924

LANGUAGE AND SPEECH, 1994, 37(2), 203 — 209

Hand and Mind: What Gestures Reveal About Thought. By David McNeill. Chicago University Press, 1992. xi, 416 pp. $34.95

Reviewed by
MICHAEL STUDDERT-KENNEDY*
Haskins Laboratories

The argument of this original and difficult book is that "gestures are an integral part of language as much as are words, phrases and sentences–gestures and language are one system" (p. 2). Gestures are instantaneous, imagistic, analog, holistic expressions of the same thought that speech renders in hierarchical, linear, digital, analytic form. David McNeill credits Adam Kendon (1972, 1980) with discovering the link between, and essential unity of, speech sounds and gestural movements; his own work elaborates this insight at the higher linguistic levels of semantics and pragmatics.

The topic of the book, then, is gestures that accompany speech, the left-hand end of what McNeill calls "*Kendon's continuum*: Gesticulation → Language-like gestures → Pantomimes → Emblems → Sign languages" (p. 37). The continuum ranges from the informal, spontaneous, idiosyncratic movements of the hands and arms that often accompany speech, to the socially-regulated, standardized, linguistic forms of a sign language, with its arbitrary (non-iconic) lexicon.

Between these poles the obligatory presence of speech declines and the linguistic properties of gestures increase. "Language-like gestures" are grammatically integrated into an utterance, as when a speaker, asked about the weather on his vacation, replies: "Well, it was [oscillating hand gesture]", where the "so-so" gesture replaces an adjectival predicate. "Pantomime" conveys its full meaning in silence or, at most, with inarticulate onomatopoeia; also, in pantomime, sequences of gestures can form a unit, as they can in a sign language, but cannot in gesticulation. "Emblems" conform to standards of well-formedness, a language-like property that gesticulation and pantomime lack: in England, the palm-front V-sign is Churchill's "Victory!", the palm-back V-sign is a sexual insult. (For an amusing cross-class confusion in emblem dialects, see Collett, Marsh, and O'Shaughnessy, 1979, p. 229, where Margaret Thatcher appears in an Associated Press Photo, making the palm-back V-sign at a moment of electoral triumph.)

The contrast between the two ends of Kendon's continuum, between spontaneous gesture and conventional sign, epitomizes McNeill's notion of the process by which an utterance evolves in a speaker's mind. Spontaneous gesture reveals the primitive stage of an utterance, global, unsegmented, non-hierarchical, from which its conventional representation in speech unfolds: hierarchical, segmented, linear. The inner symbols of the primitive stage are private, idiosyncratic, closed to social influence; the end stage is public, grammatical, socially regulated. McNeill supposes that the primitive

* Address correspondence to Michael Studdert-Kennedy, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695.

stage of a sentence in a conventional sign language, such as American Sign Language (ASL), consists of global images no less than in a spoken language. But these images cannot escape into gesture, because the public end state has preempted the gestural channel.

The microevolution of an utterance may, in its turn, epitomize the macroevolution of linguistic system from primitive gesture. To support this speculation, McNeill reports the results of an undergraduate thesis at the University of Chicago by Ralph Bloom (1979). Bloom videotaped adults, who had no knowledge of a sign language, while they recounted a traditional fairy story to an adult viewer without speaking. Striking changes took place over the course of a session. For example, within 15 minutes or so, one story teller had developed a system with many of the standard properties of a spoken or signed language: segmentation; compositionality (of both signs and propositions); a lexicon (including three types of pronoun, one of them an abstract spatial pronoun similar to those of ASL); paradigmatic opposition (as when "King" and "Queen" shared a hand circling the head for "crown", but were distinguished by iconic gestures for "has-muscles" *vs.* "has-breasts"); ergativity (for example, the incorporation of a noun into a verb, as when the movement of threading a needle is made with a hand shaped for holding a thread); sign or "word" order (usually SV or SVO); standards of well-formedness; and fluency. With regard to the last, not only did the storyteller begin to streamline signs, stripping them to their essential features, but the viewer became so comfortable that he began to formulate questions in the new "language".

