BEHAVIORAL AND BRAIN SCIENCES (2004) **27,** 491–541
Printed in the United States of America

Prelinguistic evolution in early hominins: Whence motherese?

Dean Falk
Department of Anthropology, Florida State University, Tallahassee, FL 32306-7772
dfalk@fsu.edu **http://www.anthro.fsu.edu/people/faculty/falk.html**

Abstract: In order to formulate hypotheses about the evolutionary underpinnings that preceded the first glimmerings of language, mother-infant gestural and vocal interactions are compared in chimpanzees and humans and used to model those of early hominins. These data, along with paleoanthropological evidence, suggest that prelinguistic vocal substrates for protolanguage that had prosodic features similar to contemporary *motherese* evolved as the trend for enlarging brains in late australopithecines/early *Homo* progressively increased the difficulty of parturition, thus causing a selective shift toward females that gave birth to relatively undeveloped neonates. It is hypothesized that hominin mothers adopted new foraging strategies that entailed maternal silencing, reassuring, and controlling of the behaviors of physically removed infants (i.e., that shared human babies' inability to cling to their mothers' bodies). As mothers increasingly used prosodic and gestural markings to encourage juveniles to behave and to follow, the meanings of certain utterances (words) became conventionalized. This hypothesis is based on the premises that hominin mothers that attended vigilantly to infants were strongly selected for, and that such mothers had genetically based potentials for consciously modifying vocalizations and gestures to control infants, both of which receive support from the literature.

Keywords: bipedalism; brain size; chimpanzees; foraging; gestures; hominins; infant riding; motherese; prosody; protolanguage

1. Introduction

One of the most fascinating puzzles to confront evolutionary biologists has to do with *Homo sapiens*' ability for speech. Why are we the only animals that talk? How and when did our ancestors begin to formulate and spew forth segmented bits of air into meaningful sequences, and what behaviors led to the earliest language (protolanguage)? In order to formulate hypotheses about the evolutionary underpinnings that *preceded* the first glimmerings of speech in early hominins, this article synthesizes findings from infant and child development, psychology, primatology, and anthropology.

It is widely recognized that acquisition of vocal language is scaffolded onto the special sing-song way in which parents vocalize to their infants, known as "baby talk" or *motherese* (Dooling 1974; Ferguson 1977; Hirsh-Pasek & Golinkoff 1996; Hirsh-Pasek et al. 1987; Karmiloff & Karmiloff-Smith 2001; Monnot 1999; Snow 1972; 1998; 2002). As detailed below, the worldwide practice of directing musical speech toward human babies provides a temporary framework or scaffold that, among other functions, facilitates their eventual comprehension and production of speech. Nevertheless, one school of thought views the main feature that distinguishes motherese from adult-directed (AD) speech, namely tone of voice or prosody, as a component of a primate gesture-call system that is totally separate from language. Burling (1993), for example, notes that "tone of voice amounts to an invasion of language by something that is fundamentally different"(p. 30). However, because motherese is the medium in which infants around the world initially perceive and eventually process their respective languages, an analysis of its features may elucidate the prelinguistic foundations of the protolanguage(s) evolved by early hominins. Instead of separating prosody from language, then, the view developed below is that parental prosody is not only an integral component for propagating language today, it also formed an important substrate for the natural selection of protolanguage in early *Homo.* In addition to focusing on infant-directed (ID) communications of parents, clues for modeling the evolution of prelinguistic behaviors are also gleaned from examining the processes by which infants acquire languages.

Although there is a robust literature on the vocal aspects of motherese, few workers have appreciated the important parallel roles of mother-infant interactions in visual, gestural, and tactile domains. For example, infant-directed communications from mothers of 3- to 4-month-old infants are frequently accompanied by exaggerated facial expressions that have precursors in other primates and that signal

DEAN FALK is the Hale G. Smith Professor of Anthropology at Florida State University and Honorary Professor of Human Biology at the University of Vienna. She is the author of over 100 publications that focus on early hominids, brain evolution, comparative neuroanatomy, primate behavior, and cognitive evolution. Research on cranial blood flow and australopithecine endocasts led Falk to develop the "radiator hypothesis" of brain evolution and to question the conventional interpretations of certain fossils. Falk is the author of *Braindance: Revised and Expanded Edition* (2004). In 2003, she received the Austrian Cross of Honor for Science and Art, 1st class.

© 2005 Cambridge University Press 0140-525X/04 $12.50

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

affiliation and invitation for contact (e.g., raise eyebrows, eyebrow flash, smile, nod, bob head backward) (Dissanayake 2000). With this in mind, mother-infant interactions that encompass visual, vocal, gestural, and tactile communication are compared below in chimpanzees and humans in order to identify the probable nature of the mother-infant interactions that characterized early hominins.[1]

Hominins are believed to have spent much of their prehistory in fission-fusion communities that foraged daily for food, which entailed mothers traveling in the company of dependent offspring and a small number of other individuals (Nishida 1968; Stanford 1998). Around the time of the australopithecine/early *Homo* transition, maternal pelves that had been modified to accommodate bipedalism became subject to an emerging trend for increasingly large brains (Falk 1998; Falk et al. 2000), which eventually caused a selective shift toward females that gave birth to relatively helpless infants (Small 1998). Consequently, the ability of babies to cling actively to their mothers was lost in hominins (Ross 2001). Similar to some anthropoid mothers that live under difficult foraging circumstances (Fuentes & Tenaza 1995; Lyons et al. 1998), these mothers are hypothesized to have adopted postnatal foraging-related changes in maternal care, which included periodically putting their infants down beside them in order to obtain and process food. As a result, the incidence of distal mother-infant gestural communications increased (Tomasello & Camaioni 1997) and prosodic (affective) vocalizations became ubiquitous to compensate for the reduction in sustained mother-infant physical contact.

The *"putting the baby down" hypothesis* focuses on events that *preceded* the emergence of protospeech, and is in keeping with the *continuity hypothesis* that the biological capacity for language evolved incrementally within the hominin line (Armstrong et al. 1994; King 1996): "differences between human language and nonhuman primate communication are only quantitative and . . . these differences may be accounted for by gradual shifts in abilities due to changing selection pressures – perhaps in the ability to create . . . communicative utterances (Gibson 1990) or to donate information to others" (King 1996, p. 193).

According to the *discontinuity hypothesis,* on the other hand, language appeared suddenly, without phylogenetic links to earlier communication systems (Burling 1993). This latter hypothesis views "language backward through the lens of contemporary linguistic theory rather than in the context of how evolution operates" (Callaghan 1994, p. 359). Most evolutionary biologists, however, believe that reproductive fitness (an individual's production of viable offspring) is the driving force behind evolution and that, whether it proceeds gradually or rapidly, most "evolutionary change occurs in the context of what is already in place as a result of prior selective pressures" (Callaghan 1994, p. 359). The present paper is grounded squarely on this premise. Thus, contemporary motherese is viewed as the result of prior selective pressures, the nature of which is explored in the following sections. Since language acquisition today is universally scaffolded onto motherese, it is argued that selection for vocal language occurred after early hominin mothers began engaging in routine affective vocalization toward their infants, a practice that characterizes modern women, but not relatively silent chimpanzee mothers. Below, it is shown that human infants are "primed" to learn their native languages by the particular flavor of motherese to which they are exposed. As well, data are presented that strongly suggest that this universal practice and its associated ontogenetic unfolding of language acquisition in human infants is genetically driven. For all of these reasons, "positing a phylogenetic discontinuity between primate vocal communication and speech seems to [be] an unnecessarily complicating assumption in the absence of more compelling evidence" (Armstrong et al. 1994, p. 358).

2. Mother-infant interactions in chimpanzees and humans

2.1. Mother-infant communication in chimpanzees and bonobos

Because common chimpanzees (*Pan troglodytes*) and the less-studied bonobos (*Pan paniscus*) provide the best referential models for early hominin behavior (Falk 2000; Moore 1996), this section reviews the literature on their mother-infant interactions in order to provide background for examining the evolution of prelinguistic behaviors. As is the case for humans, the period during which infant and juvenile chimpanzees are emotionally and physically dependent upon their mothers is extended compared to monkeys. Indeed, prolongation of the various developmental stages is thought to be one of the trends that characterized the evolution of higher primates. According to this view, increased durations of dependency facilitated extended learning associated with the evolution of bigger-brained, highly intelligent, and longer-living primates (Falk 2000).

Much of what is known about the vocalizations of wild common chimpanzees has been discovered by Jane Goodall and her colleagues (Goodall 1986). Many emotional states of chimpanzees are obviously similar to those of humans, and are expressed in a variety of easily recognizable facial expressions (Preuschoft 2000; Preuschoft & van Hooff 1995; Schmidt & Cohn 2001) that, in turn, are frequently linked with particular vocalizations. Chimpanzees produce vocalizations by alternating the sizes and shapes of their mouths and resonating cavities, and "facial expressions play a key role in close-up communication between chimpanzees" (Goodall 1986, p. 119), which may be related to the fact that, at about the age of 3 months, infants show "a sudden intense interest for the mother's face" (Plooij 1984, p. 142).

