The role of gesture and mimetic representation in making language the province of speech

Article · January 2012

DOI: 10.1093/acprof:oso/9780192632593.003.0009

2 authors:

Susan Goldin-Meadow
University of Chicago
347 PUBLICATIONS **22,576** CITATIONS

SEE PROFILE

David Mcneill
University of Chicago
159 PUBLICATIONS **12,868** CITATIONS

SEE PROFILE

In Michael C. Corballis and Stephen Lea (eds.), 1999. Evolution of the Hominid Mind, pp. 155-172. Oxford University Press.

THE ROLE OF GESTURE AND MIMETIC REPRESENTATION IN MAKING LANGUAGE THE PROVINCE OF SPEECH[1]

Susan Goldin-Meadow and David McNeill

Why, in all cultures in which hearing is possible, has language become the province of speech (the oral modality) and not of gesture (the manual modality)? This question is particularly baffling given that humans are equipotential with respect to language-learning---if exposed to language in the manual modality, that is, to a signed language, children will learn that language as quickly and effortlessly as they learn a spoken language (Newport and Meier 1985; Petitto 1992). Thus, on the ontogenetic time scale, humans can, without retooling, acquire language in either the manual or the oral modality. Why then, on an evolutionary time scale, has the oral modality become the channel of choice for languages across the globe?

Intuitively, one might suppose that the oral modality triumphed over the manual modality simply because it is so good at encoding messages in the segmented and combinatorial form that human languages have come to assume. We suggest, however, that this is not the case. In fact, the manual modality is just as good as the oral modality at segmented and combinatorial encoding. As a result, there would be little to choose between on these grounds.

Rather, we suggest that the oral modality assumed the segmented and combinatorial code not because of its strengths but to compensate for its weaknesses. The oral modality is *not* well suited to conveying messages mimetically, even though that function is also important to human language. This function is, however, very well served by the manual modality. The manual modality consequently assumes the role of mimetic encoding, in the form of spontaneous gestures found to accompany speech in all cultures (Feyereison and de Lannoy 1991; McNeill 1992), leaving segmented and combinatorial encoding by default to speech.

This argument rests on several assumptions. The first is that the manual modality is as adept as the oral modality at segmented and combinatorial encoding. We describe data that support this assumption in our review of the structural properties of signed languages, both conventional and idiosyncratic. The second assumption is that mimetic encoding is an important aspect of human communication, well served by the manual modality. We describe data supporting this assumption in our review of the gestures that are spontaneously produced along with speech. We end with a discussion of the advantages of having a language system that contains both a mimetic and a segmented/combinatorial code (*see also* Donald, this volume), and of the role that gesture might have played in linguistic evolution.

SEGMENTED AND COMBINATORIAL ENCODING IN THE MANUAL MODALITY

We begin by describing the properties of communication systems in the manual modality that are codified and that are used, and learned, as native languages by deaf people---conventional sign languages that have been transmitted from one generation to the next over *historical time*. We then turn to idiosyncratic gesture systems invented over two different time spans. The first are the gesture systems invented by deaf children who have not been exposed to conventional sign language to communicate with the hearing individuals around them, that is, idiosyncratic gesture systems developed over *ontogenetic time*. The second are the gestures that hearing individuals create in an experimental situation when asked to communicate using their hands and not their mouths, that is, idiosyncratic gesture systems developed over *experimental time*.

Segmentation and combination in a manual system developed over historical time

Sign languages of the deaf are autonomous languages, independent of the spoken languages of hearing cultures (Bellugi and Studdert-Kennedy 1980; Klima and Bellugi 1979; Lane and Grosjean 1980). Despite the fact that they are processed by the hand and the eye and not the mouth and the ear, sign languages have the essential properties of segmentation and combination that characterize all spoken language systems. For example, American Sign Language (ASL) is structured at the level of the sentence (i.e., syntactic structure: Liddell 1980; Lillo-Martin 1986; Padden 1983), at the level of the sign (i.e., morphological structure: Klima and Bellugi 1979; Newport 1981; Schick 1990; Supalla 1982), and at the level of sub-sign, and meaningless, elements akin to phonemes (i.e., 'phonological' structure: Coulter 1990; Lane *et al*. 1976; Liddell and Johnson 1986; Padden and Perlmutter 1987; Sandler 1986; Stokoe 1960; Wilbur 1986).

As in all spoken languages, the signs of ASL combine to create larger wholes, that is, sign sentences. ASL sentences have a basic or canonical sign order (Subject-Verb-Object), with other orders possible when one of the constituents is fronted and marked for topic (Fisher and Gough 1978; Friedman 1976; Liddell 1980; Padden 1983). Moreover, the signs that comprise the sentences of ASL are themselves composed of meaningful components, morphemes. Like spoken languages, ASL has grammatical markers that serve as inflectional and derivational morphemes, that is, systematic changes in form internal to the sign associated with changes in meaning (Klima and Bellugi 1979; Supalla and Newport 1978).

Unlike spoken languages, however, the form-meaning pairs that comprise the morphology of ASL are not produced in a linear string but are instead produced simultaneously. For example, the ASL verb 'ask both' is composed of two parts simultaneously produced: 'ask' which involves moving the index finger away from the chest area and bending it as it moves, and 'both' which involves reduplicating the motion. The sign 'ask both' is therefore produced by superimposing the grammatical morpheme 'both' on the uninflected form of 'ask,' resulting in reduplication of the basic outward bending movement, once directed to the left and once to the right (Klima and Bellugi 1979).

The morphemes of ASL are produced simultaneously. Nevertheless, they appear to have psychological integrity as isolable parts. For example, children acquiring ASL produce the meaningful parts of signs (the morphemes) in isolation and prior to combining them into

composite wholes (Newport 1981; Supalla 1982) despite the fact that the parts do not appear in isolation in their input. Thus, sign language, when developed within a community and passed down from generation to generation, is characterized by a system of segmented units that combine in rule-governed fashion.

