stage 3 depends, in our interpretation, upon the naturally logical foundations of spontaneous production of sounds in stage 1 and their exploration and elaboration within the available acoustic space in stage 2. Westermann's (this volume) modeling suggests that categories can emerge from vocal exploration and self-perception, even without influence from other ambient vocalization. Our reasoning about the importance of exploration as a foundation for development of new categories of action is emphasized also in an important literature on early motor development focused on hand, arm, and leg movements (Thelen, 1981, 1994, 1995). In both vocal and more general motor development, it appears then that new categories of action can appear by self-organization if the infant engages in self-monitored physical exploration.

The formation of new categories appears to be critical to further development of complex communication. Every aspect of language depends on the ability to learn new discrete sound categories. And crucially, these categories must be free of specific

function or meaning, because they must constitute a repertoire that can be recombined to form meaningful units (lexical items and sentences) of unlimited number. This principle of language, obvious for generations, has been referred to in recent formulations as "discrete infinity" (Chomsky, 1986; Hauser et al., 2003). It should be emphasized that what the infant accomplishes in stage 3 is not, however, the development of discrete phonological units of a mature sort. These sounds are not phonemes, nor allophones, nor phonetic features of the mature sort found in human languages. They are instead embryological precursors to such elements, infraphonologically significant forerunners that manifest the infant's ability to form categories from the raw material of elaborated, nonstereotyped sounds, categories that have no predetermined function but are free to be utilized in new ways.

Infraphonological Stage 4 (The Canonical Stage): Emergence of Canonical Syllables
Canonical syllables are relatively mature and well formed in the sense that words in spoken languages can be composed of them. They manifest characteristics that are found in the vast majority of syllables in languages, and without them it would not be possible to have indefinitely large lexicons. Canonical syllables emerge under systematic control by the fifth to the tenth month of human life. They consist typically of a nucleus with the supraglottal vocal tract open rather than at rest (i.e., a vowel-like sound), produced in modal or normal phonation, along with at least one margin (i.e., a consonant-like supraglottal articulation). Importantly, the transition between the margin and the nucleus must be smoothly and quickly articulated (nominally within 120 milliseconds).

Occasional canonical syllables can be heard in human infants long before the canonical stage (e.g., in gooing), but in such cases they appear to occur as accidents of the exploratory elaboration of more primitive vocalizations. Our quantitative approach to identification of the canonical stage is based on indications of repetitive occurrence of well-formed syllables. Parents prove to be excellent judges of the onset of canonical babbling (Oller et al., 2001) and are capable of listing the syllable types that infants produce under systematic control once the canonical stage is under way. They systematically ignore the great bulk of the rich raw material of which each canonical syllable is composed (i.e., they ignore within-category variation, which at this stage is still considerable) and instead focus on repetitive features that indicate syllabic control.

That canonical syllables form a crucial, naturally logical foundation in the development of human languages is not controversial. Languages require canonical syllables to form words, and after the onset of the canonical stage in an infant, several months typically pass before meaningful words are produced consistently so that they can be identified by parents (Oller et al., 1998). There is no credible report that we know of that indicates canonical syllable control by any nonhuman primate at any point in time.

Canonical syllables are not, however, the starting point on the developmental path to human language as has been suggested by MacNeilage (1998). The three prior stages have been found to occur in every one of the scores of infants we have studied in longitudinal research; further, the events of the earlier stages are naturally logical prerequisites to canonical syllables (Oller and Griebel, 2008). Canonical syllables require coordinated phonation *and* supraglottal articulation, developed to primitive extents in stages 1 and 2. Further, canonical syllables are vocal *categories* formed of both phonation and systematically well-timed supraglottal articulations: Vocal categories are first formed out of phonatory distinctions without well-timed articulations at stage 3. Canonical syllables are, thus, formed from components that are developed in three prior stages that constitute a naturally logical sequence of increasing complexity and increasing precision of vocal action.

MacNeilage's view also fails to recognize the naturally logical precedence of phonatory over supraglottal articulatory development. He assumes phonation need not be developed first. However, phonation is the primary sound source in syllables—without phonation, supraglottal articulatory movements produce low-intensity sounds that are ill suited to forming the nuclei of syllables. Articulatory movements during syllable production are largely perceived because phonation provides a carrier for formant transition information. Consequently, systematic development of phonatory capabilities comes in stage 1, providing a basis upon which systematic supraglottal articulatory movements can be perceived when they appear in subsequent stages.

Early Infrasemiotic Development: Emergence of Free Expressivity
Perhaps the most notable development in functional flexibility during the first six months of life is that each category developed at stage 3 is produceable in multiple, sometimes opposite, circumstances of affect and that the same sort of functional flexibility is seen with canonical syllables when they emerge at stage 4 (Oller, 1981, 2000; Scheiner et al., 2002).

The significance of the occurrence of particular sound categories with varying affect cannot be undercut by the argument that variations of affect may be (at least partly) a product of varying general states of arousal or external environmental conditions. Fixed signals, by definition, do not allow variation of the connection between affect and vocal signals—for example, an aggressive signal must be negative, an affiliation signal must *not* be negative, and an exultation signal must be positive. Consequently, empirical illustration of the production of the same sound category in differing conditions of affect on different occasions provides a conclusive illustration of functional flexibility.

To provide such illustration, we conduct cross-classification analysis and find that, for example, squeals are used sometimes to express complaint (as indicated by facial expression) but on other occasions to express exultation or delight (as indicated by

broad smiling). On other occasions the same vocalization type is used with a totally neutral face in circumstances that can be interpreted as pure vocal play. All these types of cross-classification variations apply to the other vocal types as well—growls and full vowels are also used with multiple expressions (Oller et al., 2003).