Goodall (1986) notes that vocal communication of chimpanzees is far more complex than previously appreciated, and has classified 34 discrete calls along with the emotions with which they are associated. She also observes that chimpanzee listeners learn much from the sequences of vocalizations that pass back and forth between individuals. (For example, the *screaming* of an adult followed by *squeaks* and then *pant-grunts* indicates to a distant chimpanzee that an aggressive interaction has occurred and that the victim has relaxed and approached the aggressor.) Chimpanzee calls are distinguished (with presumably more difficulty for human than chimpanzee listeners) from an acoustically graded continuum. Thus, the *hoo* is an isolated but distinctive part of the *whimpering* sequence:

> The single hoo may be uttered several times in succession, but each vocalization is made separately; as a hoo sequence starts to rise and fall in pitch and volume, and when each sound is produced in temporally rapid succession, it grades into the whimper. The hoo is uttered by both an infant and (much less often)

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

> his mother when they need to reestablish physical contact – when, for example, the infant wants to ride on his mother's back during travel or when she reaches to retrieve him from a situation she perceives to be dangerous. (Goodall 1986, p. 129)

In addition to *hoos,* several other calls are used by infants as well as older chimpanzees, including *screams* (mothers recognize those of their infants); *whimpering* (most commonly heard in infants, especially during weaning); and *tantrum screams* (which occur in older infants that have been rejected during weaning). Plooij (1984) discusses several additional calls that are emitted by common chimpanzee infants including *effort-grunts, staccatos,* and *uh-grunts.* Because chimpanzees are unable to cling properly for the first two months of life, they are as helpless as human neonates and must be carried and supported on the ventral side of their mothers' bodies (Plooij 1984). Significantly, maternal support for chimpanzee infants varies, is related to their whimpering, and is crucial for infant survival:

> Some mothers supported and carried their babies almost continuously from shortly after birth whereas others restricted themselves to the minimum necessary not to lose their baby. Consequently, during locomotion over greater distances (=travel) babies from the first group were safe; they rarely whimpered or screamed. Babies in the second group, on the other hand, whimpered frequently when loosing their grip on the mother's hair, dangling from only one or two of their four limbs. . . . The maternal support is of vital importance to the baby. *Without it, the baby would surely fall off and may die* (Plooij 1984, p. 45, emphasis mine).

The structure and contextual use of vocalizations of bonobos have been investigated in the wild (Bermejo & Omedes 1999, p. 355). Voices of bonobos are higher pitched than those of common chimpanzees (Kano 1992), and their utterances appear to be more structured and flexible and to always occur in the context of facial expressions, gestures, and tactile communication (Bermejo & Omedes 1999). In bonobos, *peep* sequences are among the most important vocalizations, and *croaks, muffled barks,* and *panting laughs* are used mainly by young individuals. *Peep yelps* and *peeps* that may escalate into *screams* are given by infants that are prevented from nursing, accompanied by intense pouts. Bonobos also produce choruses in which individuals echo each other's calls, and seem to be trading information about emotions and intentions during aggressive confrontations that involve vocalizations, which led de Waal (1997) to suggest that bonobos appear to engage in more language-like exchanges of information about their internal states than do common chimpanzees. Although de Waal did not claim that bonobos talk, they seem, at least, to have a latent ability to learn names, as shown by a study in which two human-enculturated bonobos were able to learn to comprehend English words for novel objects with few exposures to the novel items, an ability that did not require visual contact with items during acquisition of their names (Lyn & Savage-Rumbaugh 2000). In this context it is interesting that, although many believe that apes do not imitate vocally (Fitch 2000), recent spectrographic and statistical analyses reveal that the well-known bonobo Kanzi produces distinct vocalizations for "banana," "grape," "juice," and "yes" (Taglialatela et al. 2003).

2.1.1. Infant-directed vocalizations of common chimpanzees. Mothers of infant chimpanzees are notoriously shy (McGrew 1992) and, except for *hoos,* calls that are specifically directed by mothers to their infants are rarely mentioned in the literature. The few other maternal ID calls noted by Goodall (1986) include *replies to screams* of their infants "even if the child is out of sight" (p. 131), and *soft barks* or *coughs* given in mild rebuke to weaning infants that begin to suckle after throwing temper tantrums (p. 576). Chimpanzee mothers have also been reported to emit soft vocalizations while examining their infants (Nicolson 1977). Maestripieri and Call (1996) note that, when they occur, ID vocalizations of chimpanzee mothers, such as *hoos* and *whimpers,* are similar to the vocalizations produced by their infants. It is significant that one of the few circumstances under which chimpanzee mothers routinely produce ID vocalizations is in conjunction with foraging and travel. For example, *hoos* are uttered to retrieve infants for travel, and "soft grunts may be exchanged when . . . two or more familiar chimpanzees, especially family members, are foraging or traveling together. Typically one individual grunts when he pauses during travel, or when he gets up to move on. . . . Thus these grunts function to regulate movement and cohesion" (Goodall 1986, p. 131).

2.1.2. Infant-directed vocalizations of bonobos. Bermejo and Omedes (1999) note that bonobo mothers in the wild are very sensitive to *screams* of their infants and emit *barks* or *hiccups* during alarm situations, which elicit immediate responses from offspring. Similar to common chimpanzees, bonobo mothers have also been observed vocalizing in order to retrieve infants for travel:

> Nevertheless, the mother often carries her offspring during travel until it is at least three or four years old. The signal initiating this kind of transportation is the mother's vocalization. Then, after walking a short distance, up to 6 m, she will stand with one foot slightly lifted, the sole facing toward the rear, in a stationary walking position. There she will stand, waiting for the juvenile to run after and jump onto her back. (Kano 1992, p. 164)

Bonobos are thought by some to be more intelligent than common chimpanzees, partly because of their relatively greater success at learning nonvocal, humanlike language (Savage-Rumbaugh 1984; Savage-Rumbaugh et al. 1998). Compared to common chimpanzees, the human-enculturated bonobo Kanzi accompanies many gestures with spontaneous vocalizations that "appear to be voluntary and used intentionally to draw attention to Kanzi and to what he wants" (Savage-Rumbaugh 1984, p. 408). Although his adoptive mother (Matata) anticipated and aided Kanzi's developing locomotor activities, there is no indication that she vocalized during these ID gestures. In sum, although both bonobos (Kano 1992) and common chimpanzees have rich vocalization systems, there is little evidence that mothers engage in a significant amount of ID vocalization, in stark contrast to the case for humans.

2.1.3. Infant-directed gestures of common chimpanzees. Although chimpanzees use gestures involuntarily to express moods, as well as intentionally to call attention to themselves or to deliver imperatives, their repertoire of gestures anticipates but fails to achieve the sophistication of that acquired by a typical 1-year-old human child (Tomasello & Camaioni 1997). Tomasello and Camaioni point out three characteristics of natural gesturing in chimpanzees that differ from gesturing in human infants: (1) Chimpanzee gestures are almost exclusively dyadic (used to attract attention

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

to oneself) instead of mostly triadic (used to attract attention to an outside party), (2) their gestures remain largely imperative without developing declarative or referential elements, and (3) most chimpanzee gestures involve physical contact between the signaler and recipient (i.e., they are not distal). Significantly, the two triadic exceptions noted for chimpanzee gestures by Tomasello and Camaioni (1997) appear to be similar to *request* and *offer* gestures of human mother-infant pairs (Messinger & Fogel 1998, see sect. 2.2.1), not only physically, but also motivationally (i.e., used to request food and to seek positive social contact).

For chimpanzees, ID gestural communication appears to be much richer than ID vocal communication. A newborn common chimpanzee is licked and groomed by its mother immediately after birth, and bouts of maternal ID grooming increase in duration during the first year of its life (Goodall 1986). Plooij (1984, Appendix A) documents a rich repertoire of ID gestural and kinesic behaviors toward developing infants in chimpanzee mothers from Gombe related broadly to carrying, cradling, nursing, weaning, play, traveling, and acquisition of motor skills. ID gestures have also been noted for captive mothers (Nicolson 1977), two of which spent considerable amounts of time examining their young infants. One cradled her infant and "kissed" it on the mouth (p. 541). Captive mothers also frequently patted their infants' heads and backs. The captive mothers seemed to test and encourage their infants' developing motor skills by giving them "walking lessons" (pp. 541–42). The first cross-fostered chimpanzee schooled in American Sign Language for the Deaf (ASL), Washoe, has even been reported to mold her adopted son Loulis's hands in the form of a sign (Fouts et al. 1989). (It is important to note, however, that gesturing should not be confused with sign language because it lacks the complex grammar and arbitrariness found in the latter [Karmiloff & Karmiloff-Smith 2001].)

Some of the most interesting ID gestures of chimpanzee mothers have been observed in conjunction with feeding. Mothers begin sharing solid food with infants when they are about 5 months old, and have been observed snatching leaves that were not part of their normal diet from their infants' mouths (Goodall 1986). In addition to teaching infants which foods are palatable, Goodall believes this sort of ID intervention serves to reinforce traditional food preferences in chimpanzee communities. Along these lines, it is fascinating that at least some chimpanzee mothers from the Taï forest, Ivory Coast, have anecdotally been reported to teach their offspring to use implements such as rocks to crack open nuts that have been placed on anvils (Boesch 1991; Boesch & Boesch-Achermann 1991).