Segmentation and combination in a manual system developed over ontogenetic time

Not only is segmentation and combination characteristic of communication in the manual modality when that communication has been conventionalized within a community, but it is also a salient feature of manual communication systems invented within a single generation by a deaf child of hearing parents. Deaf children exposed from birth to a conventional sign language such as ASL acquire that language following stages comparable to those of hearing children acquiring a spoken language (Caselli 1983; Hoffmeister and Wilbur 1980; Kantor 1982; Newport and Meier 1985). However, 90 per cent of deaf children are not born to deaf parents who can provide early exposure to conventional sign language. Rather, they are born to hearing parents who, not surprisingly, speak to their children. Unfortunately, it is extremely uncommon for deaf children with severe to profound hearing losses to acquire spontaneously the spoken language of their hearing parents and, even with intensive instruction, their speech is very likely to be markedly delayed (Conrad 1979; Mayberry 1992; Meadow 1968). In addition, unless hearing parents send their deaf children to a school in which sign language is used, the children are not likely to be exposed to a conventional sign system.

Despite their lack of a usable model of conventional language, deaf children of hearing parents do manage to communicate and do so by means of a self-created system of gestures called 'homesign' (Fant 1972; Lenneberg 1964; Moores 1974; Tervoort 1961). Most interesting to our concerns is the fact that the homesign systems invented by individual deaf children are characterized by a variety of language-like properties, including segmentation and combination. Rather than mimetically display a scene, the child conveys the message using segmented gestures combined into a rule-governed string. For example, rather than going over to the cookie jar and pretending to remove the cookie and eat it, the child will point at the cookie and then jab her hand several times toward her mouth, effectively conveying 'cookie-eat'. Moreover, the gesture strings generated by each of the deaf children can be described in terms of very simple 'rules.' The rules predict which semantic elements are likely to be gestured and where in the gesture string those elements are likely to be produced (Feldman *et al*. 1978; Goldin-Meadow and Feldman 1977; Goldin-Meadow and Mylander 1984, 1990). Thus, the gesture systems have sentence-like structure.

In addition to structure at the sentence level, each deaf child's homesign system also has structure at the word level. Each gesture is composed of a handshape and a motion component, and the meaning of the gesture as a whole is determined by the meanings of each of these parts (Goldin-Meadow *et al*. 1995). For example, a child moves his hand shaped like an O in a short motion arcing downward to request the experimenter to lay a penny down flat. The O-handshape represents 'roundness' (of the penny) in this gesture and in the child's entire corpus of gestures, and the short-arc motion represents 'putting down,' again across the entire gesture corpus. When produced together within a single gesture, the component parts combine to create the meaning of

the whole, 'putting down roundness.' In addition to combining components to create the stem of a gesture, one deaf child also altered the internal parts of a gesture (the number of times a motion is performed, and the placement of the gesture) to mark the grammatical function of that gesture, in particular, to distinguish between a noun role and a verb role (Goldin-Meadow *et al*. 1994). For example, when using a 'twist' gesture as a noun, the child tended to produce the twisting motion only once and in neutral space (near the chest area); in contrast, when using the 'twist' gesture as a verb, the child produced the twisting motion several times and extended it toward (but not on) the object to be twisted. Thus, the parts of a gesture vary as a function of its role in discourse.

Interestingly, the structure found at the sentence and word levels in each of the deaf children's gesture systems could *not* be traced back to the spontaneous gestures that their hearing parents produced when talking to them (Goldin-Meadow and Mylander 1983, 1984; Goldin-Meadow *et al*. 1994, 1995). The systems thus appeared to be generated in large part by the children themselves. It is consequently of great interest that these self-created gesture systems contained the properties of segmentation and combination, properties that characterize all naturally evolving language systems, be they spoken or signed.

As a caveat, we note that the creation of language by deaf children cannot be taken as a simulation of first creation of language in hominid evolution. Deaf children are developing their gesture systems in a world in which language and its consequences are pervasive. The human cultural world may therefore be necessary for an individual child to create a communication system characterized by segmentation and combination, although it does not appear to be necessary for a child to be exposed to a model of a conventional language to do so. In any case, the findings from deaf children make it clear that segmentation and combination can blossom in the manual modality over a relatively short period of time (the deaf children were all under 5 years of age when observed). In the next section, we demonstrate that these properties can arise in the manual modality even within a single experimental session.

Segmentation and combination in a manual system developed over experimental time

In an effort to determine whether gesture, if divorced from speech, would assume the properties of speech, Goldin-Meadow *et al*. (1996) asked adults who had no previous experience with sign language to describe a series of videotaped scenes using their hands and not their mouths. They then compared the resulting gestures (the gesture condition) to the gestures these same adults produced when asked to describe the scenes using speech (the gesture+speech condition).

The results showed that, in the gesture condition, the adults frequently combined their gestures into strings and those strings were reliably ordered, with gestures for certain semantic elements occurring in particular positions in the string; that is, there was structure across the gestures at the sentence level. In addition, the verb-like action gestures that the adults produced in the gesture condition could be divided into handshape and motion parts, with the handshape of the action frequently conveying information about the objects in its semantic frame; that is, there was structure within the gesture at the word level. Thus, the adults produced gestures

characterized by segmentation and combination and did so with essentially no time for reflection on what might be fundamental to language-like communication.

Interestingly, however, the adults in the gesture condition did not develop all of the properties of a natural language, or even all of the properties found in the gesture systems of the deaf children studied by Goldin-Meadow and colleagues. In particular, they failed to develop a system of internal contrasts in their gestures. When incorporating handshape information into their action gestures, they rarely used the same handshape to represent an object, unlike the deaf child whose handshapes for the same objects were consistent in form and in meaning (Singleton *et al*. 1993). Thus, a system of contrasts in which the form of a symbol is constrained by its relationship to other symbols in the system (as well as by its relationship to its intended referent) is *not* an immediate consequence of symbolically communicating information to another. The continued experience that the deaf children had with a stable set of gestures (cf. Goldin-Meadow *et al*. 1994) may be required for a system of contrasts to emerge in those gestures.