There are also clear instances where a panel of judges each independently characterizes a particular vocalization as pertaining to a particular category, for example, a squeal, based on the audio information only, and the facial expression as, for example, an exultation, based on video information only. Another vocalization judged to be of the same category (e.g., a growl) is seen uniformly as a complaint, while another is judged by the panel to be produced with neutral affect. Further, it appears that infants vary day to day, tending to use particular sound categories more heavily one way or another (with positive or negative emotion) on differing days (Oller et al., 2007). All these patterns are viewed as evidence of a kind of free expressivity (Oller, 2000). Continued quantification of free expressivity in infant vocalizations and the development of a general methodology that could be applied in nonhuman primates is a primary goal of our current work.

The natural logic we propose suggests that free expressivity with newly formed categories is a step of development that depends upon the prior step of category formation. We emphasize here that the vocal flexibility occurring in stage 3 is not merely that of random occurrence of states with unsystematically varying sounds. Rather, the flexibility seems targeted by the infant: New systematic categories are first controlled and thereafter utilized to express specific states under the infant's voluntary control.

Summary on Points of Natural Logic in Early Infant Vocal Development

In summary, then, we see naturally logical dependencies that produce necessary sequences for vocal development in at least the following ways within the first half year of life:

1. Simple, stereotyped quasivowels *precede* more variably produced quasivowels and gooing, a naturally logical progression from simple phonatory acts to more complex phonatory and primitively articulated acts.

2. Variably produced quasivowels and gooing *precede* category formation from the variably produced raw material, a sequence that appears to be based on a naturally logical learning dependency (category formation is apparently achieved through active exploration initiated with quasivowels and gooing).

3. Relatively less complex categories (such as squeals and growls), involving phonatory manipulation only, are formed through vocal exploration *before* the more complex categories of canonical babbling, where both phonatory action and supraglottal articulation must be coordinated.

On the infrasemiotic side, after categories are developed, they are recruited to serve multiple communicative functions. Thus, as new flexible potential signals appear in the repertoire of infants, they provide the naturally logical foundation for multifunctional expression with those signals.

All these sorts of developments are infrastructural foundations for capabilities of vocalization that come later in the first year of life. For example, vocal imitation of categories developed by the infant (including canonical syllables) comes later than category formation because there is no possibility of systematically recruiting the production of a voluntary category to match a sound produced by another person without the prior establishment of a capability to produce the category spontaneously. Indeed, clear evidence of infant vocal imitation occurs only after the first half year of life (Kessen et al., 1979). The ability to perform vocal imitation has long been noted as a logical precursor to all other aspects of spoken language and has recently been reemphasized in Hauser, Chomsky, and Fitch (2003). Notably, the authors do not emphasize the apparently logical precursors to imitation that we have outlined here. One possible reason for this omission is that very few of the individuals involved in research on evolution of language have to the present taken note of the data from longitudinal research on infant vocal development, and consequently they have not considered the sequence of early vocal events that appears to be universal in human infants.

Features of Natural Logic as an Interpretive Framework

Interleaving Development of Naturally Logical Capabilities

The developmental sequence indicated above is seen in interleaving of growth of fundamental capabilities in the domains of signal and functional capability, a pattern of overlap in time that continues throughout development. To illustrate interleaving, notice first that emergence of spontaneous production in a particular signal domain—for example, phonation as in infraphonological stage 1—is followed by elaboration of the signal through systematic vocal exploration of variations in phonation in stage 2, and the creation of categories of primarily phonatory nature in stage 3. Notice also that later, control of a new signal domain, movement of the jaw during phonation, begins to emerge, and again proceeds through the same three stages: First coordination of jaw movement and phonation occurs in spontaneous production, then again later in systematic exploration, and finally in the creation of canonical syllabic categories in stage 4. Thus, the occurrence of spontaneous production does not end when the first evidence of vocal exploration begins, but rather growth of the capabilities is interleaved with spontaneous production leading the way and vocal exploratory elaboration following as each new domain of potential

signaling (and possibly each new signal type) comes under systematic control by the infant. The growth of these capabilities (spontaneous production, exploration, and category formation) is interleaved across domains of signal control throughout infant vocal development.

Interleaving is also evident in the relation between signal and functional flexibility. Signal flexibility emerges in successive stages (1–4), and functional flexibility emerges in increasingly complex ways repeatedly at each stage to take advantage of the signal developments as they become available. New signals are recruited to serve existing functions as the signals become available, and the communicative functions that can be served become more elaborate as development proceeds, presumably in part because there are new signals available to serve them.

The interleaving of developments in different domains across years implies that the stages of signal and functional flexibility indicated above do not specify discrete begin and end points for each vocal capability. Instead the stages are characterized by increases in *degree* of capability in the designated domains, and development in those domains is interleaved across time. The stages of vocal development provide a heuristic overview of ordering, but the processes they entail show overlapping developmental schedules.

Interleaving is seen in many domains of development. MacWhinney (1982) has pointed out that in the development of productive syntax, rote learning of word strings and analysis of word strings as individual words both play important roles, with rote learning repeatedly leading the way and analysis following for each new item or set of items (words and word strings) that enter the child's repertoire. Interleaving of processes (rote learning and analysis) is evident, since rote learning does not cease its growth when analysis begins but continues expanding to provide a foundation for new analysis possibilities throughout the early development of syntax.

The Natural Logic Alternative to Preformationism: An Analogy to Illustrate Interleaving

Our interpretive approach is distinctly nonpreformationist—we do not assume that mature vocal categories are preordained by an innate language endowment (see, e.g., Pinker, 1994). Key advantages of the natural logic approach over an innatist approach can be illustrated by considering an analogy with embryology. When the seed that will become a tree germinates, it has no trunk, no branches, no leaves, and no bark. The seed does not possess *preformed* mature structures such as these but only precursors to them. One of the first things that is noticed by the casual observer of the tree's growth is that a shoot appears, with no branches and no trunk. The shoot is a precursor to a tree trunk and to every other structure that the tree will possess above ground. As the tree begins to grow, the shoot thickens, and smaller shoots begin to emerge at the leading edge. These are precursors to branches. Eventually,

the base shoot expands so much that we feel comfortable calling it a trunk, and the smaller shoots expand and diversify until we feel comfortable calling them branches. When the tree is mature, it has many new structures both internally and externally, including leaves and bark, that are not at all evident in the seed stage or the shoot stage.

