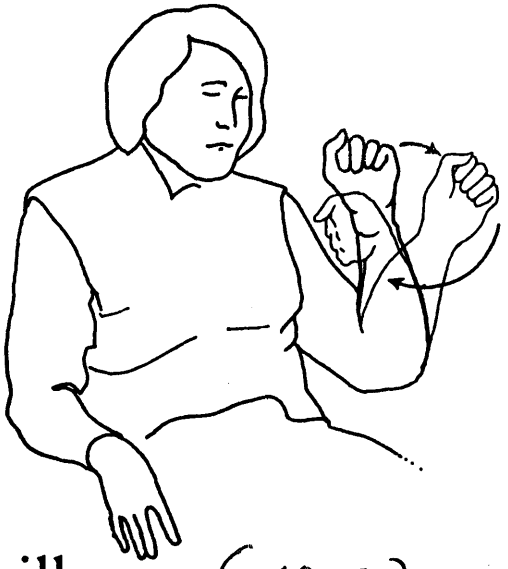
HAND AND MIND

What Gestures Reveal about Thought



David McNeill

(1992)

~~- S. 11 - 72~~ S. 11. - 35

~~- S. 365 - 394~~

The University of Chicago Press
Chicago and London

Contents

1 *Images, Inside and Out*

If language was given to men to conceal their thoughts, then gesture's purpose was to disclose them.

—John Napier, *Hands*

This inscription provides an appropriate place to start the book that follows. Napier's book (1980) was about the evolution, mechanics, and functioning of the human hand. He ended his account with the above comment on the hands as symbolic instruments. This book starts at the same place and asks: how *are* human thoughts disclosed in gestures?

When people talk they can be seen making spontaneous movements called gestures. These are usually movements of the arms and hands and are closely synchronized with the flow of speech. The implication of gestures that interests me is the possibility of embracing in one theoretical system two forms of expression, speech and action. Gestures and speech occur in very close temporal synchrony and often have identical meanings. Yet they express these meanings in completely different ways. Comparing speech to gesture produces an effect on our understanding of language and thought something like the effect of triangulation in vision. Many details, previously hidden, spring out in a new dimension. Rather than causing us to slice a person analytically into semi-isolated modules, taking gesture into account encourages us to see something like the entire person as a theoretical entity—his thinking, speaking, willing, feeling, and acting, as a unit.

My aim is to provide a conceptual framework that includes both gesture and language. This broader framework will show how gestures and speech are linked, and how they are different. At a minimum, the framework should explain how speech, which is linear through time, is related to the type of thinking that we see exhibited in the simultaneous gesture, thinking that is instantaneous, imagistic and global—analog rather than digital.

Gestures exhibit images that cannot always be expressed in speech, as well as images the speaker thinks are concealed. Speech and gesture must cooperate to express the person's meaning. A conception of language and gesture as a single integrated system is sharply different from the notion of a "body language"—a communication process utilizing signals made up of body movements, which is regarded by its believers as separate from and beyond normal language. This concept is the product of an excessively narrow analysis, just as is the traditional linguistic notion of a

spoken language as exclusively comprising a system of speech sounds plus a grammar.

The topic of this book is, specifically, gestures that exhibit images. With these kinds of gesture people unwittingly display their inner thoughts and ways of understanding events of the world. These gestures are the person's memories and thoughts rendered visible. Gestures are like thoughts themselves. They belong, not to the outside world, but to the inside one of memory, thought, and mental images. Gesture images are complex, intricately interconnected, and not at all like photographs. Gestures open up a wholly new way of regarding thought processes, language, and the interactions of people.

I want now to give a sketch of the major types of gesture. I will illustrate these types and give many more details in later chapters.

Types of Gesture

Iconics

Some gestures are "iconic" and bear a close formal relationship to the semantic content of speech. For example, when describing a scene from a comic book story in which a character bends a tree back to the ground, a speaker appeared to grip something and pull it back (see fig. 1.1):

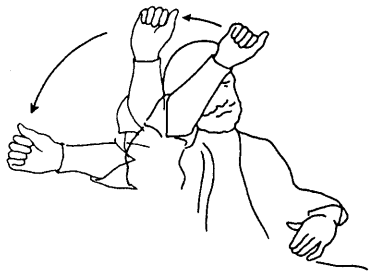


Figure 1.1. Illustration of an iconic gesture with "and he [bends it way back]." The gesture exhibits the same act referred to in speech.

(1.1) and he [bends it way back][1]

Iconic: hand appears to grip something and pull it from the upper front space back and down near to the shoulder.

1. In citing gesture examples I will use the following conventions: speech will be underscored; the extent of the meaningful part of the gesture—the stroke phase—will be shown by enclosing the concurrent segments of speech in square brackets [ ]; when the gesture is held motionlessly there will be dots (. . .); the gesture itself will be described in italics. If more than one stroke takes place, they will be numbered. Silent hesitations will be shown with slanted lines (/), and filled hesitations with a rendition of the sound. On occasion, I will also show the gesture preparation and retraction phases. In these cases, to avoid a clutter of brackets, I will show the onset of the preparation with a left bracket ([), the stroke in **boldface,** and the end of the retraction with a right bracket (]).

The speaker (not a habitual comic book reader) had been given a comic book as part of an experiment on narrative (see Marslen-Wilson, Levy, and Tyler, 1982). The gesture reveals not only the speaker's memory image but also the particular point of view that he had taken toward it. The speaker had the choice of playing the part of the agent or the tree. He was "seeing" the event from the viewpoint of the agent performing the act—otherwise his hand would not have taken the form of a grip—rather than the viewpoint of the tree undergoing the act. In the latter case, we would expect a gesture in which the arm moves back but without the grip.