Thus when gesture is called upon to fulfill the communicative functions of speech, it immediately takes on the properties of segmentation and combination that are characteristic of speech. The appearance of these properties in the adults' gestures is particularly striking given that these properties were *not* found in the gestures that these same adults produced when asked to describe the scenes in speech. In contrast to the gesture condition, when the adults produced gestures in the gesture+speech condition, they rarely combined those gestures into strings, and rarely used the shape of the hand to convey object information within a gesture (Goldin-Meadow *et al*. 1996). In other words, they did not use their gestures as building blocks for larger units, either sentence or word units. Rather, they used their gestures to holistically and mimetically depict the scenes in the videotapes, as speakers typically do when they spontaneously gesture along with their talk, a topic to which we now turn.

MIMETIC ENCODING IN THE MANUAL MODALITY

We have shown that segmentation and combination are properties that appear in manual communication whether it was developed over a long or a short period of time. Thus, the manual modality can serve as a medium for language, suggesting that the capacity for creating and learning a linguistic system is modality independent. However, communication in the manual modality does not always assume language-like properties. When speakers use their hands to gesture as they talk, those gestures do not take on the analytic properties characteristic of speech (McNeill 1992).

Gesture conveys meaning differently from speech

In contrast to verbal behavior which is assumed to be closely tied to a speaker's thoughts, nonverbal behavior, including gesture, has traditionally been assumed to reflect the speaker's feelings or emotions (Wundt 1900/1973; see Feyereisen and de Lannoy 1991 for a review of studies focusing on gesture as a reflection of emotion and attitude). Recently, however,

researchers who have focused on the hand gestures speakers produce while talking (e.g., Kendon 1980; McNeill 1985, 1987, 1992) have argued that gesture can convey substantive information and, as a result, can provide insight into a speaker's mental representations. Children too produce gestures (Church and Goldin-Meadow 1986; Crowder and Newman 1993; Evans and Rubin 1979; Perry *et al*. 1988; Jancovic *et al*. 1975) and those gestures convey substantive thoughts that even observers not trained in gesture-coding can interpret accurately (Alibali *et al*. 1996; Kelly and Church 1996; Goldin-Meadow and Sandhofer 1996; Goldin-Meadow *et al*. 1992; but see Krauss, Morrell-Samuels and Colasante 1991, who argue that gesture does not convey to listeners information above and beyond the information conveyed in speech).

Thus, gesture conveys meaning. However, it conveys meaning differently from speech. Speech conveys meaning by rule-governed combinations of discrete units, codified according to the norms of that language. In contrast, gesture conveys meaning mimetically and idiosyncratically through continuously varying forms. McNeill (1992, p. 41) lists the fundamental properties of the gestures that accompany speech as follows:

1. Gestures are global in meaning. The meanings of the parts of a gesture are determined by the whole (and not *vice versa*, as is the case in speech). Indeed, the parts of a gesture cannot really be considered isolable units, as they are dependent for their meaning on the whole. In contrast to the bottom-up structure of sentences, there is consequently a top-down structure *within a gesture*.

2. Gestures are noncombinatoric. Gestures do not combine to form larger, hierarchically structured gestures. Most gestures are one to a clause and, even when there are successive gestures within a clause, each corresponds to an idea unit in and of itself. There is, as a result, no hierarchical structure *across* gestures (though there may be other kinds of non-hierarchical structure, as in 'catchments' described below).

3. Gestures are context-sensitive. They are free to incorporate only the salient and relevant aspects of the context. Each gesture is created at the moment of speaking and highlights what is relevant. Because of gestures' sensitivity to the context of the moment, there is variability in the forms gesture takes *within a speaker*.

4. Gestures do not have standards of form. Different speakers display the same meanings in idiosyncratic ways. There is consequently variability in the forms gesture takes *across speakers*. Even when there is cross-speaker similarity, this is not because of standards but because of similarity of meaning---similar meanings engender similar gestures.

It is often easy to analyze a given gesture into parts, but these parts have a different status from the parts of sentences---they are individually constructed with meanings that are determined by the context and that percolate from the top down. For example, in describing an individual running, a speaker moved his hand forward while wiggling his index and middle fingers (McNeill 1992). The parts of this gesture gain meaning because of the meaning of the whole; the wiggling fingers mean 'running' only because we know that the gesture, as a whole, depicts someone running and not because this speaker uses wiggling fingers to mean running in any other context. Indeed, in other gestures produced by this same speaker, wiggling fingers may well have a very different meaning (e.g., indecision between two alternatives). To argue that the

wiggling fingers gesture is composed of separately meaningful parts, one would have to show that each of the three components that comprise the gesture---the V handshape, the wiggling motion, and the forward motion---is used for a stable meaning across the speaker's gestural repertoire. The data suggest that there is no such stability in the gestures that accompany speech (McNeill 1992).

Thus, the gestures that accompany speech are not composed of parts but instead have parts that derive from wholes. Moreover, they are wholes that represent by way of mimetic depiction. Because the gesture as a whole must be a good (i.e., relatively transparent) representation of its referent, the addition of semantic information to a spontaneous gesture never decreases its iconicity. Although the simultaneous occurrence of morphemes within a sign in a conventional sign language such as ASL can give that sign an iconic quality, mimetic depiction is *not* the principle underlying ASL. Thus, ASL will sacrifice iconicity if the rules of combination require it, while gesture never does. For example, the sign for 'slow' in ASL is made by moving one hand across the back of the other hand. When the sign is modified to be 'very slow,' it is made *more rapidly* because this is the particular modification of movement that denotes intensification (Klima and Bellugi 1979). As a result, modifying the meaning of the sign reduces its iconicity because the meaning of the sign as a whole is, in rule-governed fashion, made up of the meanings of the components that comprise it. In contrast, if a gesture is generated along with speech to depict something that is very slow, that gesture will be articulated particularly slowly---it would not work to increase the speed of the gesture if the goal is to convey extreme slowness (McNeill 1992).