Play is the hallmark of a young chimpanzee's life, and its frequency peaks between the ages of 2 and 4 years (Goodall 1986). Females with infants play more than other adults, which entails a good deal of ID physical activity:

> A chimpanzee infant has his first experience of social play from his mother as, very gently, she tickles him with her fingers or with little nibbling, nuzzling movements of her jaws. Initially these bouts are brief, but by the time the infant is six months old and begins to respond to her with play face and laughing, the bouts become longer. Mother-offspring play is common throughout infancy. (Goodall 1986, pp. 369–70)

Significantly, turn-taking in chimpanzees has been documented in the context of mother-infant play: "The early biting triggered the onset of mother-baby play: contingent upon when bitten, the mother started to tickle the baby and this biting-tickling grew into an alternating interaction, in which both mother and baby could take their turns" (Plooij 1984, p. 142).

As the infant matures in the wild, its mother "shapes and cushions his first interactions with other individuals" (Goodall 1986, p. 568), primarily by keeping a wary eye on the infant, which she hurries to remove from potentially harmful social situations. Although chimpanzee mothers are extremely lenient, occasionally a mother seizes her infant and drags it away, for example, if it continues to ignore her obvious signals that it is time for them to move on to a new location (p. 368). Maternal tolerance decreases during an infant's fourth and fifth years as it is weaned and forced to walk by itself. When juveniles throw temper tantrums, their mothers often give in by embracing them and allowing them to suckle. For example, after a 4-year-old son who was being weaned was rejected twice while attempting to climb onto his mother's back, he uttered terrified screams that galvanized her "into instant action, [she] rushed back and with a wide grin of fear gathered up her child and set off – carrying him" (p. 582).

2.1.4. Infant-directed gestures of bonobos. ID gestures of bonobo mothers are similar to those of common chimpanzees. Infant bonobos and common chimpanzees begin eating solid food at about the same age, although the two species differ in how they request solid food from their mothers (Kano 1992). The most observed pattern in common chimpanzees is for infants to put their mouths near their mothers' mouths. In bonobos, the most prevalent form of begging is for the offspring to touch their mothers' mouths. Under these circumstances, bonobo mothers may look away while shaking their heads as if annoyed, but they usually give up the food. As Kano (1992) summarizes, "a kind of food-sharing occurs frequently in which a juvenile approaches and snatches food from its mother or takes food directly from her mouth. The mother certainly does not dole out the food, but she lets her offspring pull and bite at it" (p. 167).

Bonobo mothers frequently play with their infants using slow-moving and gentle motions, often while resting in day nests. During play, mothers tickle with their fingers, playbite, and grab their infants. "While lying sprawled looking up, she will tickle the infant and hold its hands and feet; hanging high in space, the infant looks very happy and fortunate" (Kano 1992, p. 132). Interestingly, bonobo mothers in this position sometimes appear to be playing "airplane" with their infants (p. 165).

Based on observations of Kanzi and his mother, Matata, Savage-Rumbaugh (1984) and Savage-Rumbaugh et al. (1998) suggest that bonobo mothers foster the emergence of intentional communication skills in their infants by responding to their gestures for aid as they move independently from place to place. Matata monitored Kanzi's acrobatics closely when he was 4 to 11 months old and "would nearly always raise a foot or arm toward Kanzi and shove him toward the object he had been trying to reach . . . Kanzi, like human infants, began to signal his desired intent to go to a particular location and to look back and forth between his locomotor goal and his mother" (Savage-Rumbaugh 1984, p. 405). Such gestures and visual checking appeared rather suddenly when Kanzi was 10 months old, and he then began to "ask" his mother to pick him up, and to

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

help him reach a particular place. At the same age, Kanzi's half-brother, Akili, also signaled his desire for help getting from one place to another to Matata. Shortly after Kanzi began signaling his intentions, he spontaneously began to point by touching objects with an extended index finger. Although common chimpanzees may use an extended hand to refer to things, use of an extended index finger is rare (Butterworth 1997; Savage-Rumbaugh 1984).

It must be kept in mind, however, that Kanzi is a bonobo that was enculturated by humans (Savage-Rumbaugh et al. 1998), rather than mother-reared in a more natural setting, which has important implications for learning to engage in social interactions that focus attention on a third entity and, indeed, the development of triadic gestures (Tomasello & Camaioni 1997; Tomasello et al. 1993). Unlike mother-reared chimpanzees, enculturated chimpanzees imitatively learn actions upon objects in a manner similar to young children, an ability that appears to be scaffolded onto *socialized attention,* which is acquired by interacting with humans (Tomasello et al. 1993). Tomasello et al. argue that such "broadly based skills of social cognition are a prerequisite to the acquisition of language skills" (p. 1702).

By the time the wild bonobo is 6 months old, it starts to move around the periphery of its mother. If the infant attempts to go far away, however, the mother will bar its way with her hand and resume carrying it. Mothers continue to carry their offspring during travel until they are 3 to 4 years old. Similar to common chimpanzees, when it is time to move within or from trees, bonobo mothers assume a posture and wait for their infants to jump on their backs. When the infant gets close, its mother may extend her hand toward it (Kano 1992).

2.1.5. Chimpanzee and bonobo laughter. According to Goodall, laughing that somewhat resembles human laughter is heard during play sessions. Although most laughter results from physical contact such as tickling, it also occurs during chasing play. Because they play more frequently, infants laugh more than adults. "Sound spectrograph analysis shows a change from steady exhaled sound, to chuckle-like pulsed exhaled sound, to 'wheezing' laughter" (Goodall 1986, p. 130). Sonagrams have also been collected of short series of rhythmic *panting laughs* in wild bonobos, which are the only bonobo vocalizations that are clearly associated with only one context, namely play (Bermejo & Omedes 1999).

Provine (1996; 2000) notes that chimpanzee laughter has the sound and cadence of a handsaw cutting wood, and differs from that of humans in the way that sounds are typically related to the airstream. The vowel-like notes of human laughter (e.g., "ha") "are performed by chopping a single expiration, whereas chimpanzee laughter is a breathy panting vocalization that is produced during *each* brief expiration and inspiration" (Provine 1996, p. 40). Chimpanzee laughter also lacks the vowel-like notes that typify human laughter. In other words, unlike the norm for humans, chimpanzees breathe in and out as they produce a breathy, panting laughter. (In a personal communication, however, Phillip Tobias noted that the late Louis Leakey had a marvelous belly laugh that was vocalized on both the exhale and the inhale, an anecdote which shows that the classic human "ha-ha" laugh is a central theme around which variation occurs.) Provine suggests that chimpanzee-like laughter was present in the common ancestor of apes and humans. If so, it would have been an important component of mother-infant communication in early hominins.

2.2. Motherese in humans

Human infants discover how rhythm organizes their native languages between birth and 2 months of age (Karmiloff & Karmiloff-Smith 2001). In most cultures, learning to process the rhythms of speech is facilitated by the special way in which infants are addressed, known variously as motherese, *musical speech* (Trainor et al. 2000), or infant-directed (ID) speech. In ID speech, intonation contours around phrases are exaggerated, as are stress patterns within words and sentences. Many repetitions and questions with rising intonations are used. The following examples provide a feel for the exaggerated stressed syllables (in capitals) that typify motherese (see also Wheeldon 2000):

> Aren't YOU a nice BAby? Good GIRL, drinking all your MILK. Look, look, that's a giRAFFE. Isn't that a NICE giRAFFE? DOGgie, there's the DOGgie. Ooh, did you see the lovely DOGgie? (Karmiloff & Karmiloff-Smith 2001, p. 47)

Infants' preference for ID as opposed to adult-directed (AD) speech increases during the first several months of life (Cooper et al. 1997), and ID speech is used most intensively with 3- to 5-month-old infants, although it persists until around 3 years -of-age (Stern et al. 1983). Six-month-old hearing and deaf infants also show greater attention and responsiveness to ID than to AD Japanese Sign Language (Masataka 1998).

Despite several "flawed" studies to the contrary, Monnot (1999) marshals strong support for the hypothesis that ID speech that is characterized by a simplified vocabulary, more repetition, exaggerated vowels, higher overall tone, wider range of tone, and slower tempo is a universal trait among modern humans. Pitch and rhythmic structure comprise two main dimensions of singing and music, as well (Dissanayake 2000). The singing of lullabies and playsongs to infants is also universal (Trehub et al. 1993), conveys meaning that is emotional rather than linguistic, and has acoustic features that are similar to ID speech: "For both playsongs and lullabies the tempo was slower, there was relatively more energy at lower frequencies, inter-phrase pauses were lengthened, and the pitch and jitter factor were higher" (Trainor et al. 1997, p. 383). From the beginning, then, babies everywhere are predisposed to respond to certain maternal vocalizations that function as unconditioned stimuli that alert, please, soothe, and alarm the infant (Fernald 1994). The *universalist hypothesis* also specifies that ID speech contributes initially to infant emotional regulation, then to socialization, and finally to the acquisition of speech in a sequential, age-appropriate manner (Monnot 1999; Trainor et al. 2000).

