The notion of a biologically determined link between input and output of an animal's communication system is a commonplace of ethology. Seventy years ago Huxley (1914) remarked that the courtship rituals of the great-crested grebe must have evolved by selection of perceptually salient patterns from the bird's repertoire of motorically possible gestures. More recently, behavioral geneticists have sought evidence of perceptuomotor genetic coupling in breeding studies of crickets and grasshoppers. The conclusion from several such studies over the past ten years is that generating and perceiving mechanisms are polygenically determined and mutually adapted, but not genetically coupled (Barlow, 1981).

The perceptuomotor relation is even more complicated in those animals, such as certain species of songbird, that have to learn their species' song. In such cases, the bird often seems to be endowed with an auditory template that permits it to select its own species song from the many others that it may hear (Marler and Peters, 1977). The template may be no more than a skeletal structure that gets filled out by exposure to a particular dialect. The bird then has to discover—by gradual development from subsong muttering to full song—the motor patterns necessary to reproduce that dialect.
The analogy with the human infant learning to speak is obvious. The parallel may be fruitful because a capacity for vocal imitation is quite rare—peculiar, it seems, to songbirds, certain marine mammals, and humans. Thus, imitation is a species-specific behavior that calls for a specialized relation between the products of perceptual analysis and corresponding motor controls. Here we should note that a child’s reproduction of speech, though extraordinarily precise (as the persistence of dialects attests), is not mimicry; it differs from mynah bird or parrot imitations on two counts. First is the fact that, although the acoustic resemblance between a mynah bird’s utterances and its human models may be remarkably close (Klatt and Stefanski, 1974), the two are produced in quite different ways. Nottebohm (personal communication) has described an African Gray Parrot who learned phrases in German, American English, and an Irish brogue, but showed no tendency (as would a human) to produce, say, an English phrase with a German accent. One reason for the difference is presumably that while the bird imitates speech (as it would a fire siren or a lawnmower) in analog fashion, the human divides speech into segments (consonants and vowels) for purposes of articulation.

Moreover, a phonetically identical reproduction of one speaker’s utterance by another is usually quite different in acoustic structure from its model. Thus a young child who might, by appropriately adjusting its vocal tract length, imitate at least the formant pattern of adult vowels quite accurately, does not do so. Rather, it automatically normalizes the adult model and scales its reproduction to its own vocal tract size and shape (Lieberman, 1980). In other words, a phonological system, that is, a pattern of relations among segments, is implicit in the child’s speech from the first. All this argues that the motor organization of speech may be mediated by distinctive perceptual processes—in fact, by phonetic rather than general auditory perception.

What evidence do we have for the operation of such a distinctively phonetic mode of perception? And what can we say about its nature? From a variety of possible lines of evidence (Liberman and Studdert-Kennedy, 1978), let us briefly consider studies of (1) audiovisual adaptation and (2) duplex perception. The studies combine to demonstrate “online” perceptual dissociation of speech and sound.

Roberts and Summerfield (1981) started from a standard speech adaptation paradigm. In this paradigm listeners are asked to identify members of a synthetic speech continuum before and after prolonged exposure to (that is, adaptation with) a syllable drawn from one or other end of the continuum. If the synthetic continuum runs from, say, /be/ to /de/, with the boundary between the two phonemes set roughly in the middle, the effect of hearing several dozen repetitions of a particular syllable, say, /be/, is to reduce the number of /be/ responses and so shift the boundary toward the /be/ end of the continuum. Adaptation with /de/ shifts the boundary in the opposite direction toward /de/.

The novel twist introduced by Roberts and Summerfield (1981) was to exploit an audiovisual effect discovered by McGurk and MacDonald (1976). McGurk (see also MacDonald and McGurk, 1978; Summerfield, 1979) showed that if subjects viewed a videotape of a face uttering one syllable, say /ge/, while listening to a loudspeaker play in synchrony a different syllable, say, /be/, they often reported a percept different from either of the syllables presented—in this example, /de/. Roberts and Summerfield used precisely this pairing (visual /ge/, auditory /be/) as an adapting stimulus on a /b-d/ synthetic continuum. Of their twelve subjects, six reported hearing the audiovisual adaptor most of the time as either /de/ or /de/, four as /kle/, one as /fle/, one as /ma/. Not one reported hearing the auditory signal that was actually presented, namely, /be/. Yet, for every one of the twelve, the effect of adaptation was that of /be/—a significant reduction in the number of /be/ responses, and so a shift of
the boundary toward the /be/ end of the continuum. Thus despite each subject's conscious phonetic percept of an introral stop, his auditory system was appropriately adapted by the labial stop to which it had been exposed. The procedure effectively dissociated the perception of sound from the perception of speech, demonstrating that the phonetic percept is neither auditory nor visual, but abstract.

