9

What Are the Uniquely Human Components of the Language Faculty?

Marc D. Hauser and W. Tecumseh Fitch

9.1 Introduction

From a biologist's perspective, language has its own particular design features. It is present in virtually all humans, appears to be mediated by dedicated neural circuitry, exhibits a characteristic pattern of development, and is grounded in a suite of constraints that can be characterized by formal parameters. Thus, language has all the earmarks of an adaptation, suggesting that it could be fruitfully studied from a biological and evolutionary perspective (Bickerton 1990; Deacon 1997; Jackendoff 1999; Lenneberg 1967; Lieberman 1984; Pinker and Bloom 1990). On the other hand, linguistic behaviour leaves no fossils, and many characteristics of language appear unique to our species. This suggests both that the phylogenetic approach (constructing adaptive narrative that captures the timing and functionality of language evolution in our species) and the comparative approach (using data from other species to gain perspective on characteristics of our own) will be fraught with difficulty.

We tend to agree that the construction of historical narratives of language evolution are too unconstrained by the available data to be profitable at present, especially since many plausible scenarios have already been exhaustively explored (see e.g. Harris 1996). At best, this practice provides a constrained source of new hypotheses to be tested; at worst it degenerates into fanciful storytelling. In contrast, we claim that the comparative approach to language has been and will continue to be a powerful approach to understanding both the evolution and current function of the language faculty. Our purpose in this chapter is to review the current state of the art in comparative studies of the faculty of language, focusing specifically on the sensory-motor system involved in the production and perception of acoustic

signals. We show that comparative research is an extremely valuable source of information in biolinguistics, allowing us to isolate and study those componenets of the language faculty inherited from our non-human ancestors. We also point out that such work is logically necessary before claims of human uniqueness and/or language-specific evolution can be made.

Despite the wide variety of theoretical and empirical perspectives in the study of human language, everyone agrees that language is a complex entity, incorporating a variety of interacting subsystems, from neurobiological and cognitive to social and pragmatic. As a research strategy, especially one aimed at uncovering similarities and differences across species, it appears necessary first to fractionate the 'language faculty' into a set of relevant subsystems and then explore which are uniquely part of the human capacity to acquire language. It is clear that some aspects of the language faculty must be unique to our species as no other theoretical stance can account for the fact that chimpanzees raised identically to humans cannot attain full language competence, despite their impressive achievements (Savage-Rumbaugh et al. 1993). At a minimum, all the subcomponents necessary for language might be present in chimpanzees, without being adequately interconnected. More likely, we think, a number of critical systems are not fully present in chimpanzees, though a significant proportion of the necessary machinery is in place. It is also possible that most of the subsystems of language have been modified to suit language in the course of human evolution, and thus are unique in a more limited sense. From this viewpoint, the lexicon would be an exception that proves the rule. Despite its obvious homology with memory systems in non-humans, the huge number of words that every child learns dwarfs the capabilities of the most sophisticated non-humans. This suggests that, despite a broadly shared neural basis, even the lexicon has undergone some special modifications in humans. In view of the currently available data, all of these possibilities seem at least reasonable, and a priori commitments to one viewpoint or the other seem premature.

We hope that the points made thus far are rather obvious and uncontroversial. Somewhat more controversially, we argue that the study of language must proceed by a detailed, comparative approach to each of the mechanisms contributing to language, and of their interactions and interface conditions. The most obvious reason concerns empirical research into the neural systems that contribute to language, much of which is limited to animals due to both practical and ethical considerations. To the extent that component subsystems are shared with animals, only there can we study them

with the full panoply of modern neuroscientific techniques. A second reason concerns communicative function. Language functions (among other things) as a system of communication for members of our species, although some have argued that important aspects of language did not initially evolve for communicative purposes (Hauser et al. in press). During human evolution, it was minimally necessary for the faculty of language to coexist with the extant vocal communication system. Further, the evolving language faculty probably co-opted certain aspects of the pre-existing communication system (e.g. parts of the phonetic system) for its own use. This constraint makes it important to have a clear sense of what the pre-linguistic hominid communication system was like, an understanding that can only come from the comparative study of communication. Finally, and most critically, it is logically necessary to study animals before making claims about human uniqueness, or positing that a particular subsystem or mechanism evolved 'for' language. Although this point also seems obvious to us, it has traditionally been ignored or misunderstood. In summary, animal studies are a necessary component of the biological study of the language faculty, allowing us to discover by a process of elimination those mechanisms that are unique to human language, as well as to study exhaustively those language-related mechanisms that are shared.

In this chapter we will concisely review comparative research that has specifically focused on mechanisms believed important for human language, many of which were or are posited to be uniquely human and/or specific to language. Because much of this work has focused on speech production and perception, we will focus on these areas. It is, of course, clear to us that speech and language are logically and sometimes practically separable (written and signed languages providing two clear examples) and might well have had independent evolutionary trajectories (Fitch 2000*a*). Nonetheless, all human societies use speech as the primary input/output system for language, and the constraints of the phonetic interface surely played a role in the evolution of the language faculty in its broad sense. Furthermore, the empirical grounding of speech science, combined with its measurable behavioural manifestations, has allowed the comparative study of speech and vocal production/perception to progress far beyond that of other aspects of comparative language research (e.g. syntax or semantics). This should not be taken as an indication that the comparative study of these latter topics is impossible, but simply that other research areas have lagged behind that of animal speech research. We expect that many of the same

theoretical perspectives and empirical tools that have been applied in the arena of speech can be fruitfully extended to other aspects of language as well (Hauser et al. in press).