Given that gesture and speech convey meaning differently, it is possible for the meanings expressed in each of the two modalities to complement one another, creating a richer picture than the view offered by either modality alone. For example, when describing Granny's chase after Sylvester in a cartoon narrative, a speaker said 'she chases him out again' while moving her hand as though swinging an umbrella (McNeill 1992). Speech conveys the ideas of pursuit and recurrence while gesture conveys the weapon used during the chase. Both speech and gesture refer to the same event, but each presents a different aspect of it. As a second example, a speaker who may not be able to convey a particular meaning in speech may well still be able to express that meaning in gesture. At a certain stage in the acquisition of mathematical equivalence, a child may explain that she solved a problem such as 4+3+5=__+5 by adding all of the numbers on both sides of the equation (e.g., she says, 'I added the 4, the 3, the 5, and the 5 and got 17'), never commenting on the fact that the equal sign divides the equation into two parts. However, in her gestures, the same child manages to convey just this notion (e.g., she produces a sweeping gesture under the 4, the 3, and the 5 on the left side of the equation, and the same sweeping gesture under the blank and the 5 on the right side of the equation; Perry *et al*. 1988). Thus, gesture conveys aspects of equivalence that are not found anywhere in the child's speech (Alibali and Goldin-Meadow 1993a; Goldin-Meadow *et al*. 1993). In this way, gesture expands on the representational possibilities offered by the codified spoken system.

Gesture and speech form an integrated system

Despite the fact that gesture and speech represent meaning in different ways, the two modalities form a single, integrated system. Gesture and speech are integrated both semantically and temporally. For example, a speaker produced the following iconic gesture when describing a scene from a comic book in which a character bends a tree back to the ground (McNeill 1992): He grasped his hand as though gripping something and pulled the hand back. He produced this gesture as he uttered the words 'and he bends it way back.' The gesture was a concrete description of precisely the same event described in speech, and thus contributed to a semantically coherent picture of a single scene. In addition, the speaker produced the 'stroke' of the pulling-back gesture just as he said 'bends it way back.' The gesture was consequently synchronized with speech (see also Kendon 1980).

Even when gesture and speech convey different information, the two modalities adhere to the principles of gesture-speech integration described by McNeill (1992). Consider, for example, a child asked to explain why she thinks an amount of water has changed after it has been poured. The child says the amount is different because 'the glass is tall' while indicating the width of the glass in her gestures (Church and Goldin-Meadow 1986). Although this child is indeed expressing two different pieces of information in gesture and speech, she is nevertheless describing the same object in the two modalities. Moreover, the timing of the gesture-speech 'mismatch' also reflects an integrated system. The child produces the width gesture as she says 'tall,' thus synchronously expressing her two perspectives on the glass.

Further evidence that gesture-speech 'mismatches' reflect an integrated system comes from two sources. First, children begin to convey different information in speech and its accompanying gesture for the first time *after* gesture and speech have become temporally synchronized, that is, mismatch appears for the first time after gesture and speech have been integrated into a single system (Butcher and Goldin-Meadow 1997). Second, children who produce many gesture-speech mismatches when explaining their solutions to a given task have been found to be in a transitional state with respect to that task; that is, they are particularly ready to learn the task (Church and Goldin-Meadow 1986; Perry *et al*. 1988). If gesture and speech were independent of one another, their 'mismatch' would be a random event and, as a result, should have no cognitive consequence whatsoever. The fact that mismatch is a reliable index of a child's transitional status suggests that the two modalities are, in fact, *not* independent of one another (see Alibali and Goldin-Meadow 1993b; Goldin-Meadow *et al*. 1993).

In addition, speech sometimes takes the form that it does precisely because of a prior gesture, thus making it clear that gesture and speech are part of the same communication system. For example, in his initial utterance, a speaker produced a gesture to the right for Sylvester and a second gesture to the left for Tweety. In a subsequent utterance, the speaker then used only pronouns in his speech, relying on gesture to disambiguate the utterance. The speaker said 'so he knows that he's gonna come and get him,' producing a gesture to the left, the space associated with Tweety, just as he uttered the first 'he.' The first 'he' is therefore understood as a reference to Tweety, and the two other pronouns are understood in relation to this starting point---a starting point that depends crucially on gesture and its synchronization with speech.

Gesture and speech therefore form an integrated system, with gesture providing a representational format that complements the format found in speech. The fact that gesture is found along side of speech in all of the spoken languages that have been examined thus far (cf.

McNeill 1992) suggests that the mimetic representation that gesture offers language may be essential to human communication. Humans can, of course, speak without gesturing (we can talk with our hands folded or even bound). Nevertheless, gesture is pervasive in human talk. We often produce gestures in situations where no observer is present to appreciate the output of the act. For example, speakers gesture when talking on the telephone or when they have their backs turned to their listeners (e.g., Rimé 1982). Even more compellingly, children who have been blind from birth and have never experienced the communicative value of gesture produce gestures along with their speech (Iverson and Goldin-Meadow 1997). Thus, gesture is a robust component of human communication, supplementing the analytic code that speech offers with a code that is analog and mimetic. We turn next to a discussion of the advantages that a language system with both a segmented/combinatorial and a mimetic code can bring to human communication.

SOME ADVANTAGES OF HAVING A MIMETIC CODE ALONG WITH A SEGMENTED/COMBINATORIAL CODE

Corballis (1989, p. 500) describes the benefits of a generative system based on categorical elements for human language and thought: 'Generativity is a powerful heuristic, for it allows us to describe, represent, or construct an enormous variety of composites, given only a relatively small number of building blocks and rules of construction.' At the time same time, however, Corballis (1989) notes the limitations of generativity. A generative system becomes unworkable if the number of units in the system is too large; moreover, the relatively small number of units required to make the system manageable also makes it difficult to capture subtle distinctions. These distinctions may be more easily expressed via an analog representational format. For example, a verbal description of the shape of the east coast of the United States is likely, not only to be very cumbersome, but also to leave out important information about the coastline (Huttenlocher 1973, 1976). It is just this information that can easily be captured in a mimetic gesture tracing the outline of the coast.

