The goal of our current program of research is to understand the unique features of face-to-face dialogue, the features that have no counterpart in written text (e.g., Bavelas, Hutchinson, Kenwood, & Matheson, 1997). Linell (1982) has outlined several differences between the two kinds of language use (see Table 1). Here, we will focus on one of the distinguishing features, namely, the observation that the speaker's speech behavior is continuously accompanied and supplemented... by various non-verbal signals, which means that the verbal message as such is often much less explicit than in writing; referents may be pointed to, interpretations may be made more precise and complex through gestures, facial expressions, tone of voice, etc.

After all, the use of an utterance in a normal situation involving face-to-face interaction is not an isolated speech act; it is part of a comprehensive communicative act which comprises the use of both verbal means (speech) and non-verbal means (gesticulation, etc.). The message is conveyed, or shown, in several ways simultaneously, and the role played by spoken language cannot be properly understood without taking into consideration the whole communicative act. (Linell, 1982, p. 6)

Yet, most studies of face-to-face dialogue have either focused entirely on what is audible (words and prosody), or they have treated visible acts as a separate, nonverbal channel of communication identified only by what it is not. We propose that language in face-to-face dialogue is composed of both audible and visible acts and that although these acts can be separated analytically, they are completely interwoven in performance: "The human organism is one integrated whole... it is not to be thought of as a box, carrying various independent pairs of terminals labeled 'ears,' 'eyes,' 'nose,' et cetera" (Cherry, 1957, pp. 127-128).

Many readers will recognize that our proposal is only the next step in a historical trend toward including some nonverbal elements in language use. The importance of nonverbal as well as verbal elements in face-to-face dialogue has been emphasized by many theorists and researchers, starting in 1955 with the Natural History of an Interview project (cf. Leeds-Hurwitz, 1987) and including scholars such as Birdwhistell (1960, 1966, 1970); Pike (1967, 1972); Scheff (1968); Ekman and Friesen (1969b); Kendon (1972, 1980, 1983, 1985); Burton-Jones (1972); Weiner, Devoe, Rubinow, and Geller (1972); Slama-Cazacu (1976); Duncan and Fiske (1977, 1985); Poyatos (1980); Scherer (1980); McNeill (1985, 1992); Goodwin and Goodwin (1986); Sanders (1987); Leeds-Hurwitz (1989); Bavelas, Black, Chovil, and Mullett (1990a, chap. 6); Friidlund (1991a); Streeck and Knapp (1992); and Clark (1996, chap. 6). In spite of these advocates, there has been little theoretical elaboration and only isolated experimental tests. We believe it is now possible to begin to sketch out a more specific theory and to review the evidence already available for it, which may in turn suggest new directions for research in the area. Our theory is definitely
Table 1  
Differences Between Written and Spoken Language

<table>
<thead>
<tr>
<th>Written Text</th>
<th>Face-to-Face Dialogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is a persistent, static &quot;object&quot;:</td>
<td>Is ephemeral, &quot;dynamic&quot;:</td>
</tr>
<tr>
<td>It can be reread any time.</td>
<td>It cannot ordinarily be reviewed.</td>
</tr>
<tr>
<td>It seldom requires a rapid response.</td>
<td>The participants must respond immediately, &quot;online.&quot;</td>
</tr>
<tr>
<td>Consists of discrete, separate symbols:</td>
<td>Is virtually continuous:</td>
</tr>
<tr>
<td>Words are easily separated.</td>
<td>Words and other acts merge.</td>
</tr>
<tr>
<td>Text is organized spatially.</td>
<td>Dialogue is organized temporally.</td>
</tr>
<tr>
<td>Is relatively context free:</td>
<td>Is highly dependent on context:</td>
</tr>
<tr>
<td>It uses only words and punctuation.</td>
<td>It uses face and hand gestures, as well as prosodic features.</td>
</tr>
<tr>
<td>The words are highly explicit.</td>
<td>The words can be less explicit.</td>
</tr>
<tr>
<td>Text is monologue and solitary.</td>
<td>Dialogue is a &quot;social interplay.&quot;</td>
</tr>
<tr>
<td>There is no immediate reader.</td>
<td>There is an addressee present.</td>
</tr>
<tr>
<td>The writer and reader are in different places.</td>
<td>The participants are in the same setting.</td>
</tr>
<tr>
<td>Text must often be addressed to a general audience.</td>
<td>Dialogue can draw on the setting and the ongoing conversation.</td>
</tr>
<tr>
<td>Is acquired as secondary socialization:</td>
<td>Is acquired as primary socialization:</td>
</tr>
<tr>
<td>Literacy is learned institutionally (in schools).</td>
<td>Dialogue is learned interpersonally (at home).</td>
</tr>
<tr>
<td>It is taught with explicit, conscious norms.</td>
<td>It is practiced rather than explicitly taught.</td>
</tr>
<tr>
<td>The norms are more standardized, with less variation.</td>
<td>The norms are freer, with more variation.</td>
</tr>
</tbody>
</table>

Source. The information in this table has been adapted from Linell, 1982, pp. 6-10.

Note. Some of the above features of written text do not apply to computer-mediated communication, especially if both parties are online at the same time.

and (c) some other communicative body movements such as nodding, shrugging, or whole-body depictions of an event or action as part of telling about it (e.g., demonstrating someone’s appearance while referring to it verbally, Clark & Gerrig, 1990).

We explicitly exclude other nonverbal acts such as behaviors that occur when the individual is (or believes that he or she is) alone; movements too small to be visible to addressees, even though they may be measurable by electromyography (e.g., Cacioppo, Petty, Losch, & Kim, 1986); involuntary or reflexive acts such as breathing, sneezing, or sleeping (even when they occur in the presence of another); voluntary acts that accomplish an other-than-communicative function, such as drinking coffee, scratching a bite, or turning a page in a book; and static features such as overall physical appearance or resting positions of the legs or arms; see also Bavelas (1990). Indeed, rapidity is one of the main characteristics of the acts we are including; they are typically of the same brief duration as the words and phrases with which they are tightly synchronized—rarely more than a few seconds and often much less. We will describe below some empirical methods for establishing whether a particular kind of nonverbal behavior can be treated as a visible act of meaning. Here, it is crucial to stress that in making distinctions within the vast domain of nonverbal behaviors we do not mean to imply that the behaviors we have excluded here (such as the actions of solitary individuals, self- or object-adaptors, EMG activity, appearance, or static postures) should be ignored by other researchers who will remain interested in them for a variety of important reasons not directly related to talk in dialogue.

Equally important, we make no exclusive claim over the behaviors that we are including because they often serve other functions as well (Bavelas, 1994). Many human actions serve more than one function, and this is particularly true for language acts; for example, words and phrases serve syntactic, semantic, and pragmatic functions. Gestures may serve both intra- and interpersonal functions (Krauss, Chen, & Chawla, 1996); facial displays may express emotion and also be part of a communicative act (Ekman & Friesen, 1969b). One advantage of a functional over a taxonomic approach is that functions are not mutually exclusive, whereas classifications are (Bavelas, 1994, pp. 202-204; Bavelas, 1998, pp. 187-188). Therefore, any evidence we offer here that gestures or facial displays in dialogue are part of an integrated message with words is not intended to (and cannot logically) exclude other functions. For example, sometimes a gesture slightly precedes the related word. When this happens, the gesture may be serving all three of the following functions: (a) expressing the speaker’s thought-image with a temporal advantage over speech because the gesture is also an image, (b) helping the speaker find the word (i.e., self-prompting), and (c) marking a word search for the addressee or even seeking the addressee’s help in finding the word (i.e., other-prompting). These
different functions (and the theories that propose them) are orthogonal, not incompatible.