9.2 Speech Production

A basic understanding of the physics and physiology of speech was attained more than fifty years ago (Chiba and Kajiyama 1941; Fant 1960; Stevens and House 1955), providing a necessary first step for adequate analysis and synthesis of speech, and thus paving the way for the major advances in speech perception that followed (Liberman 1996). Surprisingly, the ethological study of mammalian vocal communication proceeded in the opposite direction, with an almost exclusive focus on the perception of signals and little understanding of their production. Extremely basic questions concerning animal sound production have only recently become the focus of concerted research (Fitch 1997; 2000*a*; Goller and Larsen 1997; Nowicki and Capranica 1986; Owren and Bernacki 1988; Suthers et al. 1988). Such information is important both for the practical reason that adequate analysis and synthesis of animal signals requires a solid understanding of how they are produced, and because the evolution of communication systems is characterized by a constant interplay between signal production and perception (Bradbury and Vehrencamp 1997; Hauser 1996). As we describe below, the physics of the production mechanisms can lead to certain types of information being available in signals, providing the conditions for subsequent selection of perceptual mechanisms to access this information. New perceptual mechanisms can then create selection pressures back on production mechanisms, to conceal, enhance, or exaggerate such acoustic cues. Thus, an adequate understanding of the evolution of acoustic communication systems requires mature theories of both production and perception.

9.2.1 *Source–Filter Theory of Animal Vocal Production*

A central insight of modern speech science is that human vocal production can be broken down into two components, termed source and filter (Titze 1994). This source–filter model has more recently been generalized to vocal production in other terrestrial vertebrates (Fitch and Hauser, in press) and holds true for virtually all vertebrates whose production mechanisms are

well understood. The *source* (typically the larynx) converts the flow of air from the lungs into an acoustic signal. Typically this source signal is periodic (it has a fundamental frequency which determines its perceived pitch) and has energy at many higher frequencies (harmonics). This sound then propagates through the vocal tract (including the oral and nasal cavities). The air contained in the vocal tract, like any tube of air, possesses multiple resonances at which it can oscillate, termed *formants*. Formants act as band-pass filters, allowing energy to pass through at their centre frequency, and suppressing frequencies higher or lower than this. Together, all the formants of the vocal tract create a complex, multi-peaked *filter* function (a formant pattern), which forms the acoustic basis for much of speech perception. Thus vocal production is typically a two-part process: the production of an acoustic signal by the source and the subsequent spectral shaping by the formants constituting the vocal tract filter. In humans, and the other mammals studied thus far, these two components are independent. Thus, properties of the source (such as pitch) can be varied independently of the filter (formants), and vice versa.

9.2.1.1 *The sound-producing source*

From a comparative perspective, the larynx is a conservative structure (Harrison 1995; Negus 1949). The original function of the larynx was to protect the airway by acting as a gatekeeper to the respiratory system. This function is preserved in all tetrapods. The larynx in mammals also typically produces sounds: it contains paired vocal cords which are set into vibration by air flowing through the glottis. The rate of vibration, which can be modified voluntarily via changes in vocal fold tension, determines the voice pitch. This sound-generating function evolved later than the more basic gatekeeper function of the larynx, and must coexist within the constraints imposed by it. Nonetheless, interesting modifications of the mammalian larynx exist, including such phenomena as vocal membranes for high-pitch phonation and laryngeal air sacs (Fitch and Hauser 1995; Gautier 1971; Kelemen 1969; Mergell et al. 1999). However, the cartilaginous framework, innervation, vasculature, and musculature of the larynx are essentially invariant in mammals. It seems likely that the 'dual-use' constraint following from a single organ serving the dual functions of airway protection and vocalization is one reason for the relative conservatism of the mammalian larynx (Fitch and Hauser in press). Conveniently, the conservatism of the larynx allows us to apply the insights of the theory of human vocal fold

vibration, termed the myoelastic-aerodynamic theory (Titze 1994), to laryngeal function in other mammals.

One interesting type of variation in laryngeal anatomy, extreme among mammals, is in its relative size: howler monkeys have a massive larynx and hyoid complex, nearly the size of their head (Schön Ybarra 1988), while the hugely enlarged larynx of male hammerhead bats fills the entire chest (Schneider et al. 1967). Less extreme, but nonetheless impressive, hypertrophy appears to have evolved independently in many mammalian lineages, particularly among males (Frey and Hofmann 2000; Hill and Booth 1957). Because the lowest frequency producible by the vocal cords is determined by their length (Titze 1994), the primary function of a large larynx is the production of loud, low-pitched calls. Any force selecting for low voices (e.g. female mate choice or aggressive encounters with competitors) will select for lengthening of the vocal folds and accompanying enlargement of the larynx. This is nicely illustrated in humans: at puberty the male voice drops in pitch due to a testosterone-dependent enlargement of the larynx and concomitant lengthening of the vocal folds (Kahane 1978); this doubling of vocal cord length leads explains why men's voices are lower than women's (Titze 1994). In addition to providing an excellent example of convergent evolution between humans and animals, laryngeal dimorphism in our species also shows that laryngeal size has no critical impact on speech production.

Thus, the presently available comparative data indicate that the human larynx does not differ from that of most other mammals in ways obviously relevant to speech production. The aspects of human speech that involve the larynx (control of voicing and pitch) are almost certainly built on a phylogenetically ancient set of mechanisms shared with other mammals (although the possibility remains that humans have finer control over these functions than other mammals: Lieberman 1968*a*). In contrast, the human vocal tract is strikingly different from that of other mammals.

9.2.1.2 *The vocal tract filter*

The vocal tract includes the pharyngeal, oral, and nasal cavities. Their size and shape determine the complex formant pattern of the emitted sound. A central puzzle in the study of speech evolution revolves around the fact that human vocal tract anatomy differs from that of other primates: the human larynx rests much lower in the throat. This fact was recognized more than a century ago (Bowles 1889). In most mammals, the larynx can be engaged into the nasal passages, enabling simultaneous breathing and swallowing

of fluids (Crompton et al. 1997). This is also true of human infants, who can suckle (orally) and breathe (nasally) simultaneously (Laitman and Reidenberg 1988). At about the age of 3 months, the larynx begins a slow descent to its lower adult position, which it reaches after 3–4 years (Lieberman et al. 2001; Sasaki et al. 1977; Senecail 1979). A second, smaller descent occurs in human males at puberty (Fitch 1999). A comparable 'descent of the larynx' occurred over the course of human evolution.