Categorization, which is at the heart of a segmented and combinatorial code, is the grouping of elements as alike and, as such, necessarily creates its complement---uncategorized elements. In general, there are two kinds of uncategorized elements: those that are distinguishable but are categorized in the same way (i.e., elements that fall within a codified category), and those that are not categorized at all (i.e., elements that fall outside of the codified categories for that language). A mimetic code is able, at least in principle, to capture both types of uncategorized elements. In fact, we find that the mimetic characteristics of gesture not only allow it to capture both, but allow it to do so alongside the segmented and combinatorial code provided by speech.

Capturing information within a codified category

As an example of how gesture can convey distinctions not captured by a linguistic category, a child participating in a conservation task (Church and Goldin-Meadow 1986) explained his belief

that an amount of water had changed after it had been poured from a tall, thin glass into a short, wide dish as follows: He said 'it's different because this one's fatter than this one' while varying the distance between his thumb and fingers to indicate, first, the exact width of the dish and, next, the exact width of the glass. The child's speech arbitrarily categorized the two containers into more fat and less fat, while his gestures indicated not only the relative widths of the containers but also the precise value of each container on the width dimension. Although it would have been possible to explicitly describe the width of each container in English (e.g., 'it's different because this one's 5 inches and fatter than this one which is 3 inches'), such a statement detracts from the focus of the sentence (which is the relative widths of the objects, not their exact values). Moreover, the child is not likely to be able to describe the widths of the containers in inches, although he clearly knows, at some level, precisely how much the containers vary in width.

Conveying information not found in any codified category

Mimetic encoding fills in where categorization reaches its limit. Exactly where this limit is and thus where the filling in occurs varies across languages. The way manner is conveyed in Spanish and English is a good example. Manner is *how* a motion is performed. Contrasts in manner are illustrated by English verbs of locomotion: 'walk,' 'run,' 'stroll,' 'limp,' 'hop,' 'sidle,' etc., all involve motion with the feet but differ in how the motion is done. Slobin (1996) notes that while manner can be encoded in Spanish, Spanish speakers rarely include manner in their motion descriptions. Indeed, McNeill and Duncan (1997) have confirmed that, when asked to describe a cartoon, Spanish speakers infrequently convey manner information in their speech. However, they do express manner in their gestures. For example, one Spanish speaker described Sylvester's ascent up a drainpipe without mentioning in speech how the cat accomplished the feat; at the same time, the speaker's hands continuously displayed Sylvester's clambering style of ascent throughout the spoken description. Thus, Spanish speakers can use gesture to expand the resources of their codified system.

Downplaying unwanted information within a codified category

Gesture can also be used to focus attention away from an element that is obligatorily encoded in speech. In contrast to Spanish, the codified categories English offers a speaker come equipped with manner as an obligatory semantic component (cf. the locomotion verbs cited above). The spoken code therefore does not easily allow a speaker to omit manner information. Gesture, however, can be used to focus attention away from the manner of motion---it can be used to trim manner when it is not part of the speaker's communicative focus. For example, an English speaker describing Sylvester's descent down the drainpipe said 'and he rolls down the drain spout' while plunging his hand straight down. The speaker's words convey manner information (rolling), but his gesture downplays that semantic component by focusing exclusively on the path of motion. English speakers can, of course, use gesture to highlight manner, as in the following example. Once again to describe Sylvester's trajectory, the speaker said 'but it rolls him out' while arcing his two hands to the left with his fingers wiggling. The speaker thus reinforced the

rolling manner conveyed in speech with a gesture mimetically displaying the motion that propelled Sylvester along the path. In these ways, the mimetic properties of gesture allow it to modulate the verb semantics of the spoken code to fit the speaker's immediate communicative needs.

Creating impromptu coherence across sentences

A mimetic code is, by definition, transparently related to its referents and, as a result, can be fashioned on the spur of the moment and still be understood. Links that might otherwise be difficult to make within the bounds of a codified system can therefore be made via the mimetic properties of gesture. For example, discourse context is not richly categorized by the spoken language system. Gesture, however, offers resources to fill the gap. A 'catchment' displays the cohesive relations that bind a gesture to the larger discourse. Catchments are conveyed mimetically by gesture features (e.g., handshape, locus in space, hand choice) that recur across successive gestures (Kendon 1972). Such recurrence links a given gesture to earlier (anaphoric) and later (cataphoric) gestures that share the same feature(s). As a consequence of this recurrence, gesture has the ability to create larger discourse units.

For example, in a series of four gestures, a speaker consistently used her left hand to represent Sylvester who was climbing up a pipe, and her right hand to represent a bowling ball that was rolling down the pipe. In addition, across the same four gestures, the speaker consistently placed the hand that represented Sylvester below the hand representing the bowling ball. Finally, and again across the four gestures, the speaker made her hands play symmetrical roles when describing events before the bowling ball was dropped, and asymmetrical roles when describing events after the drop. Thus, hand choice (left *vs*. right), spatial configuration (up *vs*. down), and the relationship between the hands (symmetrical *vs*. asymmetrical) functioned together to unify these four gestures and the utterances they accompanied into a cohesive discourse unit---a unit that was not displayed in the accompanying speech.

In sum, a mimetic code offers its own advantages. Having a mimetic code provides a way for speakers to combine into one representation the specific advantages of both the categorical and the mimetic modes of representation. Equipped with a mimetic code, speakers can express information that is indistinguishable within a codified category (without diluting the category itself), they can express information that is not captured in any codified category, they can downplay unwanted information that comes along obligatorily within a codified category, and they can create larger units, particularly discourse units, that are not easily captured within the codified system. Mimetic encoding thus offers speakers a way of enhancing their categorical spoken system, a system which has the advantages of generativity but the disadvantages of stodginess. Mimetic encoding offers speakers flexibility so that their communicative needs of the moment can be met---and gesture provides an ideal medium with which to accomplish mimetic encoding.