Even with all of these qualifications, the set of behaviors we will be examining is a large and significant one. In face-to-face dialogue, many visible behaviors act with words to convey meaning. Speakers spontaneously emphasize, particularize, embellish, or replace words with their facial displays, gestures, and other depictions. For example, a participant in one of our dyadic experiments was describing her brother's experiences in a busy Italian train station where he was uncertain which train to board: "He's trying to decide which one to go to, y'know." At the point indicated by italics, her parallel hands waved from right to left several times; these depicted the moving trains he was choosing from. Simultaneously, her eyes widened and moved from right to left but were not synchronized with the hand movement; this portrayed her somewhat confused brother looking from train to train, trying to decide. Her intonation was light and even, and she was smiling as if to verify her listener's understanding. Thus, she created a scene in which her hands were the trains and her eyes were her brother looking at the trains. Her smile was not part of this scene but instead was addressed directly to her addressee and, in concert with the verbal y'know (without rising intonation), invoked or presumed the other's appreciation of the situation. This excerpt was one brief moment in a series of similar enactments. Nor is it an unusual or virtuoso performance; close scrutiny of face-to-face dialogue routinely reveals many instances at least as complex and elegant.

The same fine-grained analysis is appropriate for listeners, who not only contribute audibly with yeah and no but also constantly convey their reactions visibly (Bavelas, Coates, & Johnson, in press), for example by simple attentiveness, mimetic facial displays (Chovil, 1991), or even mirroring the speaker's gestures. Bavelas et al. (in press) have shown that when speakers were telling a close-call story and when, in the experimental condition, their addressees could not react, their stories were less well told and fell flat at what should have been the dramatic ending.

Our final general point is that these visible acts of meaning are inseparable from the words with which they occur at the moment and with which they form an integrated message. As participants in conversation, we hardly notice which information comes from where. Indeed, authors such as Streeck and Knapp (1992) have concluded that "the classification of communicative behavior as either 'verbal' or 'nonverbal' is misleading and obsolete" (p. 3); they use the term visual features. Birdwhistell (1968, 1970), Leeds-Hurwitz (1989), Sanders (1987), and Scherer (1980) have characterized communication as a multichannel process. Pike (1972, p. 113) described units he called behaviouremes, which can be composed of verbal or nonverbal acts—or both; he proposed an integrated synthesis of such events, as did Poyatos (1980).

McNeill (1985, 1992) postulated that gestural and verbal production originate in the same computational stage and are two parts of the same (linguistic) process. Slama-Cazacu (1976) called the outcome of this process mixed syntax, in which a nonverbal act can convey the subject, verb, object, adjective, adverb, or other part of a sentence. In one of her examples, the gesture conveys the imperative verb in a sentence composed of both words and gesture: "A film director, talking about the Number Five spotlight placed on a baulstraded platform, tells the electrician: 'Five balcony!' [gesture as if switching on the light: 'Switch it on!]" (Slama-Cazacu, 1976, p. 222). More recently, Engle and Clark (1995) and Clark (1998, p. 156) have called these "composite signals, the artful fusion of two or more methods of signaling."

The following sections contain conceptual and empirical criteria for identifying the visible acts of meaning being used by the interlocutors in face-to-face dialogue. We specify four defining characteristics, each of which builds on the previous: (a) Visible acts of meaning are sensitive to a sender-receiver relationship; (b) these acts are symbolic, that is, they are being used to stand for something else; (c) their meaning can be explicated or demonstrated contextually; and (d) these acts are always integrated with the accompanying words, whether their meaning is redundant or nonredundant with the words.

**VISIBLE ACTS OF MEANING ARE SENSITIVE TO A SENDER-RECEIVER RELATIONSHIP**

The first characteristic of these acts is that they occur primarily in dialogue and are affected by the visual availability of a receiver. For reasons to be discussed at the end of the section, it is not necessary to show that these acts never occur without a receiver. However, the act should be substantially reduced or modified when no receiver is available.

The empirical paradigm was first developed by Cohen and Harrison (1973) for hand gestures. They established that hand gestures were significantly more frequent when people were giving directions in person rather than over an intercom. Cohen (1977) used three groups (face-to-face, intercom, and practicing alone) and again found a significantly higher rate of gestures in the face-to-face condition.

Several studies of facial displays, particularly smiles, have demonstrated that these are also sensitive to whether they will be received by another (see Chovil, 1997, for a detailed review). Kraut and Johnston (1979) conducted four field studies demonstrating that people were more likely to smile when they were with or facing other people than when alone or facing away. Smiling was relatively independent of whether a good or bad event had occurred; for example, bowlers tended to smile depending on whether they were facing their companions not
whether they had thrown a good ball. Fernandez-Dols and Ruiz-Belda (1995) studied smiling during Olympic medal awards ceremonies. The gold medalists were presumably happy throughout the ceremony, but they smiled almost exclusively during the social phase of the ceremony.

In a study of preschoolers, aged 3½ to almost 6, Schneider and Josephs (1991) found that the frequency of smiles increased with the possibility of eye contact with an experimenter. The children who played a game alone (while the experimenter sat at another table with her back turned) exhibited fewer smiles than children who played the game with the experimenter beside them. In addition, the children beside the experimenter were more likely to look toward the experimenter when she turned away. As will be seen in the next study, this looking might be interpreted as seeking a receiver.

Jones and Raag (1989) found that 1½-year-olds glanced and smiled at their mothers (or a stranger) more often than at their toys and even more when the other person was attentive. As these authors concluded,

A demonstration that very young children smile more during social interaction than during solitary play would surprise no one. However, in our study the smiles in question did not originate in social interaction. Rather, the children presumably smiled for reasons connected with their toy play but took the trouble to turn around and direct the facial expression (italics added) at their mothers. The dramatic decline in smile-plus-glance sequences when mother was not attending indicates that very young subjects were well aware of whether... the communicative channel to their social partners was open or closed. (p. 815)

This finding has been replicated by Jones, Collins, and Hong (1991) with even younger infants. Mothers’ attentiveness affected the frequency of smiles of 10-month-old infants, suggesting that the infants monitored the possibility that the display would be received. In other words, mere presence was not sufficient to elicit a display: The receiver had to be looking at the infant as well.

Returning to adults, other types of facial displays have also been shown to be affected by social variables. Brightman, Segal, Werther, and Steiner (1975) found that participants’ faces portrayed their reactions to sweet or salty sandwiches only when they ate them in the visual presence of others, that is, in a group setting where they could see each other. There was no change in facial display for the participants who ate alone—even for the highest levels of tastants (110% salt or 136% sugar).

The precise moment of visual availability during a 4-second interval was varied in an experiment by Bavelas, Black, Lemery, and Mullett (1986). Participants observed at close range as the experimenter apparently dropped a heavy TV monitor on an already injured finger. They were significantly more likely to produce and to hold mimetic, pained facial displays when the experimenter made eye contact than when he did not.

