Cost and complexity: Selection for speech and language

John L. Locke*

Department of Speech–Language–Hearing Sciences, Lehman College, City University of New York, 250 Bedford Park Boulevard West, Bronx, NY 10468, USA

Received 18 August 2007; received in revised form 20 December 2007; accepted 21 December 2007
Available online 6 January 2008

Abstract

The handicap principle has been applied to a number of different traits in the last three decades, but it is difficult to characterize its record, or even its perceived relevance, when it comes to an important human attribute—spoken language. In some cases, assumptions regarding the energetic cost of speech, and the veracity of linguistically encoded messages, have failed to recognize critical aspects of human development, cognition, and social ecology. In other cases, the fact that speech contains honest (physiological) information, and tends to be used honestly with family and friends, has been overlooked. Speech and language are functionally related but they involve different resources. Individuals can increase the attractiveness of their speech, and of more stylized vocal and verbal performances, without enhancing linguistic structure or content; and they can modify their use of language without significant changes in the physical form of speech. That its production costs are normally low enables speech to be produced extravagantly in bids for status and mating relationships, and in evolution, may have allowed its content—linguistic knowledge and structure—to become complex.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: Speech; Language; Evolution; Handicap principle

1. Introduction

Darwin (1859) recognized that some male traits, such as the peacock’s long and colorful tail feathers, were sexually selected, but offered no ideas as to why such characteristics might benefit females. Wallace (1901), on the other hand, assumed that females were attracted to certain physical traits, and males’ ability to display them, because the traits and displays were correlated with general health or vigor (Chapter 10; also see Cronin, 1991). Taking this a step further, Fisher (1915, 1930) suggested that by choosing an attractive mate, females increase the probability of having a son that is also attractive. This would have been essential to the transmission of their own genes, he argued, since attractive male traits owe their existence to the perceptual and evaluative systems of females, and these, too, would have evolved.

Little more was said about these issues until the early 1970s when Emlen (1973, p. 51) suggested, and Zahavi (1975) more formally claimed, that the more developed the male feature or signal that is favorably appraised, the greater the “handicap” it poses for its bearer. The handicap may be due to the cost of development or production, or to the risks or hardship that individuals must endure because they have the trait. Males who are fit can meet this cost because they also possess one or more other traits that enable them to compensate for the handicap. The offspring of females who choose males with the handicapping trait thus receive genes both for the handicap and the beneficial characteristics with which it is associated (Harvey and Bradbury, 1991; Kirkpatrick, 1986).

Handicap theory is relevant to physical traits like elaborate feathers as well as evolutionarily stable signals and displays. One handicap that relates to signal production is the expenditure of energy. Signals that are energetically “cheap,” according to the theory, cannot serve as “honest” or reliable indices of fitness because selection would have applied only to signals that preferentially benefited individuals of high genetic quality. These individuals would have possessed greater ability to endure energy loss than less fit competitors. Inexpensive signals

*Tel.: + 1 718 960 8033; fax: + 1 718 960 7376.
E-mail address: John.Locke@lehman.cuny.edu

0022-5193/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.12.022
could not have been selected because individuals of inferior quality could produce them.

Since it was proposed, handicap theory has been applied to numerous traits; intensively debated and evaluated (e.g. Dawkins, 1976, 1993; Grafen, 1990; Guilford and Dawkins, 1993; Johnstone and Grafen, 1993; Kirkpatrick, 1987; Maynard Smith, 1976, 1978); and, in response, explained in greater detail (Zahavi and Zahavi, 1997; see reviews in LaPorte, 2001, 2002; Miller, 1998). Handicap theory has also been applied to an important human trait—language. Although several scholars have commented on these efforts (e.g. Fitch, 2004; Knight, 1998), a number of critical issues remain to be discussed. It is important that we examine what can or must be assumed in any cost-oriented assessment of speech and language, considered separately, doing so not in an abstract way, but in relation to the actual living and speaking practices in traditional (oral) societies. Just as it is assumed that the ecology and lifestyle of traditional (e.g. hunting and gathering) societies is closer than that of industrial societies to the “environment of evolutionary adaptedness” (Franks, 2005), so do I assume that the social environment in these societies is more similar to the “environment of linguistic evolution” (for more on hunter-gatherers, see Lee and Daly, 1999).

Members of oral cultures have never seen printed words or received academic training in language, and are usually less conscious of their linguistic knowledge than members of literate societies (Goody and Watt, 1972; Linell, 2005; Locke, in press-b). Since formal instruction can affect the way people talk, these uninstructed groups are of unusual interest.

A second issue relates to the choice of behaviors to be explained or modeled. If one’s goal is to provide an evolutionary account of the traits that collectively compose human language, it is necessary to consider precursors to, or elements of, those traits, not the fully evolved faculty of language that now exists. Thus some attention must be paid to intermediate forms that have been hypothesized, including the use of isolated words and word combinations as well as controlled or elaborate vocalization without words. It is also the case, where handicap theorists have dealt with language, that there has been confusion about the nature of this important human faculty. Are the behaviors to be addressed mostly articulatory, primarily cognitive, or some combination of these?