Although some of these linguistic properties may have been modeled on English (e.g. "word" order), others evidently were not modeled on any language known to the storyteller (e.g. three types of pronoun, ergativity). Rather, they seem to have emerged automatically, shaped by pressures toward clear and expressive communication. An analogous process in ontogeny, the emergence of recursion in the signing of a deaf child of hearing parents who used no sign themselves, has been reported by Goldin-Meadow (1982).

Whatever the worth of these parallels, or of the microgenesis of an utterance itself, as models of the phylogenetic evolution of language, they emphasize two aspects of Kendon's continuum essential to McNeill's argument. First is the contrast between spontaneous gesture (gesticulation) and formal language, whether spoken or signed. Second is the central theme of the book, namely, the common origin of the contrasting modes, gesture and language, in the thought that a speaker intends to express.

Let us turn now to the empirical work on which McNeill's monograph is based. Most of the gestural examples come from quasi-experimentally induced narrative discourse, although some are drawn from TV broadcasts, videotaped conversations, and naturally observed academic discourse. In the basic experimental situation, a speaker sees a film, animated cartoon, or comic book, and then recounts its story to a listener, a genuine listener who has not seen the "stimulus", and who will later have to retell the story to a third person. Neither speaker nor listener knows that gestures are the objects of study. The entire session is recorded on audio-video tape and the tape is subjected to minute analysis. All spoken utterances are transcribed, clause by clause, together with an indication of hesitations, and of the durations of pauses, filled and unfilled, in tenths of a second. All gestures are classified, coded and transcribed, with an indication of the words, parts of

words, or pauses with which they were temporally aligned. (A 25-page Appendix includes instructions for coding and transcription, detailed enough for other researchers to replicate the procedures of McNeill and his associates.)

The four main types of gesture in McNeill's system are: iconics, metaphorics, beats and deictics. All are symbolic, in that hand and arm stand for something other than themselves, and all are closely related to the semantic and/or pragmatic aspects of the speech they accompany. Iconics are "gestures of the concrete", exhibiting more or less transparent images of their referents. The images may be redundant, that is, coexpressive, with speech – a circling hand with downward pointing index finger for a cake on a table, a rising hand for someone climbing –, or complementary, capturing an aspect that the speech misses, as when a narrator describes an old woman chasing a cat out of her house and indicates her weapon, an umbrella, not in words, but with threatening shakes of the forearm.

Complementarity is one of several gestural properties demonstrating that gesture is neither a different version of the same covert verbal plan as speech, nor a translation from speech into another modality, nor, finally, an independent visual display, a photograph as it were, of the scene that speech puts into words. Rather, "...gestures and speech are operations that have been connected *within*" (p. 33, italics in the original), each arising from the same emergent thought, separate, yet integral, and each essential to full expression of a speaker's meaning.

Another index of this relation is the use of iconic gesture to highlight what a narrator finds salient in a situation. For example, a narrator describes two attempts by a cartoon cat to climb a drainpipe, first up the outside, then, having failed, up the inside: for the first attempt, the narrator's hand rises with the palm flat, for the second, with the palm in a hollow basket shape, depicting the interiority of the path. Here, the gesture reveals the speaker's "psychological predicate" (Vygotsky, 1962) at the moment of speaking, that is, "the novel, discontinuous, unpredictable component" (p. 127), that sets her current thought off from what went before. Evidently, the aspect of the second climb that the speaker particularly wished to capture was that it was inside rather than outside the pipe. If, as Vygotsky argued, thought is a process of forming contrasts with respect to preceding context, and if gestures express these contrasts (or "psychological predicates"), then we may sometimes apprehend the full meaning of an utterance more clearly, or even only, by attending to gestures as well as speech.