Chovil (1991) created face-to-face, partition, telephone, and answering-machine conditions and found that the listener’s facial displays in response to the speaker’s close-call story occurred primarily when they would be seen by the storyteller. Facial displays were very infrequent in the two voice-only conditions, and they were virtually nonexistent when the listener was alone, listening to a dramatic story on an answering machine.

Our last line of evidence is the well-established tendency for listeners to gaze at speakers in dialogue, although speakers often look away (Argyle & Cook, 1976; Duncan & Fiske, 1977; Kendon, 1967). We would argue that listener gaze occurs in large part because what the speaker is “saying” is not just auditory but also visual. Listeners need to see speakers’ hands and faces as well as hear their words.

Two conclusions are apparent in this literature. First, there are a wide variety of methods for testing the sensitivity of nonverbal acts to the visual availability of a receiver, which encourages us to hope that more will be done. Second, the results so far also reveal many acts that show such sensitivity. These may be functioning as visible acts of meaning; they have met a necessary but not sufficient criterion. As will be seen, they must go on to show other properties as well.

Before we leave this topic, however, it is necessary to address the often-asked question: If hand and facial gestures are communicative, why do some still occur when no one would see them? In most of the above studies, the frequency was not reduced to zero when there was no visible receiver, and each of us can probably recall making gestures on the telephone or making facial displays when alone. It may be tempting to conclude, for instance, that hand gestures therefore cannot be communicative acts but must be serving some other intrapersonal function (such as lexical access; e.g., Krauss et al., 1996). This conclusion, however, presupposes a taxonomic approach, a mutually exclusive classification system in which the act must be either communicative or intrapersonal regardless of setting. As outlined above, we are proposing that these acts can have several functions, some of which might still be relevant when the individual is alone. (We would also observe that the same is true for words: For example, we sometimes exclaim to ourselves or count something out loud, even when completely alone. Yet, no one would suggest that, therefore, words are not communicative when they are used in social settings.)

This leads to the interesting empirical question of whether the acts (visible or audible) are really the same in social and nonsocial conditions. One way to test this would be to film covertly. Gilbert, Fridlund, and Sabini (1987) found that when individuals were not aware of being videotaped, the facial displays elicited by strong odors were judged much less accurately by observers than those posed for videotape. That
is, facial displays when alone were not as distinct or clear as those deliberately produced for the purpose of communication.

In gesture experiments (Bavelas, Chovil, Lawrie, & Wade, 1992), the effect was even more precise, varying with the kind of gesture that was made. Bavelas, Hagen, Lane, and Lawrie (1989) have distinguished between topic gestures, which illustrate some aspect of what the speaker is describing, and interactive gestures, which refer instead directly to the addressee and the process of having a dialogue. In these experiments, the participants knew they were being videotaped, but, in some conditions, they either did not have an addressee or were separated by a partition. The rate of topic gestures in these conditions was not much different from the face-to-face dialogue condition; we interpret these gestures as being made for the camera and videotape. However, the rate of interactive gestures dropped significantly because the camera is not an interlocutor. (Also, our informal observation was that, like the faces in Gilbert et al., 1987, the interactive gestures that did occur were less clearly formed and harder to analyze.)

As a final point, we would suggest that the term alone also needs closer examination. As just described, the future audience of a videotape was sufficient to produce selected visible acts of meaning, even when the individual was physically alone. Bavelas et al. (1986, p. 328) and Clark (1996, pp. 179-180) have suggested that if these displays are intrinsically social, they may occur when the situation is psychologically though not physically social. Obviously, the potential tautology of this proposal needs to be avoided, but it is experimentally possible to do so. Fridlund’s (1991b) participants watched a pleasant videotape sequence (cute babies, otters playing, a comedy skit, etc.) in one of four conditions that varied in the degree to which they were psychologically social. They watched the videotape (a) with a friend watching in the same room, (b) alone while a friend watched the same videotape in another room, (c) alone while a friend did an unrelated activity in another room, or (d) alone with no friend in the experiment. Reported happiness did not differ across conditions, but smiling dropped monotonically over the four sociality conditions. Pairwise comparisons revealed that there was no difference between having a friend in the room and having the friend watching in another room (condition a vs. b). Also, there was significantly more smiling when a friend was watching in another room than when the viewer was psychologically alone (condition b vs. d).

**VISIBLE ACTS OF MEANING ARE SYMBOLIC**

Returning to our criteria, an obvious defining characteristic of visible acts of meaning is that they are not just physical actions but also symbols in the simplest sense of that term: something that stands for something else (Quine, 1987, p. 763). In the example given earlier, the speaker’s parallel hand movements were not wrist exercises; they stood for the moving trains. Nor was she moving her eyes to see something; she was depicting her brother looking, at another time, in another place. Symbolic acts do not merely reveal information for an observer to infer; rather, their meanings are deliberately adopted by a speaker for an addressee. Weiner et al. (1972) pointed out the common confusion between informative and communicative nonverbal acts. For example, when a woman standing alone at the bus stop looks at her watch, this act might be informative to an observer, who could infer that she is wondering whether the bus is late. In another setting, she reminds her husband of their next engagement and looks at her watch to signal that they must leave soon—a visible act of meaning.

Several scholars have proposed that hand gestures are symbolic, for example McNeill (1985, 1992): “Gestures are not just movements and can never be fully explained in purely kinesic terms. They are not just the arms waving in the air, but symbols that exhibit meaning in their own right” (1992, p. 105). Kendon (1985) suggested that gestures are particularly useful in depicting the appearance of an event or the visual features of the referent. Clark and Gerrig (1990, pp. 766-767) included some gestures as well as words in their concept of demonstrations and stated clearly that they have referents that can be events, states, processes, or objects. Streeck and Knapp (1992) noted that speakers often use hand movements that enact or reenact some action being described.

Facial displays can also be symbolic. Bavelas et al. (1986, 1988) found evidence that motor mimicry (such as wincing when another person's pain or leaning with another's effort) conveys to the other person, “It is as if I am you, having your reaction.” More generally, Chovil (1991/1992) found a wide variety of facial displays with semantic and syntactic functions. For example, when asked to tell about a past interaction, speakers often used their facial displays (as well as words) to describe both themselves and the other person. In the latter case, their faces portrayed the reactions of another person at a different time and in a different place than the present dialogue. Thus, faces as well as hands can portray objects and events that do not necessarily exist at the moment—the essential characteristic of human symbolic activity.

**ANALOGIC CODING**

The symbolic nature of gestures and facial displays is often overlooked or rejected because their encoding is analogic (also called iconic or natural encoding), whereas encoding into words is arbitrary (also called conventional or digital). However, although the act is like what it represents, it is still not equal to what it represents. For example, when
a storyteller gesturally depicts hammering a nail, no hammer or nail is present, and no carpentry is accomplished by the depiction.