A second effect, dubbed "duplex perception" has been elaborated by A. M. Liberman and his colleagues (see, for example, Liberman, Isenberg, and Rakerd, 1981) on the basis of an effect discovered by Rand (1974): Two different percepts, one auditory, the other phonetic, arise simultaneously when the acoustic constituents of a synthetic syllable are separated and presented dichotically. Figure 1 displays a nine-step continuum of patterns sufficient to induce the effect. If the base (bottom left) is presented alone, it is usually heard as [da]; if one of the isolated transitions (bottom right) is presented alone, it is heard as a nonspeech "chirp." If the two patterns are presented dichotically in appropriate temporal alignment, the listener hears a fused syllable (either [da] or [ga], depending on which transition is presented) and, at the same time, a nonspeech chirp perceptually identical to the chirp heard in isolation. If, now, the patterns are presented for discrimination in pairs of stimuli, separated by three steps along the continuum, with instructions to attend on one series of trials to the speech percepts and on the other series of trials to the nonspeech chirps, the results are those of figure 2: For the nonspeech a more or less continuous function; for the speech, a discrimination function peaked at the phoneme boundary in the fashion typical of categorical perception. Once again, we have a dissociation of auditory and phonetic perception. (For a fuller discussion, see Studdert-Kennedy, 1982.)

The distinction between auditory and phonetic (or phonological) processes is not new (see, for example, Liberman, 1970). In fact, Donald Shankweiler and I concluded from
the results of a dichotic study over ten years ago that "while the general auditory system common to both hemispheres is equipped to extract the auditory parameters of a speech signal, the dominant hemisphere may be specialized for the extraction of linguistic features from those parameters" (Studdert-Kennedy and Shankweiler, 1970, p. 579). Subsequent work with normals and split-brain patients (some of it reviewed in the chapters by Moscovitch and by Zaidel) has tended to support this conclusion.

All that I want to do here is to raise the question of the function and possible origin of phonetic specialization. We do not know the form of a phonetic representation. Specifically we do not know how the representation is segmented—whether by syllables, phoneme-sized units, or features—but it presumably contains enough detailed information on temporal and structural relations among articulators to preserve properties of dialect. And this may provide a clue to phonetic function. For whatever the original function of dialect—whether to isolate breeding populations or to facilitate group cohesion within a territory—its biological depth is attested by the fact that every child does learn the dialect of its peers—and that this is the necessary route by which the child ultimately discovers the sound pattern, or phonology, of its language. We may surmise, then, that one function of the phonetic representation is to link the auditory structure of speech to its articulatory source. Thus the phonetic percept serves as the interface between sound and articulation.

This notion, in turn, offers a key to the origin and function of cerebral lateralization for language. The most striking property of the left hemisphere is, of course, that it alone can speak. If we add to this its dominance in manual praxis for some 90% of the population and the recent discovery of the linguistic status of American Sign Language, we may reasonably hypothesize that language was drawn to the left hemisphere because, first, that hemisphere already possessed the neural circuitry for coordination of fingers, wrists, and
hands, adapted for tool use, and second, this was precisely the type of circuitry required for unilateral coordination of larynx, velum, tongue, jaw, and lips. The specialized perceptual processes, essential to imitation and acquisition (and, in due course, perhaps the neurological substrate of syntax) would then have emerged in close association with that fine motor capacity.