Vocal, gestural, and kinesic social interactions between parents and infants serve, in part, to reinforce the latter's attention to, and eventual development of, language. Thus, parents unconsciously establish eye contact with infants and then use motherese to maintain joint attention. As parents realize infants are responding to their voices by kicking, jerking, or with coos and gurgles, they begin taking turns with the infants. Parents speak, pause for the infant response, then speak again. As Karmiloff and Karmiloff-Smith note (2001), "These 'conversations' that are initially one-sided linguistically may actually constitute an impor-

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

tant preparation for taking part in later dialogue when the toddler will be capable of using language to replace the primitive kicks and gurgles" (p. 48).

What is particularly important for this discussion is that, rather than meaning or grammar, it is the melodic and exaggerated prosodic patterns of ID speech that initially interest infants (Karmiloff & Karmiloff-Smith 2001). The melodies of mothers' speech are compelling stimuli that are effective in eliciting emotion in preverbal infants (Fernald 1994; Morton & Trehub 2001; Soken & Pick 1999) and, in addition to revealing information about mothers' feelings and motivational states, may be used instrumentally to influence infants' behaviors:

> When the mother praises the infant, she uses her voice not only to express her own positive feelings, but also to reward and encourage the child. And whether or not the mother feels anger when producing a prohibition, she uses a sound well designed to interrupt and inhibit the child's behavior. . . . *In this respect, the use of prosody in human maternal speech is similar to the use of vocal signals by some nonhuman primates.*" (Fernald 1994, p. 64, emphasis mine)

As babies mature, motherese has an important role for their development of speech. For example, English, Russian, Swedish, and Japanese mothers hyperarticulate vowels when addressing their infants (but not other adults), thus amplifying the phonetic characteristics of vowels and facilitating the phonological aspects of their infants' development (Andruski et al. 1999; Burnham et al. 2002; Kuhl et al. 1997). The fact that hyperarticulation is didactic rather than merely reflecting high emotional content is illustrated by a comparative study of pitch (fundamental frequency), affect (intonation and rhythm), and vowel hyperarticulation (vowel triangles) of mothers as they spoke to their 6-month-old infants, their pets (cats or dogs), and other adults:

> These results show that infant- and pet-directed speech are similar and distinctly different from adult-directed speech in terms of heightened pitch and affect. Interestingly, only infant-directed speech contains hyperarticulated vowels. Thus, vowel hyperarticulation does not accompany special registers simply because they differ from adult speech in pitch and affect. Rather, it seems to be a didactic device: Mothers exaggerate their vowels for their infants but not for their pets. (Burnham et al. 2002, p. 1435)

By around 10 months of age, children begin to babble in rhythms that are consistent with the prosodic structure of their language (Levitt 1993). At a fundamental level, the vocal turn-taking that develops between mothers and their babbling babies (Karmiloff & Karmiloff-Smith 2001) helps the latter grasp the "rule" that conversationalists take turns. Such "social syntax" (Snowdon 1990) may enhance infants' acquisition of other rules that are preliminary to learning the proper arrangements for elements within sentences (syntax). Infants appear to learn an important aspect of syntax, namely, the boundaries between linguistic categories such as words or phrases, through *phonological bootstrapping*, that is, by attending to the correlations between the prosodic cues of motherese (phonological features, intonation, stress, vowel length) and linguistic categories (Burnham et al. 2002; Gleitman & Warner 1982; Morgan 1986; Morgan & Demuth 1996).

By the time infants reach the single word stage (at around 17 months of age), they are becoming sensitive to the way in which different word orders convey different meanings in English (Hirsh-Pasek & Golinkoff 1996). But once infants acquire some feel for linguistic categories, how do they begin to grasp a sentence's meaning? Pinker (1987; 1994) suggests a likely mechanism is through *semantic bootstrapping* – the mapping of sounds onto mental semantic concepts such as transitive and intransitive verbs. Thus, after an infant has learned the meanings of the relevant nouns, he or she is able to infer the semantic meaning of syntactical categories from the context in which they are heard:

> Upon hearing "The boy is patting the dog," for example, the child needs to know what the words "boy" and "dog" mean before he can even start a grammatical analysis of the sentence. Then, upon seeing the accompanying action (boy touching the dog's back), the child can use this real-world situation to make the formal linguistic analysis, mapping "the boy" to the subject noun phrase, and "patting the dog" to the verb phrase containing a direct object. In other words, to get syntax under way, the child initially extracts an appropriate semantic representation for a verb by mapping the extralinguistic context onto the syntactic string and by inferring what the speaker is trying to convey. In this way, the child is able to learn that "pat" means to move your hand on something in a certain way, as he can infer from the extralinguistic context. He can also derive from the linguistic context that "pat" is a transitive verb that must take a direct object. (Karmiloff & Karmiloff-Smith 2001, p. 115)

Although Pinker's hypothesis is difficult to apply to nontransparent situations, it finds support from research that shows that most of the utterances addressed to infants in the early stages of learning a language are simple, active sentences of the type, "The boy is patting the dog" (Karmiloff & Karmiloff-Smith 2001). What is particularly important for the present discussion is that Pinker stresses the importance of nonsyntactic prosodic cues provided by motherese for semantic bootstrapping.

Recent work illustrates that motherese is also important for infants' acquisition of morphology (Kempe & Brooks 2001). In certain languages in which nouns are classified into different gender classes in ways that seem arbitrary rather than systematically form-based, diminutives are used more frequently when talking to children than adults and serve to increase the transparency of the gender markings. In these languages, learners exposed to diminutives are able to generalize gender from the diminutives' transparent suffixes to the nouns that they modify. Diminutives are important for acquisition of proper gender or case markings in Russian, Spanish, Finnish, Lithuanian, and "there is widespread agreement that the occurrence of diminutives in CDS [child-directed speech] is primarily motivated by pragmatic and semantic factors" (Kempe & Brooks 2001, pp. 251–52).

To summarize, the above studies show that motherese varies between cultures in subtle ways that are tailored to the specific difficulties inherent in learning particular languages. Additional studies on speech development in infants document the effects of prosody on syllable omission (Lewis et al. 1999) and reduction (Snow 1998); the shaping of monosyllabic utterances (Snow 2002) and words (Demuth 1996; Fee 1997); auditory memory of speech (Mandel et al. 1994); and prediction of dialogue structure (Hastie et al. 2002). As a general rule, infants' perception of prosodic cues in association with linguistic categories is important for their acquisition of knowledge about phonology, the boundaries between words or phrases in their native languages, and, eventually, syntax. Prosodic cues also prime infants' eventual acquisition of semantics and morphology. Finally, the fascinating discovery (Burnham et al. 2002) that

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

infant-directed speech contains separate elements that serve to express emotions, on the one hand, and function as didactic devices, on the other, is consistent with the view that motherese evolved incrementally from largely affective ancestral vocal communications to its present highly complex form.

2.2.1. Multimodal motherese in humans. Because communication with infants involves tactile and visual as well as auditory stimuli, interest is growing in multimodal motherese that involves gesture, facial expressions, and touching of infants in addition to vocal utterances (Dissanayake 2000; Fogel 1993). For example, studies of American and Italian mother/infant pairs suggest that ID speech is accompanied by ID bodily gestures that are relatively simple compared to gesticulations directed towards adults (Iverson et al. 1999; Shatz 1982). Compared to AD gestures, ID gestures occur less frequently, are simpler and less abstract, and function to highlight certain utterances or to attract attention to particular objects. As Italian infants' absolute numbers of both gestures and words increased between the ages of 16 and 20 months, their relative use of gestures decreased from 42% to 27%, in proportion to the sharp increase in word production (Iverson et al. 1999, p. 65). Rather than adding information to verbal communications, most ID gestures serve to reinforce the linguistic message.

ID speech is perceived visually as well as aurally by infants. Facial imitation has been reported for human neonates (Meltzoff 1988), and 3- to 4-month-old infants imitate mouth movements only when auditory and visual representations of vowels are temporally coordinated (Legerstee 1990). Four-month-old infants also prefer vowels that are presented with the visual image of the appropriate mouth shape (Kuhl & Meltzoff 1988). Similarly, 5-month-old infants prefer speech sounds that are steadily increased in amplitude when they are presented with gradually opened mouths (Mackain et al. 1983). These studies suggest that infants attend to mouth shapes that correlate with speakers' utterances (Gogate et al. 2000). Maternal speech is also tied to facial expressions at other levels (Schmidt & Cohn 2001): The muscles of facial expression participate in the mother's articulation of speech sounds (Massaro 1998) and contribute information about their meaning (Ekman 1979). Facial expressions on the part of the infant, on the other hand, provide cues about his or her attentiveness to the mother's speech. Interestingly, women appear to be more sensitive and accurate decoders of facial expressions than men (Hall 1984; McClure 2000), and infants appear to vary their facial expressions depending on the sex of the parent (Forbes et al. 2000).