I begin here by examining various aspects of human communication and sociality that have a bearing on the probable costs of speech. Then I offer two proposals that relate to the transition from ape to human behaviors, and thus to early stages in the evolution of spoken language. The first proposal envisions selection for vocal complexity in association with competitions for dominance and mates, a proposal that presupposes a particularly active role by males in relation to speech. The second proposal suggests a role for complexity at the level of meaningful utterances in the context of affiliation and coalition building, which presupposes an active role by females, and lays additional groundwork for language. Finally, I discuss possible trade-offs between speech and language, and some research that could shed additional light on complexity at the level of the brain, languages, and individuals.

2. The costs of speech

It is often claimed that speech cannot be considered a handicap because it is neither costly nor honest (e.g. Bergstrom and Lachmann, 1998; Dessalles, 2007; Hasson, 1994; Jeffreys, 2006; Knight and Power, 2006; Lachmann et al., 2001; Power, 1998; Zahavi and Zahavi, 1997, p. 223). Some theorists say that speech lacks honesty because it can be used to lie. This claim, among other things to be discussed, confuses “physical honesty” with truth or falsity in the representation of ideas. Others claim that speech—the vocal and articulatory activity associated with word production—is “cheap” because it consumes little energy, and therefore can be faked by less fit individuals.

2.1. Physical cost

There has been remarkably little research on the energetic requirements of speaking. One (pilot) study does indicate that speech produced at a normal loudness level and rate consumes only slightly more oxygen than quiet rest (Moon and Lindblom, 2003).

2.2. Developmental costs

The evidently low physical cost of speaking does not mean that the demonstration of one’s ability to engage in this activity is free of all developmental costs. When people speak, they speak a language, and this requires them to do things that were made possible by a vast number of earlier achievements. These include the learning of thousands of words, sounds and sound sequences, and stereotyped phrases over many years; the application of grammatical rules that could only have been inferred in infancy from fragmentary and variable input; and the acquisition of pragmatic principles of usage in later stages of life history (Locke, 1993; Locke and Bogin, 2006). There are individual differences here. Adolescents who speak in a way that is typical of that developmental stage were probably at least minimally fit at the time the relevant mechanisms began to develop in infancy (Lenneberg, 1967; Locke, 1993) and, because development is largely continuous, may still be.

2.3. Signals

It should also be kept in mind that speech is not merely a “signal” in the sense of a brief sensory event, or isolated call, that is rigidly designed to convey information about the signaler or his awareness of environmental changes. In hunter-gatherer camps, conversations last for hours—even days—(Lee, 1979; Marshall, 1961), but most human societies are loquacious. In Midwestern American homes,
where someone initiates a new utterance very 4.5 s (Hart and Risley, 1995), there are presumably few distinct signals that rise above this mostly continuous hum. Engineering models of communication have seized upon a particular social configuration—the dyad—in which “the speaker” conveys a message to “the listener,” who responds in kind (Weiner, 1988). The interactions featured in these models actually take place, of course, but in real life one frequently encounters gatherings of three or more socially related individuals who jointly construct topics and negotiate meaning.

2.4. Information

Some theorists have assumed that the initial benefit of language was an increase in the rate or quality of information that might have been exchanged, and have evaluated this benefit by generating possible outcomes with mathematical models (Shannon and Weaver, 1949). These efforts have produced interesting results, given certain (tacit) assumptions about the role of information in human societies, and the mapping of information onto linguistic form during intermediate and final iterations of language (Nowak and Krakauer, 1999; Nowak et al., 2000, 2001, 2002). No one doubts the ability of language to transmit information, and yet it is unclear, from the abundance of joking, teasing, and “chit-chat” that occurs in many societies, how much information is actually embedded in linguistic structure. For example, in an extensive analysis of the conversations of a young American couple in the 1960s, it was found that three-fourths of all the utterances contained no facts or impersonal information; most were classified as “relation-changing messages” (Soskin and John, 1963). Many verbal interactions can be classified as “small talk” (Coupland, 2003), and some amount of even the most semantically significant conversations is devoted to back-channel responses (“uh-huh,” “I see”) and to ritualized openings and closings (Cheepen, 1988). Qualifications such as these must be taken into account in assessing the cost of communicative behaviors and the reliability of the individuals who issue them. Of course, information about the physical and social environment can be valuable—one needs to think twice before giving it to competitors (Palmer, 1991)—but informing others can serve one’s own purposes too.1 Fitch (2004) has argued that among kin, there would have been selfish reasons to share good quality information. He therefore proposed that language emerged in a familial context. This proposal is reasonable, although it appears that similar factors operate in genetically unrelated individuals who are raised together from an early age (see Shepher, 1971)—the usual circumstance in our cooperatively breeding species (Hrdy, 2006).