Comments

Geschwind took issue with the notion that the expressive aspects of language preceded the receptive aspects. He pointed out that if a mutation occurred making it possible for a human to speak, it would not be useful, since he would have no language to speak and no one to understand him if he spoke. The opposite side of the paradox is that if a mutation occurred so that a human could suddenly understand language, he would be confronted with a society in which no one was speaking. This form of paradox is common in the evolutionary study of any capacity that depends on linked activities between individuals. Darwin tended to approach such problems by looking for a series of evolutionary steps by which the linkage might have come about. The earlier steps in the evolutionary process may look quite different from the final product. It has been suggested that this paradox, although typically not faced by theorists of the evolution of language, can be resolved (Geschwind, 1964).

Consider as an example of one possible resolution of this paradox (of course, not necessarily the correct one) the following hypothetical series of evolutionary events. Let us assume that a distant ancestor of modern humans had some very limited capacity to associate the distinctive sounds made by a particular species of animal with its visual form: A mutation has occurred conferring this ability on some particular individual. This would be useful to him, since the characteristic sound of an animal would evoke its visual form, thus enabling him to avoid danger or to hunt more effectively. This would enhance his chances of survival and the chances of those to whom he transmitted the gene, and so this new trait would spread through the population. One could speculate that at some later date, one of the humans who had inherited this ability underwent a further mutation, enabling him to imitate the sounds of an animal. (This can be recognized as a refined form of the classical “bow-wow” theory of language evolution.) Since this new mutation would be useful, it too would spread through the population. From this point on, further specializations would take place, including, for example, those that would adapt auditory perceptual capacities in one hemisphere to the particular characteristics of sound produced by the human phonatory mechanism. The hypothesized hominid could now transmit information about an animal he had heard or seen to other members of the group carrying the mutated gene that made it possible for an auditory stimulus to evoke an associated visual form.

Although this particular hypothesis depicts only one of many possible scenarios, its assumptions are not unreasonable. In particular, it places emphasis on an ability underlying the capacity to learn names. It is often asserted that this type of associative capacity is widespread among infrahuman species, but this conclusion lacks experimental support. One perhaps unique study is that of Warden and Warner (1927), who found, on testing a dog who had been prominently featured in films, that he could carry out a long list of verbal commands (that is, motor responses to spoken words). His owner claimed that he understood many object names, but his performance in selecting the named objects from a group of objects was either at or only slightly above chance.

Geschwind pointed out that there can be forms of communication between animals that may seem to involve the comprehension of names, but do not, in fact, do so. Vervet monkeys, for example, communicate in the wild (Seyfarth,
Cheney, and Marler, 1980), but the hypothesis that the cries are names of specific predators, understood by other vervets, is probably incorrect. When one vervet emits the specific cry elicited by the sight of an eagle, the other vervets look up, but this need not represent comprehension of the name—it may simply be a motor response of upward gaze triggered by a specific stimulus. To use a simple analogy, the cry made on seeing an eagle may be “Look up!” rather than “Eagle!” It should be recalled that in the Warden and Warner experiments the dog showed evidence of comprehension of commands, but not of names.

Finally, the studies of sign-language acquisition by chimpanzees may not have represented the optimal strategy for research on precursors of language. It might have been better to study auditory language and to discover whether chimpanzees could learn the names of large numbers of classes of objects. This might have made better biological sense because of the known predominance of audition in language. Furthermore, Geschwind concluded, the stress on making the chimpanzee produce language may have been misplaced if the first evolutionary step was similar to that in the briefly conjectured scenario presented earlier.

Geschwind’s general point concerning the problem posed for evolutionary theory by biologically determined behaviors linked across individuals is, of course, well taken: Human language is, in fact, simply one of a vast class of behaviors in which the activity of one organism is matched to that of another—from the food gathering of social insects to the hunting patterns of wolves and the mating of sexually dimorphic animals. Moreover, in communicative behaviors, it is perhaps futile to puzzle over whether sending or receiving came first, particularly since, as Geschwind implies, many small, interlocking mutations must underlie the behavioral complexes we now observe.

Nonetheless, the hypothesis that lateralized cerebral specialization for language was more strongly driven by cognitive pressures for increased motor-phonetic range than for specialized processes of perception is attractive for several reasons. First, studies of aphasic and split-brain patients, as well as a variety of experimental studies of normals (see the chapters by Moscovitch and Zaidel), suggest that the right hemisphere has at least some capacity for recognizing spoken words, but is largely incapable of both speaking and the perceptual phonetic segmentation apparently necessary (as speech error data suggest) for speech motor control. Second, the purely auditory perception of speech sounds seems to be comfortably within the capacities of cats, dogs, and even chinchillas (Kuhl and Miller, 1978), indicating that speech is probably well matched to the general mammalian auditory system. Third, on the assumption of some prior lateralization for manual control, adapted to the use of tools, the hypothesis goes some way toward rationalizing both the link between handedness and speech and the similarities between spoken and signed languages.