The example illustrates the close connection that exists between speech and gesture. It shows how what is depicted through gesture should be incorporated into a complete picture of a person's thought processes. The gesture movement—the "stroke"—coincided with the part of the utterance that presented the same meaning. Semantically, the sentence described bending something back while the gesture concurrently exhibited the same bending back image. Moreover, the image was from the point of view of the actor, and when we look at the active form of sentence, "he bends it way back," we see that it also implies the point of view of the actor. The passive, "it got bent way back," for example, would be more appropriate for the viewpoint of the tree. Thus, both semantically and pragmatically, in terms of focus, the gesture and utterance were parallel expressions of meaning.

Along with this kind of coexpressiveness, there is also complementarity. Speech and gesture refer to the same event and are partially overlapping, but the pictures they present are different. Jointly, speech and gesture give a more complete insight into the speaker's thinking. In the following example speech conveys the ideas of pursuit and recurrence while gesture conveys the weapon used (an umbrella); both speech and gesture refer to the same event, but each presents a somewhat different aspect of it (see fig. 1.2):

Figure 1.2. Illustration of speech-gesture complementation with "she [chases him out again]." The gesture reveals the weapon while speech conveys the action (chasing) and the idea of recurrence.

(1.2) and she [chases him out again]

Iconic: hand appears to swing an object through the air.

If we were to look only at the gesture or the speech, we would have an incomplete picture of the speaker's memory and mental representation of the scene. It is only through a joint consideration of both gesture and speech that we see all the elements: the type of action, its recurrence, the weapon, and how it was used.

Metaphorics

Other gestures are "metaphoric." These are like iconic gestures in that they are pictorial, but the pictorial content presents an abstract idea rather than a concrete object or event. The gesture presents an image of the invisible—an image of an abstraction. The gesture depicts a concrete metaphor for a concept, a visual and kinesic image that we feel is, in some fashion, similar to the concept. For example, in the following, a speaker is announcing that what he has just seen and is about to recount to the listener is a cartoon (see fig. 1.3):

Figure 1.3. Illustration of a metaphoric gesture with "it [was a Sylves]ter and Tweety cartoon." The gesture is an instance of the conduit metaphor: the idea of a genre presented as a bounded container supported by the hands.

(1.3) it [was a Sylves]ter and Tweety cartoon

Metaphoric: Hands rise up and offer listener an "object."

A particular cartoon event is concrete, but the speaker here is not referring to a particular event: he is referring to the genre of the cartoon. This concept is abstract. Yet he makes it concrete in the form of an image of a bounded object supported in the hands and presented to the listener. The gesture creates and displays this object and places it into an act of offering. This is the metaphor: the concept of a genre of a certain kind (the Topic) is presented as a bounded, supportable, spatially localizable

physical object (the Vehicle). Such metaphors can be documented in speech forms as well. In speech we say, for instance, "hollow words" or "a deep book"—implying that a word is a container, or a book has a vertical dimension (Reddy 1979; Lakoff and Johnson 1980). We also speak of the "presentation" of an idea or argument—implying that communication is over a path or conduit. The metaphor in which language, meaning, knowledge, genre, works of art, etc., are presented as a physical container into which substance is put and the whole is moved along a conduit has been called the *conduit* metaphor; we will see many examples of gestural conduit metaphors in this book. The conduit image of abstract ideas as physical containers is, for speakers brought up in the tradition of Western culture, a major source of metaphoric images. However, this image does not appear with speakers brought up in other, non-Western traditions (Chinese, for example).

Beats

A third type of gesture we term the "beat." Beats are so named because they look like beating musical time. Others have termed this gesture the "baton" (Efron 1941; Ekman and Friesen 1969)—naming it after the instrument rather than the function. The hand moves along with the rhythmical pulsation of speech (although the synchrony is not absolutely perfect; see McClave, 1991). Unlike iconics and metaphorics, beats tend to have the same form regardless of the content (McNeill and Levy 1982). The typical beat is a simple flick of the hand or fingers up and down, or back and forth; the movement is short and quick and the space may be the periphery of the gesture space (the lap, an armrest of the chair, etc.). The critical thing that distinguishes the beat from other types of gesture is that it has just two movement phases—in/out, up/down, etc. Iconics and metaphorics typically have three phases—preparation, stroke, and retraction. Of all gestures, beats are the most insignificant looking. But appearances are deceptive. Beats reveal the speaker's conception of the narrative discourse as a whole. The semiotic value of a beat lies in the fact that it indexes the word or phrase it accompanies as being significant, not for its own semantic content, but for its discourse-pragmatic content. Examples are marking the introduction of new characters, summarizing the action, introducing new themes, etc. Thus beats mark information that does not advance the plot line but provides the structure within which the plot line unfolds. With beats, events on the meta-level of the discourse can be inserted directly into the narrative, signaling that whatever refers in speech to the event departs from the narrated chain of events. Such departures can be brief and can be over in the

confines of a single word. An example is the following, a beat that accompanied a reference to the theme of an episode. The spoken utterance does not refer to a particular incident but characterizes a class of incidents, and the beat marked the word ("whenever") that signaled this reference to the discourse as a whole rather than a specific event (see fig. 1.4):

Figure 1.4. Illustrating a beat with a summing up statement, "when[ever she]." The beat coincides with the specific linguistic segment that does the summing up.

(1.4) when[ever she] looks at him he tries to make monkey noises

Beat: hand rises short way up from lap and drops back down.

Cohesives

Another kind of discourse gesture I call "cohesive" because it serves to tie together thematically related but temporally separated parts of the discourse. This function in the case of speech is called the cohesive function (Halliday and Hasan 1976). While beats highlight discontinuities in the temporal sequence, cohesives emphasize continuities. Cohesive gestures are quite eclectic about their form. They can consist of iconic, metaphoric, or pointing gestures; they can even consist of beats. Politicians, in fact, are great demonstrators of cohesive beats. Political speeches are accompanied by an incessant beat presence. The meaning of all those beats is, I think, cohesion on the meta-level. What the politician is in effect saying is: Here is a series of points that I am making, and the crucial thing about them is that each belongs to a consistent platform (hence the cohesion of beats). Certainly every politician holds that his views on the issues are individually significant while adding up to a consistent platform. The beat is accordingly the politician's gesture par excellence.