2.5. Deception

As indicated, some theorists have held that speech is unreliable because it can be used to lie, presumably an easy thing to do. But assumptions about the theoretical significance of lying, in which one person intentionally misleads another, frequently ignore social, psychological, and physiological facts about the human animal. For one thing, people are not as free to lie as has been supposed. Many individuals have difficulty controlling critical aspects of their vocal, facial, and bodily activity while lying or omitting crucial pieces of information. As a consequence, deception increases certain types of facial cues, such as pupil dilation and blinking (Riggo and Friedman, 1983). The same is true of several different vocal characteristics, including the fundamental frequency of the voice, as well as its variability and range, and the number of pauses and nonfluencies (Anolli and Ciceri, 1997; Scherer et al., 1985). Observers tend to notice these behaviors and sense that something is wrong (DePaulo et al., 2003; Sporer and Schwandt, 2007). It is also the case that lying, the presumed easy option with cheap signaling systems, may actually consume more energy than telling the truth. Deception, as an attempt to manage impressions, is a form of self-regulation—an activity that consumes mental resources (Baumeister, 1998; DePaulo et al., 2003; Gombos, 2006). In lying, one must weigh one’s words carefully, taking into account what is known about the listeners, including their present knowledge, beliefs, and goals. In habitual lying, one must also remember what has been said to whom in the past (Zuckerman et al., 1981). Thus, even where words spoken truthfully and deceptively are the same, and the articulatory energy equivalent, the cognitive demands of lying may be greater. Unless individuals are able to convince themselves that they are actually telling the truth (Trivers, 2000), attempts to manage these demands may be perceptible (DePaulo et al., 2003). Thus it may be unwise to assume that lying is either cheap, from a cognitive standpoint, or indiscernible.

2.6. Reputation

There also are issues arising from exposure and familiarity. In rhesus macaques, females signal benign intentions to other females, situated nearby, by issuing low-volume grunts and girennes. These presumptively inexpensive signals carry information that is reliable, and they tend to decrease social friction. But this is true only if the individuals interact repeatedly (Silk et al., 2000). It is also the case that in several species, males are recognized across the forest canopy from their loud calls (Delgado, 2006). These males, like the softly calling females, have a vocal history. For any of these animals to vocalize at all is to honestly communicate their continuing existence and present location to eavesdropping members of the group (Seyfarth and Cheney, 2003). But it goes beyond this. In a playback experiment, baboons responded abnormally to

1It has been argued that knowing things of relevance to others would have increased fitness since one presumably cannot know things without some amount of acquisitive ability (Dessalles, 1998, 2007).
the vocalizations of high-ranking animals if they were manipulated to imply subordination to a low-ranking animal (Cheney and Seyfarth, 2005). Human voices have histories, too. We keep mental records of statements made by others and ourselves, including promises, and these records tend to keep speakers honest. If liars seem to “get away with it” at the moment, there may be negative consequences later on, when listeners “compare notes” (Blumenthal, 1932). Over time, individuals who say things that are untrue develop a reputation for untruthfulness, are no longer believed, and are subject to social reprisal (Bailey, 1971; Heppenstall, 1971). The risks of lying, therefore, may outweigh any short-term benefits, especially in small groups (Grassly et al., 2000; Lachmann et al., 2001; Schlenker et al., 2001; Viljugrein, 1997). These ecological aspects of human life suggest that speech should not merely be seen as a signal, and that larger situational and species characteristics, which mitigate the role of energy and veracity, must be taken into account (Knight, 1998).

3. The role of competition

Although some modelers have emphasized the cooperative benefits of language (e.g. Nowak and Krakauer, 1999), the benefits of precursory forms may be difficult to appreciate outside of a competitive context. Chimpanzees are clearly capable of cooperating, but their intelligence becomes unusually evident when they compete (e.g. de Waal, 1982; Hare, 2001; Hare and Tomasello, 2004; Melis et al., 2006; Muller and Mitani, 2005). Males are inclined to compete directly (Wrangham and Peterson, 1996), whereas females tend to compete through the agency of others (Silk, 2007; Silk et al., 2006), and this pattern is also evident in modern humans. I submit that knowledge of these differences can usefully guide our thinking about the cost of speech—whether produced normally or in extravagant displays—and the complexity of meaningful utterances and language.