The acoustic significance of the descended larynx was first recognized by Lieberman and his colleagues (Lieberman and Crelin 1971; Lieberman et al. 1969), who realized that the lowered larynx allows humans to produce a wider range of formant patterns than other mammals. The change in larynx position allows us to independently vary the area of the oral and pharyngeal tubes and to create a broad variety of vocal tract shapes and formant patterns, thus expanding our phonetic repertoire. In contrast, the standard mammalian tongue rests flat in the long oral cavity, making vowels such as the /i/ in 'beet' or the /u/ in 'boot' difficult or impossible to produce because they require extreme constriction in some vocal tract regions and dilation in others. These vowels are highly distinctive, found in virtually all languages (Maddieson 1984), and play an important role in allowing rapid, efficient speech communication to take place (Lieberman 1984).

Until recently, the descended larynx was believed unique to humans. Considerable debate has centred on when in the course of human evolution the larynx descended (reviewed in Fitch 2000*a*), a debate which remains unresolved because the tongue and larynx do not fossilize. Attempts to reconstruct the vocal tract of extinct hominids must thus rely on skeletal remains, combined with tenuous assumptions about the relationship between the anatomy of the skull or hyoid bone and the position of the larynx. Recent studies of the dynamics of vocal production in non-human mammals raise serious doubts about the reliability of such reconstructions. This work, involving X-ray video of vocalizing dogs, pigs, goats and monkeys (Fitch 2000*b*), shows that the mammalian larynx is surprisingly mobile, flexibly moving up and down during vocalization. This flexibility suggests that attempts to estimate the resting position of the larynx based on skull anatomy are superfluous, since the larynx typically moves far from its resting position during mammalian vocalization. These data also indicate the existence of a gradualistic evolutionary path to laryngeal descent: speech in early hominids might have been accompanied by a temporary laryngeal retraction (as seen in other mammals during vocalization). Finally, recent

data demonstrate that humans are not unique in having a permanently descended larynx. Male red deer show laryngeal descent exceeding our own, and further lower the larynx to the sternum while vocalizing (Fitch and Reby 2001). These comparative data indicate that a descended larynx is not necessarily indicative of speech. We conclude that questions of historical timing and attempts at fossil reconstruction have been overemphasized in the literature concerning speech evolution, at the expense of detailed consideration of robust comparative data available from living animals.

9.2.2 *Communication With Formants: The Comparative Perspective*

Although the phonetic, and hence communicative, importance of formants is axiomatic in speech science, there has until recently been little discussion of formants in animal communication, and one might easily conclude (incorrectly) that formants play little role in non-human communication. Here we briefly review the literature on the communicative function of formants in animals, to provide a richer perspective on the evolution of human speech. This research not only reveals that formants are present in vocalizations and perceived by animals but suggests that communicative uses of formants have a rich evolutionary history, long preceding human evolution (Fitch 1997; Owren and Bernacki 1988; Rendall et al. 1998).

Although researchers have recognized the existence of formants in the vocalizations of non-human primates for many years (Andrew 1976; Richman 1976), little attention has been paid until recently to the information they might convey. Two types of information that might theoretically be conveyed via formants are individual identity and body size. Because the detailed shape of the oral and nasal vocal tracts vary, individuals should have slightly different formant patterns that would allow listeners to determine the identity of a vocalizer. For example, individual differences in the sizes and locations of the nasal sinuses lead directly to individual differences in the speech output spectrum (Dang and Honda 1996), and discriminant function analysis of rhesus macaque calls suggests that similar phenomena may apply in monkeys (Rendall 1996). This has led Owren and Rendall (1997) to suggest that formants could provide important cues to individual identity in primates. While plausible, this suggestion has yet to be rigorously tested, and the flexibility of the mammalian vocal tract during calling (Fitch 2000*b*) suggests caution in interpreting individual vocal tract morphology as 'fixed'.

The idea that formants in animal vocalizations convey body size information has received more empirical support. Formant frequencies are strongly influenced by the length of the vocal tract (Fitch 1997; Titze 1994). Vocal tract length, in turn, is largely determined by the size and shape of the skull, which is strongly correlated with total body size (Fitch 2000*c*). Thus vocal tract length and formant frequencies are both closely tied to body size in the species examined so far (Fitch and Giedd 1999; Fitch 1997; Riede and Fitch 1999). This linkage should hold true for most mammals (Fitch 2000*c*). Thus, a listener that can perceive formants could gain accurate information about the body size of the vocalizer. A variety of birds and mammals can be easily trained to perceive formants (Hienz et al. 1981; Sommers et al. 1992), or to perceive them spontaneously (Fitch and Kelley 2000), suggesting that the ability to perceive formants was present in the reptilian common ancestor of birds and mammals. Because body size is highly relevant to social behaviour and reproductive success in most terrestrial vertebrates, it seems likely that an initial function of formant perception was to help judge the body size of a vocalizer. Particularly in dense forest environments or in darkness, an ability to perceive body size based on acoustic cues would be highly adaptive. These data provide strong support for the notion that communication via formants has a long evolutionary history in terrestrial vertebrates. Though more comparative data are necessary to exclude the possibility that formant perception in birds and mammals is a convergent adaptation (homoplasy), the most parsimonious interpretation of current data is that formant perception represents a homologous character, present in the common ancestor of birds and mammals that lived during the Palaeozoic several hundred million years ago.