Gestural cohesion depends on repeating the same gesture form, movement, or locus in the gesture space: the repetition is what signals the continuity. The repeated gesture shows, in the most direct way, the recurrence or continuation of a theme. An example with an iconic gesture is the following, in which a speaker describing one of the cartoon episodes first made a crisscross gesture for intersecting overhead wires, interrupted herself to make a back-and-forth movement to represent a trolley car, and then went back to the crisscross gesture. The interruption came about because the speaker realized that she had better first explain how trolleys work (no longer so obvious in some places) and broke off her narrative to do this; the back-and-forth movement was part of the explanation. Such a statement ("you know the trolley") is clearly not part of the narrative story line, so a connection back to the main theme had to be arranged after the explanation had been given, and this was accomplished by the second crisscross iconic (see fig. 1.5):

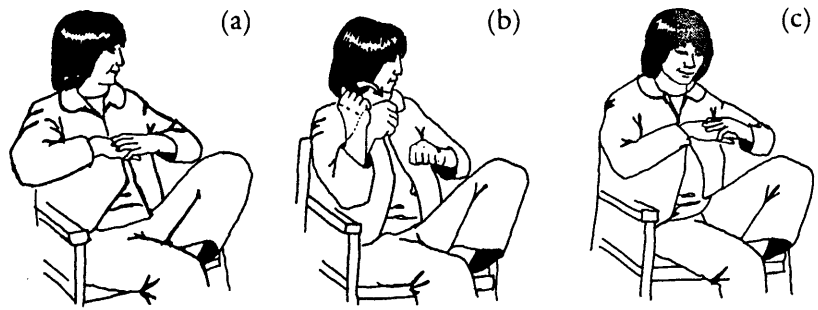


Figure 1.5. Illustrating cohesion with iconic gestures through the utilization of form. Panel (a) appeared with "[the network of wires that hooks up the cable cars]," (b) with "[you know the trolley system]," and (c) with "[right and there's a whole network of these wires]." The gesture at (c) replays the one at (a) and shows where the story resumes after the interruption at (b).

(1.5) [the network of wires that hooks up the cable cars . . . um]

Iconic: both hands, palms facing down, come together at the tips and form a crisscross.

(1.6) [you know the trolley system]

Iconic: right hand moves back and forth at side of head.

Listener: *oh across* [also gestures]

(1.7) [right and there's a whole network of these wires]

Iconic: both hands resume crisscross.

The same gesture thus occurred twice with the effect of informing the hearer where to go to get back to the story line after the interruption. The second crisscross was cohesive in that it tied together two parts of the

narrative by showing, literally, where the old theme was located. Again there is speech and gesture coexpressiveness. The sentence with the second crisscross gesture also included a cohesive item, the pronoun "these." It presupposed an earlier reference to the wires in question and likewise had the effect of sending the hearer back to the interrupted theme (Halliday and Hasan 1976). Notice that the cohesive connection appeared first in the gestural channel.

Deictics

A final type of gesture important for narrative is the familiar pointing, or "deictic" gesture. Pointing has the obvious function of indicating objects and events in the concrete world, but it also plays a part even where there is nothing objectively present to point at (McNeill, Cassell, and Levy, to appear). Most pointing gestures in narratives and conversations are of this abstract kind. The following example is from a conversation between two previously unacquainted students (see fig. 1.6):

Figure 1.6. Illustration of an abstract pointing gesture with "[where did you] come from before?" The space being pointed at is not the space where the speaker and hearer currently find themselves, but an abstract space housing an introduced reference.

(1.8) [where did you] come from before?
Points to space between self and interlocutor.

The gesture is aimed not at an existing physical place where the interlocutor had been previously, but at an abstract *concept* of where he had been before. As we know from the earlier context of the conversation, the physical locus of this place was in a different city. Although the space may seem empty, it was full to the speaker. It was a palpable space in which a concept could be located as if it were a substance. Abstract pointing gestures imply a metaphorical picture of their own in which abstract ideas have a physical locus.

Gestures Don't Convey Meaning as Language

Gestures do convey meanings and their expressiveness is not necessarily inferior to that of language. If one knows how to "read" them the gesture can convey meaning no less than language, but the method used by the gesture for doing this is fundamentally different from that of language.

Language has the effect of segmenting and linearizing meaning. What might be an instantaneous thought is divided up and strung out through time. A single event, say, somebody sitting down on a chair, is analyzed into segments: the person, the chair, the movement, the direction, and so forth. These segments are organized into a hierarchically structured string of words (for example, the sentence above). The total effect is to present what had been a single instantaneous picture in the form of a string of segments. Segmentation and linearization to form a hierarchy are essential characteristics of all linguistic systems, including languages that are not spoken at all, such as American Sign Language (ASL)—the manual language of the deaf in North America (see chap. 2). Saussure ([1916] 1959) explained that the linear-segmented character of language is a property that arises because language is unidimensional while meanings are multidimensional. Language can only vary along the single dimension of time—phonemes, words, phrases, sentences, discourse: at all levels, language depends on variations along this one axis of time. This restriction forces language to break meaning complexes into segments and to reconstruct multidimensional meanings by combining the segments in time.

Gestures are different in every way. This is because they are themselves multidimensional and present meaning complexes without undergoing segmentation or linearization. Gestures are *global* and *synthetic* and *never hierarchical*. The following sections describe these properties of gestures that make them different from language.