3.1. Duels

I have proposed elsewhere that hominin males played a distinct role in the early stages of linguistic evolution, and that vocal complexity was selected in association with competitions for dominance and mates (Locke, in press-a; Locke and Bogen, 2006). The proposal takes as a starting point the observation that females are unusually interested in two broad features of male behavior in the context of mate selection. The first is the ability of males to compete with other males (Berglund et al., 1996; Cox and Le Boeuf, 1977; Maestripieri and Roney, 2005; also see references in Locke, in press-a). The other feature is males’ ability to appeal directly to females, consistent with a general trend for males to display to potential mates, and for females to choose on the basis of these displays (Bastock, 1967). Although courtship displays and dominance contests are affected in various ways, there is frequently a vocal component. In most primate groups, males are the primary or sole producers of loud calls (Gautier and Gautier-Hion, 1977) and these calls contain reliable cues. In baboons, for example, loud calls predict age, competitive ability, and stamina as well as rank (Fischer et al., 2004; Kitchen et al., 2003), and much the same is true in other mammalian and primate species (Delgado, 2006; Fitch and Hauser, 2002; Palombit et al., 1999; Wich et al., 2003). Thus it is possible that even before the expansion of vocal capacity, female listeners received useful information from male vocalization. In our own species, there are vocal characteristics and perceptual preferences that appear to influence fitness evaluations, and others that do not do so, at least honestly. In men, status and dominance are linked to testosterone (Mazur and Booth, 1998), which tends to be higher in men with low vocal pitch (Dabbs and Mallinger, 1999; Pedersen et al., 1986); and men with low voices are judged by female listeners, from vocal samples, to be more dominant and attractive (Collins, 2000; Collins and Missing, 2003; Feinberg et al., 2006; Puts et al., 2006). Women also prefer male voices that are low in pitch (Collins, 2000; Oguchi and Kikuchi, 1997), a preference that is evidently strengthened by estrogen (Feinberg et al., 2006; Puts, 2005). Male university students with low voices report slightly more sexual partners than other men (Puts et al., 2006), and baritone opera singers report having more affairs than tenors (Wilson, 1984). In hunter-gatherer societies, men with low voices report fathering more children than men with higher voices (Apicella et al., 2007). These findings could be taken to mean that vocal pitch is a fixed trait, but of course men are able to manipulate their voices, and do so when it could alter their perceived dominance (Puts et al., 2006). An extreme form of manipulation is revealed in vocal and verbal rituals. In a broad range of human societies, adolescent and young adult males speak competitively when attempting to elevate their status (Abrahams, 1989; Kochman, 1969; Labov, 1972; also see reviews in Locke, 2001; Locke and Bogen, 2006). Frequently, these competitions take the form of verbal duels, a ritualized display of verbal talent that usually involves humorous insults, issued loudly and boisterously, typically before a mixed audience. Winning favorably affects status. Although many accounts of duels involve traditional societies or groups, these contests also occur among educated young men in modern societies (e.g. DeCapua and Boxer, 1999; Schwebel, 1997).

3.2. Extravagance

I suggest that in early stages of linguistic evolution, males who vocalized extravagant ly were able to command attention, a critical step in achieving rank or status (Chance, 1967; Chance and Jolly, 1970). In nonhuman primates, there are indications that novel vocal material attracts attention (Hopkins et al., 2007) and that novel words attract the attention of human infants (e.g. Akhtar,
2005) and adults (Bauman, 1975). In some cultures, males tend to use extravagant speech in courtship. For example, when “rapping” to women, young African-American men produce elegant and highly stylized verbal material, much of it previously prepared and stored for appropriate occasions (Abrahams and Gay, 1975; Kochman, 1969). There also are associations between verbal extravagance and status. When speaking in public, the Suya Indians of Central Brazil alter “the phonetic and rhythmic features of speech…generally by exaggerating rhythms and stressing unstressed syllables” (Seeger, 1981, p. 186). When Samoan leaders speak, they “demonstrate their skill at weaving verbal garlands of flowery words and phrases” (Holmes, 1969, p. 344). My speculation is that selection for vocal extravagance expanded the capacity to coordinate and control longer sequences of phonetic material, and that listeners who were able to evaluate these sequences preferentially received fitness information. Over time, this would have extended the perceptual “window” for speech memorial systems, contributing to the capacity to relate disparate elements—a possible precursor to the invention and use of linguistic grammars.

3.3. Fitness value of extravagant speech

The larger claim here is that “extravagant” speech increased fitness, and that some amount of vocal and proto-lexical complexity emerged independently of lexical semantic considerations. In early stages of linguistic evolution, it is possible that structured articulatory activity provided a quick and reliable way to evaluate cognitive abilities. Even if a small amount of vocal complexity were submitted to selection, this may have produced a huge increase in the capacity to make vocalization even more complex (Harvey and Arnold, 1982). Our attention is thus called to a property of signals that are ordinarily inexpensive: they can be made in a more costly way when it suits the signaler to do so (Guilford, 1997; Krebs and Dawkins, 1984). The costliness of exaggeration has been confirmed in laboratory research, which indicates that unusually loud and rapid speech does require exceptional movement, tension, and respiratory activity (Finnegan et al., 2000; Moon and Lindblom, 2003; Nelson et al., 1984). It is therefore unsurprising that in various species, including our own, correlations have been obtained between the frequency, duration, and loudness with which individuals vocalize and their rank or status (Fischer et al., 2004; Fisek and Ofši, 1970; Horn et al., 1995; Kalma, 1991; Kitchen et al., 2003; Leonard and Horn, 1995; Locke and Hauser, 1999). Of course, exaggerated speech is not merely challenging from a motoric standpoint; it also requires special kinds of skills. In traditional societies, individuals who are unusually able to manipulate vocal, articulatory, and linguistic material, and thereby to command attention, tend to enjoy higher than normal status and preferential access to positions of political power. Among the Pa’ikwéné people of the Amazon, social success and status are accorded to those with a “loudb-and-strong” way of speaking, which index a person’s health and strength. Public accusations, which are “louder and stronger” than normal speech, are considered a sign of “social and physical well-being and strength” (Passes, 2004, p. 10). Similar trends exist in modern societies (Locke, 2001; Locke and Bogan, 2006). Clearly, good talkers are different from others. In fact, some individuals are unable to speak for extended periods in public without vocal fatigue and clinical symptoms (McHenry and Carlson, 2004; Welham, 2003), and this includes adolescent males (Kelchner et al., 2006). But “honest signaling” is not merely assessed on the basis of loudness and endurance. For it is impossible to pretend a better knowledge of language than one really has, and to fake unusual skills in the delivery of speech. There are no individuals who seem eloquent, but in reality are not.