There are two key points to consider in the analogy of the tree's growth to the growth of speech capability. First, in both cases the beginning phases include *no* mature structures but only precursors to them that diversify in stages toward a mature form, with no simple discontinuous change from immature to mature—structures are not preformed (as assumed in traditional innatist proposals) but grow and diversify from precursor forms. Second, the growth in each type of structure is interleaved across time. The shoot does not stop growing when the branches begin to emerge, and the main branches do not stop growing when yet smaller branches begin to emerge or when leaves begin to form. Each stage of development of the tree involves growth at the foundations to support growth at the top. Yet there is a naturally logical precedence of the original shoot to any of its branches and of any of the main branches to any of their smaller branches or to the leaves that will emerge from them.

When we posit a natural logic for human vocal development and evolution, we advocate a view that is analogous to the natural logic of the growth of the tree. Spontaneous production of phonation is a logical foundation for all that vocal language will become, just as the shoot that first appears is a logical foundation for the tree. Neither the shoot nor the capability for spontaneous production ceases to grow after its first appearance—nothing could progress without continued growth of the foundational structures at each stage for both language and the tree. Still, the logical precedence of structural elements is clear. There is an interleaved ordering in both cases where the more foundational elements precede the later ones by necessity.

Naturally Logical Diversity in Evolution and Development

The logically necessary capabilities for speech indicated above (spontaneous production, elaboration of produced forms, category formation, etc.) are abstract, leaving considerable room for individual variation in concrete developmental patterns and in possible routes of evolution. Individual differences among infants in vocal development are notable. For example, while all infants we have studied develop phonatory categories of vowel-like sounds contrasted with squeals or growls in infraphonological stage 3, a variety of additional categories and subcategories (raspberries, ingressive sounds, subcategories of growls produced with either vocal fry or harshness, etc.) occur with considerable individual variation during the same period.

The fact that there are notable *similarities* across infants in the types of sounds that occur at particular stages is presumably a product of similarities in the physical

structures that must be manipulated to produce sound and in the relative efficiency commonly occurring sound categories possess to serve the functions of early vocal communication. We envision a Darwinian competition among various possible categories such that the categories that are most efficiently produceable and/or effective in transmitting the messages they may bear at the stage in question survive and tend to occur frequently. Good examples of less successful forms during phonatory development are ingressive sounds or sounds produced with vocal tremor—they occur occasionally but tend to be infrequent and to drop out of repertoires quickly.

Just as development shows variability consistent with the proposed natural logic, so, according to our proposal, evolution must have proceeded consistent with the natural logic along paths with a variety of options in terms of particular sound types that might have been used and particular functions they might have served at various points of hominin evolution. The logic provides outlines within which evolution was presumably constrained, but the possible concrete routes of progression consistent with the natural logic were various, and there was presumably competition among the routes such that only those that yielded efficient communicative systems survived. Oller (2000) provides speculative scenarios of evolution based on the broad outlines described here, including a discussion of the types of sounds that plausibly might have occurred as communications at various points in hominin evolution.

Natural Logic and Neurological/Physiological Foundations for Language

In accord with Darwinian thinking, we assume that the competition that led to the capabilities for vocal control seen in modern humans included natural selection for efficient neurological systems to implement the stages of natural logic in development. Hammerschmidt and Fischer (this volume) and Owren and Goldstein (this volume) both review literature on neurological foundations for vocal flexibility in humans and contrast those foundations with those observed in nonhuman primates. We shall not provide a detailed proposal here regarding how the natural logic we propose might be implemented neurologically. However, the proposal is indeed intended to imply specific mechanisms of neurological structure and development to account for the stages of the natural logic, and so we provide a brief outline of the mechanisms we envision consistent with the stages of development and the natural logic.

The primate background provides evidence to support schematic components of vocal communicative control presumed to be present in all primates, including humans. This pan-primate system includes at least the following components:

1. A vocal output pathway primarily associated with the brain stem's periaqueductal gray (PAG; see Hammerschmidt and Fischer, this volume, for references).

2. An emotional interpretation system (presumably subcortical) that derives information from various sensory modalities (Damasio, 1999; LeDoux, 1998), affording the basis for determination of (fixed) signals appropriate to particular circumstances.

3. Pathways from emotional interpretation centers that feed limbic-based control centers to initiate largely involuntary commands to the PAG (see Hammerschmidt and Fischer, this volume).

4. Fairly well-developed pan-primate auditory categorization systems presumably housed in the temporal lobe's supramarginal gyrus, allowing early speech discrimination in humans as well as remarkably similar discrimination in a variety of nonhuman species including nonhuman primates (see evidence in Eimas et al., 1971; Jusczyk, 1992; Kluender et al., 1987; Kuhl and Miller, 1978; Trehub, 1976).

In the human case, it would appear that new subsystems of control must emerge to account for each of the stages of development that we have outlined:

1. The first *new* system not present in nonhuman primates to come online at infraphonological stage 1 may be a frontal cortex control system making spontaneous phonation possible. This system could include not only excitatory but also disinhibitory properties to allow frequent spontaneous production of vocalization. Disinhibition might help account for relatively unfettered production in human infants in contrast with otherwise more strictly constrained vocalizations that usually occur in nonhuman primates, where vocalizations tend to have stereotyped form and to occur only when the immediate circumstances warrant their production. Such a new frontal control system, according to our proposal, at an early stage of human development could make spontaneous production of quasivowels possible.