Global-Synthetic

These terms refer to the relationship of parts to wholes in gestures. In language, parts (the words) are combined to create a whole (a sentence); the direction thus is from part to whole. In gestures, in contrast, the direction is from whole to part. The whole determines the meanings of the parts (thus it is "global"). In language, moreover, the relationship of words to meaning is analytic. Distinct meanings are attached to distinct words. In gestures, however, one gesture can combine many meanings (it is "synthetic"). An example that illustrates both the global and synthetic properties is the following typical iconic gesture (see fig. 1.7):

Figure 1.7. Illustrating the global-synthetic property of a gesture with "[and he's trying to run ahead of it]." The gesture has parts (trajectory, wiggling fingers), but the meanings of the parts depend on the meaning of the whole; the parts are not independently meaningful morphemes or words in a language.

(1.9) [and he's trying to run ahead of it]

Iconic: hand moves forward at chin level while fingers wiggle.

The gesture is a symbol in that it represents something other than itself—the hand is not a hand but a character, the movement is not a hand in motion but the character in motion, the space is not the physical space of the narrator but a narrative space, the wiggling fingers are not fingers but running feet. The gesture is thus a symbol, but the symbol is of a fundamentally different type from the symbols of speech.

This gesture-symbol is global in that the whole is not composed out of separately meaningful parts. Rather, the parts 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. It's

not that a gesture depicting someone running was composed out of separately meaningful parts: wiggling + motion, for instance.

The gesture also is synthetic. It combines different meaning elements. The segments of the utterance, "he + running + along the wire," were combined in the gesture into a single depiction of Sylvester-running-along-the-wire.

Noncombinatoric

Gestures also are noncombinatoric: two gestures produced together don't combine to form a larger, more complex gesture. There is no hierarchical structure of gestures made out of other gestures. This noncombinatoric property contrasts with the hierarchical structure of language. In sentences, lower constituents combine into higher constituents. With gestures, each symbol is a complete expression of meaning unto itself. Most of the time gestures are one to a clause but occasionally more than one gesture occurs within a single clause. Even then the several gestures don't combine into a more complex gesture. Each gesture depicts the content from a different angle, bringing out a different aspect or temporal phase, and each is a complete expression of meaning by itself. This situation is illustrated by the next example, taken from a film narrative (see fig. 1.8):

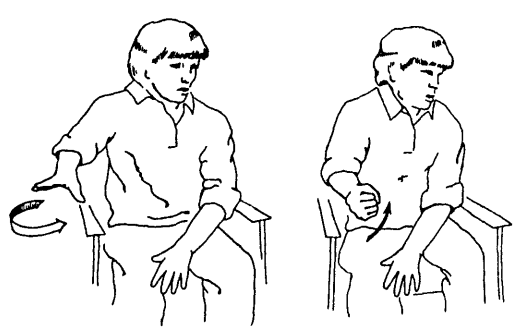


Figure 1.8. Illustrating two gestures in one clause with "[. . . and she . . .] [grabs] the knife." The gestures are two phases of an event, but do not combine into a higher level gesture.

(1.10) [. . . and she . . .][grabs] the knife
(1) (2)

(1) Hand gropes in a circle with the palm facing down and the fingers extended.

(2) Hand turns up and closes to a fist: gripping an "object."

The gestures are related but do not combine into a single higher gesture. The gestures, rather, present successive snapshots of the scene. The clause also describes this scene, but whereas the parts of the clause, "she," "grabs," and "the knife," combine to form the clause, groping plus grabbing do not combine to form a larger gesture. Far from combining, the two gestures maximized the contrast between them: the horizontal orientation and circling movement of (1) changed into the vertical orientation and closed fist of (2).

Other Nonlinguistic Properties

NO STANDARDS OF FORM. Linguistic systems impose standards of well-formedness to which all utterances must conform or be dismissed as "not English" or "not Japanese" or not whatever language the speaker is using. The standards are "the way we do it." Gestures have no equivalent to this implicit standard of form. Gestures by different individuals often are similar when the content of the gesture is similar, but this similarity is because of the content and not because individuals are conforming to standards for making the gesture. The gestures of different speakers can present the same meanings but do so in quite different forms. Moreover, the gestures of people speaking different languages are no more different than the gestures of different people speaking the same language. While their speech moves in different directions to meet linguistic standards, their gestures remain close together. This nonstandardized quality of gestures is important for theoretical purposes. Precisely because gestures are *not* obliged to meet standards of form, they are free to present just those aspects of meaning that are relevant and salient to the speaker and leave out aspects that language may require but are not relevant to the situation (linguists have discussed the problem of defining relevance, but have not availed themselves of the gesture channel as a source of evidence; see Sperber and Wilson 1987).

NO DUALITY OF PATTERNING. Another difference from language is that gestures lack duality of patterning. This is the property of true linguistic systems in which words enter into two patterns of contrast at once (Hockett 1958). The word "dog" differs from other words in meaning: here the contrasts are to "cat," "wolf," "monkey," etc. The word also differs from other words in sounds: now the contrasts are to "cog," "doll,"

"dig," etc. One pattern is a structure of meanings and the other of sounds. Duality of patterning is deeply connected to the arbitrariness of signs. Since meaning and sounds are structured separately, the two systems can be related by *arbitrary* mappings (Saussure [1916] 1959). Gestures are profoundly different from words in that they lack this duality of patterning and the associated system of arbitrary mappings. There are no separately structured systems of form and meaning in gestures. A gesture patterns in only one way, that of meaning. Kinesic form is not an independent level as sound is an independent level of language. Kinesic form in a gesture is determined by its meaning. (This is true even of beats, whose role as a universal emphasizer lacking content of its own is paralleled by its simplicity of form.) If we explain the meaning of a gesture we explain the form. Not having duality of patterning is a strength of gestures. It explains how they are able to express meanings that may be difficult to get into the verbal channel.