Not surprisingly, iconics tend to predominate in narrative, but "gestures of the abstract" (metaphorics, beats, deictics) also often occur and predominate in other genres, especially conversations and lectures. Metaphorics are no less pictorial than iconics, but the image they present is of an abstraction. For example, a type of metaphoric, called a "conduit metaphor", represents language, meaning, knowledge, art or other abstract notions as a substance, packed into a container that can be passed from one person to another. (Such metaphors are common in speech: "empty words", "deep book", "an amusing article, but not much in it", "a difficult idea to get across", and so on.) Thus, a speaker introduces a narrative with the words: "It was a Sylvester and Tweety cartoon"; as he speaks, he raises his hands as though holding a box, and then moves them apart, as though breaking it open. In another example, a speaker says: "I have a question", holding out her hand in a cup shape, as though to receive an answer.

The "cup-of-meaning" handshape seems to be a common symbol, adaptable to diverse circumstances. A speaker, referring to an event that might have happened but did not, says: "...even though one might have supposed..."; with the first four words his hands move out to the side in a cup shape, symbolizing potentiality (waiting for something to fall into them), on the fifth word ("have"), the hands snap shut onto emptiness (nothing fell after all). Or again, a speaker at an academic conference, emphasizing the importance of organization in a certain domain, moves his hand forward in a cup shape, as though carrying the domain itself, then with the word "organization", abruptly extends and spreads his fingers to form a rigid supporting armature.

Many other nicely analyzed examples – metaphors for states of mind, for dynamic processes of change, for mathematical concepts, and so on – illustrate the ease with which abstract ideas take on gestural form. Evidently, a concrete image aids the expression, and perhaps communication, of abstract thought.

The two remaining major types of gesture (beats, deictics) are not pictorial. Beats are simple two-phase movements (in/out, up/down) in which the hand moves rhythmically in time with the speech, typically taking the same form regardless of context. Beats fulfil a pragmatic function within a discourse, indicating that a word or phrase is important, not for its own semantic content, but for its contribution to the development of the narrative, or argument. Beats typically accompany a change of scene, the introduction of a new character or of a new theme.

Finally, deictics (pointing gestures) may be either concrete or abstract. The abstract function is particularly striking in light of the extensive formal use of deixis in ASL (Klima and Bellugi, 1979). Here, of course, we are concerned with the spontaneous, often idiosyncratic, informal use from which ASL forms presumably arose. A simple example comes from a conversation in which a speaker asks: "Where did you come from before?", pointing to the space between himself and his hearer. The indicated space is clearly not the actual space between the two interlocutors, but an abstract concept of the place that the hearer came from. In a more elaborate example, a narrator adopts the space in front of him as a metaphor for the plot of a story, assigning different loci to different characters and different modes of action. Appropriate pointings to one locus or another then lead the listener gesturally through the story.

Before we leave the empirical work, we must briefly review one more class of evidence supporting the hypothesis that speech and gesture arise as separate, but integral, expressions of the same thought – namely, the relative timing of the two forms: gesture and speech have a constant temporal relation. A prototypical gesture has three phases: preparation; the "stroke", or main part, of the gesture; and retraction. Only the stroke is mandatory, and only the stroke is matched, or "synchronized", with the speech at three levels: phonological, semantic and pragmatic. The stroke precedes or ends at, but never follows, "...the phonological peak syllable of speech" (p. 26). (Presumably this refers to the syllable that carries primary sentence stress in the judgment of a listener, although McNeill leaves "phonological peak" undefined.) The two channels also simultaneously express the same meaning ("semantic synchrony") and/or, where relevant, fulfil the same pragmatic function ("pragmatic synchrony").