2. At infraphonological stage 2, expansion of functions of the frontal control system could make possible production of more variable sounds (as we see, e.g., in early quasi-vowel elaborations in duration, vocal quality, and pitch). We propose that an additional new neurological foundation of stage 2 may create greater *motivation* to explore vocalizations, a motivation that seems lacking in nonhuman primates. In contrast to the differences in vocal exploration between humans and nonhuman primates, both nonhuman primates and human infants show substantial playful exploratory tendencies in nonvocal domains (see Kuczaj and Makecha, this volume; Thelen, 1995).

3. In order to form new categories from newly elaborated voluntary sounds, we suggest that a new categorization feedback loop must be established in infraphonological stage 3, between the newly formed frontal motor control system and the pan-primate auditory categorization system. This connection (perhaps instantiated at least in part in the arcuate fasciculus) is proposed to account for the self-monitoring and sensitivity to ambient vocalizations that appear to play a key role in the human capability to produce new vocal categories.

4. At infraphonological stage 4 we see strong evidence of a new neurological system allowing voluntary *coordination* of phonatory (glottal) and supraglottal articulatory actions (presumably an articulatory frontal cortex control system, to complement the phonatory one). This system may begin to emerge at stage 2 and account for the primitive articulations of gooing, but the lack of systematic coordination, especially in timing of phonation and articulation in gooing, suggests that the system must become much more active in stage 4, where fine coordination of rapid articulation with phonation is the defining feature of the canonical stage.

5. Finally, a system that allows for freely expressive production of newly controlled categories would presumably require a feedback loop between the pan-primate emotional interpretation system and the new human-specific systems controlling the development and production of new vocal categories. This loop, according to our proposal, would allow for the production of new categories to serve a variety of communicative functions.

Genetic and Epigenetic Changes and Naturally Logical Stages

It is uncertain how much genetic change specific to communication would be necessary to establish the foundational capabilities for language. The reason for the uncertainty is that genetic changes could provide naturally logical foundations not specific to communication, foundations making possible greater capabilities for seeking, learning, and problem solving. An organism with these noncommunicative foundations could nonetheless gain advantages supporting self-organization of additional capabilities specific to spoken language (Kent, 1992; Sachs, 1988).

We propose (in accord with a variety of other theorists such as Bates and Mac-Whinney, 1982, and Tomasello, 2003) that both specific genetic changes corresponding to language-specific capabilities and other epigenetic changes corresponding to capabilities that may have allowed for self-organizing development of various language-necessary capabilities have been involved in the remarkable process of hominin communicative evolution. In accord with this way of thinking, we will be required ultimately, in seeking to understand the evolution of language, to seek ways to differentiate between the genetic changes that ultimately made language possible and the epigenetic developmental processes through which the concrete form of language is built.

Natural Logic in Other Domains of Child Development

Although the term "natural logic" has not been widely used in the sense we intend for it, there are many examples in child development that are consistent with our viewpoint. Consider, for example, that infants do not develop stranger fear until

they recognize individual faces (see review in Lafreniere, 2000). The application of natural logic seems straightforward: Infants cannot be fearful of strangers until they can recognize strangers as such.

Similarly, it is straightforward that infants manifest substantial general postural control, including, for example, head and neck control, long before they are able to sit unsupported (Bayley, 1969). The latter development includes the necessity for general postural control, of which head and neck control are examples, and consequently sitting can be seen as a capability that would be impossible until basic head and neck control are established, since these are subcomponents of general postural control.

In some cases the implied natural logic of sequences of development appears to be straightforwardly applicable to evolutionary scenarios. Consider an example from the foundations for "theory of mind": Until infants can track the gaze of others, they do not acquire the capability for "alternating joint attention." In the latter case, they follow another person's gaze toward an object and then quickly shift their gaze back toward the other, sharing attention to the object with another person (Butterworth, 1996; Mundy et al., 1992; Mundy and Willoughby, 1996). The simple tracking of gaze can be seen as a necessary and logical precursor to the more complex act of gaze sharing. Gaze sharing is more complex because it involves both tracking *and* alternation.

Infant initiation of joint attention is seen when infants initiate a communicative event, pointing and engaging in alternating gaze between a designated object and another person—infants often vocalize to help initiate such an event. Initiation of joint attention represents an additional step in the process of development beyond gaze tracking and gaze sharing, and it seems logically to require the foundations of the prior developments because infants would have no reason to anticipate sharing of attention to something they see until they have experienced the following of gaze and subsequent alternation of gaze with another person.

These steps of development suggest a natural progression based on the establishment of primitive, simpler capabilities along with the building of more complex capabilities through elaboration of, or combinations of, the simpler ones. In the case of the gaze examples, comparisons with other primate species provide confirmation of the naturally logical sequence that we have described. In general, the data suggest that gaze tracking occurs in some species without training by humans but that more advanced functions of joint attention are harder to find in nonhumans and may occur only when special training is provided (Povinelli, 2000; Povinelli and Eddy, 1996; Tomasello, 1996).

Consequently, hominin evolution appears to have proceeded in such a way that alternating joint attention and initiated alternating joint attention appeared after gaze tracking was already in place. Speculation about such a sequence in evolution is of

particular interest because it seems obvious that without joint attention in fairly elaborate form, it would be impossible to develop fully referential words. The latter depend on an understanding that a word spoken (e.g., "door") can represent a shareable concept (e.g., a particular class of objects, doors). Consequently, we reason that concrete concepts can be shared between individuals through words only if these concepts have been previously shared in the absence of words.