Yet Gestures and Language Are a Single System

Gestures and language thus differ from each other on a number of fundamental dimensions. Yet they are also closely linked. The following describes some of these linkages. Such linkages imply that gestures and speech should be viewed within a unified conceptual framework as aspects of a single underlying process (see McNeill 1985b):

1. *Gestures occur only during speech.* While emblems and pantomimes may be delivered in utter silence, the gestures that are the focus of attention for us are almost invariably accompanied by speech. In about 100 hours of recorded narratives, only one gesture was made by a listener. Gesture production and adopting the role of speaker are virtually limited to the same situations. Moreover, 90% of all gestures by speakers occur when the speaker is actually uttering something. The acts of speaking and gesturing are bound to each other in time.

2. *Gestures and speech are semantically and pragmatically coexpressive.* That is, the gestures that accompany utterances also present the same or closely related meanings semantically and perform the same functions pragmatically. Iconics accompany utterances that depict concrete objects and events and fulfill a narrative function (see chap. 5). Metaphorics accompany utterances that refer to the pragmatic structure of the discourse as a whole. They present their own image of the discourse as an object or space or other physical reality. Other gesture types have their own parallels with speech (see chap. 7).

3. *Gestures and speech are synchronous.* The specific linguistic segments

that are coexpressive with the gesture are cotemporal. The most meaningful segment of the gesture is the stroke, and it lines up in time with the equivalent linguistic segment, as in the example given in (1.1) where the subject said, "and he [bends it way back]," and the stroke showed bending back during the same interval of time. Such synchrony implies that the speaker is presenting the same meaning in both channels at once. Having a shared meaning could be the basis for integrating gesture and speech into a single performance.

4. *Gestures and speech develop together in children.* Children's first gestures are concrete pointing and certain kinds of iconics. Much later they add other kinds of iconics, beats, metaphorics and, last of all, abstract pointing. In general, this progression follows the same path as the development of speech. As children's speech development moves from a largely referential focus, through descriptive elaboration, and finally to the ability to structure discourse, so their gestures develop from a largely concrete deictic emphasis, through various kinds of iconic gestures, and finally to the discourse-referring gestures—metaphors, abstract pointing, and beats. Beats do not appear at all in children much younger than 5 years and are not abundant until 11 years; yet, considered as movements, beats are the simplest of motions—just flicks of the hand. It is the discourse structure that determines that these flicks have meaning, and the development of this structure is late and the beat gesture itself thus doesn't occur.

5. *Gestures and speech break down together in aphasia.* Broca's aphasia consists of a relatively intact ability to use referring terms but a radically impaired ability to combine terms into larger grammatical units. This type of speech is often called "telegraphic." The gestures of Broca's aphasics are parallel in the sense that they are discrete and consist of abundant iconics; their gesture repertoire contains almost no metaphorics or beats. Wernicke's aphasics present the contrasting picture of fluent speech but a more or less complete loss of the ability to make coherent semantic connections. The gestures of Wernicke type aphasics are large, mobile, but devoid of interpretable meaning. In contrast to the Broca's aphasics they may have few iconics but there are abundant metaphorics and beats. Thus, the neurological damage that produces contrasting aphasic syndromes affects gestures in strikingly parallel ways.

For all of these reasons gestures and speech are most appropriately regarded as two sides of a single underlying verbal-gestural process of constructing and presenting meanings. This argument will be developed in detail over the next several chapters. Despite the fundamental character

of the differences between gestures and speech—one global and synthetic, the other linear and segmented—they are closely tied together in meaning, time, function, development, and dissolution. What we can learn about this unified process of meaning construction out of such opposite systems of symbols is the substance of this book.

Gestures and Time

The issue of how gestures and speech relate in time is crucial for understanding the system that includes gesture and speech as two parts. As shown by the phenomenon that gestures slightly anticipate speech, *gestures and speech have a constant relationship in time.* To express this relationship I will first describe the phenomenon of gesture anticipation. After that, I will describe three rules for gesture synchrony. That gestures could both anticipate and synchronize with speech is only seemingly a paradox, as I will explain.

Gesture Anticipation and Its Meaning

A prototypical gesture passes through three phases (Kendon 1980). There is first the preparation for the gesture: the hand rises from its resting place and moves to the front away from the speaker. Then there is the stroke, the main part of the gesture: the hand moves backward from the preparation phase and ends up near the shoulder. Finally there is the retraction, the return of the hand to quiescence: the hand falls back to the rest position. The preparation and retraction phases are optional but the stroke is essential. If there is a preparation phase, however, we can examine where it occurs. This phase regularly anticipates by a brief interval the coexpressive linguistic segment(s) (Kendon 1972, 1980). The "bends it way back" example cited earlier is such a case. Here is that example again now with the preparation phase indicated:

(1.11) he grabs a big [oak tree and he **bends it way back**]
(1) (2)

(1) Preparation phase: hand rises from armrest of chair and moves up and forward at eye level, taking on grip shape at same time.
*(2) **Stroke phase:** hand appears to pull something backwards and down, ending up near the shoulder.*

There is no explanation of the movement of the hand at (1) other than to get ready to perform the stroke at (2). The image of grasping and pull-

ing was already taking shape when the speaker was saying "oak tree" in the previous clause. This can be taken to show that the speaker was formulating the next utterance while still producing the previous one.

The anticipation of speech by gesture is important evidence for the argument that gestures reveal utterances in their primitive form: there is a global-synthetic image taking form at the moment the preparation phase begins, but there is not yet a linguistic structure with which it can integrate. The gesture in (1.11) began in the previous clause but could integrate with speech only with the words "bends it way back." One could argue that the full sentence was planned in advance, during the preparation phase, and this is what started off the gesture preparation, but this argument is actually quite weak. I will consider it and several other counterarguments at the end of the chapter.

For the moment, I will only emphasize that gesture preparations regularly anticipate their coexpressive speech. I next will introduce what appears to be a paradox: the gesture in (1.11) also synchronized with speech.