3.4. Duets

The second proposal, which envisions a prominent role for females, is that selection for utterance complexity occurred at the level of meaningful utterances in the context of affiliation and coalition building. As indicated earlier, female primates tend to handle competition somewhat differently than males. In a number of species, females spend more time than males huddled together in small groups, building and reinforcing social relationships by manual grooming (Walters, 1987). This activity enhances female fitness. Females with strong social ties are more likely than others to achieve coalition support. This elevates rank, which can favorably affect reproductive success (Silk, 2007). Thus it appears that there may be an indirect association between grooming and fitness, mediated by relationships and rank. Dunbar (1996) proposed that when group sizes increased, new pressures to compete elevated the role of coalitions. This caused a change in the means of servicing relationships from manual grooming, which is usually performed on a one-to-one basis, to a vocal form (“vocal grooming”) that could reach many individuals simultaneously. Although Dunbar referred to contact calls, the low-volume vocalizations that frequently accompany grooming, girneys and lipsmacks, seem more relevant. These sounds bear some resemblance to speech (Locke, 2008), and they, like grooming itself, are produced more frequently by females (Blount, 1985). There are parallels in our own species. When women talk to other members of their sex, they usually allow less physical distance between them, and make better eye contact, than men (Aiello, 1977; Argyle and Dean, 1965), and they are more likely to disclose personal information (Derlega et al., 1993). Thus, in referring to “duets,” I do not refer to the highly audible alternating calls and notes of mated pairs, as in gibbons and thrushes, but the bouts of more subdued and intimate vocalization by women. It has been observed, for example, that two women who have shared an experience in the past may relate it to other females, and
to themselves, *jointly*. These co-speakers repeat or paraphrase each other’s comments, talk simultaneously, and complete each other’s turns. In order to make their separate contributions they also “interrupt” each other. At one time, there was a tendency to interpret all interruptions as clashes, but these interruptions are not clashes. They are collaborations (Bennett, 1981; Coates, 1991; Falk, 1980). There is frequently a unifying topic, too—other people. Studies of women across educational levels, cultures, and at least five centuries of social history agree that adult women are far more likely than men to gossip, that is, to privately disclose intimate information about others and themselves (Bischoping, 1993; Capp, 2003). Gossip serves an affiliative function (Emler, 2001), and women are particularly likely to come together, and gossip, in times of stress (Taylor et al., 2000; Wilson, 1985). This is unusually clear in adolescence, when females who are opposed or perceptibly slighted by a competitor frequently respond by privately discussing the rival’s behavior with peers (Björkqvist et al., 1992; Burbank, 1994; Crick and Bigbee, 1998; Galen and Underwood, 1997; Hess and Hagen, 2006; Owens et al., 2000a, b). As for the benefits of doing so, it is appropriate to recall the reproductive advantages of close female relationships in other primates, and the primary means of servicing those relationships—manual grooming. For in some human cultures, membership in “gossip networks” is required for mutual aid. Only women who are willing to gossip are eligible for membership (Bott, 1971). Thus in female duets, as in male duets, we find a link between social factors that enhance fitness and a capacity for complex verbalization. Since gossip is about the behavior of people, a subject of exceptional complexity (Whiten and Byrne, 1997), any discussion of their behavior is likely to be extremely complex (Pinker and Bloom, 1990). I assume that verbal duets would have given the evolution of utterance complexity, extended narratives, and linguistic grammars a good push.