Moreover, Grice (1957) questioned the usefulness of a distinction based on how an act is encoded and replaced it with a distinction between having a meaning in a natural or a nonnatural sense. A cough might mean bronchitis; this is a natural meaning. In contrast, a stylized cough used to interrupt discreetly is a nonnatural meaning:

This question about the distinction between natural and nonnatural meaning is, I think, what people are getting at when they display an interest in a distinction between "natural" and "conventional" signs. But I think my formulation is better. For some things which can [nonnaturally mean] something are not signs (e.g., words are not), and some are not conventional in any ordinary sense (e.g., certain gestures) [italics added]. (p. 379)

That is, unlike emblems, conversational gestures are spontaneously improvised. They are not conventional, yet they have a nonnatural meaning. Grice (1957, 1967/1989) went on to propose that the appropriate criterion for meaning is the speaker’s intention to mean something, rather than how he or she chooses to do so. In a similar vein, Clark (1996) distinguished between the various methods of signaling and the act or signal itself, emphasizing that most signals are composites of his three methods.

We propose that when nonverbal behaviors are used nonnaturally as visible acts of meaning, their distinguishing features will be both different and clearer than when they are natural. Building on Clark and Gerrig (1990), we pointed out that facial displays used to depict meaning are both nonliteral and selective (Bavelas & Chovil, 1997). They are nonliteral in the sense that the individual is not necessarily claiming to have the reaction being depicted. For example, the speaker may be depicting someone else’s reaction or a reaction that he or she had at some earlier time. They are selective in the sense of being partial portrayals rather than faithful reproductions. For example, a speaker may depict his or her earlier surprise by a stylized raising of eyebrows without all of the other words and actions that had actually occurred at the time.

Ekman, Hager, and Friesen (1981) described differences between posed and spontaneous facial expressions. Recall also that Gilbert et al. (1987) found that spontaneous facial displays in response to unpleasant, neutral, or pleasant odors were hard to interpret accurately, whereas the observers' accuracy improved significantly when participants had posed their reactions for videotaping. Leslie Jarmon has provided us with a video clip of a serendipitous example she filmed at a building site:

One woman is talking to another when a loud noise startles her. She then laughs at herself andreenacts her startle response twice, as part of retelling the incident with the other woman. The second and third (nonnatural) startles are much clearer and have more of the classic features of a startle response (eyebrows up, eyes wide, hands jerking out) than does the first, spontaneous or natural startle.

Bavelas, Black, Chovil, Lemery, and Mullet (1988) focused on a particular form of motor mimicry, namely, ducking or leaning. When a storyteller described and enacted having to duck to her right to avoid being hit, all but one of the addressee’s mimicry is to their left, that is, in the same direction as the storyteller. This direction is called reflection symmetry, and further studies confirmed it is more readily decoded as empathic than leaning in the opposite direction would be. Thus, a closer look at how the act is encoded may confirm its symbolic function.

**CONCRETE-ABSTRACT IS A CONTINUUM**

Even when they are recognized as symbols, gestures and facial displays are often seen as limited to the portrayal of concrete objects and actions (e.g., gestures point at things or depict physical size, shape, and motion; faces depict one’s current emotional reaction). However, visible acts of meaning can also be abstract or metaphorical.

McNeill (1992) distinguished between gestures about concrete referents, which he called iconic gestures (e.g., depicting a climbing action or the shape of an object), and gestures about abstract referents, or metaphorical gestures (e.g., balancing cupped palms to weigh two alternative possibilities). In an earlier analysis, McNeill and Levy (1982) implied a continuum with three degrees of abstract symbolization in gestures, which we still find useful. In the first level of iconic gesture, the hands portray hands in relation to an object that is not present (e.g., the hand turns an imaginary key). In the second level of iconic gesture, the hands do not portray hands but another object or action (e.g., the hand depicts the steep angle of a path). In metaphorical gestures, the hands depict abstract intangibles by means of metaphors (e.g., a hand flipped over the shoulder can locate the past as metaphorically behind us).

We propose here that facial displays are also both iconic and metaphor. What we call iconic facial displays portray or enact literal reactions (e.g., disgust in response to something inedible). The first level of iconic facial display would be portraying one’s own facial reaction (whether currently, in the past or future, or hypothetically). Portraying someone else’s reaction, as in motor mimicry, would be the second iconic level. The difference is that at the first level, the face is still one’s own face, whereas at the second level, it portrays someone else’s face. Facial metaphors convert the literal meaning of a facial reaction to an...
abstract one by using it in a metaphor. For example, squinting suspiciously at an implausible story is a metaphor based on squinting as a means of peering more closely. Or, when told about someone's social faux pas, listeners often react by pulling their eyebrows together in a pained expression, although the event was only metaphorically painful. Some facial displays have become extremely abstract; for example, eyebrow movements used as syntactic stressors (Chovil, 1991-1992; Ekman, 1979) are probably a metaphor that draws on the startle reaction.

In sum, analogic encoding is a flexible and powerful system for the symbolic representation of abstract as well as concrete meaning in conversation.

THE MEANING OF THESE VISIBLE ACTS CAN BE EXPLICATED OR DEMONSTRATED IN CONTEXT

New theoretical approaches often require new methods. The usual approaches to scoring or coding nonverbal acts, such as treating them as physical actions to be tallied or as a separate channel of communication, are obviously incongruent with our theory (Bavelas, 1994). A method compatible with our meaning-based approach must meet two requirements. First, it must assess, directly or indirectly, the informational meaning of the act, and second, it must do so by keeping the message in its immediate linguistic context, that is, by focusing on the message as a whole and treating the visible act(s) in question as part of an integrated audible and visible package. We can think of two approaches that satisfy these criteria. Either we as expert analysts can demonstrate high interjudge reliability for our explication of the meaning of the act, or we can demonstrate from the response of the addressee that he or she understood the act. These two alternatives will be considered in turn.

EXPLICATING MEANING

Our main approach has been to demonstrate that we can reliably express the meaning of facial displays and gestures in words using both the form of the act and its immediate linguistic context. Verbal paraphrasing permits us to treat all of the components of the message in a familiar lingua franca. This is a long-standing, although often implicit, practice in previous literature. For example, Weiner et al. (1972) described circling gestures—a slow continuous series of circular movements of any part of the hand including the arm which is hypothesized to indicate nonspecificity, globality, or generality of the verbal components in the communication. In words, this gesture is posited as indicating "I mean more than the specific words I have used," or "The specific referent term is only one general attribute of the event I'm describing." (p. 211)

Notice how much easier it is to visualize the gesture and to understand its meaning with the verbal paraphrase than it was from the initial physical description. Ekman (1985) identified one type of smile (listener response smile) as "equivalent to the 'mm-hmm,' 'good,' and head nod it often accompanies" (p. 157). Brunner (1979) discussed how both nonverbal and vocal backchannels convey the message "Yes, you are understood. Proceed." (p. 733). Siama-Cazacu (1976) has been most explicit about this approach, as seen in her example given above, which we would paraphrase as "Switch on the Number Five balcony light!"

Verbal paraphrasing is not, however, simply a matter of looking in a nonverbal dictionary. These acts are analogous to polysemous words; they do not have a single, context-free meaning. Indeed, it is in exactly this respect that they are most wordlike. Clark and Clark (1977) emphasized that to treat words as if they have one and only one sense is incorrect for three reasons: First, almost all words are polysemous—they have more than one sense. Second, there are expressions called idioms, like kick the bucket meaning "die," which do not get their meaning directly from the words they contain. And third, some words can be used for conveying senses never before associated with them. This phenomenon will be called lexical creativity. (p. 444)

Moreover, as any thesaurus shows, there are also many different words for expressing the same meaning. We propose that both of these characteristics (many meanings for one word and many words for one meaning) are true for visible acts of meaning as well. There are no one-to-one relationships; the meaning of both audible and visible communicative acts is highly dependent on context.