Synchronization

I will give three "rules" governing how speech and gesture synchronize. Oddly, gestures both anticipate and synchronize with speech. This is not, however, the paradox that it may seem. Anticipation and synchronization refer to different phases of the gesture. The synchrony rules refer to the stroke phase: anticipation refers to the preparation phase. It is only the stroke of the gesture that is integrated with speech into a single smooth performance, but the preparation for the stroke slightly leads the coexpressive speech, as we have seen (also see table 4.2).

PHONOLOGICAL SYNCHRONY RULE. The synchrony rule at this level is that the stroke of the gesture precedes or ends at, but does not follow, the phonological peak syllable of speech (Kendon 1980). In other words, the stroke phase of the gesture is integrated into the phonology of the utterance. For example, the stroke phase of the "bends it way back" gesture was the hand pulling back on an imaginary tree and it ended at the word "back." This word was the phonological peak of the utterance. The speaker will temporarily cease moving his hand when phonological synchrony threatens to break down, so strong is the urge to keep the gesture and speech together. In an example cited by Kendon (1980), there was a downward stroke followed by a static post-stroke hold. The effect of the hold was to maintain the "umbrella" hand shape of the gesture until the

phonologically most prominent part of the utterance could be reached (from Kendon 1980, see fig. 1):

(1.12) this patient has been a problem so far as a history is

concerned uh y'know [a] [very formal one] uh or any

(1) (2)

(1) Stroke: umbrella hand moves sharply down.
(2) Hold: umbrella hand posture held statically.

The stroke phase had taken place prematurely during the low intensity part of the utterance, but the speaker held the gesture until the phonological peak caught up.

SEMANTIC SYNCHRONY RULE. Semantic synchrony means that the two channels, speech and gesture, present the same meanings at the same time. The rule can be stated as follows: if gestures and speech co-occur they must cover the same idea unit. The term "idea unit" is meant to make provision for synchronized speech and gestures where the meanings complement one another. In the "and he bends it way back" example, the idea unit was a character seizing a tree and bending it back. The gesture, in being made with one hand, depicted the unique information that the tree was fastened at one end. It thus complemented speech by making this information specific, and speech and the gesture co-occurred. It is theoretically possible to combine gestures and speech that have different and unrelated meanings, but to do so is difficult and I have seen no spontaneous gestures that conflict with semantic synchrony. Some gestures and/or utterances are so vague that it is hard to say if they really present the same idea unit, but there are no examples of speech presenting one idea unit and gesture another. For instance, we do *not* find such pairs as "he bends it way back" with a gesture that depicts throwing something or flying or smoking a cigarette.

Three complications arise for semantic synchrony: (1) pauses, (2) multiple gestures, and (3) gestures that correspond to more than one clause. How is semantic synchrony preserved in the face of these phenomena?

1. *Pauses.* A semantically coexpressive gesture stroke will continue through the pause, thus showing that the semantic structure of the interrupted speech remains intact (Kendon 1980). Despite the interruption in

the flow of speech, semantic synchrony is preserved. Kendon gives this example:

(1.13) they wheel a big table in
[with a with a . . . (1 sec.) . . .] cake on it

Iconic: a series of circular motions with the hand pointing down and index finger extended—outlining the cake on the table.

The stroke phase continued until the word "cake," as if the gesture had to keep discharging the idea of a cake. "Cake" was still uttered with peak phonological stress, however, demonstrating that the phonological unit also was intact, and speech and gesture were integrated despite the long pause.

2. *Multiple gestures.* A small but significant proportion of clauses have two or more distinct gestures accompanying them. How do these cases of multiple gestures preserve semantic synchrony? Basically, they work the same way as single gestures when they complement speech. Each gesture covers the idea unit from a different angle. Thus, for each gesture the accompanying speech is coexpressive and there is no breakdown of semantic synchrony. The earlier example of multiple gestures in one clause (fig. 1.8) preserved semantic synchrony with each of its gestures ("[. . . and she . . .] [grabs] the knife"). The first groping gesture is coexpressive with the clause, as is the second grabbing gesture. The successive gestures depicted the temporal phases of the event, and are both synchronous on the semantic level.

3. *Multiple clauses.* Some strokes present information that is ultimately unpacked into more than one clause. Often, in this situation, the stroke is held, as in (1.12), but in some cases the hand relaxes and returns to a rest position after the first clause. The following is an instance:

(1.14) so he ig[nites himself and] flies off out the window
(1) (2)

(1) Stroke: hand quickly moves upward and forward, opening at the same time.
(2) Retraction: hand relaxes and falls back to armrest of chair.

The stroke illustrates both igniting (the hand opening) and flying off out the window (the hand moving up and out). Only the igniting-and-opening combination is truly synchronous. The rest of the stroke's semantic content (moving up and out) anticipates its coexpressive verbal description ("flies off out the window"). The phenomenon is the reverse

of having multiple gestures in one clause: here we have one gesture and multiple clauses. Semantic synchrony would be violated if during the second clause the speaker proceeded to make a new gesture or if the second clause were not an unpacking of the stroke at (1) but some other, altogether different clause. However, neither of these breakdowns of synchrony occurred. Retraction has no significance in itself, it is semantically neutral; and the second clause was semantically a continuation of the stroke at (1). Thus examples like (1.13) do not violate semantic synchrony. In fact, such examples show an important further observation: the images exhibited in gestures might require multiple clauses to be unpacked into speech.