GESTURE AND LINGUISTIC EVOLUTION

Given that there are advantages to having a mimetic code as well as a segmented and combinatorial code, we can now understand why speech has taken over language across the globe. While both the manual modality and the oral modality are able to assume a segmented and combinatorial representational format, only the manual modality is well suited to a mimetic representational format. Mimetic representation thus falls to the manual modality, leaving segmented and combinatorial representation to the oral modality. Having segmented structure in the oral modality as we currently do leaves the manual modality free to co-occur with speech and to capture the mimetic aspects of communication along with speech. Thus, our current arrangement allows us to retain, along with a segmented representation and *in a single stream of communication*, the imagistic aspects of the mimetic that are so vital to human communication. The alternative arrangement---in which the manual modality would assume the segmented code and the oral modality would serve the mimetic functions---has the disadvantage of forcing the oral modality to be unnaturally imagistic in form (although see Haiman 1985 for evidence that the oral modality does exhibit some iconic properties). If our hypothesis is correct, speech became the predominant medium of human language not because it is so well suited to the segmented and combinatorial requirements of symbolic communication (the manual modality is equally suited to the job), but rather because it is not particularly good at capturing the mimetic components of human communication (a task at which the manual modality excels).

This speculation about the importance of maintaining a vehicle for mimetic representation along with speech raises an interesting question about sign language. In sign, it is the manual modality that assumes the segmented and combinatorial form essential to human language. Can the manual modality at the same time also be used for holistic and mimetic expression? In other words, do signers gesture along with their signs and, if not, how is the mimetic function filled? One possibility is that the mouth movements associated with particular sounds might assume the mimetic function for signers. Although such movements have frequently been observed in fluent signers (cf. Padden 1990), as far as we know, no work has been conducted to investigate whether these behaviors (or any others, for that matter) serve for sign the mimetic function that gesture serves for speech.

Mimetic encoding as we have used the term is reminiscent of, but narrower than, Merlin Donald's notion of mimesis. Donald (1991, p. 16; *see also* Donald, this volume) defines mimetic representation as the ability to use the body to mime, or re-enact, events. This function is part of a wider mimetic culture, underlying ritual, play, acting, and sport. Donald reconstructs the emergence of human cognitive and cultural systems as an evolution in which mimesis is a key stepping-stone through a series of four stages: the *episodic* (shared by all apes), the *mimetic* (the system of *Homo erectus*, the creator of sophisticated stone tool industries), the *mythic* (the evolution of language and of narrative modes of thinking of our own *Homo sapiens* culture), and the now-dawning *theoretic* (moving toward effortless propositional representations). Each stage is characterized by its own memory stores (including external memories), representation modes, technical potential, and limits. Donald speculates, as did Condillac in the eighteenth century, that a mimetic stage was a pre-speech style of cognitive representation. It supported a language of some kind in which the main principle of expression was mimesis. The form of such a language,

both Condillac and Donald argue, would naturally have been gestural (see also Armstrong, Stokoe, and Wilcox 1995; Corballis 1992; Hewes 1973; Kendon 1974; and Wescott 1974, who express similar views). Gradually, the mimetic stage and its language of gestures gave way to the mythic stage, and its language of sounds. Out of this evolutionary process came narrative discourse and oral language as we know it. Donald emphasizes that the new stage absorbed the old, the representational forms of the mimetic system being reformatted into mythic ones based on spoken linguistic models. The outcome is the old mimetic system encapsulated within the new mythic system. The crucial point in Donald's model is that gesture and speech evolved separately and successively.

We agree that mimesis is an important mode of representation, one that could well have been a way-station along the route to language as we know it. However, we disagree that the mythic system embodied in speech replaced the mimetic system found in gesture, or that the two representational systems continued to develop independently of one another. Rather, we suggest that the spoken system we have today evolved hand-in-hand, as it were, with gesture. Indeed, there is no evidence in modern-day gesture to suggest that speech and gesture evolved separately. Donald himself finds evidence for the independence of gesture and speech in emblems (Ekman and Friesen 1969), gestures that have conventional paraphrases or names and that can be used, often without speech, as if they were spoken words (e.g., the 'okay' sign can be used without speech in American culture to mean 'things are fine'). While emblems may well be independent of speech (but see Kendon 1995 for arguments against this claim), they are a very small part of a speaker's repertoire of gestures. Iconic and metaphoric gestures of the sort we have focused on here constitute the bulk of the gestures speakers spontaneously produce---and these gestures are not at all independent of speech. Donald (1991, p. 223) notes that 'in some situations gestures may actually override a linguistic message' and suggests that the ability of gesture and speech to, at times, go their separate ways is evidence that the two constitute separate systems (*see also* Donald, 1993, p. 744). However, as described earlier, even when gesture and speech convey different messages, the two modalities adhere to the principles of gesture-speech integration described by McNeill (1992)---that is, they form a single unified system, with gesture assuming the mimetic functions and speech the mythic.

In sum, we have provided evidence suggesting that the manual modality is as good as the oral modality at segmented and combinatorial encoding---the manual modality assumes such a format whenever it is required to take on the full burden of communication. Why then do all hearing cultures place language in the oral modality? We have suggested that it is because the segmented and combinatorial format is not sufficient to capture all the essential components of human communication---a mimetic and imagistic format is needed to integrate discourse and get beyond the limits of categorical thinking that underlie the segmented code. Having a mimetic code alongside a segmented and combinatorial code creates a composite communication system that not only is generative but also is responsive to the context-specific communicative needs of human speakers. Such an integrated system retains the virtues of categorical generativity, while avoiding the unworkability of an over-refined linguistic code. A mimetic code is therefore needed to realize the advantages of the categorical code.

It is, moreover, the manual modality---and not the oral modality---that is particularly well suited to mimetic representation. As a result, the manual modality takes over the mimetic aspects of human communication, leaving the analytic aspects by default to speech. Under this scenario,

the mimetic and linguistic sides of language evolved together, producing a single system. Our current-day arrangement therefore allows the simultaneous production of both formats, making possible the flexibility and scope of human language.

REFERENCES

Alibali, M. W., Flevares, L., and Goldin-Meadow, S. (1996). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? Journal of Educational Psychology (in press).

Alibali, M. W. and Goldin-Meadow, S. (1993a). Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind. Cognitive Psychology, **25**, 468-523.

Alibali, M. W. and Goldin-Meadow, S. (1993b). Modelling learning using evidence from speech and gesture. Proceedings of the 15th Annual Conference of the Cognitive Science Society, **1993**, 203-8.