9.2.3 *Convergent Evolution: The Descent of the Larynx in Non-Humans*

As mentioned earlier, the permanently descended larynx in humans represents an important difference between humans and our primate relatives, highly relevant to speech production. For many years researchers believed that this trait was uniquely human (Lieberman 1984; Negus 1949). Recent comparative studies demonstrate otherwise: at least two deer species have a descended larynx (red and fallow deer: Fitch and Reby 2001), which is pulled down to its physiological limit during vocalizations, substantially surpassing laryngeal descent in our own species. Other species have simi-

larly lowered and/or lowerable larynges, including lions, tigers, and other members of the genus *Panthera* (Peters and Hast 1994; Pocock 1916; Weissengrüber et al. in press), as well as koalas (Sonntag 1921); we confine our discussion mainly to deer. The descent of the larynx in deer is clearly not an adaptation to articulate speech, but its dynamic retraction during vocalization strongly suggests that it serves a vocal function. Why does the larynx descend in these species?

Detailed audio-video analysis of deer vocalizations demonstrates that laryngeal retraction lowers formant frequencies, as predicted by acoustic theory (Fitch and Reby 2002). One possible function of this formant lowering might be to increase the propagation of sounds through the environment, since atmospheric absorption is more pronounced for high frequencies. However, when a sound source is close to the ground (less than a metre, as with deer), the interference with reflections from the ground can actually weaken low-frequency transmission. This, along with behavioural data, suggests that formant lowering does not aid sound propagation in deer.

A more likely hypothesis is that laryngeal retraction serves to exaggerate the size of the vocalizer, a form of 'bluffing' that would be valuable in animals that often vocalize at night and in dense foliage, as do deer. Once perceivers use formants as a cue to size, the stage is set for deception: any anatomical mechanism that allows a vocalizer to evade the normal constraint linking body size and vocal tract length enables a smaller animal to duplicate the formant pattern of a larger individual by elongating its vocal tract, thus exaggerating its apparent size. Male red deer have partially evaded the constraint linking skull size to vocal tract length by evolving a highly elastic thyrohyoid ligament (which binds the larynx tightly to the hyoid skeleton in most mammals). Combined with powerful laryngeal retractor muscles, this allows stags to extend their vocal tracts far below the normal position, by about a third of their body length. The impressive roars thus produced have very low formants, serving to intimidate rivals (Clutton-Brock and Albon 1979) and attract females (McComb 1991) and creating an 'arms race' where all males without the trait will be out-competed. Finally, this sets the stage for the next round of perceptual evolution. This 'size exaggeration' hypothesis for laryngeal descent in deer is consistent with the available behavioural and acoustic data, and with data on vocal tract elongation in other taxa (Fitch 1999).

Because the common ancestor of deer and humans did not have a descended larynx, laryngeal descent in these species represents an example

of convergent evolution. There is obviously no guarantee that the descent of the larynx in each lineage occurred for the same reasons. However, since the size exaggeration hypothesis is based on physical and physiological principles that are common to all mammals it also provides a plausible alternative explanation for the initial descent of the larynx in our own species. By this argument, the permanently descended human larynx might have evolved early in the hominid lineage (e.g. in australopithecines), serving a size exaggeration function, long before the advent of language. The increased phonetic potential allowed by this arrangement may have lain dormant for millennia (as it still does in red deer) before being exapted for use in spoken language by later hominids. Consistent with this hypothesis, the initial descent of the human larynx, which happens in infants, is followed by a second descent which occurs at puberty, but only in males (Fitch and Giedd 1999). This second descent does not increase the phonetic abilities of teenaged boys, but probably serves a function similar to that in deer: increasing the impressiveness of the adult male voice via size exaggeration (Fitch and Giedd 1999; Ohala 1984).

9.2.4 *Speech Production: Conclusions and Future Directions*

The data reviewed in this section indicate that researchers interested in the mechanisms underlying human speech production can gain important insights from the study of vocal production in other animals. Far from being unique to humans, communication via formant frequencies appears to be an ancient characteristic antedating the origin of humans. Communication via formants originally functioned for size perception or individual identification, not for transmitting sophisticated linguistic messages. Most non-human mammals lower the larynx during vocalization, suggesting that the unusual descended larynx in our species probably evolved gradually, based on a pre-adaptive flexibility in larynx position in mammals. Furthermore, several non-human species show a permanent descent of the larynx which evolved convergently with humans, and a likely explanation for descent in these species might apply to humans as well. Thus, certain key aspects of speech are likely built upon an ancient foundation that can be fruitfully studied from a comparative perspective.

A broader conclusion that is that the importance of changes in the hominid vocal periphery has historically been overemphasized. Future work should focus on changes in the neural mechanisms underlying speech

production (e.g. Deacon 1997; Fitch 2000*a*; Lieberman 2000; MacNeilage 1998). There are at least two important candidate mechanisms for the role of critical adaptations in the evolution of spoken language: vocal imitation and hierarchical composition. Although vocal imitation is not uniquely human (it is seen in most songbirds and a number of marine mammals), it is obviously critical for the acquisition of large open-ended vocabularies, and is not shared with other non-human primates (Janik and Slater 1997). Thus, the neural basis and evolutionary history of vocal imitation should be a focus of future research (Studdert-Kennedy 1983; Hauser et al. in press). Second, speech requires a flexible and powerful ability to recombine small acoustic units (phonemes and syllables) into larger composites (words and phrases); this is the only way that an open-ended vocabulary of readily discriminable vocalizations can be created. Again, recombination of small units into larger units is seen in other animals (Hauser 1996), but not to the same degree as in human speech (MacNeilage and Davis 2000). The comparative study of the neural bases of these abilities, both cortical (Deacon 1997; MacNeilage 1998) and subcortical (Lieberman 2000), will be an important source of new information relevant to the evolution of spoken language.