PRAGMATIC SYNCHRONY RULE. The rule here says that if gestures and speech co-occur they perform the same pragmatic functions. Pragmatic synchrony implies that speakers are limited to one pragmatic reference at a time. However, this does not prevent speakers from making other, semantic references in the same utterance. In the earlier example of a conduit gesture (fig. 1.3), the utterance mentioned the genre of the cartoon: "it was a Sylvester and Tweety cartoon." This semantic reference was used to identify the genre of the upcoming narrative. The gesture depicted something quite different, a bounded object. The utterance thus had one reference and the gesture a different metaphoric one, but utterance and gesture came together on the *pragmatic* level of presenting the narrative genre. The utterance did this by identifying the main characters of the story and the gesture did it by presenting the story as an object in its own right. Coexpressiveness in this case arose on the level of pragmatic meaning, not on the level of semantic meaning. Pragmatic synchrony has no exceptions so far as I am aware, even though, again, it is theoretically possible to combine gestures and sentences that perform different functions (for instance, one could combine "and drops it down the pipe" with a conduit gesture that presents the cartoon as an object, although we never find such chimerical monstrosities). Again, the effort of making the gesture have one form of pragmatic content and the utterance a different one appears to be too great for this kind of mismatch to occur during running speech.

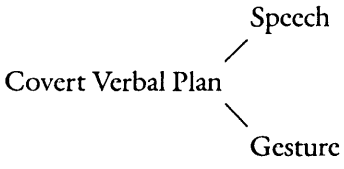
Alternative Hypotheses

The hypothesis of this book is that gesture and speech arise from a single process of utterance formation. The utterance has both an imagistic side and a linguistic side. The image arises first and is transformed into a com-

plex structure in which both the gesture and the linguistic structure are integral parts. The early signalling of the image in the preparation phase is one piece of evidence in support of this theory. The integration of the stroke with the utterance itself is another piece of evidence. That is, the image that is signaled in gesture preparation is tightly linked a moment later with the articulation of speech and the expression of semantic and pragmatic content. It is tightly linked despite its fundamentally different character from speech as a symbol. If the image is the primitive form of the utterance and the stroke is part of the utterance performance, we can account for the progression: a preparation phase that is separate from speech, and then the stroke phase that is integrated into speech.

Nonetheless, it is not difficult to imagine alternative hypotheses to this proposition. I have already briefly mentioned one—viz., the hypothesis that the full sentence is planned in advance during the gesture preparation phase—and I will now discuss it and several others. Such a method seems appropriate whether or not the reader also has thought of these hypotheses, since replying to them will clarify the hypothesis of a single process underlying speech and gesture.

THE SHARED PROCESS IS VERBAL. This is one version of the hypothesis of an advance sentence plan. According to it spontaneous gestures, along with speech, are generated from a common plan and this is the verbal plan of the utterance:



I can think of four replies to this hypothesis. First, from this theory one cannot explain the global-synthetic form of gestures. They should be linear-segmented if they are produced from a covert verbal plan.

Second, where covert linear-segmented verbal plans have been postulated (e.g., Sternberg et al. 1978) there has been an experimental procedure in which speakers were repeating verbal materials verbatim from memory; this is not planning the content of the utterance but planning only a succession of phonetic shapes. Such a verbal plan is precisely one that excludes all basis for gestures.

Third, from this theory we cannot explain the division of meaning between the gesture and speech channels as in (1.2). If there is a meaning in the gesture channel, it must have come from the covert verbal plan and

should also be present in the speech channel, and to the same degree of specificity.

Finally, if gestures undergo a less complex transformation than speech after their shared computational stage, the global-synthetic image that gestures reflect can be regarded as closer to the verbal plan at an early stage of its development. This is, in fact, how I interpret gestures: gestures make manifest the utterance's primitive form, and *this* is the "verbal" plan of the utterance at an early phase of its internal development, while it is still in the form of imagery.

For all these reasons, it seems sound to conclude that speech and gesture are coexpressive manifestations of a single underlying process. The underlying process is equally speech and gesture, and there is a subsequent evolution of expressive action with outputs in both channels concurrently. The channels, moreover, have a constant relationship in time, with the gesture manifesting the primitive stage of the shared process and speech its final socially presentable stage.

THE GESTURE TRANSLATES THE SENTENCE. This is another version of the advance sentence plan hypothesis. It runs as follows. The visual-actional medium of the gesture has its own qualities, but these do not characterize the psychological structure of the sentence. There are in fact two psychological structures. The sentence medium has one characteristic set of qualities (linear and segmented), and the gesture medium has another set (global and synthetic or imagistic). The gesture is parasitic on the sentence. When a gesture occurs, this argument goes, the linear-segmented qualities of the sentence are translated into the imagistic qualities of the gesture. This does not mean that there is a shared process in which both sets of qualities exist as integral parts. On the contrary, it implies that there is computation of the sentence, then a separate computation of the gesture.

There are several replies to this hypothesis:

First, there are very few cases where speech and gesture are repaired. Rather, what typically happens is that speech is repaired and the gesture that accompanied the utterance being repaired is repeated with the repair in exactly the same form. The gesture, therefore, must not have been the output of the sentence. (This argument is due to Sotaro Kita.)

Second, there is complementarity. In the "she chases him out again" example, the gesture and speech did not manifest the same information. There was mutual complementarity in that the utterance conveyed aspects of the scene that the gesture did not convey, and vice versa. Such examples of mutual complementarity can be multiplied many fold.

Assume that gestures are copies of sentences. Then we have a situation that we can diagram as follows:

Speech → Gesture

However, there is no possibility of complementarity with this arrangement. Everything in gesture must have come from speech. Thus if a meaning appears in gesture, it must appear in speech as well. The same argument applies to the opposite diagram, with speech copying gesture (a more appropriate hypothesis in any case, in view of the anticipation of speech by gesture):

Gesture → Speech

In this case everything in speech must have come from gesture, but this also is contradicted by the phenomenon of mutual complementarity.

The DAF experiments in chapter 10 demonstrate that complexity representation spontaneously shifts from the speech to the gesture channel. The Speech → Gesture model is ruled out by this shift, since if the speech channel is blocked by DAF there is no way the gesture channel can take over its complexity if gesture is a copy of speech. Thus, complementarity of speech and gesture implies that gestures are coexpressive of meaning. Since they are so closely linked to speech, they are a manifestation of meaning that is different from but closely connected to the meaning as conveyed in spoken form.