### 3.5. Song

Much of what I have said about the evolution of speech may also have applied to song. Darwin (1871) was among the first to identify a possible relationship between musical sounds and language, although he associated himself with a position, expressed nearly a hundred years earlier, “that the first language among men was music, and that, before our ideas were expressed by articulate sounds, they were communicated by tones, varied according to different degrees of gravity and acuteness” (Burnett, 1774, p. 469, italics in original). Darwin himself saw a specific connection between music and extravagant verbal displays, noting that all performers, whether impassioned orators, bards, or musicians, use “varied tones and cadences” to excite the strongest emotions in their listeners, as he presumed their “half-human ancestors [had] aroused each other’s ardent passions, during their mutual courtship and rivalry” (1871, p. 337). After Darwin, surprisingly little was written about the evolution of music (or language) until recent years, when several proposals appeared almost simultaneously (e.g. Dunbar, 2003; Masataka, 2007; Merker, 2000; Mithen, 2005; Patel, 2008). An extensive review and analysis of critical issues has been provided by Fitch (2006). He sees merit in the hypothesis that song preceded (or co-evolved with) speech, as does Brown (2000, 2001), who proposed the existence of “musilanguage,” an ancestral stage that was neither explicitly linguistic nor musical but embodied putative precursors of both language and music: lexical tone, combinatorial formation of small phrases, and expressive phrasing principles. From this common core, according to Brown, a richer capacity for emotional meaning evolved, producing music, as well as a more complex capacity for referential meaning, leading to language. It is interesting, in the light of Brown’s emphasis on core abilities, that some traditional societies have *song duels* (Brenneis and Padarath, 1975; Herndon and McLeod, 1980; Hoebel, 1964; McLean, 1965; Mirsky, 1937; Solomon, 1994; Travassos, 2000; Weyer, 1932) that resemble verbal duels in at least two ways. First, as with verbal duels, the participants in song duels are typically men who use their wits to insult each other, humorously, before a responsive audience, the outcome affecting their social standing. Second, although verbal duels are spoken and not sung, they are frequently performed with memorized material, and with altered prosody and rhythmicity (Abrahams, 1970c; Fortier, 2002; Seeger, 1981). This increases their similarity to song duels. It also causes verbal duels to intergrade with *poetic duels* (Bowen, 1989; Mathias, 1976), and *sung poetic duels* (Caraveli, 1985; Doukanari, 1997), and what must surely be the epitome of rhythmic performances, *drum duels* (Elliott, 1960). In view of their shared properties, it is perhaps unsurprising that singing and speaking recruit many of the same neural resources (Callan et al., 2006; Özdemir et al., 2006).

### 4. The complexity of language

The proposed role of females thus leads us to another observation about the handicap hypothesis. When it is applied to “speaking,” another factor is often lurking in the wings, uncommented, like the proverbial elephant in the room. It is language. Speech never occurs without language—the words and grammar that are known to native speakers. This may explain theorists’ conflation of the two (Zahavi and Zahavi, 1997), but speech and language are not identical traits, and each may contribute
something different to utterance complexity. This property—complexity—has not readily yielded to precise characterization (Adami, 2003), though there seems to be some agreement that complex systems generally have “a large number of parts that interact in a nonsimple way” (Simon, 1962), or have “a large number of interlocking parts” (Hübler, 2007). Twenty years ago, Bonner (1988) wrote that evolution “usually progresses by increases in complexity,” mainly because these increases improve efficiency and reproductive success and are therefore selected. In the case of birds, complexity of songs is sometimes defined simply as the number of phrases per trill or notes per song. Selection may have operated here, for there is evidence that male birds that can learn and use complex songs are usually more virile and vigorous than males with simpler songs (Leitão et al., 2006).

Surely no evolved behavior has more interlocking parts than human languages, but scholars have had difficulty thinking of a selection-based account for complex linguistic grammars. One psychologist, Premack (1985), suggested that it would be a challenge “to reconstruct the scenario that would confer selective fitness on recursiveness” (p. 281), an important component of grammatical systems (Hauser et al., 2002). Linguists have also had difficulty imagining how complex grammars could have been selected for any communicative benefits they would have afforded. Lightfoot (1991) expressed difficulty in seeing how specific syntactic patterns such as subjacency could have enhanced individuals’ ability to communicate. This does not mean that a selection-based account of communicative benefits is impossible, of course, and in Lightfoot (1991) one encounters a tempting possibility. An adaptive account “for each of the subparts” of grammar would be unlikely, he wrote, but “the whole package may have been adaptive” (1991, p. 69). Newmeyer (1991) also recognized that universal grammar, which has a large number of “interlocking parts,” may never have a specific explanation, but argued that the totality of the grammar may.

4.1. Ornamentation

Where speech can be exaggerated, and produced in unusually skilled ways, much the same is true of the linguistic material that is expressed in speech. In traditional societies, group leaders are those who use an exceptionally broad range of words, and phrases that are “ornamented” by material from outside the more limited repertoires of their listeners (Abrahams, 1970a,b, 1989; Garrett, 1993; Sherzer, 1990; Strathern, 1971). The “ornaments” include archaic or esoteric language, metaphors, metonymy, formulas, riddles, and special prosodic patterns as well as over-learned phrases (Bauman, 1975; Comaroff, 1975; Gossen, 1976; Malinowski, 1935). In the Trobriand Islands of New Guinea, Malinowski (1922) noted that rhetorical “power” was achieved through the use of “archaisms, mythical names, and strange compounds, formed according to unusual linguistic rules.” (p. 432)