9.3 Speech Perception

Our ears are bombarded with sound. However, when we hear spoken language, as opposed to sounds associated with either human emotion (e.g. laughter, crying) or music, different neural circuits appear to be engaged. The fact that specialized and even dedicated neural circuitry is recruited for speech perception is certainly not surprising, especially when one considers the evolution of other systems of communication. Exploration of this comparative database reveals that the rule in nature is one of special design, whereby natural selection builds, blindly of course, adaptations suited to past and current environmental pressures. Thus, by looking at the communicative problems that each organism faces, we find signs of special design, including the dance of the honey bee, electric signalling of mormyrid fishes, the song of passerine birds, and the foot drumming of kangaroo rats. The question of interest in any comparative analysis then becomes which aspects of the communicative system are uniquely designed for the species of interest, and which are conserved. In the case of speech perception, we

know that the peripheral mechanisms (ear, cochlea, and brainstem) have been largely conserved in mammals (Stebbins 1983). The focus of this section is to inquire which components of speech perception are mediated by a specialized phonetic mode, and which by a more general mammalian auditory mode. Evidence that non-human animals parse speech signals in the same way that humans do provides evidence against the claim that such capacities evolved for speech perception, arguing instead that they evolved for more general auditory functions, and were subsequently coopted by the speech system.

9.3.1 *Categorical Perception and the History of the 'Speech Is Special' Debate*

In the 1960s, Liberman and his colleagues (reviewed in Liberman 1996) began to explore in detail the mechanisms underlying human speech perception. Much of this work was aimed at identifying particular signatures of an underlying, specialized mechanism. An important early candidate mechanism was highlighted by the discovery of categorical perception.

When we perceive speech, we divide a continuously variable range of speech sounds into discrete categories. Listening to an artificially created acoustic continuum running from /ba/ to /pa/, human adults show excellent discrimination of between-category exemplars, and poor discrimination of within-category exemplars, a phenomenon termed 'categorical perception'. When first discovered, this phenomenon seemed both highly useful in speech perception and specifically tailored to the speech signal. This fact led Liberman and colleagues to posit (before any comparative work was done) that categorical perception was uniquely human and special to speech. To determine whether the mechanism underlying categorical perception is specialized for speech and uniquely human, new methods were required, including subjects other than human adults. In response to this demand, the phenomenon of categorical perception was soon explored in (1) adult humans using non-speech acoustic signals as well as visual signals, (2) human infants using a habituation procedure with the presentation of speech stimuli, and (3) animals using operant techniques and the precise speech stimuli used to first demonstrate the phenomenon in adult humans (Harnad 1987). Results showed that categorical perception could be demonstrated for non-speech stimuli in adults, and for speech stimuli in both human infants and non-human animals (reviewed in Kuhl 1989). Although

the earliest work on animals was restricted to mammals (i.e. chinchilla, macaques), subsequent studies provided comparable evidence in birds (reviewed in Hauser 1996). This suggests that the mechanism underlying categorical perception in humans is shared with other animals, and may have evolved at least as far back as the divergence point with birds. Although this finding does not rule out the importance of categorical perception in speech processing, it strongly suggests that the underlying mechanism is unlikely to have evolved forspeech. In other words, the capacity to treat an acoustic continuum as comprising discrete acoustic categories is a general auditory mechanism that evolved before humans began producing and perceiving the sounds of speech.

9.3.2 *Beyond Categorical Perception*

The history of work on categorical speech perception provides both a cautionary tale and an elegant example of the power of the comparative method. If you want to know whether a mechanism has evolved specifically for a particular function, in a particular species, then the only way to address this question is by running experiments on a broad array of species. With respect to categorical perception, at least, it appears that the underlying mechanism did not evolve for processing speech. We cannot currently be absolutely confident that the underlying neurobiological mechanisms are the same across species, despite identical functional capacity (Trout 2000). Nonetheless, a question arises from such work: What, if anything, is special about speech, especially with respect to processing mechanisms? Until the early 1990s, animal scientists pursued this problem, focusing on different phonemic contrasts as well as formant perception (reviewed in Trout 2000; Hauser 2002); most of this work suggested common mechanisms, shared by humans and non-human primates (for a recent exception, see Sinnott and Williamson 1999). In the early 1990s, however, Kuhl and colleagues (1991; 2000) published intriguing comparative results showing that human adults and infants, but not rhesus monkeys, perceive a distinction between so-to-speak *good* and *bad* exemplars of a phonemic class. The good exemplars or *prototypes*, functioned like perceptual magnets, anchoring the category, and making it more difficult to distinguish the prototype from sounds that are acoustically similar; non-prototypes function in a different way, and are readily distinguished from more prototypical exemplars. In the same way that robins and sparrows, but not penguins or flamin-

gos, are prototypical birds because they carry the most common or salient visual features (e.g. wings for flying, small beaks) within the category bird, prototypical phonemes consist of the most common or salient acoustical features. Although there is controversy in the literature concerning the validity of this work in thinking about the perceptual organization and development of speech (Kluender et al. 1998; Lotto et al. 1998), our concern here is with the comparative claim. Because Kuhl failed to find evidence that rhesus monkeys distinguish prototypical from non-prototypical instances of a phonetic category, she argued that the perceptual magnet effect represents a uniquely human mechanism, specialized for processing speech. Moreover, because prototypes are formed on the basis of experience with the language environment, Kuhl (2000) further argued that each linguistic community will have prototypical exemplars tuned to the particular morphology of their natural language. We consider this work to be a an elegant example of the comparative method, especially with respect to testing animals before claiming a uniquely human speech processing mechanism.