Orderly Sequences in Language Development that Have Already Come to Be Viewed as Corresponding to Evolutionary Sequences

Simple Concatenative Syntax Precedes More Advanced Grammar

Parallel with the sort of reasoning that has been pursued regarding joint attention, there is increasing acceptance of the idea that logical sequences of events in nonevolutionary domains can help to specify evolutionary sequences. Bickerton, for example, has suggested that there existed in hominin history a stage of "protolanguage" in which only the simplest sort of syntax existed (Bickerton, 1981, 1990). Words were combined in telegraphic style, with no inflections (i.e., grammatical markers on root words) or bound morphemes, and constructions were short. His reasoning draws support from pidgin languages that serve as media of exchange among peoples who share no full-fledged language. He points out that such systems are important and communicatively valuable as simple languages of trade even though they include no inflections or bound morphemes. Consequently, it seems reasonable, in accord with Bickerton's reasoning, to suggest that hominin evolution included a protolanguage stage that resembled pidgin languages.

Developmental patterns in humans in the second and third years of life conform to the patterns that would be expected based on the idea that the protolanguage-to-full-fledged-language sequence is required by natural logic. Children go through a telegraphic communication stage during which early words are produced in a way that has much in common with pidgin languages (Bloom, 1970; Brown, 1973). Only thereafter do bound morphemes and other inflectional phenomena take hold. The sequence makes perfect logical sense, because bound morphemes and inflectional elements provide fine-tuning for the syntactico-semantic relations among individual words produced in sequence. What use would such grammatical devices be in the absence of individual root words? And so it seems reasonable to posit a naturally logical sequence in which simple word concatenation syntax precedes the grammar of bound morphemes and other inflectional phenomena and to infer that any developmental or evolutionary sequence that could ever reach a level of grammar would have to have also reached a level of word concatenation syntax or protolanguage.

Bickerton's reasoning suggests that protolanguage could have represented a stable evolutionary stage prior to the appearance of grammar. The reasoning seems to be supported by (1) normal developmental patterns as suggested above, (2) by the fact that there exist language disorders where grammatical capabilities are severely disrupted but where concatenation syntax is present (Johnston, 1988; Leonard and Schwartz, 1985) and, (3) by the fact that apes that learn signs seem to be capable of simple concatenation syntax but not grammar (Gardner and Gardner, 1969; Terrace et al., 1979). These considerations suggest two genetically determined stages in hominin evolution, presumably a pattern of interleaved growth of capabilities with concatenation syntax leading true grammar.

Words Precede Simple Syntax

In a similar line of reasoning, it can be noted that the occurrence of utterances consisting of single truly referential words precedes concatenation syntax in modern children and must have done so in hominin evolution. There exists the illusion of an exception to this principle in the case where children produce short multiword utterances by rote; however, it has become clear that in such cases children have misanalyzed what they have heard in adult speech, such that a short phrase or other short word sequence from the adult language has been understood by the child as an unanalyzed whole (Bloom, 1970). Later in development these unanalyzed types have to be reanalyzed by the child as multiple words that can function independently. As in the case of telegraphic speech, there is stage stability in children, such that words are produced one at a time for a notable period, after which concatenation in simple syntax begins.

This pattern suggests the plausibility of a "word stage" of evolution, where ancient hominins spoke to each other only one word at a time and presumably with a small vocabulary compared to vocabularies of modern languages or even pidgin languages. While the advantages of such communication fell far short of the advantages of modern language, a "word stage" could surely have had enormous communicative benefit, even with no syntax at all.

The small size of presumed vocabularies at this proposed evolutionary stage also corresponds to strong evidence from modern human childhood. Children learn and use single-word utterances for months before beginning to use productive word combinations. And further, vocabulary size shows both a strong positive predictive value with regard to the appearance of early concatenative syntactic forms and further positive predictive value with regard to the emergence of more advanced grammatical forms (Bates, 1996). It seems, then, necessary for children to amass a critical number of words that can be used individually before syntax can take hold. These facts are consistent with a natural logic whereby combinatorial syntax of any kind presupposes control of the individual words that syntactic operations combine.

The notion of interleaving of developmental stages and processes introduced above is also relevant to evolution. The idea of a naturally logical sequence for emergence of primitive syntax, as we propose it, includes the interleaving of the ability to produce and understand individual words with the logically subsequent ability to analytically produce and understand strings composed of words (MacWhinney, 1982, 1988). The processes are overlapped, and the naturally logical sequence is repeatable at increasingly higher levels of complexity as the lexicon grows and the syntax itself becomes more complex. In accord with our general view that development and evolution have important parallels, we propose that interleaving of word and (analytical) multiword stages must have occurred also in the evolution of human language.

Canonical Babbling Precedes Large-Scale Vocabulary

By a similar line of reasoning to that proposed for primitive syntax, it is arguable that canonical babbling (which implies the production of well-formed syllables especially notable in reduplicated sequences such as "baba" or "dada") precedes individual word development in a naturally logical way (Oller, 1978, 1980, 2000; Oller and Griebel, 2005). The argument is supported empirically by the following observations:

1. In normally developing infants, the appearance of a substantial vocabulary of conventional words is always preceded by the development of canonical babbling.

2. Nonhuman primates do not appear to be able to produce voluntarily controlled canonical syllables.

3. The vast majority of words in natural languages consist of canonical syllables.

This evidence has been accorded prominence in the writings of MacNeilage, who argues for the emergence in ancient hominins of specific neural mechanisms to support canonical babbling as a foundation for speech (MacNeilage, 1998; MacNeilage and Davis, 1990, 1993).