4.2. Complexity as fitness

If our ancestors began to compete vocally, either by adding new elements or by uttering familiar elements in novel sequences, communication systems would have diversified, perhaps doing so independently of systems responsible for lexical and grammatical invention. I would speculate that these actions created the capacity for complexity at other levels of vocal production and processing, expanding cognitive support systems, and laying the groundwork for utterances that approach the rate, duration, and intricacy of sentences in modern languages. Zahavi (1975) suggested that a handicap could be viewed “as a kind of a test imposed on the individual” (p. 207). Where vocal complexity is concerned, the first test may have come early in development. There is evidence that infants who produce complex (syllabic) vocalizations tend to be more robust than others (Oller, 2000), and these infants may receive more favorable evaluation and treatment (Locke, 2006). I am aware of no evidence that demonstrates a direct connection between linguistic complexity and fitness in adulthood, but an association between the ability to speak well and intellectual capacity is clearly believed. Among the Igbo of southeastern Nigeria, for example, oratorical ability “is directly equated with intelligence and success” (Finnegan, 1967, p. 22). The Barundi people of central Africa associate verbal ability with “successful cleverness” (Peek, 1981, p. 22). In Sierra Leone, “it is noticeable how strongly the Limba connect intelligence and speaking” (Finnegan, 1969, p. 75).

Research carried out in modern societies has also documented associations between intelligence and vocabulary size—a possible link to utterance complexity—and between each of these variables and the use of rare words (for references see Locke, in press-a). This makes it possible for speakers to reveal their intelligence in brief social displays. One example is the courting male’s “chat-up line” (Bale et al., 2006), but in a range of cultures around the world there are specialized verbal art forms that are used competitively.

4.3. Folk IQ tests

These forms appear to qualify as “conditional handi- caps” (Kirkpatrick and Barton, 1997) or, as I have called them elsewhere, “folk IQ tests” (Locke and Bogin, 2006). For example, in the poetic song duels of Malta, men improvise complex metaphorical insults. “Due to the

3Ultimately, of course, language would have enabled assessments of the “sender’s” mind. From that point forward, as Miller (2000) has said, the greater cost of language or speech would not have been the energy that went into its formulation and execution, but “the cost of thinking of something verbally expressible that will impress another human.” (p. 360).
complexity of this sung poetry,” wrote Herndon and McLeod (1980, p. 147), “those who attempt it must be extremely intelligent.” In African-American communities, specialized forms of verbal art, including “rap” and “the dozens,” flower in adolescence, permitting evaluations of sexually mature males’ “underlying intelligence, verbal ability, speed, strength, agility, and endurance” (Kochman, 1971, p. 19). Among the Chamula Indians of southern Mexico, the ability to verbally duel “serves as a fairly dependable index of a man’s social maturity, linguistic competence, intelligence, and political potential” (Gossen, 1974, p. 106). In Samoan speaking contests, participants intentionally create distractions in order “to test the powers of concentration” of opposing orators, as well as their “tenacity and intelligence” (Holmes, 1969, p. 346).

5. Trade-offs

I have suggested that both speech and grammar contribute to utterance complexity, but do they do so equally? In the avian literature, it is evident that there are different cues to attractiveness, and that these cues interact in complex ways (Badayev and Hill, 2003; Badayev et al., 2002; Candolin, 2003; Coleman et al., 2004; Johnson, 1999; Van Doorn and Weissing, 2004). Handicap theorists, however, have generally assumed that only a single kind of display is available to signalers (Johnstone, 1995; Johnstone and Grafe, 1993). As a result, there has been little attention to complex, multi-component displays, and therefore, to the possibility that receivers react to combinations of cues, and that senders manipulate these combinations. But human communication systems are composed of multiple components and levels, and in speaking, component systems appear to compete with each other. One component, such as syntactic or lexical complexity, can interfere with another component, such as phonological processing or control (Bock, 1982; Connine and Titone, 1996; Masterson and Kamhi, 1992). In other cases, trade-offs involve competition between some aspect of language and an external behavior. For example, even a simple competing task (e.g. walking, finger tapping) can significantly reduce the fluency, rate, sentence length, and grammatical complexity of any ongoing speech (Hiscock et al., 1985; Kemper et al., 2003). This is especially true where performance on the task selectively recruits the (left) cerebral hemisphere that is typically dominant for speech and language, as does the right hand (Hicks, 1975; Lewis and Christiansen, 1989).