To further investigate the comparative claim, Kluender and colleagues (1998) attempted a replication of Kuhl's original findings, using European starlings and the stimuli used in Kuhl's original work: the English vowels /i/ and /I/, as well as the Swedish vowels /y/ and /ʉ/. These vowels have distinctive prototypes that are, acoustically, non-overlapping. Once starlings were trained to respond to exemplars from these vowel categories, they readily generalized to novel exemplars. More importantly, the extent to which they classified a novel exemplar as a member of one vowel category or another was almost completely predicted by the prototypical acoustic signatures of each vowel, as well as by the exemplar's distance from the prototype or centroid of the vowel sound. Because the starlings' responses were graded, and matched human adult listeners' ratings of *goodness* for a particular vowel class, Kluender and colleagues concluded, *contra* Kuhl, that the perceptual magnet effect is not uniquely human, and can be better explained by general auditory mechanisms.

In contrast to the extensive comparative work on categorical perception, we have only two studies of the perceptual magnet effect in animals. One study of macaques claims that animals lack such capacities, whereas a second study of starlings claims that animals have such capacities. If starlings perceive vowel prototypes but macaques do not, then this provides evidence of an analogy or homoplasy. Future work on this problem must focus on whether the failure with macaques is due to methodological issues

(e.g. differences in exposure to speech prior to training) or to differences in sensory-motor capacities that are indirectly (e.g. starlings are vocal mimics whereas macaques show no such evidence) or directly linked to recognizing prototypical vowels. If macaques lack this capacity while starlings have it, then our evolutionary account must reject the claim concerning uniqueness, but attempt to explain why the capacity evolved at least twice, once in the group leading to songbirds and once in the group leading to modern humans; again, we must leave open the possibility of a difference in the actual neurobiological mechanisms underlying the perceptual magnet effect in starlings and humans.

9.3.3 *Spontaneously Available Mechanisms for Speech Perception in Animals*

To date, when a claim has been made that a particular mechanism X is special to speech, animal studies have generally shown that the claim is false. Speech scientists might argue, however, that these studies are based on extensive training regimes, and thus fail to show what animals spontaneously perceive or, more appropriately, *how* they actually perceive the stimuli. They might also argue that the range of phenomena explored is narrow, and thus fails to capture the essential design features of spoken language (Trout 2000). In parallel with work on other cognitive abilities (e.g. number, tool use, food: Hauser 1997; Hauser et al. 2000; Santos et al. in press), we have been pushing the development of methodological tools that involve no training and can be used with animals or human infants, thereby providing a more direct route to understanding which mechanisms are spontaneously available to animals for processing speech, and which are uniquely human. Next, we describe several recent experiments designed to explore which of the many mechanisms employed by human infants and children during the acquisition of spoken language are spontaneously available to other animals.

A powerful technique for exploring spontaneous perceptual distinctions is the habituation/dishabituation procedure. Given the variety of conditions in which our animals live, each situation demands a slightly different use of this technique. The logic underlying our use of the procedure for exploring the mechanisms of speech perception is, however, the same. In general, we start by habituating a subject to different exemplars from within an acoustic class. A response is scored if the subject turns and orients in the direction of the speaker. Once habituated, as evidenced by a failure to orient,

we present test trials consisting of exemplars that deviate in some specified way from the training set. A response to the test stimuli constitutes evidence for perceptual discrimination, while the failure to respond (i.e. transfer of habituation) constitutes evidence for perceptual clustering or grouping across habituation and test stimuli.

9.3.3.1 *The role of rhythm in discriminating human languages*

The first comparative habituation/dishabituation experiment on speech perception (Ramus et al. 2000) explored whether the capacity of human infants both to discriminate between, and subsequently acquire two natural languages is based on a mechanism that is uniquely human or shared with other species. Though animals clearly lack the capacity to produce most of the sounds of our natural languages (see previous section, 'Speech production'), and are never faced with the natural problem of discriminating different human languages, their hearing system is such (at least for most primates: Stebbins 1983) that they may be able to hear some of the critical acoustic features that distinguish one language from another. To explore this problem, we asked whether French-born human neonates and cotton-top tamarin monkeys can discriminate sentences of Dutch from sentences of Japanese, and whether the capacity to discriminate these two languages depends on whether they are played in a forward (i.e. normal) or backwards direction; given the fact that adult humans process backwards speech quite differently from forward speech, we expected to find some differences, though not necessarily in both species. For neonates we used a non-nutritive sucking response, whereas for tamarins we used a head orienting response.

Neonates failed to discriminate the two languages played forward.[1] Rather than run the backwards condition with natural speech, we decided to synthesize the sentences and run the experiment again, with new subjects. One explanation for the failure with natural speech was that discrimination was impaired by the significant acoustic variability imposed by the different speakers. Consequently, synthetic speech provides a tool for looking at language discrimination, while eliminating speaker variability. When

[1] Strictly speaking, when subjects fail to dishabituate in the test trial following habituation, one cannot conclude that subjects have failed to discriminate. Specifically, and in contrast to psychophysical experiments that uncover just noticeable differences (JNDs), the habituation/dishabituation technique only reveals meaningful or salient differences (JMDs); even though two stimuli may not be considered meaningfully different, they may nonetheless be discriminable under different testing conditions.

synthetic speech was used, neonates showed discrimination of the two languages, but only if the sentences were played in the normal, forward direction. In contrast to the neonates, tamarins showed discrimination of the two languages played in a forward direction, for both natural and synthetic sentences. Like neonates, they also failed to discriminate Dutch from Japanese when the sentences were played backwards. More recent work (Tincoff et al. in prep) shows that tamarins can discriminate two other languages differing in rhythmic class (Polish and Japanese), but not two languages from the same rhythmic class (English and Dutch).