Gogate et al. (2000) studied multimodal motherese involving vocal, gestural, and tactile stimuli in European, American, and Hispanic mother/infant pairs representing three developmental ages: prelexical (5–8 months), early-lexical (9–17 months), and advanced-lexical (21–30 months) infants. Mothers were asked to teach novel names for two brightly colored puppets (dubbed *chi* and *gow*) and two verbs (*pru* meaning leap, and *flo* meaning shake) to their infants by any means they would normally use. Nearly 100% of the mothers' communications were multimodal, with mothers tailoring their productions to the infants' lexical development when specifically teaching words. Mothers spoke the target words synchronously with moving the puppets and touching their infants with them ("auditory-visual-tactile synchrony") in decreasing frequencies from earlier to later developmental stages. This suggests that mothers' trimodal coordination "highlights word-referent relations for infants on the threshold of lexical development" (Gogate et al. 2000, p. 890). Mothers of advanced-lexical infants, on the other hand, were more likely to name objects and actions when the object remained static or was held by the infant. Further, "the decrease in maternal use of temporal synchrony . . . appears to be well-timed with infants' at 14 months increased ability to detect word-referent relations without temporal synchrony on the basis of object motion alone. . . . In addition, mothers' naming of objects or actions with static objects seems well adapted to older infants' ability to glean word-referent relations on their own" (p. 891).

Messinger and Fogel's fascinating 1998 study demonstrates how vocalizations combined with certain gestures become increasingly intentional or instrumental rather than emotionally induced as infants mature, and supports the opinion that intentional gestures were important during language evolution (Corballis 2002; Rizzolatti & Arbib 1998). In Messinger and Fogel's study, smiling, gazing at mothers, and manual gestures (with and without accompanying vocalizations) were analyzed in 11 infants between 9 and 15 months of age as they played with their mothers, several times a month. Gestures were coded as *requests* when either mother or infant extended an arm toward an object, pointed to it, or made a palm-up gesture in a context that indicated a desire for the partner to give the object to the requester; and scored as *offers* when either gave an object he or she was holding to the other. When vocalizations accompanied gestures, approximately 96% of them did not involve recognizable words, that is, they were nonverbal. Interestingly, the proportion of infant requests involving vocalizations rose with age, showing "that as infants approach 15 months of age, they use the behavioral precursors of speech instrumentally to communicate their desire for objects," and these "infant vocalizations increased the instrumental tone of infant gestures, particularly because the vocalizations were not related to either gazing at mother or infant smiling" (Messinger & Fogel 1998, p. 587). Infant offers, on the other hand, did not rise significantly with age, but were more likely to involve smiling and gazing at the mother. Thus, "in offering objects to mother, infants appeared to share and create positive social contact" (p. 586). It appears that infants increasingly use vocalizations with requests to compensate for the fact that they are more ambiguous than manual offers (p. 584), and that "in so doing, they may be combining linguistic topics (the object referred to) with comments (the request gesture) in a manner that presages more complex language use" (Rome-Flanders & Cronk 1995). This important study suggests that development of intentional manual gestures in infants is accompanied by increased use of vocalizations that precede the production of actual words.

The gestures studied by Messinger and Fogel (1998) were triadic rather than dyadic (see sect. 2.1.3). The request and offer gestures were also imperative ("Take this!" or "Give me that!") rather than declarative (informing another about an outside entity), and carried out at a distance from the partner (distal). As such, these gestures were representative of the earliest intentional gestures of developing human infants, which are preparatory to acquisition of referential gesturing:

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

> Developing human infants' earliest gestures are triadic and distal, and they produce gestures for declarative purposes soon thereafter. Soon after that, they produce a totally novel kind of gesture, the referential gesture, which is clearly learned through imitation and understood bidirectionally and conventionally from the beginning. (Tomasello & Camaioni 1997, p. 19)

Although Tomasello and Camaioni emphasize the primacy of the visual-gestural modality for language evolution, Messinger and Fogel's research suggests that vocalization was the crucial factor that facilitated evolution of the abstract, instrumental aspects of speech. In any event, the discovery that mother-infant multimodal (vocal plus gestural) communication contains separate elements that serve to enhance social contact, on the one hand, and to allow infants to communicate their desires instrumentally, on the other, is concordant with the view that multimodal motherese evolved incrementally from largely affective, multimodal ancestral communications to its present, more complex, form.

2.2.2. Mother-infant laughter in humans. Laughter is predominantly an involuntary behavior that usually occurs in social situations, is associated with high intensity affect, and lasts less than two seconds (Nwokah et al. 1999). Provine (1996, p. 41) underscores the social and emotional aspects of laughter: "Mutual playfulness, in-group feeling and positive emotional tone – not comedy – mark the social settings of most naturally occurring laughter." In adults, most laughter seems to punctuate speech, for example, by occurring after a spoken phrase. For this reason, speech has been interpreted as having priority over laughter for accessing the vocalization channel (Provine 1993).

Bachorowski et al. (2001) propose that laughter influences listeners through acoustic properties that affect attention, arousal, and emotional responses. A listener's attention is tweaked by laughter because of learned positive emotional responses that have been conditioned as a result of repeated pairings of laughter with positive affect. Although Bachorowski et al.'s (2001) research is on young adults, their hypothesis is attractive in light of the fact that infants usually begin to laugh between the ages of 14 to 16 weeks, often during positive interactions with their mothers, and "laughter, smiles and other gestures by the baby reinforce the mother's behavior (tickling, for example) and regulate the duration and intensity of the interaction" (Provine 1996, p. 39). Interestingly, women produce significantly more song-like bouts of laughter than men, who produce significantly more grunt-like laughs (Bachorowski et al. 2001).

But what about laughter that is directed toward infants? Interactions between 13 American mothers and their infants were scored for maternal laughter from videotapes that were taken periodically as infants grew from 4 weeks to 2 years of age (Nwokah et al. 1999). Particular attention was given to co-occurrences of speech with laughter (speech-laughs) in mothers, which were coded for vowel elongation, syllabic pulsation, breathiness, and pitch change. Compared to the near absence of speech-laughs in AD laughter (Provine 1993), speech co-occurred in approximately 19% of the total number of ID laughs that were analyzed, with the figure for individual mothers ranging from 5% to 50%. In most speech-laughs, speech and laugh began simultaneously and incorporated prosodic, affective, repetitive rhythmic features that typify vocal motherese.

Production of speech sounds entails alterations in breathing and manipulation of the respiratory apparatus, which means that important changes in both the vocal tract and respiration were required before hominins could begin speaking (Provine 2000). Because apes and humans both engage in laughter that is constrained by breathing, comparative studies of this behavior provide clues about the nature of those changes. In addition to information about the anatomical and physiological evolution of respiration in hominins, studies of laughter also illustrate give-and-take turn-taking ("social syntax"; see Snowdon 1990) between mothers and very young infants. The nearly identical mother-infant tickling/laughter bouts of chimpanzees, bonobos, and humans provide some of the best evidence for the continuity hypothesis with respect to the evolution of mother-infant communication. Despite the similarities in these bouts, however, the breathing and vocalizations that they entail differ fundamentally between apes and humans, and walking upright appears to have been *the* critical event in the respiratory/vocal transition that accompanied not only the evolution of laughter, but also of speech (Provine 2000).

3. Prelinguistic evolution in early hominins

3.1. The role of bipedalism and loss of infant clinging

Two features related to development in chimpanzees and humans differ in profound ways that are important for formulating hypotheses regarding the prelinguistic substrates of language. *Difference 1:* Although infants of both taxa exhibit remarkable similarities in the sequence and timing of various developmental phenomena (e.g., helplessness at birth, distress at separation from mother, disappearance of blind rooting responses, production of social faces, and fear of strangers [Plooij 1984]), landmarks related to control of posture and locomotion (pushing off, sitting and standing without support, creeping on all fours, and walking bipedally (Plooij 1984) appear much later in humans than in chimpanzees. *Difference 2:* Unlike chimpanzee mothers, human mothers continually produce affectively positive vocalizations to their infants. Below, it is reasoned that this first difference between humans and chimpanzees is associated with the evolution of bipedalism and the subsequent trend for brain size increase in late australopithecines/early *Homo* (Falk et al. 2000), and that the second derived from an initial evolution of prosodic and instructional vocalizations in early hominin mothers. Further, it is hypothesized that these differences are related, that is, that the prelinguistic substrates for protolanguage began to evolve from ID vocalizations similar to those of chimpanzees as brain size started to increase in bipedal hominins.

But how? To explore this question, one must address the definitive trait that makes a hominin a hominin, namely, bipedalism. Many candidates (summarized in Falk 2000) have been proposed as the main advantage (or selective pressure) that led to bipedalism including: freeing of the hands to carry things (food, water, babies) or to make tools; increased ability to see predators and game over tall grass or to reach higher to pick food from trees; better stamina in running after game and hunting; and enhancement of sexual signals (genital displays). An important advantage of bipedalism is that upright hominins were more efficient at

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

keeping cool because they had reduced areas of skin exposed to the intense solar radiation (Wheeler 1988) that would have presented a thermal liability for later australopithecines/early *Homo,* which dovetails with the *radiator hypothesis* of brain evolution (Falk 1990; 1992a; 1992b). Although consensus is lacking about the causes of bipedalism (and how long it took to become fully achieved), one thing is certain: Fossil evidence shows that by the time hominins left Africa to begin colonizing the rest of the world (around two million years ago), they did so using fully developed bipedal gaits.