Third, to copy speech into gesture means not only reproducing content but radically changing the form of utterance as well. Speech and gesture are representations different in form. This is most apparent in the difference between gestures and manual signs in a true linguistic system such as ASL. I will devote the entire second chapter to the contrast between spontaneous gesture and signs. For the present, I will just state that ASL signs are segmented, combinatoric, context-stable, etc., while gestures are the opposite on each dimension. It is not that gestures are uninfluenced by conventions, but that the conventions that influence them are the conventions of social life in general, not specific gesture conventions. Thus, if gestures were copies of speech, they would make the context-stable unstable, the segmented global, the combinatoric synthetic, and so on. To suppose that gesture is a copy of speech in a different medium simply overlooks the enormous complexity of such a change. Instead, I believe that gesture and speech manifest different stages of an evolution of the utterance.

Finally, the kinesic medium is pliable and has syntactic potential. This point is shown by the phenomenon of sign languages and will also be

fully described in chapter 2. It is abundantly clear that the gesture medium can adopt the linguistic properties of finished utterances, which are so obviously lacking from spontaneous gestures. Linguistic properties can actually be seen emerging in gestures when a speaker is required to use gesture as his only channel of communication (Bloom 1979, described in chap. 2). If the global-synthetic properties of gestures are not inherent in the visual-kinesic medium, they must be explained in some other way. A plausible account is that they are imposed on it by a stage of processing in which the representations are imagistic. This is the primitive stage of the utterance—imagistic, analogic, global, and synthetic.

The four arguments jointly lead to the conclusion that gestures are forced to be as they are by the form of thought that exists at the moment the gesture is triggered, and this representation is imagistic. The deep time evolution of utterances carries global-synthetic imagery into a final stage of socially regulated coded symbols.

GESTURES ARE INDEPENDENT VISUAL DISPLAYS. According to this hypothesis, the gesture is just a separate display exhibited along with the utterance. It is like a photograph or drawing. And just as holding up a real photograph would have nothing to do with the process of utterance generation, so performing a gesture could have nothing to do with it either. The first point in reply is that, unlike a photograph, a gesture is something the speaker is creating while speaking. Second, the gesture is very closely connected to speech temporally, semantically, and pragmatically, all of which suggest a coordination between the gesture and the utterance that is quite different from presenting a photograph. Third, when meaning is divided between a gesture and sentence, it is a true division. It is not that the gesture is called up, as a photograph might be held up, to repair an otherwise interrupted message. We don't observe, for example,

so he . . . [*gesture to complete the idea*]

but rather,

[so he chases him out again]

Gesture shows the means.

These points reduce to one reply: gesture and speech are operations that have been connected *within*. This is the sense in which they are parts of a single process.

Moreover, there is a fundamental difference between gestures and photographs. Gestures are necessarily schematic, and are to a degree that

photographs usually are not. More crucially, a gesture is structured by meaning, while a photograph is not; the meaning of a photograph is something we—the observers—bring to it whereas the meaning of a gesture is the foundation of its construction.

A further difference between a gesture and photograph, however, is that a photograph is disconnected from its original context in time. This disconnection is indeed the raison d'être of the photograph. "The most popular use of the photograph is as a memento of *the absent*" (Berger 1972, 180, emphasis added). A gesture, in contrast, exists only at a given moment and this moment must be included in the meaning of the gesture. We can video a gesture, but we also video its moment. The video itself is a photograph of sorts and is disconnected from *its* context in time, but the gesture, the subject of the video-photograph, is a real entity that exists only in context. A photograph records scene x at moment y, and this very fact is its source of significance, as Berger explains: the photograph refers to the past moment and to the recording of it. It means "I have decided that seeing this is worth recording" (Berger 1972, 179). We can then look at it with comprehension at other moments z, q, or a. The $y \neq z$, q, or a difference is what separates photos from gestures. A gesture also has its moment y, but can be comprehended only *at* y (and if we record it, we replay y too). Performing it at moments z, q, or a is play-acting, or example-citing.

THE GESTURE IS THE TIP OF THE ICEBERG. According to this hypothesis, the global-synthetic gesture is the surface manifestation of what is, underneath, a *sign*—a gesture with language-like properties.[2] If this is the case, it is a mistake to infer from the gesture a form of thought that is global and synthetic. The overt gesture is just the tip of the iceberg. The rest of the iceberg is more like language, with segmentation of complex ideas into elements, compositionality, standardization of form, distinctiveness, arbitrariness, and the rest of the properties of signs as opposed to gestures. However, this hypothesis is empirically unsound. If the observed global-synthetic gesture is just the tip of the iceberg and under the "sea of speech" is a more language-like sign, then by "raising up" the iceberg we should expose more and more language-like properties. That is, a simple gesture might be global-synthetic, but what of more complex gestures? They might be expected to be more independent of speech and display more linguistic properties of their own. However, this is not true. More complex gestures, gestures utilizing two hands

2. Susan Goldin-Meadow and Jenny Singleton pointed out this counterargument.

in different roles or gestures with several internal components, do not bring up latent linguistic structure and are still tied to speech. Complex gestures are global-synthetic all the way down, to keep to the iceberg metaphor. All high-complexity gestures, indeed, show global-synthetic-meaning properties. Thus, I also feel justified in rejecting the fourth hypothesis.

Conclusion

Thus, gestures are not the product of a linear-segmented verbal plan, not translations of speech into visual-kinesic form, not like photographs, and not the tip of a linguistically structured iceberg. They are closely linked to speech, yet present meaning in a form fundamentally different from that of speech. My own hypothesis is that speech and gesture are elements of a single integrated process of utterance formation in which there is a synthesis of opposite modes of thought—global-synthetic and instantaneous imagery with linear-segmented temporally extended verbalization. Utterances and thoughts realized in them are both imagery and language.