A proviso regarding the relation between canonical babbling and word use is important: Some words (although very few relative to the vast size of human vocabularies) in natural languages consist of noncanonical syllables. The formal definition of "canonical syllables" includes the specification that each such syllable must possess at least one supraglottally articulated consonant. However, some words either have no consonants at all or the only consonants they have are glottal, and thus these words do not require supraglottal articulation. Examples of such noncanonical syllables that constitute words in English are "he" and "ah." Children, when they begin to talk, sometimes include a small number of words with noncanonical form. The existence of noncanonical word forms does not, however, undercut the basic logic of the argument that canonical syllable production is a prerequisite to full-scale vocabulary development. Only a very small number of words can be constructed without

articulated consonants, and so there is a practical (a naturally logical) requirement that canonical syllables precede *large-scale* vocabulary development as occurs in children. Phonologies abide by a "particulate principle" that affords the possibility of recombination of syllables to produce lexicons of virtually unlimited magnitude (Studdert-Kennedy, 2000).

In accord with the interleaving characteristic of developmental patterns, we see continued increases in the numbers of canonical syllables that children can command in production as the lexicon grows. New words often require new syllables to produce them, and consequently canonical syllable development must continue to support lexical development for many months after the first appearance of words composed of canonical syllables.

The naturally logical sequence of canonical syllable production followed by word production appears to be dependent, then, on large vocabulary size—very small-scale vocabulary development can occur prior to development of canonical syllables, but large-scale development of vocal lexicons cannot occur in the absence of canonical syllables.

Oller (2000) has pointed out that the earliest meaningful and contextually flexible vocal communications in ancient hominins could have been noncanonical in type. The fact that infants only occasionally use noncanonical forms as words suggests the possibility that the communicative value of contextually flexible noncanonical forms in ancient hominins may have been primarily of a simple illocutionary sort, with each form corresponding to an act such as requesting, rejecting, affiliating, and so on, a pattern of usage that also occurs in noncanonical early word usage by modern children. More advanced semantic communications, where words are used to reference concepts with multiple possible illocutionary forces that can be invoked on different occasions, would have occurred later in evolution, according to our reasoning, following from the observations that semantically laden communication also tends to occur later than simple illocutionary communication in modern child development (Bates et al., 1979). Whatever the situation may turn out to be in terms of the relation between the appearance of semantically laden words and canonical or noncanonical forms, the general point at stake here seems to have been largely unchallenged: Development of canonical syllable control must have played a major role in establishing foundations for full-scale vocabulary evolution in hominins.

The Need for an Explicit Enterprise to Develop Natural Logic as an Explanatory Framework

Thus, in at least three widely publicized cases (simple concatenation syntax precedes grammar, single words precede simple concatenation syntax, and canonical syllable

control precedes significant vocabulary growth), there appears to be general acceptance of the idea that certain logical relations implied by ordered human development can be reasonably adopted in the construction of evolutionary scenarios. Acceptance of these ideas is consistent with the notion that natural logic provides the connection accounting for parallelism between development and evolution.

Bickerton (1990), Oller (2000), and Jackendoff (2002) have all supplied stage models for evolution of language that are, at least in part, built upon an idea of naturally logical sequencing. Given that these ideas seem to be at least tacitly accepted in many quarters, and explicitly posited in others, it seems important to take the notion of natural logic further. What is needed is a general theory of the natural logic of vocal communication. Issues of natural logic have often been invoked in speculations about why infant development proceeds as it does or about why evolution must have proceeded as it does, but the speculations have tended to invoke ad hoc constraints (Prince and Smolensky, 1993) or ad hoc possible solutions in a communication problem space (Elman et al., 1996).

The first author has argued elsewhere (Oller, 2005) in favor of a fundamental replacement for the ad hoc approach, an overarching theory of natural logic in the emergence of communicative systems. The study of natural logic should become a general effort to characterize possible sequences of evolution (and development) in vocal communication away from the primate background and toward systems of communication with greater power. An initial model of communicative natural logic is provided in Oller (2000). The model outlines properties of capability required to be developed for language, starting with contextual flexibility (both signal and functional flexibility) in the simple forms detailed in this chapter, and forming foundations for such further sequentially developed capabilities as recombinability of syllables, vocal imitation, learning of conventional words, formation of truly semantic word categories with illocutionary flexibility, formation of propositions (simple syntax), and finally steps of more advanced grammatical development.

If we are to succeed in applying a naturally logical model to language evolution, it would seem critical that we start at the very beginning of the break between the hominin line and our nonhominin primate ancestors. We ask then, what were the first logical steps of vocal communication upon which all others must have been built? As we have argued here, a tentative answer is found in interpretation of the patterns of vocalization of the human infant, beginning well before canonical syllable control, and providing foundations for all that vocal language becomes.

Acknowledgments

References

Bates E. 1996. *From preverbal communication to grammar in children*, Waseda University International Conference Center, Tokyo.

Bates E, Benigni L, Bretherton I, Camaioni L, Volterra V. 1979. *The Emergence of Symbols: Cognition and Communication in Infancy*. New York: Academic Press.

Bates E, MacWhinney B. 1982. Functionalist approaches to grammar. In *Language Acquisition: The State of the Art*, ed. E Wanner, LR Gleitman, pp. 173–218. Cambridge, England: Cambridge University Press.

Bayley N. 1969. *Bayley Scales of Infant Development: Birth to Two Years*. New York: Psychological Corporation.

Becker ML, Buder EH, Ward JP. 2003. Spectrographic description of vocalizations in captive *Otolemur garnetti*. *International Journal of Primatology* 24: 415–46.

Bickerton D. 1981. *Roots of Language*. Ann Arbor, MI: Karoma.

Bickerton D. 1990. *Language and Species*. Chicago: University of Chicago Press.

Bloom L. 1970. *Language Development*. Cambridge, MA: MIT Press.

Boliek CA, Hixon TJ, Watson PJ, Morgan WJ. 1996. Vocalization and breathing during the first year of life. *Journal of Voice* 10: 1–22.

Brown R. 1973. *A First Language*. London: Academic Press.

Buder EH, Chorna L, Oller DK, Robinson R. In press. Vibratory regime classification of infant phonation. *Journal of Voice*.