In any event, the issue, for whatever interest it may have, is certainly not foreclosed. A telling question is perhaps whether, in communicative behaviors, motor lateralization does generally precede perceptual lateralization. Petersen et al. (1978) have reported neural lateralization for the perception of species-specific vocalizations in Japanese macaques. But whether these animals also display motor lateralization is not known.
Concluding Comments

Electrical Stimulation Mapping

How are we to interpret the stimulation data? In particular, what is to be made of the association reported by Ojemann between disruptions in the ability to mimic sequences of orofacial movements and in the ability to identify stop consonants, evoked by stimulation at frontal, temporal, and parietal perisylvian sites? These findings seem to identify a cortical region that links speech perception to motor control in a fashion broadly consistent with a motor theory of speech perception (Liberman et al., 1967).

But the interpretation is not simple, since these results are in sharp disagreement with lesion data; deficits in phoneme and word identification are associated with damage to posterior temporal lobe (Wernicke's area), not to frontal (Broca's) or other perisylvian areas. Of course, many lesions have short-term effects quite different from their long-term effects; and, perhaps, if stimulation were sustained for a much longer time, the effects would be different. However, Darwin and his colleagues (Ettlinger, Teuber, and Milner, 1975, p. 132) have described a similar deficit in phoneme identification after resection of left face motor cortex—in patients, moreover, who were not aphasic—so that the transience of the stimulation is not likely to be relevant.
Perhaps the most plausible interpretation, at present, is that the task of phoneme identification calls for organization of a motor response in short-term memory by a process quite distinct from those of normal syllable or word perception and production. The fact that blockage may be induced by stimulation in orofacial areas not normally involved in speaking even suggests that the effect is essentially non-linguistic. Two responses, apparently identical in form, can be produced by different cortical systems if they have different functions. What we may have here, then, is a double dissociation between tasks and sites of lesion or stimulation; the normal linguistic task of lexical access is disrupted by posterior stimulation/lesion, while the metalinguistic mimicry (or production to command) of a meaningless stop consonant is blocked by stimulation/lesion in areas normally engaged for the organization of nonspeech orofacial movements.

In any event, the revival and refinement of the electrical stimulation technique promises a rich new source of data on localization. Like the grosser Wada sodium amytal test (Branch, Milner, and Rasmussen, 1964), stimulation mapping is validated by its success in predicting the effects of surgical resection. The next few years are likely to see a process of mutual accommodation and refinement in localization as the results from stimulation mapping and the new lesion scanning techniques (described by Benson) come in.

Relations between Lateralizations for Language and for Motor Control

The stimulation mapping results just discussed raise the general issue of the relation between movement control and language. The fact that both sequencing of movements (of hand and mouth) and language functions are disturbed by lesions of the left hemisphere does not necessarily mean that the two phenomena are causally related by some principle of economy, or efficiency, of processing. If this were so, we would expect individuals in whom language dominance and handedness are dissociated (for example, left-handers with left-hemisphere language) to display some manual or language deficit—which they do not. Moreover, as the patients of Darwin and his colleagues (Ettlinger, Teuber, and Milner, 1975) show, an individual may display disrupted orofacial motor sequencing, yet still be able to speak perfectly—and vice versa. In short, the familiar link between language and handedness is no more than a correlation; each function is the product of something else.

The “something else” may well be a general capacity of the left hemisphere for precise positioning and sequencing of movements, rooted in processes evolved for tool use, following the adoption of bipedalism (Lovejoy, 1981), some 3 1/2 million years ago. Speech and language seem to have exploited neural networks, already in place, to develop the rapid, informationally dense signaling system that an efficient communication system requires (compare Kimura, 1979; Studdert-Kennedy, 1981). To what extent speech and hand control still draw on shared neural organizations and mechanisms is, of course, an empirical issue (see, for example, Kinsbourne and Hicks, 1979; Lomas and Kimura, 1976; Lomas, 1980; Kelso, Tuller, and Harris, 1982). But it can hardly be an accident that the only articulatory system with a precision and agility to rival that of the vocal apparatus is in the hand—and that the hand, too, is an executor of language.