The fossil record also reveals that anatomical changes that broadened and shortened the pelvis and reshaped the birth canal began occurring well before this exodus. These changes, together with the subsequent trend for increasingly large brains that began in late australopithecines/early *Homo* (Falk 1998; Falk et al. 2000), would have made parturition progressively more difficult. The evolutionary solution to this dilemma is that, today, women give birth sooner, that is, before infants' heads are too big to pass through the birth canal, which results in neonates that are relatively undeveloped. This is why human babies reach landmarks related to posture and locomotion later than ape infants (Difference 1), and it is why they are unable to ride clinging to their mothers' bodies. The trend for increasingly difficult parturition was well underway in *Homo* by 1.6 million years ago, as indicated by the comparatively modern body proportions, narrow pelvis, and approximately 900 cm^3 cranial capacity of the famous Nariokotome skeleton from Kenya (WT 15000), which suggests that this youth's female relatives would have been subject to difficult deliveries of relatively undeveloped neonates (Walker & Leakey 1993).

Unlike the infants of many prosimian species that are frequently parked in nests or trees, unweaned infants of monkeys and apes are rarely parked for any length of time but, instead, ride clinging to the fur on their mothers' chests or backs (Ross 2001). In the infrequent reports of infant-parking in lieu of riding in higher primates (e.g., occasional instances in pig-tailed langurs, Mentawai Island langurs, Hanuman langurs, patas monkeys, and talapoins), mothers either place their infants on the ground or leave them alone in tree crowns before moving away (Fuentes & Tenaza 1995). Apparently, these unusual instances of baby parking in anthropoids occur where there are few natural predators and free the mother "from the potential energetic cost of carrying the infant" (Fuentes & Tenaza 1995, p. 173). It is important to emphasize, however, that infant parking is extremely rare in anthropoids; riding in which the infant does the clinging is the norm. For this reason, riding was presumably present in the ancestor of all anthropoids and, although energetically costly to the mother, may have been strongly selected for because it prevented exposure of parked infants to parasites (in nests), predation, and infanticide (Ross 2001). Observations of parking and riding across the primate order suggest that once riding had evolved it was "difficult to lose . . . [and] the only lineage in which riding has been lost . . . is that leading to *Homo sapiens*" (Ross 2001, p. 765).

The occasional reports of anthropoid mothers parking or putting down their young infants are almost always in the context of maternal foraging, which is significant because foraging was a primary means by which early hominins made their living. Since chimpanzee mothers and contemporary women in hunting and gathering societies (who use baby slings) usually forage for food with their infants attached to their bodies, one might assume that early hominin mothers did too. In this context, it is relevant to consider the interaction of maternal foraging and infant-riding in a higher primate species that, like humans (Leutenegger 1972), produces relatively large infants. Mother squirrel monkeys (*Saimiri sciureus*), for example, normally carry infants that are less than 17 weeks old on their shoulders and backs, after which time the infants, having grown to between one-third and one-half the mother's size, move about on their own (Lyons et al. 1998).

Experimental evidence reveals that squirrel monkey mothers stop carrying their infants at earlier ages and spend more time foraging when food is relatively scarce and difficult to find, although they do not decrease the amount of time they nurse (Lyons et al. 1998). For their part, infants living under harsh foraging circumstances make frequent unsuccessful efforts to ride on their mothers compared to infants living under more optimal conditions. Under difficult conditions, mother squirrel monkeys focus their energy on obtaining enough calories to feed themselves and to nurse their infants. Thus, "by rescheduling some transitions in development (carry → self-transport), and not others (nursing → self-feeding), mothers may have partially protected infants from the immediate impact of an otherwise stressful foraging task" (Lyons et al. 1998, p. 290). Similar postnatal foraging-related changes in maternal care have been reported for free-ranging gelada baboons (Barrett et al. 1995), long-tailed macaques (Karssemeijer et al. 1990), and yellow baboons (Altmann 1980).

Although it is the mothers that bear the burden of their infant's weight during infant carrying, it is the infants that usually do the hanging-on in anthropoids, with the exception of humans. Thus, because chimpanzee infants develop motor skills relatively rapidly compared to human babies (again, Difference 1), they are able to cling to the mother's furry belly after 2 months of age (Plooij 1984) and to shift to her back for travel as they grow heavier. During the first weeks of life, however, it is the mothers themselves that support and cling to infants, frequently in response to their distress *whimpers* or *hoos.* Human babies, on the other hand, are born extremely helpless and never develop the ability to cling unaided to their mothers' (unfurry) bellies or backs. This observation is corroborated by literature that documents a strong grasping reflex in human neonates (Halverson 1937a; 1937b; 1937c). For example, the ability of a young infant to support its weight by clinging with one hand decreases from monkeys to chimpanzees, and is apparently extremely limited in human infants despite the fact that they are born with strong vestigial grasping responses (Halverson 1937a). However, even if human babies had the ability to cling to their mothers' bellies, it would be difficult for mature human infants to ride unaided for extended lengths of time on backs that are habitually oriented vertically rather than horizontally. Infant carrying is therefore entirely up to the human mother (or substitute) and, as any mother will attest, growing babies soon become heavy.

Although contemporary hunters and gatherers do not provide exact models for our hominin ancestors, groups such as the Ache, !Kung san, and Efe pygmies offer clues that may help us to formulate hypotheses about the lives of Plio-Pleistocene hominins, including how mothers may have cared for infants (Small 1998). As a general rule, care-

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

taking of infants in most non-Western cultures is physically engaging, with demand feeding, close contact with infants during the day, and sleeping with them at night being the norm. In order to go about their business with freed hands, contemporary women from most of the world's cultures use slings to secure their babies onto their backs or hips, or onto the bodies of older siblings (Small 1998). These habits may seem strange to Westerners that value and nurture independence in very young infants, and thus may permit them to cry for extended periods or to sleep in separate rooms. The cross-cultural ethnographic evidence pertaining to baby slings reinforces the suggestion by Zihlman (1981) and others that baby slings, perhaps made from vegetal matter, may have been among the first nonlithic tools that were invented.

In what contexts would infant riding have suffered its setback in hominins (Ross 2001), and what would have replaced it before the invention and general use of baby slings? Did evolving hominin mothers revert to the prosimian adaptation of parking their babies far away for extended periods of time while they foraged, despite the threats from parasites, predators, and (possibly) infanticidal males? Probably not. For one thing, parking infants would have severely constrained travel distances for lactating mothers, since comparative primatological and ethnographic data suggest that infants would have required frequent nursing bouts throughout the day (Plooij 1984; Small 1998). Instead, as documented above for a number of anthropoids, early hominin mothers may have engaged in foraging-related changes in maternal care. Unlike chimpanzee mothers, by the time early hominins had evolved into habitual bipeds that bore relatively helpless young, it would have been adaptive for them to adopt a "putting the baby down" strategy in which mothers periodically put their infants down to release their hands (and energy) for foraging nearby. That way they could keep their babies within eyesight and, when ready to move on, simply pick them up and go.

3.1.1. Using vocalizations to "keep in touch." Infant parking is a rare event in monkeys, apes, and non-Western human cultures. When it does occur, infants are usually distressed by the unusual situation of being separated from their mothers (Ainsworth et al. 1978; Lamb et al. 1985), which is frequently conveyed by whimpering or crying. Parked infant pig-tailed langurs, for example, "cry" by emitting high-pitched squeals intermingled with low-pitched guttural sounds (Fuentes & Tenaza 1995), while infant rhesus monkeys produce a plaintive series of *coos* when separated from their mothers (Small 1998). Infant chimpanzees whimper and scream loudly if they begin to fall from their mothers' chests while traveling (Plooij 1984). Crying is qualitatively different in human babies, consisting of rhythmic patterns of vocalizations that entail short, breathy expirations alternating with long intakes of air (Frodi 1985). Human crying makes use of the lungs and vocal apparatus much as laughter does; and Provine (2000) notes that "although laughter and cying are considered polar opposites of the emotional spectrum, they are neurologically linked and share the features of tearing and rhythmic vocalization" (p. 187). By around 3 months of age, human infants develop the ability to modulate their cries to express different emotions such as anger, pain, and frustration (Marler et al. 1992; Small 1998); and, like babbling, crying may be a precursor to language (Small 1998).

Although crying is universal in human infants, the degree to which it is manifested varies with culture. In cultures where babies spend most of their hours in close physical contact with adult caregivers, infants engage in relatively little crying; whereas in cultures that encourage infants to gain independence by leaving them alone for much of the time (e.g., much of America) babies cry considerably more (Small 1998). Small believes that crying of infants today is little changed from when it first evolved in hominins as a means for communicating infants' needs. Furthermore, crying and parental sensitivity to it are adaptive traits because they:

> evolved to serve the infant's purposes: to assure protection, adequate feeding, and nurturing for an organism that cannot care for itself. By definition, crying is designed to elicit a response, to activate emotions, to play on the empathy of another. . . . The caretaker has also evolved the sensory mechanism to recognize that infant cries are a signal of unhappiness, and thus be motivated to do something about it. (Small 1998, p. 156)

It is noteworthy that crying increases the strength of the grasping reflex in human infants (Halverson 1937a), which is consistent with experimental research on American infants which suggests that the major reason that infants cry is to reestablish physical contact with separated caregivers (Small 1998; Wolff 1969).

Presumably, early hominin babies were no happier at being separated from their mothers than are anthropoid infants today, and would have been increasingly likely to vocalize distress during the period of evolution when active infant riding was lost and babies were put down periodically so that mothers could forage. It is also reasonable to assume that the crying of their infants would have produced aversive stimuli for early hominin mothers, as it does for contemporary monkey (Small 1998), chimpanzee (Plooij 1984), and human (Small 1998) mothers.