Buder EH, Oller DK, Magoon JC. 2003. *Vocal intensity in the development of infant protophones*. Proceedings of the XVth International Congress of Phonetic Sciences, ed. MJ Solé, D. Recasans, J Romero, pp. 2015–2018. Adelaide, Australia: Casual Productions.

Butterworth G. 1996. *Species typical aspects of manual pointing and the emergence of language in human infancy*, Waseda University International Conference Center, Tokyo.

Chomsky N. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Damasio A. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

de Saussure F. 1968. *Cours de Linguistique Générale*. Paris: Payot.

Eimas PD, Siqueland E, Jusczyk P, Vigorito J. 1971. Speech perception in infants. *Science* 171: 303–6.

Elman JL, Bates EA, Johnson MH, Karmiloff-Smith A, Parisi D, Plunkett K. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Gardner RA, Gardner BT. 1969. Teaching sign language to a chimpanzee. *Science* 165: 664–72.

Haeckel E. 1866. *Generelle Morphologie der Organismen: 2 volumes*. Berlin: Georg Reimer.

Hauser M, Chomsky N, Fitch WT. 2003. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–79.

Jackendoff R. 2002. *Foundations of Language*. Oxford: Oxford University Press.

Johnston JR. 1988. Specific language disorders in the child. In *Handbook of Speech-Language Pathology and Audiology*, ed. NJ Lass, LV McReynolds, JL Northern, DE Yoder, pp. 685–715. San Diego, CA: Singular Publishing Group.

Jusczyk PW. 1992. Developing phonological categories from the speech signal. In *Phonological Development: Models, Research, Implications*, ed. C Ferguson, L Menn, C Stoel-Gammon, pp. 17–64. Parkton, MD: York Press.

Kent RD. 1992. The biology of phonological development. In *Phonological Development: Models, Research, Implications*, ed. C Ferguson, L Menn, C Stoel-Gammon, pp. 65–89. Parkton, MD: York Press.

Kessen W, Levine J, Wendrich K. 1979. The imitation of pitch in infants. *Infant Behavior and Development* 2: 93–9.

Kluender KR, Diehl RL, Killeen PR. 1987. Japanese quail can learn phonetic categories. *Science* 237: 1195–7.

Koopmans-van Beinum FJ, Clement CJ, van den Dikkenberg-Pot I. 2001. Babbling and the lack of auditory speech perception: A matter of coordination? *Developmental Science* 4: 61–70.

Koopmans-van Beinum FJ, van der Stelt JM. 1986. Early stages in the development of speech movements. In *Precursors of Early Speech*, ed. B Lindblom, R Zetterstrom, pp. 37–50. New York: Stockton Press.

Kuhl PK, Miller JD. 1978. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America* 63: 905–17.

Kwon K, Buder EH, Oller DK. 2006a. *Contextual flexibility in precanonical infant vocalizations: Its role as a foundation for speech*, International Child Phonology Conference, Edmonton, AB, Canada.

Kwon K, Buder EH, Oller DK, Chorna LB. 2006b. *The role of fundamental frequency characteristics in classifying prebabbling vocalizations*, Annual meeting of the American Speech-Language-Hearing Association, Miami, FL.

Lafreniere PJ. 2000. *Emotional Development: A Biosocial Perspective*. Belmont, CA: Wadsworth Press.

LeDoux J. 1998. *The Emotional Brain*. London: Weidenfeld and Nicolson.

Leonard LB, Schwartz RG. 1985. Early linguistic development of children with specific language impairment. In *Children's Language*, ed. KE Nelson, pp. 291–318. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lorenz K. 1951. Ausdrucksbewegungen höherer Tiere. *Naturwissenschaften* 38: 113–6.

MacNeilage PF. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21: 499–546.

MacNeilage PF, Davis BL. 1990. Acquisition of speech production: Frames then content. In *Attention and Performance XIII: Motor Representation and Control*, ed. M Jeannerod, pp. 453–76. Hillsdale, NJ: Lawrence Erlbaum Associates.

MacNeilage PF, Davis BL. 1993. Motor explanations of babbling and early speech patterns. In *Changes in Speech and Face Processing in Infancy: A Glimpse at Developmental Mechanisms of Cognition*, ed. B de Boysson-Bardies, S Schonen, P Jusczyk, PF MacNeilage, J Morton. Dordrecht: Kluwer Academic Publishers.

MacWhinney B. 1982. Basic syntactic processes. In *Language Development, Vol. 1: Syntax and Semantics*, ed. S Kuczaj, pp. 73–136. Hillsdale, NJ: Lawerence Erlbaum Associates.

MacWhinney B. 1988. Competition and teachability. In *The Teachability of Language*, ed. R Schiefelbusch, M Rice, pp. 63–104. New York: Cambridge University Press.

McCune L, Vihman M, Roug-Hellichius L, Delery K, Gogate L. 1996. Grunt communication in human infants (*Homo sapiens*). *Journal of Comparative Psychology* 110: 27–37.

Mundy P, Kasari C, Sigman M. 1992. Joint attention, affective sharing, and intersubjectivity. *Infant Behavior and Development* 15: 377–81.

Mundy P, Willoughby J. 1996. Non-verbal communication, joint attention and social-emotional development. In *Emotional Development in Atypical Children*, ed. M Lewis, M Sullivan, pp. 65–87. New York: Wiley.

Nooteboom SG. 1999. Anatomy and timing of vocal learning in birds. In *The Design of Animal Communication*, ed. MD Hauser, M Konishi, pp. 63–110. Cambridge, MA: MIT Press.

Oller DK. 1978. Infant vocalization and the development of speech. *Allied Health and Behavioral Sciences* 1: 523–49.

Oller DK. 1980. The emergence of the sounds of speech in infancy. In *Child Phonology, Vol. 1: Production*, ed. G Yeni-Komshian, J Kavanagh, C Ferguson, pp. 93–112. New York: Academic Press.