Of particular interest here is the optional preparatory phase. For example, a narrator describes a comic book character in action: "...he grabs a big [oak tree and he *bends it*

way back]". (The square brackets enclose the words accompanying the gesture.) The preparatory phase begins with "oak": the speaker's hand rises up and forward at eye level, taking on a grip shape. Over the stroke phase (italicized words), the hand appears to pull something back and down toward the shoulder, ending at the phonological peak with the word "back". Now, since the preparatory phase has no function other than to prepare for the stroke, we can infer that the image of bending the tree back was already taking shape when the speaker was saying "oak tree". Evidently, then, the speaker's thought, the "minimal idea unit", or "starting node", from which both utterance and gesture grow had already taken global form during the preparatory phase of the gesture, before either the gestural stroke or the linguistic structure that would jointly express the thought had begun to emerge.

In such analyses as this (of which the book contains many dozens) we see with what subtlety McNeill builds, from empirical data, a theory of the relation between thought and its expression. His theory carries forward the work of Vygotsky (1962, 1986), deepening its biological roots and deliberately challenging (although quite without polemic) the programmatic (non-empirical), mechanical models favored by information-processing psycholinguists (e.g., Levelt, 1989).

For McNeill the starting node of an utterance (what we called above its "minimal idea unit" or "psychological predicate") is its "growth point", a metaphor from embryology with dynamic implications that its alternatives lack. The growth point is a small deviation, a minor salience, among the disordered fragments of images and linguistic categories, the residue of immediately preceding thoughts, from which utterance and gesture assemble themselves, thus assuring some degree of sequential coherence. The deviation does not arise tautologically from thought itself, but from extra-cognitive "...disruptive forces – motivation, emotion, and a future orientation" (p. 239) within the speaker. Here McNeill quotes Vygotsky: "Thought is not begotten by thought; it is engendered by motivation, i.e., by our desires and needs, our interests and emotions. Behind every thought there is an affective-volitional tendency, which holds the answer to the last "why" in the analysis of thinking" (Vygotsky, 1986, p. 252).

The assembly of utterance and gesture, set in motion by the "affective-volitional tendency", is a self-organizing process through which global order arises from local interactions among the disordered fragments, a theoretical process "...inspired by self-organizational models in developmental neurobiology (von der Malsburg and Singer, 1988)". Because a growth point consists of both images (which will surface as gestures) and linguistic categories (which will surface as words or phrases), the process of self-organization is a dialectic between gesture and language. The relation between the two modes changes during the few seconds of utterance-gesture formation (an interval that McNeill calls "deep time"): Thought, in its primitive stage more imagistic than analytic, emerges from deep time into real time, and so into existence, as a synthesis of the two modes. Essential to the synthesis is the underlying rhythmic pulse, the point of temporal convergence at which speech and gesture are integrated. With each pulse we, speaker and listener, gain momentary access to the "endless braid" (p. 237) of thought that constitutes a life.

Perhaps I have now said enough to give the reader a sense of what this book is about, and of where it stands in the field of contemporary cognitive psychology. I have

touched on its main themes, but have omitted many important topics, including those covered by two fascinating final chapters on the development of gesture in children and on the cerebral control of gesture, as evidenced by studies of aphasic and split-brain patients.

I have also omitted discussion of two chapters, reporting various experimental tests of the theory. One of these chapters describes several ingenious experiments, with delayed auditory feedback and other techniques, that support aspects of the self-organization model. The other chapter addresses questions that I found myself asking many times as I labored through the early chapters: Do gestures have any communicative function? Do listeners really use them to pick up information?

A partial answer comes from an experiment in which the "stimulus" for a speaker's narration of a cartoon film was not the film itself, but an experimenter's narration of the film, in which certain gestures had been deliberately mismatched to their accompanying speech. For example, the experimenter first established his left side as the locus of Sylvester, the wicked cat, and his right side as the locus of Tweety Bird, the good canary. Later, describing how Sylvester made a grab for Tweety the experimenter used anaphoric pronouns of which the verbal context made the reference unambiguous ("he lunges for him", where he = Sylvester), but accompanied these words with a gesture of his right hand (= Tweety) lunging to the right. The mismatch was subsequently reflected in the subject's verbal account: she initially reported in words what she had seen in gesture, while making a gesture matched to what she had heard in words; then she corrected herself, repairing her speech and repeating her original, correct gesture. Here, then, not only did gesture affect the subject's representation of the event, but it affected her initial recounting of it in words, while the words she heard determined the form of her gesture. The outcome not only demonstrates that gesture can convey meaning, but illustrates the unconscious integration of speech and gesture into a single "cross-modal" representation.