What could hominin mothers have done to discourage separated babies from crying? For one thing, they may have used a strategy commonly employed by contemporary Western women, that is, inducing infants to fall asleep before putting them down. One way to do this would have been to nurse infants because, if they resembled modern babies, "an infant who is fully fed or fatigued is likely to be quiet, if not actually sleepy" (Halverson 1937a, p. 381). Early hominin mothers may also have used other tactile strategies to soothe babies before putting them down, for example, cradling, and rocking – the latter being a coevolved "rhythmic, temporally patterned, jointly maintained" interaction between mothers and infants (Dissanayake 2000, p. 390). (Perhaps the human habit of rocking babies to sleep is effective because it produces a gentle barrage of stimuli that mimics physical contact with the mother.) The very act of placing babies in horizontal positions may also have encouraged them to sleep, as suggested by experiments which show that captive chimpanzee infants that are left horizontally in cradles most of the day sleep more than wild infants that are carried semi-upright by mothers (Plooij 1984). In addition to these tactile strategies, hominin mothers may also have used rhythmic, temporally patterned vocalizations to lull infants to sleep: precursors of the first lullabies (Dissanayake 2000).

What about instances in which hominin infants refused to sleep and, instead, fussed and cried when mothers put them down? Perhaps early hominin mothers then responded "voice to voice." Already accustomed to regulating

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

older infants' travel with vocalizations as chimpanzee mothers do today, early hominin mothers may have elaborated calls from their vocal repertoires into affectively positive, rhythmic melodies as a means, not only to lull them to sleep, but to reassure them that "mommy is near" when they were awake (a kind of vocal rocking[2], or non-tactile way of "keeping in touch"). In a sense, then, prosodic utterances would have become disembodied extensions of mothers' cradling arms. This suggestion is consistent with the fact that singing to human infants to provide comfort and ease unhappiness is a derived practice that appears to be cross-culturally universal (Trainor et al. 1997). It is also consistent with the finding that a "squealing baby, in fact, can be stopped dead in its vocal tracks by a sudden stream of baby-talk" (Small 1998, pp. 145–46).

The argument that mother-infant communication shifted away from being based almost exclusively on direct physical contact between the signaler and recipient (as baby clings to mother) to being distal (when baby is regularly put down) also applies to gestural communication. For example, while most chimpanzee gestures involve physical contact between the signaler and the recipient, the earliest gestures of developing humans do not, that is, like vocal communications, they have become distal (Tomasello & Camaioni 1997). Facial expressions are believed to have been important during the evolution of speech (Schmidt & Cohn 2001), and would have enhanced communication between hominin mothers and their nearby babies. Putting infants down may also have had a significant impact on the development of certain circular and imitative self-teaching devices (Baldwin 1906; Piaget 1952) that are hypothesized to have been uniquely associated with the evolution of symbolic communication in higher primates, especially humans (Gibson 1986; 1990; 2001; Parker 1993; 1996). For example, a secondary circular reaction (Piaget's 3rd stage) occurs in babies that are 3 to 5 months when they persistently focus on the contingent behavior between their hands and inanimate objects (Parker 1993) and "the midline supine posture . . . focuses the infant's eyes on both hands" (Parker 1993, p. 318). The fact that the "putting the baby down" hypothesis entails continuity in the evolution of prelinguistic vocalizations of early hominins from the vocalizations of ape ancestors does not mean that gestural communication is not, or was not, an important complement to speech-based communication (Armstrong et al. 1994; Corballis 1999; 2002; Hewes 1973; King 1996; Rizzolatti & Arbib 1998; Tomasello 1999; Tomasello & Camaioni 1997).

3.2. The broader evolutionary context

3.2.1. The emergence of protolanguage from prelinguistic behaviors. Just as ID speech of women first expresses emotions and engenders them in infants, and later becomes instrumental in socializing and influencing their behaviors (Fernald 1994; Monnot 1999), the prosodic ID vocalizations of hominin mothers would have taken on less emotional and more pragmatic aspects as their infants matured. As is true for human babies toward the end of their first year, prosodic (and gestural) markings by mothers would have helped early hominin infants to identify the meanings of certain utterances within their vocal streams (semantic bootstrapping, Pinker 1987; 1994). Over time, words would have emerged in hominins from the prelinguistic melody (Fernald 1994, p. 65) and become conventionalized. The prosodic elements of prelinguistic vocalizations would have contributed not only to hominins' eventual semantic grasp of utterances, but also to their acquisition and shaping of numerous sensitivities (phonology, boundaries between utterances, monosyllabic utterances, syntax, dialogue structure, and auditory memory for vocal utterances) that, ultimately, became entailed in linguistic evolution.

That said, speculation abounds about the precise nature of protolanguage. For example, it has been suggested that the earliest language might have had nouns and verbs, but lacked affixes, functional categories (Heine & Kuteva 2002), and true syntax (Newmeyer 2002). Whatever the exact configuration of protolanguage, however, certain conjectures about its emergence are relevant for the discussion of prelinguistic evolution. Thus, protolanguage is thought to have been relatively simple grammatically (Heine & Kuteva 2002), essentially pragmatic in nature (Givon 1979), and may have developed in early *Homo* "directly from the requirements of group foraging . . . and instruction of the young" (Bickerton 2002, p. 209). Although foraging is emphasized here as the context in which prelinguistic behaviors were initially selected, it is worth noting that the mother-infant dyad is fundamentally social and that, consistent with Dunbar's (1993) emphasis of selection of language for "vocal grooming":

> As soon as protolanguage had achieved the necessary critical mass (some dozens or perhaps a few hundred meaningful symbols, whether oral or manual is immaterial to the present argument) it was undoubtedly co-opted for a variety of social purposes, which in turn contributed to its further expansion. (Bickerton 2002, p. 209)

Thus, instead of remaining static over time (*uniformitarianism*), once protolanguage appeared, it presumably continued to evolve in a socially meaningful, dynamic, changing, and directional manner (Newmeyer 2002).

The "putting the baby down" hypothesis is based on two fundamental premises. First, hominin mothers that attended vigilantly to their infants would have been strongly selected for; and, second, those mothers would have had a genetically based potential for modifying their vocal and gestural repertoires to shape and consciously control the behaviors of their offspring. The first premise is widely acknowledged to be the case for a variety of primates (and, indeed, other mammals), including monkeys (Small 1998), chimpanzees (Goodall 1986; Plooij 1984), and people (Small 1998). Not all primate mothers are equally attentive to their infants, however, and a "natural experiment" on a mother-infant chimpanzee pair at Gombe supports the suggestion that selection may have intensely favored early hominin mothers who developed a strategy for monitoring infants that lost the ability to cling to their bodies during travel, as well as infants that vocalized their distress upon becoming separated:

> Madam Bee had raised two infants successfully when one of her arms was paralyzed during a presumed polio-epidemic. . . . The two infants that were born afterwards died within a few months. I had the occasion to make observations on the first of these two infants: Bee-hind. Her body was full of wounds and scratches, so she must have fallen repeatedly. Whenever her mother moved about without supporting her, she whimpered and screamed continuously. (Plooij 1984, pp. 45–46)

Just as there is a good deal of variation in the degree to which healthy chimpanzee mothers living in the wild support and carry their infants (Plooij 1984), variation in the at-

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

tention provided to infants by hominin mothers would have provided the raw material upon which natural selection operated. As detailed above, humanlike crying and mothers' sensitivity to it probably evolved in early hominins to assure protection, adequate feeding, and nurturing for babies that could not care for themselves. If the hypothesis presented here is correct, hominin babies were increasingly put down, in which case maternal visual attention to gesture and facial expression would also have acquired high selective valiance. As noted by Schmidt and Cohn (2001), the fitness effects of maternal attention to facial expression of infants "are potentially great, considering the intense social and nutritional needs of the infant, as well as possible risks associated with lack of maternal attention, including failure to thrive, physical danger, and at the extreme, death from neglect or abandonment" (p. 12).

The second premise that early hominin mothers would have had a genetically based potential for modifying vocalizations and gestures consciously to control infants is consistent with recent studies that suggest that pitch discrimination is highly heritable (Drayna et al. 2001), that the volumes of gray matter in Broca's and Wernicke's language areas of the brain are highly heritable (Thompson et al. 2001), and that the orofacial motor sequencing upon which speech depends is under strong genetic control (Lai et al. 2001). Thus, in humans, a point mutation in one gene (FOXP2 on chromosome 7) severely disrupts the ability to select and sequence fine movements of the mouth and tongue (a praxic problem) that are necessary for articulate speech (Lai et al. 2001). Affected individuals tend to garble pronunciation, put words in the wrong order, and have trouble comprehending grammar and speech sounds, including sentences. Although the exact function of FOXP2 is unknown (it may help to regulate embryonic development), this gene appears to be necessary for the development of normal spoken language (Lai et al. 2001), and may have been a target of selection during recent human evolution (Enard et al. 2002).