Thus, although one might be able to enumerate and articulate all possible meanings for a word, it is not possible to select the particular way in which a word is being used in the absence of surrounding context. This is a familiar principle in linguistics:

The occurrence of a unit (e.g., a sound, word) is partly or wholly determined by its context, which is specified in terms of the unit's relations, i.e., the other features with which it combines as a sequence. The everyday sense of the term [context] is related to this, as when one "puts a word in context," in order to clarify the meaning intended, as in dictionary entries. Providing a context in this way is referred to as contextualization. Words, it is suggested, have meaning only when seen in context. (Crystal, 1991, pp. 78-79)

In defining a highly polysemous word such as so, most dictionaries sort out the many different meanings by contextualizing them in illustrative sentences. Similarly, pronouns cannot be interpreted out of their
context, especially in natural dialogue. The same is true for larger units as well (e.g., “Can you drive?” could be a factual question about ability to drive or an indirect speech act about willingness to drive). The necessity for contextualization is a central tenet of our integrated message model.

When explicating the meaning of a visible act of meaning, we use its immediate linguistic context, which consists of not only its precise place in the conversation but also all of the co-occurring linguistic acts (words, prosody, facial displays, and gestures). For example, two participants had been asked to plan a meal of disliked foods (Chovil, 1991-1992, p. 184): One of them squinted her eyes and wrinkled her nose while she said “Basic steamed white rice.” In the context of this task, we would paraphrase her facial display as meaning “I don't like [basic steamed white rice].” In many cases, the package is more complex; recall the Italian train station excerpt above, which we would paraphrase as conveying the following meaning: “He's looking at the different trains moving in different directions, trying to decide which train he should take—you know what that's like.”

The first procedural requirement for reliably explicating messages in face-to-face dialogue is high-quality video equipment for both filming and repeated replay. These are essential because, valuable as video records are, they are intrinsically inferior to the participants’ view. Video records are two dimensional, whereas gestures occur in three dimensions. The images of face and gesture are smaller (especially in split-screen) and less immediate than they are face-to-face. All of these problems are compounded when there are technical limitations such as inadequate lighting, poor resolution, small screen, or black-and-white recording.

Even when, in spite of these technical obstacles, the analyst can apprehend the speaker's meaning in real time, many more viewings are needed to articulate what was said and how. There are two reasons for this. First, as outsiders, we are overhears rather than addressees (Schober & Clark, 1989) and therefore lack the mutual grounding that the original interlocutors achieved in the course of their dialogue (Clark & Wilkes-Gibbs, 1986). Schober and Clark (1989) showed a significant difference in the apprehension of verbal material by addressees versus overhears; the same should be true of visible acts of meaning. Second, unlike the interlocutors, the analysts must articulate metalinguistically the meaning of these acts; the participants simply had to respond appropriately and move on. This is directly analogous to the verbal case in which a recipient can understand exactly what the speaker means by “I wish he were here” but does not need to articulate metalinguistically how this meaning was achieved (e.g., that the verb is in subjunctive mood and that the referents for he and here must be found in previous context). In our experience, the effort required for microanalysis is more than repaid by the impressive regularity and precision of behavior when viewed at this level.

Our research has consistently demonstrated high reliability for the interpretation of meaning. To identify interactive hand gestures (Bavelas, Chovil, Coates, & Rau, 1995; Bavelas et al., 1989, 1992), we use a paraphrase of the meaning of the gesture to reach a decision about which of two functions the gesture is serving (i.e., a topic gesture referring to the topic of conversation vs. an interactive gesture referring to the addressee). About a dozen different analysts, who have varied widely in their knowledge of gestures, have achieved reliabilities averaging 90% for whether gestures were topical or interactive. This decision can only be made by explicating the meaning of the gesture, even though the specific paraphrase itself was not the final product. Bavelas et al. (1992) have also reliably assessed the degree of redundancy of a gesture with the accompanying verbal phase, which requires a precise explication of the meaning of each.

Faces at first appear much more difficult. However, studies by members of our research group illustrate the ways in which such analysis can be implemented as well as the high reliabilities that can be achieved after careful training and viewing. Chovil (1989, 1991-1992) developed the first system for explicating the meaning of facial displays in face-to-face dialogue. She obtained and analyzed almost 1,200 facial displays from pairs of participants engaged in a variety of conversational tasks. To test the reliability of her system, an independent analyst viewed approximately 20% of these displays, using the following procedure for each display. The first step was repeated viewing of each display in its linguistic context. Operationally, linguistic context means (a) the particular moment in the conversation—what has been said previously, what they are talking about in general, and where they are in their talk (e.g., an explanation, correction, digression, etc.) and (b) the simultaneous meaningful audible and visible acts (the accompanying words, intonation, and gestures). Again, this is exactly analogous to the way in which we understand words. To know the meaning of the word run at a particular point in a conversation, it is necessary to know its exact linguistic context. Then the analyst explicated what the facial display was depicting, for example, looking quizzical (by narrowing the eyes) or emphasizing a particular word (by flashing the eyebrows).

Based on this interpretation, the analyst classified the display according to its general linguistic function. There were five mutually exclusive and exhaustive functions: Speakers made syntactic displays, which conveyed grammatical information; speakers also made semantic displays, which were of two kinds—redundant or nonredundant, depending on whether they added information to the words; there were also nonredundant semantic displays by the listener (who was, by definition, not speaking, so there could be no redundant listener displays);
finally, both speakers and listeners made adaptors (such as blinking or licking the lips), which were not acts of meaning and not analyzed further. For the three semantic functions, the classification was mediated by the analyst's paraphrase (a quizical look might be explicated as "I'm not sure I understand you"). Syntactic displays did not admit paraphrasing because they do not have semantic content; they were described as, for example, stressing a particular word or phrase. Independent agreement on the five general functions was 97% for over 200 sampled displays.

Next, the analyst assigned the display to a more specialized function, again based on its meaning. For example, syntactic displays were divided into emphasers, underliners, question markers, story announcements, and so forth. The redundant semantic speaker displays included personal reactions, portrayals (such as the narrator portraying his brother looking at the trains), facial shrugs, agreements, and so forth. The nonredundant semantic speaker displays included personal reactions (e.g., the speaker's opinion of basic steamed white rice), thinking/remembering displays, metacommunication, and so forth. Listener displays conveyed information such as understanding or personal reactions (e.g., surprise). Reliability for assignment to specific function was 89% for the sample.

Thus, Chovil's system is a highly reliable one for describing facial displays at the level of meaning, and it reveals the wide variety of ways in which the face is used during dialogue. Two limitations of the present system should be noted here. First, although a variety of tasks were used, it is possible that new tasks might generate some further specific functions. Second and more important, smiles were not included, although they could be added relatively easily (for example, by building on Ekman's, 1985, description of 18 different kinds of smiles). Anyone who learned Chovil's system first would be able to elaborate it for new kinds of data.