This material raises a question: what is the relationship, if any, between vocal displays and linguistic content? In a study of golden-collared manakins, experimentally administered testosterone increased several different types of audible wing-flap displays while retarding the growth of the “content,” that is, the colorful plumage that is normally seen during the displays (Day et al., 2006). In the case of human language, differences between verbal performance and linguistic knowledge have been exposed by endocrine research. Testosterone appears to increase verbal assertiveness and the disposition to perform or display verbally (Locke and Bogin, 2006) but may slow the acquisition of linguistic knowledge, at least in early stages of development (Baron-Cohen et al., 2004; Lutchmaya et al., 2002). Testosterone also delays responses on language processing tasks in adulthood (Christiansen and Knussmann, 1987; Van Goozen et al., 1994). Complementary relations between performance and knowledge are also exemplified, to some degree, in comparisons of the sexes. For example, women usually score the same as—or only marginally better than—men on paper-and-pencil tests of linguistic knowledge (Hyde and Linn, 1988), but do less well in various types of verbal competition and verbal performance (Friedley and Manchester, 1985; Manchester and Friedley, 2003).

The relative ease of ordinary sound making is not a “problem” for evolutionary linguistic theory. Rather, it may be precisely this property that allowed speech to be exaggerated and displayed competitively, and enabled verbally facile individuals to assume conditional handicaps, thus to succeed unusually in the quest for dominance and mating opportunities. Interdependencies of the components of language, and of speech and language, encourage us to consider the possibility that exaggerations increased vocal complexity and, with the advent of symbolic capacity—an issue beyond our scope here—heightened verbal complexity, building up the perceptual, memorial, and control systems that would be needed for intricate linguistic grammars. It was the low cost of ordinary speech, I suggest, that enabled linguistic grammars to become so complex.

6. Further research

Several of the suggestions offered here may be testable at the level of brains (and minds), languages, and individuals. There is evidence that pragmatic and grammatical functions dissociate clinically, as in Asperger syndrome (Baron-Cohen et al., 2005). This is consistent with laboratory findings on normal subjects, discussed earlier, that exceptional or exaggerated performance in one area of language occurs at the expense of performance in another area of language. I hypothesize further dissociations between pragmatic functions, such as obedience to principles of cooperation (Grice, 1975; Levinson, 1983), and categories of verbal performance, including joking, riddling, and arguing, which may have played a role in bids for dominance in evolution, as they do today, particularly in juveniles (Blacking, 1961; Dundes et al., 1970; Gossen, 1976; Haring, 1985; Maranda, 1976; McDowell, 1985).

A second area of needed investigation relates to linguistic complexity. Most languages have far more complex systems than are required for communication. They typically possess more phonemes than the lexicon requires; more words (including synonyms) than speakers need to classify their physical environment and express their ideas; and more grammatical rules than are necessary
to convey the information customarily transmitted in conversations and narratives. I have claimed that low-cost benefits of vocal complexity contributed to the evolution of grammatical capacity, along with parallel changes in cognitive and symbolic capacity. Some years ago it was suggested that high complexity in one area, e.g., phonology, co-exists with low complexity in other areas, e.g., morphology or syntax, such that the total complexity is more or less constant across languages (Hockett, 1958). It has also been assumed that there is no such thing as “simple” or “complex” languages. Recently, these assumptions have been revisited (Sampson, 2007), and they deserve additional attention.

In humans, the voice contains information about age, sex, and, to some degree, physical stature (Fitch and Hauser, 2002), and speech contains a great deal of additional information about education, social class, ethnicity, and geographical origin (Coulmas, 1998), but exactly how much information about fitness is honestly conveyed by each of the numerous elements that are embedded in linguistic utterances will require a great deal of additional research—at this point the surface has barely been scratched. The same can be said of verbal performances. As we saw, verbal duels share contextual, functional, and physical properties with song duels. It would be interesting to know more about the perceptual and memorial consequences of any prosodic and rhythmic alterations that add a sense of musicality to verbal performances. Research should also be conducted to identify the variations of voice and speech that attract and hold the attention of males and females, the actual costs of these variants, and other physical and behavioral traits with which they may be correlated. It would be helpful, additionally, to know if individuals who are unusually quick, fluent, and rhythmic in verbal monologues (e.g. “rappers”), and responsive in verbal contests (e.g. “duelers”), possess more linguistic knowledge—or the skills to acquire it—than their peers. Clearly, individuals who score high on language tests are not necessarily better performers, but there is a problem: most standardized tests of language are sensitive to behaviors acquired in formal programs of instruction in literacy and language arts, whereas success in verbal performance requires no such training (Locke, in press-b). Before research in this area can go forward, fundamental problems of measurement must be resolved.

Acknowledgments

I wish to acknowledge comments on the ideas offered here, and on earlier drafts of the manuscript, by Cathy Flanagan, Bjorn Lindblom, and Michael Studdert-Kennedy, and helpful criticisms by Tecumseh Fitch and two anonymous reviewers.

References


Bock, J.K., 1982. Toward a cognitive psychology of syntax: information processing contributions to sentence formulation. Psychol. Rev. 89, 1–47.


Finnegan, R., 1969. Attitudes to speech and language among the Limba of Sierra Leone, Odu n.s. 2, pp. 61–77.


Franks, B., 2005. The role of ‘the environment’ in cognitive and evolutionary psychology. Philos. Psychol. 18, 59–82.