Fascinating research on language acquisition in hearing and deaf subjects strongly suggests that, rather than being "hard-wired" to process only vocal language, humans are genetically predisposed to detect aspects of the temporal and distributional regularities which correspond to prosodic and syllabic levels of signed or spoken languages (Petitto 2000). Thus, while certain aspects of abstract grammatical patterning of natural languages may, indeed, be hard-wired in our species (Donald 1993; Pinker & Bloom 1990), Petitto offers a persuasive argument that language acquisition is nevertheless neurologically plastic and biologically flexible because it can be acquired and expressed easily via the hands or tongue. (This is not meant to deny the primacy of vocal over sign languages. All normal people acquire speech; relatively few learn sign languages.) The dominant mode in which natural language is expressed is determined largely by infants' biological circumstances (e.g., hearing, deaf) (Petitto 2000), while the particular flavor of language that they learn (e.g., Chinese, English) is clearly a product of their cultures.

Just as certain referential calls of vervet monkeys (Cheney & Seyfarth 1990) and over 30 discrete calls of chimpanzees from Gombe (Goodall 1986) are produced and interpreted similarly by members of their respective social groups, protolinguistic utterances of early hominins would have become conventionalized across their groups. But how could the cultural propagation of specific utterances that resulted from a genetically driven propensity to produce natural protolanguage have happened? Although a review of the extensive literature on social transmission in nonhuman primates is beyond the scope of this paper, it is interesting to consider how protocultural innovations that arose in foraging contexts were socially transmitted, primarily by mothers and youngsters, in at least one species. As is well documented for the innovations of sweet potato-washing and wheat-washing that "were invented" by a female Japanese macaque named Imo (Kawai 1965), the process of propagation of new behaviors may have gone through two stages: In the initial "Period of Individual Propagation" (Kawai 1965, p. 5), novel behaviors are transmitted between youngsters, and from them to older females and siblings. After the behaviors became fixed (adult males being the last to acquire them), a second "Period of Pre-cultural Propagation" (Kawai 1965, p. 8) ensues in which infants learn the behaviors from their mothers and the practices are thus passed to future generations.[3] If one applies this model to early hominins, once bipedal mothers began using vocalizations to reassure and instruct their infants, processes similar to those documented for Japanese macaques could have facilitated the use, sharing, and understanding of utterances between youngsters and from youngsters to their mothers. As youngsters matured into adults and these utterances became fixed across all members of groups (conventionalized), new generations of infants would begin acquiring the vocalizations from their mothers. This is one example of how individually developed "words" could have come to be shared. It is also worth mentioning that the calls of different groups of chimpanzees are now thought to have different cultural dialects (Gibbons 1992; Mitani & Brandt 1994; Mitani et al. 1992), which is consistent with the possibility that multiple dialects of protolanguage may have eventually arisen.

3.2.2. What's in a name? Although the exact nature of protolanguage is (I believe) unknowable, one may at least speculate about the referents for the first protolinguistic words (or, rather, their English equivalents). Many workers assume that naming was the basic protolinguistic vocal behavior (Harnad 1996a; Horne & Lowe 1996); that a study of the origin of names is a study of the origin of symbolic categories (Harnad 1996b); and that naming was eventually transformed into language by "enhancing the ability of hominids to comment on and think about the relationships between things and events, that is, by enabling them to articulate and communicate complex thoughts" (Armstrong et al. 1994, p. 354). But what concrete categories would the very first names have referred to? Possible answers include "kinfolk, tribesmen, enemies, foods, predators, weather conditions, tools, places, discomforts, [and] dangers" (Harnad 1996b). With respect to the kinfolk category, recent research on the English word "Mama" (Goldman 2001; MacNeilage 2000; Tincoff & Jusczyk 1999) is particularly relevant for the "putting the baby down" hypothesis. According to MacNeilage (1998; 2000), "Mama" is an example of two successive cycles of a *pure frame* (i.e., utterances generated by mandibular oscillation alone, with the tongue held still), each of which begins with a consonant and ends with a vowel, which MacNeilage believes probably typified earliest speech. A study of 75 infants of less than 6 months of age revealed that babies be-

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press

gan producing "Mama" at a modal age of 2 months, usually as part of a cry (Goldman 2001). The results showed that some infants uttering "Mama" appeared satisfied if a favorite caretaker approached and paid attention to them, whereas others also needed to be picked up. Another study revealed that, by the time infants are 6 months of age, they understand that the word "Mama" specifically refers to "my Mom" (rather than to any woman), which suggests that they have begun to form a lexicon with sounds that are linked directly to socially significant people (Tincoff & Jusczyk 1999). Thus, it does not seem unreasonable to suggest that the equivalent of the English word "Mama" may well have been one of the first conventional words developed by early hominins. After all, wouldn't maturing prelinguistic infants, then as now, be inclined to put a name to the face that provided their initial experiences of warmth, love, and reassuring melody?

4. Concluding thoughts

Motherese has provided a rich source of information for this discussion, which is appropriate since it is the *only* available model for elucidating how humans universally acquire spoken languages today, and therefore may have acquired them in the past. The behaviors of primate (including human) mothers, of course, are pivotal for perpetuating their genes (and their offsprings') into future generations. The central thesis regarding motherese is that bipedal mothers had to put their babies down next to them periodically in order to go about their business, and that prosodic vocalizations would have replaced cradling arms as a means for keeping the little ones content. It is not a stretch to suggest that such vocalizing (and the elaboration of distal gestures) would have had strong selective value. It is reasonable to speculate that by the time individuals across social groups began to originate and conventionally share simple instructive utterances, protolanguage was in the process of emerging from the prelinguistic melody. Whatever its precise nature, however, protolanguage and the other languages that eventually evolved would, forever after, retain some of that melody. Thus, rather than being totally separate from language (Burling 1993), tone of voice represents a signature from its very origin that, as transpired for the cosmic microwave background signature left over from the Big Bang, should be recognized and investigated.

It is hoped that readers will consider the ideas developed in this article as possible alternatives to suggestions that language could not have emerged from an earlier primate communication system (Burling 1993; Hurford 2002); that it was evolved primarily for internal thought and only applied secondarily to communication with conspecifics (Burling 1993); and that the Upper Paleolithic record of artwork indicates it evolved only recently (Davidson & Noble 1989). That said, the precise role of gesture during prelinguistic evolution and the exact nature of the first language are likely to remain academic bones of contention until we get the time machine. In the final analysis, however, at least the suggestion that true syntactic language probably did not evolve until after the emergence of the genus *Homo* around 2 million years ago (Corballis 2002; Rizzolatti & Arbib 1998) rings true to many, if not most, workers.

ACKNOWLEDGMENTS

I thank Robin Dunbar, Michael Corballis, Bill Parkinson, and numerous referees for helpful comments. Ann Powell is gratefully acknowledged for library research.

NOTES

1. Although human fathers engage in motherese (or parentese), chimpanzee fathers are unrecognized in the wild and adult males interact relatively little with infants. The parental focus of the present analysis is therefore on females.

2. At least one familiar western lullaby reifies the concept of lullabies as substitutes for physical rocking. Interestingly, it also refers to infants falling from trees (or off traveling mothers?). One might therefore suggest (somewhat fancifully) that this lullaby soothes a primordial fear retained from the time when hominin mothers still slept with infants in tree nests (cradles of boughs), as chimpanzee mothers do. Primary stresses are capitalized and underlined; secondary stresses are underlined (modified slightly from Trainor et al. 1997, p. 388):

ROCK-a- // bye / *ba*- // by / *ON* the // tree / *top* ///
WHEN the // wind / *blows* // the / *CRA*- / dle / will / *rock* ///
WHEN the // bough / *breaks* // the *CRA*- / dle /will / *fall* ///
And /// *DOWN* / will / come / *Ba*- // by / *CRA*- / dle / and / all ///

3. Another invention of female Japanese macaques that applies to a natural (rather than provisioned) food was not propagated in this way. Nakamichi et al. (1998) report that 11 free-ranging adult females pulled grass roots from the ground, carried them to a river, and washed them (sometimes on flat stones), but that this behavior did not propagate to most of the group. Six of the animals were from one matriline, and two others were a mother and her adult daughter. The authors speculate that root-washing did not spread widely for several reasons: Roots are eaten only during a brief period of the year; carrying is not common among macaques in natural environments; and pulling long roots from the ground would have been difficult for juveniles. One might therefore conclude that, to become conventionalized, an invented behavior must be possible during much of the year, feasible for juveniles as well as adults, and utilize anatomical substrates that are widely available.

Open Peer Commentary

Prelinguistic evolution and motherese: A hypothesis on the neural substrates

Francisco Aboitiz[a] and Carolina G. Schröter[b]

[a]*Departamento de Psiquiatría y Centro de Investigaciones Médicas, Escuela de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile;* [b]*Facultad de Medicina, Universidad de Chile, Santiago, Chile.*

faboitiz@puc.cl

Abstract: In early hominins, there possibly was high selective pressure for the development of reciprocal mother and child vocalizations such as proposed by Falk. In this context, temporoparietal-prefrontal networks that participate in tasks such as working memory and imitation may have been strongly selected for. These networks may have become the precursors of the future language areas of the human brain.

Falk proposes a hypothesis on prelinguistic evolution in early humans that is based on the development of prosodic vocal communication between mothers and their babies. Vocal communication was used as a mechanism to compensate for the lack of physical

https://doi.org/10.1017/S0140525X04000111 Published online by Cambridge University Press