These results allow us to make five points with respect to studying the 'speech is special' problem. First, the same method can be used with human infants and non-human animals. Specifically, the habituation/dishabituation paradigm provides a powerful tool to explore similarities and differences in perceptual mechanisms, and avoids the potential interpretive problems associated with training. Second, animals such as cotton-top tamarins not only attend to isolated syllables as previously demonstrated in studies of categorical perception, but also attend to strings of continuous speech. Third, given the fact that tamarins discriminate sentences of Dutch from sentences of Japanese in the face of speaker variability, they are clearly able to extract acoustic equivalence classes, a capacity that comes online a few months after birth in humans (Jusczyk 1997; Oller 2000). Fourth, because tamarins fail to discriminate sentences of Dutch from sentences of Japanese when played backwards, their capacity to discriminate such sentences when played forward shows that they must be using specific properties of speech as opposed to low-level cues; the capacity to discriminate languages falling between rhythmic classes, but not within, adds support to this claim. Fifth, because the tamarins' capacity to discriminate Dutch from Japanese was weaker with synthetic speech, it is possible that newborns and tamarins are responding to somewhat different acoustic cues during this task. In particular, newborns may be more sensitive to prosodic differences (e.g. rhythm), while tamarins may be more sensitive to phonetic contrasts. Future research will explore this possibility.

9.3.3.2 *Speech segmentation and the implementation of statistical learning mechanisms*

A real-world problem facing the human infant is how to segment the continuous acoustic stream of speech into functional units, such as words and phrases. How, more specifically, does the infant know where one word ends

and another begins? Since periods of silence occur within and between words, and since stress patterns might only help with nouns (*Look at the ball!*'), what cues are available to the child?

A recent attempt to tackle this problem builds on early intuitions from computational linguistics, and in particular the possibility that infants extract words from the acoustic stream by paying attention to the statistical properties of a given language (Harris 1955). For example, when we hear the consonant string *st* there are many phonemes that we might expect to follow (e.g. *ork*, *ing*), but some that we explicitly would not expect (e.g. *kro*, *gni*). Saffran et al. (1996) tested the hypothesis that infants are equipped with mechanisms that enable them to extract such statistical regularities from a particular language. Eight-month old infants were familiarized for two minutes with a continuous string of synthetically created syllables (e.g. *tibudopabikudaropigolatupabiku* . . .), with no pauses between syllables. Within this continuous acoustic stream, some three-syllable sequences always clustered together, whereas other syllable pairs occurred only occasionally. To determine whether infants would extract such statistics, they were presented with three types of test items following familiarization: *words* consisting of syllables with a transitional probability of 1.0, *part-words* where the first two syllables had a transitional probability of 1.0 while the third syllable had a transitional probability of 0.33, and *non-words* where the three syllables were never associated (transitional probability of 0.0) in the familiarization corpus. Based on dozens of comparable studies on human infants, Saffran et al. predicted that if the infants have computed the appropriate statistics, and extracted the functional words from this artificial language, then they should show little to no orienting response to familiar words, but should show interest and an orienting response to both the part-words and the non-words. Results provided strong support for this hypothesis. They further show that infants are equipped with the capacity to compute conditional statistics. And it is precisely these kinds of computation, together with others, that might help put the child on the path to acquiring a language. Is the capacity to compute such statistics uniquely human and, equally important, special to language?

Saffran and collegues have excluded the 'special to language' hypothesis by showing that, at least for transitional probabilities, the same kinds of result hold for melodies, patterns of light, and motor routines (Hunt and Aslin 1998; Saffran et al. 1999). A different approach comes from testing non-human animals.

Several studies of pigeons, capuchin monkeys, and rhesus monkeys demonstrate that, under operant testing conditions, individuals can learn to respond to the serial order of a set of approximately eight to ten visual or auditory items (Orlov et al. 2000; Terrace et al. 1995; Wright and Rivera 1997). These results show that at least some animals, and especially some primates, have the capacity to attend to strings of items, extract the relevant order or relationship between items, and use their memory of prior responses to guide future responses. In addition to these data, observations and experiments on foraging behaviour and vocal communication suggest that non-human animals also engage in statistical computations. For example, results from optimal foraging experiments indicate that animals calculate rates of return, sometimes using Bayesian statistics, and some animals produce strings of vocalizations such that the function of the signal is determined by the order of elements (Hailman and Ficken 1987; Zuberbühler 2002). Recently, studies by Savage-Rumbaugh and colleagues (1993) suggest that at least some human-reared bonobos have some comprehension of speech and, specifically, attend to the order in which words are put together in a spoken utterance. Together, these studies suggest that, like human adults and infants, non-human animals are equipped with statistical learning mechanisms.

Hauser et al. (2001) used the original Saffran et al. (1996) material in order to attempt a replication with cotton-top tamarins of the statistical learning effects observed with human infants. The procedure was the same as that used with human infants, with two exceptions. Unlike human infants, who were exposed to the familiarization material for two minutes and then presented with the test items (in association with a flashing light), we exposed the tamarins in their home room to twenty-one minutes of the familiarization material on day 1 and then, on day 2, presented individuals located in a soundproof chamber with one minute of the familiarization material followed by a randomly presented set of test items.

Like human infants, tamarins oriented to playbacks of non-words and part-words more often than to words. This result is powerful, not only because tamarins show the same kind of response as human infants, but because the methods and stimuli were largely the same, and involved no training.

In terms of comparative inferences, our results on statistical learning should be treated somewhat cautiously because of subtle differences in methods between species, the lack of information on where in the brain

such statistics are being computed, and the degree to which such computations can operate over any kind of input (i.e. visual, motoric, melodic). Methodologically, the tamarins received more experience of the familiarization material than did the infants. We provided the tamarins with more input because we were unsure at the time that they would even listen to such synthetic speech, much less orient to it. Nonetheless, future work must establish how much experience is necessary in order to derive the appropriate statistics, and how the properties of certain statistics are either learnable or unlearnable by both humans and non-humans. For example, recent work (Newport et al. in preparation) suggests that both human adults and adult tamarins can learn about non-adjacent statistical relationships, but that the relevant perceptual units may differ between species; during these tasks, humans apparently extract units at the level of the phonemic tier (consonants and vowels), while tamarins extract at both the syllabic and phonemic tier, with the latter restricted to vowels as opposed to consonants. It is now important to ascertain whether human infants are more like tamarins or human adults, and the extent to which different kinds of statistical computation may or may not play a significant role in language acquisition. It is, of course, also important to ascertain which of these computational abilities are uniquely human and uniquely evolved for the purpose of language processing as opposed to other cognitive problems.