Oller DK. 1981. Infant vocalizations: Exploration and reflexivity. In *Language Behavior in Infancy and Early Childhood*, ed. RE Stark, pp. 85–104. New York: Elsevier North Holland.

Oller DK. 2000. *The Emergence of the Speech Capacity*. Mahwah, NJ: Lawrence Erlbaum Associates.

Oller DK. 2005. The natural logic of communicative possibilities: Modularity and presupposition. In *Modularity: Understanding the Development and Evolution of Natural Complex Systems*, ed. W Callebaut, D Raskin-Gutman, pp. 409–34. Cambridge, MA.

Oller DK, Buder EH, Nathani S. 2003. *Origins of speech: How infant vocalizations provide a foundation.* Miniseminar for the American Speech-Language Hearing Association Convention, Chicago.

Oller DK, Eilers RE, Basinger D. 2001. Intuitive identification of infant vocal sounds by parents. *Developmental Science* 4: 49–60.

Oller DK, Griebel U. 2005. Contextual freedom in human infant vocalization and the evolution of language. In *Evolutionary Perspectives on Human Development*, ed. R Burgess, K MacDonald, pp. 135–66. Thousand Oaks, CA: Sage Publications.

Oller DK, Griebel U. 2008. The origins of syllabification in human infancy and in human evolution. In *Syllable Development: The Frame/Content Theory and Beyond*, ed. B Davis, K Zajdo, pp. 29–62. Mahwah, NJ: Lawrence Erlbaum Associates.

Oller DK, Levine S, Eilers RE, Pearson BZ. 1998. Vocal precursors to linguistic communication: How babbling is connected to meaningful speech. In *The Speech/Language Connection*, ed. R Paul, pp. 1–23. Baltimore, MD: Paul H. Brookes.

Oller DK, Nathani Iyer S, Buder EH, Kwon K, Chorna L, Conway K. 2007. *Diversity and contrastivity in prosodic and syllabic development*. In Proceedings of the International Congress of Phonetic Sciences, ed. Trouvain J, Barry W, pp 303–308. Saarbrucken, Germany: University of Saarbrucken.

Pinker S. 1994. *The Language Instinct*. New York: Harper Perennial.

Povinelli DJ. 2000. *Folk Physics for Apes: The Chimpanzee's Theory of How the World Works*. Oxford: Oxford University Press.

Povinelli DJ, Eddy TJ. 1996. What chimpanzees know about seeing. *Monographs of the Society for Research in Child Development* 61: 191.

Prince A, Smolensky P. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. New Brunswick, NJ: Rutgers University Press.

Sachs Y. 1988. Epigenetic selection: An alternative mechanism of pattern formation. *Journal of Theoretical Biology* 134: 547–59.

Scheiner E, Hammerschmidt K, Jürgens U, Zwirner P. 2002. Acoustic analyses of developmental changes and emotional expression in the preverbal vocalizations of infants. *Journal of Voice* 16: 509–29.

Snowdon C. 2004. Social processes in the evolution of complex cognition and communication. In *Evolution of Communication Systems: A Comparative Approach*, ed. DK Oller, U Griebel, pp. 131–50. Cambridge, MA: MIT Press.

Snowdon CT, Elowson AM, Rousch RS. 1997. Social influences on vocal development in New World primates. In *Social Influences on Vocal Development*, ed. CT Snowdon, M Hausberger, pp. 234–48. New York: Cambridge University Press.

Snowdon CT, Hodun A. 1981. Acoustic adaptation in pygmy marmoset contact calls: Locational cues vary with distance between conspecifics. *Behavioral Ecology and Sociobiology* 9: 295–300.

Stark R, Bernstein LE, Demorest ME. 1993. Vocal communication in the first 18 months of life. *Journal of Speech and Hearing Research* 36: 548–58.

Stark RE. 1978. Features of infant sounds: The emergence of cooing. *Journal of Child Language* 5: 379–90.

Stark RE. 1980. Stages of speech development in the first year of life. In *Child Phonology, Vol. 1*, ed. GY Komshian, J Kavanagh, C Ferguson, pp. 73–90. New York: Academic Press.

Stark RE, Rose SN, McLagen M. 1975. Features of infant sounds: The first eight weeks of life. *Journal of Child Language* 2: 205–22.

Studdert-Kennedy M. 2000. Evolutionary implications of the particulate principle. In *The Evolutionary Emergence of Language*, ed. C Knight, M Studdert-Kennedy, JR Hurford, pp. 161–76. Cambridge, England: Cambridge University Press.

Terrace HS, Petitto LA, Sanders RJ, Bever TG. 1979. Can an ape create a sentence? *Science* 206: 891–902.

Thelen E. 1981. Rhythmical behavior in infancy: An ethological perspective. *Developmental Psychology* 17: 237–57.

Thelen E. 1994. Three-month-old infants can learn task-specific patterns of interlimb coordination. *Psychological Science* 5: 280–5.

Thelen E. 1995. Motor development: A new synthesis. *American Psychologist* 50: 79–95.

Tinbergen N. 1951. *The study of instinct*. Oxford: Oxford University Press.

Tomasello M. 1996. *The Gestural Communication of Chimpanzees and Human Children*, Waseda University International Conference Center, Tokyo.

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

Trehub S. 1976. The discrimination of foreign speech contrasts by infants and adults. *Child Development* 47: 466–72.

Zlatin-Laufer MA, Horii Y. 1977. Fundamental frequency characteristics of infant non-distress vocalization during the first 24 weeks. *Journal of Child Language* 4: 171–84.

Zlatin M. 1975. Preliminary descriptive model of infant vocalization during the first 24 weeks: Primitive syllabification and phonetic exploratory behavior. Progress Report, National Institutes of Health Research Grants.