Coates (1991) also used paraphrasing to identify the audible and visible acts involved in signaling spontaneous ironic humor (i.e., irony or sarcasm in dialogue). Coates treated all meaningful audible and visible acts occurring at a particular moment as a whole. Using repeated viewings and the linguistic context as did Chovil, the independent analyst had to decide first whether the message was mock or serious, for example, whether "Gee, thanks!", said with a smile, was sincere or sarcastic. Agreement for this decision was 100%. Then she had to select a particular meaning from seven serious alternatives and seven mock alternatives, for example, serious excitement, naivety, horror, or politeness; mock excitement, naivety, horror, or politeness. Agreement on the particular meaning was 88%. Again, independent analysts agreed well on the specific meaning of visible acts in their linguistic context.

The above methods of analysis suggest that it is possible to cope reliably with the apparently qualitative and open-ended nature of a paraphrased or explicated meaning. The reader may have wondered at the outset how it is possible to assess agreement for two or more paraphrases—does "It's great!" mean the same as "I love it"? However, as seen in the above examples, agreement on the exact wording of the paraphrase may not be needed. Chovil's (1991-1992) and Coates's (1991) systems finessed the problem by using a judgment just above the level of the paraphrase (although it is worth noting that the paraphrases of the two independent analysts were often virtually identical). Our identification of interactive gestures (Bavelas et al., 1992, 1995) was also based on meaning, but reliability was assessed at a more abstract level. In all three systems, the goal was to identify the conversational function of the acts, and the paraphrases were a means to that end. However, if raters were not interpreting the facial display or gesture in the same way, they would be much less likely to agree on its function, even though it was not necessary to examine the precise words they used to express that interpretation.

The paraphrasing issue would become important for researchers whose primary interests are in the interpretations themselves. Then it would be necessary to demonstrate directly that two independent observers have explicated the same meaning from a message. The key question would become, given their verbal paraphrases, What is a match? (We should reemphasize here that the same problem holds for paraphrasing verbal material. If we asked readers to paraphrase a complex sentence rather than repeating it word-for-word, two individuals could produce paraphrases that say essentially the same thing but do not use the same words.) One solution to this problem would be a multiple-choice technique in which the second analyst sees the entire array of paraphrases that the first analyst generated from the data set and must choose the one that best fits each particular instance. If one were interested in testing the equivalence of completely open-ended paraphrases by two analysts, the solution would be to use a second set of independent judges for the paraphrases themselves, their task would be to decide "What is a match"—that is, whether any given pair of paraphrases is effectively equivalent.

Another potential concern is whether this kind of scoring is particularly vulnerable to bias. The bias that might be introduced by awareness of hypothesis or condition can and must be eliminated or assessed by standard procedures (including "blind" scorers). However, Bavelas et al. (1992) found that degree of awareness of hypothesis did not affect degree of interjudge agreement on interactive gestures. Some readers might be concerned about whether the training itself is a kind of bias, that is, whether highly trained decoders are fundamentally different from naive decoders. We think not. Based on some of the evidence below,
we assume that trained decoders are tapping into their everyday ability to make sense of what other people say and do. As noted at the outset, training is required primarily to deal with the technical setting (e.g., repeated viewings of a video rather than actual interaction) and to elicit metalinguistic responses, which are not common in everyday interaction. People can follow a conversation perfectly well without verbalizing all of their understandings about the conversation; the training is in how to verbalize.

**DEMONSTRATING THAT RECIPIENTS USE THE INFORMATION**

It is necessary and encouraging to demonstrate that analysts can reliably explicate the same, often subtle meanings from momentary hand and facial gestures. A different question is what ordinary recipients make of these acts, whether they are actually extracting the information that we, as experts, say is there (Krauss, Morrel-Samuels, & Colasante, 1991). We describe here some behavioral approaches to verifying that recipients have received and processed the meaning that we propose is in these acts; see Kendon (1994) for a detailed review of the evidence that hand gestures communicate.

Shea and Rosenfeld (1976) found that raters could reliably guess whether learners doing a word selection task gave a correct or incorrect answer by observing nonverbal reactions of the teachers. This suggests that the teacher's nonverbal messages conveyed "That is a correct answer" or "That is a wrong answer." Rosenfeld, Shea, and Greenbaum (1979) also found that specific actions were associated with teacher responses that followed right and wrong answers of learners.

In a series of studies aimed at testing whether motor mimicry was a communicative act, the meaning that naive observers inferred from various forms of motor mimicry was examined in some detail. In one study, Bavelas et al. (1986) showed naive decoders videotapes of observers' facial reactions to an apparently serious injury and found that mimetic responses such as wincing were reliably rated as concerned and aware. In another series of studies, Bavelas et al. (1988) asked a large number of respondents how they would interpret specific mimetic actions that were either enacted for them or presented in photographs. There was consistent agreement on the meaning of particular forms of bodily mimicry.

In studies of equivocation, Bavelas et al. (1990a, chap. 7) developed another method for obtaining decodings of meaning from naive judges. First, we created experimental situations in which participants would spontaneously describe something on a continuum. For example, the condition of a car that was for sale could range from good to poor. Then we played these messages for naive judges and asked them to place the message on a scale ranging from very good to very poor, depending on what they thought the message said about the condition of the car. Reliability for these decoded meanings was very high, even for messages that might be considered incongruent or bizarre.

Finally, arguably the most appropriate assessment of these acts is the effect of a visible act of meaning on the immediate addressee in face-to-face interaction. Such acts are, after all, generated for the benefit of and in collaboration with the addressee vis-à-vis and not for third parties who view them later. As noted at the beginning of this section, we would expect an overhearer effect (Scholz & Clark, 1989) for visible as well as audible acts. That is, outside observers should always be less accurate than addressees. We know of one study that examined experimentally the understanding of addressees and two others that established it by microanalysis of the actual interaction.

Graham and Argyle (1975) asked students at Oxford or the University of Milan to describe two-dimensional shapes with or without gestures and then tested the accuracy with which their addressees could draw the figure. Accuracy scores were significantly higher (for both cultural groups) when the speaker gestured than when he could not. The effect was even greater when the figures being described were more difficult to encode verbally.

To study interpersonal conflict, Camras (1977) filmed the spontaneous interaction of pairs of kindergarten children to whom she had given a caged gerbil that only one child could play with at a time. Camras analyzed not only verbal acts but also facial displays, especially aggressive or threatening expressions (e.g., face forward, with lips pressed together and brows lowered or oblique). She found significant effects of such facial displays on the other child's actions. For example, if one child tried to take the gerbil and the child in possession responded by looking aggressive, then the first child was more likely to stop and wait longer before trying again; if there was no aggressive facial display, the other child tried again sooner.

In another study, Bavelas et al. (1995) identified 12 different functions that interactive gestures could serve and then analyzed our full data set of 464 such gestures for the function each served in its particular context. Next, we predicted what the response of the addressee should be, if the interactive gesture were functioning as we hypothesized. Finally, a new group who were completely blind to our hypotheses microanalyzed and classified the immediate response of the addressee to a random sample of 88 of the gestures. Our predictions were strongly confirmed; for example, the gestures seeking help with a word search elicited help from the addressee, even though no request had been made verbally.

Thus, there are several highly reliable methods for establishing the meaning of hand and facial gestures in their immediate linguistic context, whether they use expert analysts or naive recipients. We hope
that more methods and many further studies will be added in the future.