Armstrong, D. F., Stokoe, W. C., and Wilcox, S. E. (1995). Gesture and the nature of language. Cambridge University Press, New York.

Bellugi, U. and Studdert-Kennedy, M. (ed.) (1980). Signed and spoken language: Biological constraints on linguistic form. Verlag Chemie, Deerfield Beach, Fla.

Butcher, C. and Goldin-Meadow, S. (1997). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In Language and gesture: Window into thought and action, (ed. D. McNeill). The New Lexington Press, San Francisco, Ca. (in press.)

Caselli, M. C. (1983). Communication to language: Deaf children's and hearing children's development compared. Sign Language Studies, **39**, 113-44.

Church, R. B. and Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. Cognition, **23**, 43-71.

Conrad, R. (1979). The deaf child. Harper and Row, New York.

Corballis, M. C. (1989). Laterality and human evolution. Psychological Review, **96**, 492-505.

Corballis, M. C. (1992). On the evolution of language and generativity. Cognition, **44**, 197-226.

Coulter, G. R. (1990). Emphatic stress in ASL. In Theoretical issues in sign language research, Vol. 1. Linguistics (ed. S. D. Fisher and P. Siple), pp. 109-25. University of Chicago Press, Chicago.

Crowder, E. M. and Newman, D. (1993). Telling what they know: The role of gesture and language in children's science explanations. Pragmatics and Cognition, **1**, 341-76.

Donald, M. (1991). Origins of the modern mind: Three stages in the evolution of culture and cognition. Harvard University Press, Cambridge, MA.

Donald, M. (1993). Précis of 'Origins of the modern mind: Three stages in the evolution of culture and cognition'. Behavioral and Brain Sciences, **16**, 737-91.

Ekman, P. and Friesen, W. (1969). The repertoire of nonverbal behavioral categories. Semiotica, **1**, 49-98.

Evans, M. A. and Rubin, K. H. (1979). Hand gestures as a communicative mode in school-aged children. The Journal of Genetic Psychology, **135**, 189-96.

Fant, L. J. (1972). Ameslan: An introduction to American Sign Language. National Association of the Deaf, Silver Springs, MD..

Feldman, H., Goldin-Meadow, S., and Gleitman, L. (1978). Beyond Herodotus: The creation of language by linguistically deprived deaf children. In Action, symbol, and gesture: The emergence of language, (ed. A. Lock), (pp. 351-414). Academic Press, New York.

Feyereisen, P. and de Lannoy, J.-D. (1991). Gestures and speech: Psychological investigations. Cambridge University Press.

Fischer, S. and Gough, B. (1978). Verbs in American Sign Language. Sign Language Studies, **18**, 17-48.

Friedman, L.A. (1976). The manifestation of subject, object, and topic in American Sign Language. In Subject and topic, (ed. C. Li), (pp. 125-148). Academic Press, New York.

Goldin-Meadow, S. and Feldman, H. (1977). The development of language-like communication without a language model. Science, **197**, 401-3.

Goldin-Meadow, S. and Mylander, C. (1983). Gestural communication in deaf children: The non-effects of parental input on language development. Science, **221**, 372-4.

Goldin-Meadow, S. and Mylander, C. (1984). Gestural communication in deaf children: The effects and non-effects of parental input on early language development. Monographs of the Society for Research in Child Development, **49**, 1-121.

Goldin-Meadow, S. and Mylander, C. (1990). Beyond the input given: The child's role in the acquisition of language. Language, **66**, 323-55.

Goldin-Meadow, S. and Sandhofer, C. (1996). The role of gesture in communicating a child's thoughts to an adult onlooker. (Under review.)

Goldin-Meadow, S., Wein, D., and Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. Cognition and Instruction, **9**, 201-19.

Goldin-Meadow, S., Alibali, M. W., and Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. Psychological Review, **100**, 279-97.

Goldin-Meadow, S., Butcher, C., Mylander, C., and Dodge, M. (1994). Nouns and verbs in a self-styled gesture system: What's in a name? Cognitive Psychology, **27**, 259-319.

Goldin-Meadow, S., Mylander, C., and Butcher, C. (1995). The resilience of combinatorial structure at the word level: Morphology in self-styled gesture systems. Cognition, **56**, 195-262.

Goldin-Meadow, S., McNeill, D., and Singleton, J. (1996). Silence is liberating: Removing the handcuffs on grammatical expression in the manual modality. Psychological Review, **103**, 34-55.

Haiman, J. (1985). Iconicity in syntax. John Benjamins, Amsterdam.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. Current Anthropology, **14**, 5-24.

Hoffmeister, R. and Wilbur, R. (1980). Developmental: The acquisition of sign language. In Recent perspectives on American Sign Language, (ed. H. Lane and F. Grosjean). Erlbaum, Hillsdale, NJ.

Huttenlocher, J. (1973). Language and thought. In Communication, language and meaning: Psychological perspectives, (ed. G. A. Miller), pp. 172-84. Basic Books, New York.

Huttenlocher, J. (1976). Language and intelligence. In The nature of intelligence (ed. L. B. Resnick), pp. 261-81. Erlbaum, Hillsdale, NJ.

Iverson, J. and Goldin-Meadow, S. (1996). What's communication got to do with it: Gesture in blind children. Developmental Psychology, **33**, 453-467.

Jancovic, M. A. Devoe, S., and Wiener, M. (1975). Age-related changes in hand and arm movements as nonverbal communication: Some conceptualizations and an empirical exploration. Child Development, **46**, 922-8.

Kantor, R. (1982). Communicative interaction: Mother modification and child acquisition of American Sign Language. Sign Language Studies, **36**, 233-82.

Kelly, S. and Church, R. B. (1996). Can children detect conceptual information conveyed through other children's non-verbal behaviors? Cognition and Instruction (in press).

Kendon, A. (1972). Some relationships between body motion and speech. In Studies in dyadic communication (ed. A. Siegman and B. Pope), pp. 177-210. Pergamon Press, New York.