9.4 The Future of Comparative Studies

We have argued that a crucial component for discovering how the subsystems underlying speech production and perception evolved is to explore whether such mechanisms operate in other species. Our results show that many of the subsystems that mediate speech production and perception are present either in our closest living relatives or in other, more distantly related species; the work on speech perception also integrates nicely with work on computational issues, including statistical mechanisms for extracting the relationships between abstract variables in a sequence (Hauser et al. 2001; Hauser et al. in press). As a result, we argue, such mechanisms did not evolve for speech production or perception, but for other communicative or cognitive functions. We conclude here with a few comments about the connection between the neurosciences and behavioural studies of speech and language.

Are our verbal abilities unique or not? If we had to place a wager, we would bet that humans share with other animals the core mechanisms for speech perception. More precisely, we inherited from animals a suite of perceptual mechanisms for listening to speech—ones that are quite general, and did not evolve for speech. Whether the similarities across species represent cases of homology or convergence (homoplasy) cannot be answered at present and will require additional neuroanatomical work, tracing circuitry and establishing functional connectivity. What is perhaps uniquely human, however, is our capacity to take the units that constitute spoken and signed language, and recombine them into an infinite variety of meaningful expression (Hauser et al. in press). Although many questions remain, we suspect that animals will lack the capacity for recursion, and their capacity for statistical inference will be restricted to items that are in close temporal proximity. With the ability to test animals and human infants with the same tasks, with the same material, we will soon be in a strong position to pinpoint when, during evolution and ontogeny, we acquired our specially designed system for spoken language.

One direction that is likely to be extremely productive, in terms both of our basic understanding of how human infants acquire a language and of how the brain's representational structure changes over time, is to use nonhuman animals as models for exploring the specific effects of experience on acoustic processing. A major revolution within the neurosciences over the last ten or so years has been the discovery of remarkable plasticity in the adult brain, influenced by experience (Recanzone 2000). This revolution actually started earlier, driven in part by the magnificent findings on some songbird species and their capacity to learn new songs each season (reviewed in Nottebohm 1999). More recent work on mammals (rats and primates) has shown that when an individual engages in repetitive motor routines, or is repeatedly presented with sounds falling within a particular frequency range, the relevant cortical representations are dramatically altered. Similar kinds of effect have been suggested for language acquisition in human infants (Kuhl 2000), as well as for patients suffering from phantom limb (Ramachandran and Blakeslee 1998).

This evidence for cortical plasticity suggests experiments providing animal subjects with specific 'linguistic' experience and then testing for reorganization of perceptual sensitivity. For example, consider the results on tamarins showing a capacity to distinguish two different languages from two different rhythmic groups (i.e. Dutch and Japanese). Studies of human

infants suggest that whereas natives of one rhythmic group (e.g. French) can discriminate sentences of their own language from sentences of another language within the same rhythmic group (e.g. Spanish), infants exposed to a language that falls outside this rhythmic group can not discriminate French from Spanish. To test whether this rapidly developing selectivity follows from general auditory principles or from a specialized speech mechanism that is uniquely human, we can passively expose animals to one language over a period of weeks or months, and then explore whether such experience influences their capacity to discriminate this 'native' language with other languages, or the capacity to make fine-grained discriminations within the exposed language. Similarly, it is possible selectively to expose captive primate infants at different stages of development, and thereby determine whether there are critical periods for responding to such exposure. These results can then form the basis for further studies exploring the neurophysiology underlying behavioural or perceptual changes.

It is apparent to us (Hauser et al. in press), and many other scientists (Nowak et al. 2002), that the comparative approach will be a critical branch of empirical research into the nature of the human language faculty. At a minimum, comparative research will play the necessary if somewhat negative role of determining, by process of elimination, which components of language are *not* uniquely human or specific to language. More positively, we can expect that the comparative study of brain function, evolution, and development will provide the basis for a future theory of the neural implementation of the language faculty. Such research will combine with detailed behavioural study of animal capabilities to provide insights into the neural and behavioural mechanisms that were present at the evolutionary divergence between chimps and humans, which the evolving language faculty incorporated and elaborated. We foresee an iterative process in which studies on animals help to fractionate the language faculty naturally, 'cleaving nature at its joints', thus providing insight into how brains produce and process sounds, into how genes build brains, and eventually into the specific genetic changes that were necessary for the evolution of the language faculty.

FURTHER READING

Classics in the evolution of speech that provide a starting point for all further discussion include Lenneberg (1967) and Lieberman (1975; 1984). Lieberman's early work on non-human primate vocal production also provides an early example of the value of comparative work in understanding the evolution of speech—work

that was far ahead of its time (Lieberman 1968; Lieberman et al. 1969). A rarely quoted gem is Nottebohm (1976). For a more broad-ranging and revealing comparative analysis of the parallels between birdsong and speech, a nice introductory article is by Doupe and Kuhl (1999).

At a higher level, important ongoing work in the evolution of phonology is provided by MacNeilage's work (1998), which provides a Darwinian framework for understanding basic phonological distinctions (e.g. consonant and vowel, place of articulation) based upon the basic function and motor control of the jaw. This work provides a nice bridge between the low-level aspects of speech considered in this chapter and more theoretical issues at the heart of linguistics.