Nonetheless, my doubts concerning communicative function remain. Evidently gestures *can* convey information, and perhaps often do. Yet no less often gestures express essentially the same information as speech, and are therefore redundant. Moreover, listeners over radio or telephone do not seem to be especially handicapped. On the other hand, speakers over the telephone often gesture very much as they do in face-to-face conversation. And this suggests that gestures, even if not "cognitive necessities" (p. 259), may nonetheless facilitate verbal expression, as indeed McNeill argues.

I came to see, however, as I worked through the book, that the communicative function of gestures is largely peripheral to McNeill's enterprise. What is central, once again, is the hypothesis that gesture and language are part of a single system, a hypothesis for which the book marshals a mass of compelling evidence. For it is from this insight that McNeill has developed a unique, empirically based, and biologically driven theory of the relation between thinking and speech. His approach is a refreshing change from the machine models that dominate cognitive psychology in the current social climate of the military-industrial complex. Moreover, by viewing the "primitive stage" of thought as imagistic, and gesture as its natural mode of expression, McNeill throws light on one of the most remarkable discoveries of modern linguistics: the equivalence, in function and abstract form, of spoken and signed language.

In conclusion, let me strongly recommend a book that I initially found tedious and difficult, but ended by admiring and enjoying. The difficulty and tedium stemmed partly from the novel descriptive categories, couched in an utterance-gesture notational system that I did not know, partly from the meticulous density of the argument, and partly from my uncertainty as to where all the dry *explication de geste* was leading. My admiration and enjoyment grew as I gradually apprehended the originality of McNeill's theory and the scope of its implications for the evolution of thought and language.

(Received: March 30, 1994)

REFERENCES

BLOOM, R. (1979). *Language Creation in the Manual Modality: A Preliminary Investigation.* Bachelors thesis, Department of Behavioral Sciences, University of Chicago.

GOLDIN-MEADOW, S. (1982). The resilience of recursion: A Study of a communication system developed without a conventional language model. In E. Wanner and L. Gleitman (eds.), *Language Acquisition: The State of the Art* (pp. 51—77). Cambridge, U.K.: Cambridge University Press.

KENDON, A. (1972). Some relationships between body motion and speech. In A. Siegman and B. Pope (eds.), *Studies in Dyadic Communication* (pp.177—210). New York: Pergamon Press.

KENDON, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (ed.), *The Relation between Verbal and Nonverbal Communication* (pp. 207—227). The Hague: Mouton.

KLIMA, E., and BELLUGI, U. (1979). *Signs of Language.* Cambridge, MA: Harvard University Press.

LEVELT, W. J. M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: M.I.T. Press/Bradford Books.

MORRIS D., COLLETT, P., MARSH, P., and O'SHAUGHNESSY, M. (1979). *Gestures: Their Origins and Distribution.* New York: Stein and Day.

VON DER MALSBURG, C., and SINGER, W. (1988). Principles of cortical network organization. In P. Rakic and W. Singer (eds.), *The Neurobiology of Neocortex* (pp.69—99). Chichester, U.K.: John Wiley.

VYGOTSKY, L.S. (1962). *Thought and Language* (E. Hanfmann and G. Vakar, Trans.) Cambridge, MA: M.I.T. Press.

VYGOTSKY, L.S. (1986). *Thought and Language* (A. Kozulin, ed.). Cambridge, MA: M.I.T. Press.