**VISIBLE AND AUDIBLE ACTS OF MEANING ARE INTEGRATED AND MAY BE EITHER REDUNDANT OR NONREDUNDANT**

Our emphasis on the importance of the simultaneous linguistic context for understanding the meaning of visible as well as audible acts makes it clear that we consider these actions to be completely integrated with each other. These acts form a single, often complex package to be treated as a whole (Bavelas et al., 1990a; chaps. 6, 7; Birdwhistell, 1968, 1970; Clark, 1996; Engle, 1998d; Engle & Clark, 1995; Leeds-Hurwitz, 1989; McNeill, 1985, 1992; Pike, 1972; Poyatos, 1980; Sanders, 1987; Scherer, 1986; Slama-Cazacu, 1976; Streeck & Knapp, 1992).

We should emphasize one last time that the subset of acts being considered here is limited to those that are integrated with the immediate talk-in-progress. We are interested in the momentary convergence of a particular combination of words, intonation, stress, gesture, and face to convey the speaker's meaning to an addressee. As will become clear, because of our dual emphasis on both meaning and momentary acts, we are referring to different phenomena than are researchers who study separate channels (e.g., for nonverbal leakage or clues to deception; Ekman & Friesen, 1969a, 1974). The latter use nonlinguistic analysis by experts or detection by receivers, and they usually study units that are more global (e.g., total amount of asymmetrical smiling; Ekman, Friesen, & O'Sullivan, 1988). Those approaches are better seen as orthogonal rather than opposed to our own.

What may be the best evidence of the precise integration of audible and visible acts is easily demonstrated: Gestures are usually coordinated to verbal syntax. McNeill (1985) observed that "gestures synchronize with parallel linguistic units [and] almost never cross clause boundaries" (pp. 360-361). A speaker says, "My Newfoundland dog has a huge head" and (at the point italicized) gesturally encompasses an area about the size of a basketball. Notice, first, that the gesture specifies the adjective *huge* as being at the outer limit of the potential range for dogs. Second, if the speaker makes this gesture spontaneously, the hands will reach the peak of the gesture precisely with the adjective *huge*, which the gesture specifies. To accomplish this, the preparatory phase (e.g., lifting the hands from the lap) had to be well underway earlier in the sentence, which suggests a high degree of coordination and integration. Indeed, it is difficult to suppress this coordination. That is, it takes deliberate effort to shift the gesture to an earlier, irrelevant word (e.g., *my*) or to withhold it until the sentence has been completed. In either case, the result is so bizarre as to be humorous—even though the difference in timing is less than a second. Similarly, dubbed films (e.g., English for the original French) usually render gestures meaningless because the dubbing changes the spoken order of parts of speech: A gesture describing the object of a verb appears before the verb in French (both verbally and gesturally), but this synchrony is destroyed when the English translation puts the object after the verb.

Engle's (1998a, 1998b, 1998c, 1998d; Engle & Clark, 1995) intensive analyses of multimodal signals (speech, gestures, diagrams, and object demonstrations) have provided several lines of evidence that iconic and indexical conversational gestures are integrated with speech. For example, the gesture and immediately accompanying speech segment were coexpressive, referring to the same underlying referent:

For all but one of the 108 nonverbal signals, a co-expressive speech segment could be found within [a] two intonation unit time window. . . . In stark contrast to communicative nonverbal signals, in 14 of [the] 17 noncommunicative cases, no co-expressive speech was present. (Engle, 1998d, pp. 323-324)

Moreover, the gestures were also consistent with the coexpressive speech, although they were sometimes complementary rather than duplicating the speech. One implication of these findings is that timing is a metacommunicative tool that speakers use to signal what is in the same integrated unit (Engle, 1999).

Sometimes the simultaneous audible and visible elements of a message, taken separately, might appear to contradict each other. However, as Sanders (1987) pointed out, receivers integrate these apparent contradictions at the level of overall meaning (rather than at the level of components or physical source). For example, Sanders noted that Bugental, Kaswan, and Love (1970) found positively valued utterances paired with a negatively valued facial expression and vocal qualities were judged by respondents to be sarcastic. Negatively valued utterances paired with positively valued nonverbal displays were judged to involve joking. Thus, these inconsistent pairs of utterances and nonverbal displays received a single unitary interpretation distinct from the interpretation of either constituent (italics added), not a preference for one rather than the other of two discrete messages. (Sanders, 1987, p. 142)

These interpretations are consistent with Engle's (1998d) proposal that, following Grice's (1967/1989) cooperative principle, both speaker and addressee assume that "all signals in a particular composite signal are intended to be treated from the start as contributing to a single, unified interpretation" (Engle, 1998d, p. 321).

Bavelas et al. (1990a; Bavelas, Black, Chovil, & Mullett, 1990b) tested the hypothesis that participants integrate audible and visible information, using brief real messages that were true, false, or equivocal.
Each message, originally delivered in face-to-face interaction in response to a question, was converted into seven different versions:

1. an edited written version (with disfluencies and prosodic features removed);
2. an unedited written transcription;
3. an audiotaped version with the latency between question and message edited out;
4. an audiotaped version with the latency left in;
5. a videotaped version with the latency edited out;
6. a videotaped version with the latency left in (what the addressee actually saw and heard); and
7. a videotaped version with no sound.

Naive decoders, each randomly assigned to only one version, rated the meaning of each message on a continuum of meaning (e.g., what the message said about the condition of a car or the quality of a class presentation, on a continuum from very good to very poor). Agreement among raters was very high, and there was no effect of version on decoded meaning. Even for equivocal and deceptive messages, the written, audiotaped, and videotaped versions of a message all conveyed the same meaning to independent groups of decoders. If words, paralinguistic features, and visible acts potentially send different messages (especially when deceiving), decoders should have received different messages depending on which they had access to.

REDUNDANCY AND NONREduDANCy

One question that still arises is the extent to which the information conveyed through one act overlaps with information conveyed through other acts: Are facial and hand gestures merely redundant with words? Note that, in the examples already given in this article, some visible acts conveyed their own specialized information, whereas others were fairly redundant with the accompanying words and phrases. These examples illustrate that redundancy is a continuum, ranging from completely redundant (e.g., saying “It goes in a circle” while drawing a gestural circle) to completely nonredundant (e.g., saying “It goes like this” while drawing the circle).

Before considering the empirical evidence, we need to clarify the meaning and function of redundancy in language. The term redundant is often taken to mean unnecessary; in other words, the redundant information merely duplicates what is already there. As Birdwhistell (1970) noted, however, the term was given a new technical meaning by Shannon and Weaver (1949), namely, to describe signs or behaviors that serve to reduce the ambiguity of a message. In information theory, redundant acts are not viewed as unnecessary but rather as serving to increase the likelihood of a correct decoding of the meaning. Cherry (1957) pointed out the implications for spoken versus written language:

When we speak to a friend, we carefully construct our words and phrases, building in redundancy, as we judge necessary for him to understand; with speech this is a running affair, because we are watching and listening to his reactions, and redundancy may be put in, in a changing manner, moment by moment. . . . Writing must make up for the lack of gesture or stress, if it is to combat ambiguity, by introducing redundancy through a wider vocabulary and closer adherence to grammatical structure. (p. 120)

As Linell (1982) pointed out, it is characteristic of written language to be more explicit than spoken language. In our view, explicit means that the redundancy occurs only verbally, whereas less explicit means that the redundancy often shifts to other-than-verbal means (e.g., a gesture or prosodic emphasis on a particular word). Thus, when a perfectly clear face-to-face dialogue is transcribed, the text is usually hard to follow and often makes the speakers seem disfluent and inarticulate because they were using visible rather than audible redundancies, which disappear in transcription.