American Sign Language

The discovery of ASL, of its dual structure (formational and syntactic), its elaborate morphology, its complex embeddings of movements, and its underlying abstract forms—in fact, of the whole linguistic panoply that sets ASL on a par with spoken languages—encourages us to see the relations among
movement control, handedness, and language function as a key to the perceptuomotor origins of language. We may even look to the instantiation of language in two different perceptuomotor modalities as the thin end of a wedge that may ultimately permit us to strip the modality from our descriptions and isolate language at the intersection of sign and speech. That intersection may well prove to be the common core of movement control from which a specialized syntactic capacity has evolved.

In the phonetic domain, specialized processes seem to be restricted to speech, and this is not surprising. The hands are capable of many more differentiable gestures than the vocal tract, and the eye can perhaps handle a finer and more varied range of information than the ear. We have perhaps had to evolve specialized mechanisms for extracting articulatory-phonetic information from speech sounds, but sign languages found mechanisms for reading signs already available in the visual system. If we had evolved for visual communication and were faced with the task of finding an auditory substitute for signs, we would probably find it impossible. The visual system readily adapts to a written alphabet, to finger-spelling or to the complex, interwoven structure of a sign language; but, as the history of attempts to substitute a sound alphabet for speech has shown (Studdert-Kennedy and Liberman, 1963), the auditory system is incapable of handling any sounds but those of speech at a rate fast enough for effective linguistic communication.

Granted the specialized phonetic and syntactic capacities vested for spoken languages in the left hemispheres of most individuals, we are led to wonder how sign language will prove to be localized. From Bellugi (see part IV) we already know that both fluent and nonfluent ASL aphasias result from left-hemisphere lesions, but we do not yet know the details of localization. Particularly interesting in this regard is Neville’s (1977, 1980) evidence for evoked potentials in the auditory cortex of native signers in response to flashes of light. This points to extreme plasticity, as well as to neurological constraint on localization.

Evidently, within the limits of the left hemisphere, the human cortex is not rigidly wired. We depend on experience to establish the wiring. A normal infant, exposed to speech, develops the specialized perceptuomotor and linguistic systems for which the species is adapted; born deaf, and exposed to a sign language, the infant develops an alternative system no less readily. The concept of localization tempts us to think solely in terms of hardware, when, in fact, neural activity also forms around modes of operation shaped by the environment.

Does Structure Determine Function?

A recurrent controversy concerned the assumption implicit in studies of localization (whether by aphasic deficit or by stimulation mapping) that structure determines function. Suppose that we had ideal maps of brain activity, with resolution in space and time as sharp as we wanted, what could we then learn about how language is processed in the brain? The neural circuitry alone will not give us the program—any more than the circuitry of a computer reveals the range of programs that it can implement. We would have to work from known input/output relations, figuring out the functions necessary to yield those relations and comparing them with the circuits. Moreover, if computers are any guide, the relation of a particular deficit to a particular symptom can be far from obvious. If we are to learn about language processing from studies of aphasia, the form-function relations will have to be very much closer in humans than they are in computers. The gap between form and function is likely to be narrowest in the peripheral processes, widest in central or “higher” processes. It is the cognitively impenetrable functions, especially the more peripheral ones, that we are most likely to learn about from neuroanatomy and neuropsychology.
Finally, we should remember that there may be as much to be learned from computational models as from explicit neural speculations, even if these are based on sound physiological data. For example, if it is found that a particular cell fires whenever some event is perceived, we cannot conclude that this firing causes the percept. If perception depends on belief (as speech perception, for example, may depend on listening in a speech mode), and if the cell’s firing only occurs after belief and stimulus are integrated, then the cell’s activity is an index of perception, not a cause of it. In other words, correlations between neural activity and input/output events may tell us nothing about the processes that produce or follow from these events.