Kendon, A. (1974). Kinesics, speech, and language. In Language origins, (ed. R. W. Wescott), pp. 181-4. Linstok Press, Silver Spring, MD.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In Relationship of the verbal and nonverbal communication (ed. M. R. Key), pp. 207-28. Mouton, The Hague.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversation. Journal of Pragmatics, **23**, 247-79.

Klima, E. and Bellugi, U. (1979). The signs of language. Harvard University Press.

Krauss, R. M., Morrell-Samuels, P., and Colasante, C. (1991). Do conversational hand gestures communicate? Journal of Personality and Social Psychology, **61**, 743-754.

Lane, H., Boyes-Braem, P., and Bellugi, U. (1976). Preliminaries to a distinctive feature analysis of handshapes in American Sign Language. Cognitive Psychology, **8**, 263-89.

Lane, H. & Grosjean, F. (1980). Recent perspectives on American Sign Language. Erlbaum, Hillsdale, NJ.

Lenneberg, E. H. (1964). Capacity for language acquisition. In The structure of language: Readings in the philosophy of language (ed. J. A. Fodor and J. J. Katz), pp. 579-603. Prentice-Hall, New Jersey.

Liddell, S. (1980). American Sign Language syntax. Mouton, The Hague.

Liddell, S. and Johnson, R. (1986). American Sign Language compound formation processes, lexicalization, and phonological remnants. Natural Language and Linguistic Theory, **4**, 445-513.

Lillo-Martin, D. (1986). Two kinds of null arguments in American Sign Language. Natural Language and Linguistic Theory, **4**, 415-44.

Mayberry, R. I. (1992). The cognitive development of deaf children: Recent insights. In Child Neuropsychology, Vol. 7. Handbook of Neuropsychology (ed. S. Segalowitz and I. Rapin), pp. 51-68. Elsevier, Amsterdam.

McNeill, D. (1985). So you think gestures are nonverbal? Psychological Review, **92**, 350-71.

McNeill, D. (1987). Psycholinguistics. Harper and Row, New York

McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.

McNeill, D. and Duncan, S. D. (1997). Growth points in thinking-for-speaking. In Language and gesture: Window into thought and action (ed. D. McNeill). The New Lexington Press, San Francisco, Ca. (in press.)

Meadow, K. (1968). Early manual communication in relation to the deaf child's intellectual, social, and communicative functioning. American Annals of the Deaf, **113**, 29-41.

Moores, D. F. (1974). Nonvocal systems of verbal behavior. In Language perspectives: Acquisition, retardation, and intervention (ed. R. L. Schiefelbusch and L. L. Lloyd), (pp. 377-417). University Park Press, Baltimore, MD.

Newport, E. L. (1981). Constraints on structure: Evidence from American Sign Language and language learning. In Minnesota Symposium on Child Psychology, Vol. 1 (ed. W. A. Collins), pp. 93-124. Erlbaum, Hillsdale, NJ.

Newport, E. L. and Meier, R. P. 1985). The acquisition of American Sign Language. In The cross-linguistic study of language acquisition, Vol. 1. The data (ed. D. J. Slobin), (pp. 881-938). Erlbaum, Hillsdale, NJ.

Padden, C. (1983). Interaction of morphology and syntax in American Sign Language. Unpublished doctoral dissertation, University of California at San Diego.

Padden, C. and Perlmutter, D. (1987). American Sign Language and the architecture of phonological theory. Natural Language and Linguistic Theory, **5**, 335-75.

Padden, C. (1990). Rethinking fingerspelling. Signpost, October 2-4. International Linguistics Association, University of Durham.

Petitto, L. A. (1992). Modularity and constraints in early lexical acquisition: Evidence from children's early language and gesture. In Minnesota Symposium on Child Psychology, Vol. 25 (ed. M. Gunnar), pp. 25-58. Erlbaum, Hillsdale, NJ.

Perry, M., Church, R. B., and Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. Cognitive Development, **3**, 359-400.

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal variables. European Journal of Social Psychology, **12**, 113-29.

Sandler, W. (1986). The spreading hand autosegment of American Sign Language. Sign Language Studies, **50**, 1-28.

Schick, B. S. (1990). The effects of morphological complexity on phonological simplification in ASL. Sign Language Studies, **66**, 25-41.

Singleton, J. L., Morford, J. P., and Goldin-Meadow, S. (1993). Once is not enough: Standards of well-formedness in manual communication created over three different timespans. Language, **69**, 683-715.

Slobin, D. I. (1996). Two ways to travel: Verbs of motion in English and Spanish. In Grammatical constructions: Their form and meaning (ed. M. Shibatani and S. A. Thompson), (pp 195-219). Oxford University Press.

Stokoe, W. C. (1960). Sign language structure: An outline of the visual communications systems. Studies in linguistics, Occasional papers No. 8.

Supalla, T. (1982). Structure and acquisition of verbs of motion and location in American Sign Language. Unpublished doctoral dissertation, University of California at San Diego.

Supalla, T. and Newport, E. (1978). How many seats in a chair? The derivation of nouns and verbs in American Sign Language. In Understanding language through sign language research (ed. P. Siple), (pp. 91-132). Academic Press, New York.

Tervoort, B. T. (1961). Esoteric symbolism in the communication behavior of young deaf children. American Annals of the Deaf, **106**, 436-80.

Wescott, R. W. (1974). The origin of speech. In Language origins (ed. R. W. Wescott), pp. 103-23. Linstok Press, Silver Spring,

Wilbur, R. (1986). Interaction of linguistic theory and sign language research. In The real world linguist: Linguistic applications for the 1980s (ed. P. C. Bjarkman and V. Raskin), pp. 166-182. Ablex, Norwood, NJ.

Wundt, W. (1900/1973). The language of gestures. Mouton, The Hague.

---

[1] This work was supported by Grant No. RO1 DC00491 from the National Institute on Deafness and other Communication Disorders and Grant No. R01 HD31185 from the National Institute of Child Health and Human Development to S. Goldin-Meadow; and Grant No. RO1 DC10561 from the National Institute on Deafness and other Communication Disorders and grants from the Spencer Foundation to D. McNeill.