On the other hand, belief-contingent processes are not themselves beyond physiological investigation; Ojemann has observed changes in cortically evoked potentials from the language areas as a function of the instructions that a subject thinks he is following. The fact is that there are solid evolutionary grounds for expecting at least some degree of fit between form and function, and localization studies serve to place important constraints on models of how the neural circuitry works. Moreover, we would be unwise to reject a priori the possible contributions of neurology to an understanding of behavior—even of a behavior as complex as language. The history of the past fifty years offers many instances of advances in our understanding of behavior that have sprung from advances in our understanding of the behavior’s physiological underpinnings.

For example, classical discussions of emotion often argued that emotions could not be understood on a physiological basis. One reason for this conclusion was that psychologists accepted natural language classifications of emotions and assumed homogeneity where there was none. However, a series of neurological findings, including the effects of destructive lesions, epilepsy in certain locations, and brain stimulation in human and other animals, demonstrated that certain specific emotions can be separated from the earlier gross categories and assigned specific locations in the brain.

Other examples come from vision. The classic paper of Lettvin et al. (1959) showed that visual analysis of form, at a very elementary level, might be carried out quite differently than had been thought; and the later studies of Hubel and Wiesel on cat cortex carried this revolution still further. Another instance is provided by color vision. Great advances were made by purely behavioral studies of input/output relations, but when it became possible to study the physiology of the cones and to measure the quantity of certain pigments in these cells, our understanding of color vision was revolutionized. In fact, the new knowledge of circuitry and biophotochemistry suggested new behavioral experiments that would not have been thought of before.

Similarly, in language, it would hardly have been possible to discover, by linguistic analysis, that commands for eye movements and commands for trunk movements might be dissociated in comprehension, as Geschwind reported earlier. This finding raises the possibility that certain modes of comprehension may have preceded others in the course of evolution and suggests new experiments to test this hypothesis.

In short, we do not have to define the behavior, do all the necessary behavioral studies, build a model of the input/output relations, and only then try to discover how the circuitry instantiates the models. Instead, we can look to an interaction between studies of behavior and studies of neural circuitry in which each style of research constrains and guides the other: Behavioral studies point to specific questions about the circuitry, and new understanding of the circuitry leads to new experiments on behavior.

**Autonomy, Separability, and Specialization**

The concept of autonomy is viewed quite differently by linguists, psychologists, and neurologists. For the linguist, au-
tonomy of syntax or phonology, for example, simply refers to the fact that the formal apparatus of primitive terms and relations needed to describe the two subsystems are distinct and, indeed, incommensurable with respect both to one another and to other modes of cognition. Pylyshyn’s notion of cognitive impenetrability is related to this linguistic view: Language—or, at least, certain subsystems within it—is incommensurate with, and therefore impervious to, the presumed propositional apparatus of general cognition (knowledge, beliefs, goals, intentions).

However, as soon as we consider how such formally distinct processes might be instantiated in the nervous system, the sharp distinctions begin to blur. Brain damage rarely, if ever, affects language alone. The recall of familiar sequences, the timing and sequencing of motion patterns, the ability to change set, the capacity to combine knowing and doing, all contribute to linguistic performance, and all may be damaged by brain lesions—or, for that matter, may be manipulated in psycholinguistic studies. Of course, this does not mean that separable subsystems within language do not exist; it means only that they are not readily isolated in practice. This serves to emphasize that all studies of brain activity in language function will be of dubious value until we can increase our knowledge of neural circuitry and place intelligent constraints on neural models of language.

Much of the difficulty arises from the fact that we are dealing with a system. Systems are conceptually recalcitrant because, by definition, they consist of parts that are both separable and connected. In other words, full autonomy of language, or its subsystems, is neurologically implausible. What we are more likely to find is a collection of specialized subsystems that more or less correspond to the linguist’s descriptions, but are ultimately an integral part of the larger system. The most valuable contribution of neurology may not, in the end, be to validate linguistic description so much as to offer an approach to the deeper evolutionary question of how language relates to nonlanguage.

The task for the biologist is to derive the properties of language from the properties of its components. In other words, his task is to understand how language as a system emerged from some novel combination of more primitive, nonlinguistic mechanisms. For, as Jacob (1977, p. 1165) remarks of evolution: “It is always a matter of using the same elements, of adjusting them, of altering here or there, of arranging various combinations to produce new objects of increasing complexity. It is always a matter of tinkering.”