The use of more than one act to convey an idea helps to clarify the exact meaning being conveyed. A common function of facial and gestural acts is to particularize categories given verbally; they act as unspoken phrases modifying the general terms that appear in words. For example, a narrator recalled “and then the Coyote tries to hypnotize the Road Runner” while (at the words italicized) using her hands and face to illustrate the particular kind of hypnotic action depicted in the cartoon, which was an exaggerated, “show-biz” style (both hands pointed out, with waving fingers, accompanied by a rather demented look). From the words alone, the style could have been inaccurately taken to be another stereotype of hypnosis (e.g., swinging a watch like a pendulum) or no hand action at all.

Our research on the redundancy of audible and visible communicative acts includes analyses of both facial displays and gestures. Using the system described in the previous section, Chovil (1991-1992) found that of the 405 semantic displays by speakers, 60% were to some degree redundant with verbal content, whereas 40% conveyed information that was not in the accompanying words. In addition, all listener displays were, by definition, nonredundant; they permitted the listener to interject a comment without becoming the speaker. Syntactic displays were also usually nonredundant with words.

Bavelas et al. (1992) developed a system for measuring the degree of redundancy between a gesture and the verbal parts of its accompanying phonemic clause. The meanings conveyed by the words and by the gesture were itemized separately and then compared for overlap, with the priority given to words. That is, if the words conveyed “small-sized,”...
the hands can easily depict a concrete spatial relationship between two objects or a metaphorical relationship between two levels of a hierarchical organization. Facial displays cannot readily convey any of the above information but are particularly eloquent for conveying information about people’s reactions (past or present, self or other), for metaphors based on these reactions, and for very rapid syntactic movements, such as eyebrow markers.

In speculating about different advantages, however, we do not mean to fall back into equating function with physical source. Spontaneous face-to-face dialogue is wonderfully “opportunistic” (to borrow a term from Schober & Clark, 1989, p. 229; see also Clark, 1996, pp. 167, 299). The interlocutors use whatever works at the moment, and they are quite inventive about interchangeability. If their hands are not free, they point with their head or shift to words; if they cannot find the precise word, they invent a composite of a poor word plus a visible depiction; if their addressee cannot see them, they put their reaction into words or sounds rather than depicting it facially. Phillips and Bavelas (in press) have recently shown that, even with the absence of nonverbal turn signals in computer-mediated communication, the interlocutors quickly invented ways to use spacing and timing to organize their dialogue; their improvised methods were more efficient than externally imposed turn markers. Similarly, in face-to-face dialogue, speakers choose spontaneously and opportunistically among the audible and visible alternatives available, and listeners will accurately decode the resulting message with little regard for where the information came from physically. As analysts, we can divide their acts into verbal and nonverbal (which usually means into audible and visible) just as we can divide their words into parts of speech. The participants probably do not make these distinctions; they send and receive integrated messages.

NEW DIRECTIONS

We have proposed here that a subset of nonverbal behaviors—especially the interlocutors’ moment-by-moment hand and facial gestures—are visible acts of meaning and should be treated as part of natural language in face-to-face dialogue. In the traditional approach, the physical source of a message determines its conceptual status. Meaningful acts that happen to be written or spoken are deemed language or verbal communication, whereas tightly coordinated acts that are visible bodily movements are relegated to a different domain (the so-called nonverbal channel) with different functions. Given these assumptions, the two domains are almost always analyzed separately and often in different scholarly disciplines. Physical characteristics determine theoretical status. If we take a more abstract, functional approach, then
physical characteristics are no more important than the number or shape of letters in a word. The interesting questions become, What are the functions of the (whole) message? and How do the parts accomplish those functions? That is, we can choose to start with the assumption of integration rather than with the assumption of separateness.

We have set out four criteria (and related evidence) for distinguishing between visible acts of meaning and other nonverbal behaviors and for treating them instead as integrated with verbal language acts. Because one of our goals is to generate further research, we will summarize these criteria with a focus on the future, by reemphasizing the new kinds of research questions we have suggested here.

1. Studies of sensitivity to a receiver can now go beyond simple visual availability due to the medium of communication (e.g., telephone vs. face-to-face conditions). Also, several studies have shown that visual availability can be manipulated or assessed as a momentary phenomenon within an interaction. The sensitivity of senders to these micromoments is a more precise and appropriate demonstration than are monolithic condition effects because they suit the time scale of visible acts of meaning. Furthermore, presence versus absence of a receiver can be treated as a psychological rather than a literal physical variable (e.g., the effects of an imaginary audience). On the dependent variable side, we can examine not just the frequency (presence or absence) of the acts but how their form changes when they are communicative. For example, what are the precise differences in how the hand is used when one is literally tracing a route on a map, or retracing it gesturally to prompt oneself, or tracing it to give directions for an addressee?

2. The symbolic nature of visible acts of meaning could be better understood by focusing more often on metaphoric facial and hand gestures rather than solely on those that depict a concrete referent. Such research would require using different eliciting tasks or stimuli as well as being more sensitive to when visible acts of meaning are metaphoric rather than literal. We have also implicitly raised questions about the utility of distinguishing between visible and audible symbols on the basis of how they are encoded, especially if they are treated by speakers and addressees as an integrated whole. It may be far more fruitful to study similarities (e.g., similarities between metaphoric gestures and figurative language or similarities between pointing and anaphora) than to divide them into verbal and nonverbal means of signaling.

3. We spent a great deal of time on the evidence that the meaning of these visible acts can be explicated or demonstrated in their linguistic context because this evidence raises important issues about traditional criteria for objectivity in research. Some critics would refer to our approach as interpretive, hermeneutic, or subjective with the implication that only the physical act—and not its meaning—can be scientifically studied. The answer to this criticism is high interanalyst reliability, and it is in fact often higher for the interpretation of meaning than for the counting of physical movements. The hard work of achieving such reliability is fully repaid by the way in which it liberates us to study human communication in the fullest sense—at the level of meaning.

4. The nature of redundancy, nonredundancy, and integration is only beginning to be understood. We understand how verbal syntax integrates the sequential parts of an utterance, but we have little knowledge of a communicative syntax that integrates the co-occurring audible and visible acts. Clearly, timing, meaning (e.g., coexpressivity), and indexical expressions ("like this") play a role. However, we need to know more about timing (and mis-timing), about highly nonredundant or even inconsistent meanings of components, about the rate and use of indexical expressions, and especially about other devices that hold even very complex meanings together. If, like the interlocutors, we start with the assumption of integration, we may begin to appreciate and investigate the ways in which they achieve what Clark (1996, p. 156) called their "artful fusion."

NOTES

1. Readers familiar with the first author’s earliest work (Watzlawick, Beavin Bavelas, & Jackson, 1967) will notice several differences in the approach to nonverbal communication—as one would hope after 30 years of research.

2. For this reason, we will not be discussing the actions that Efron (1941/1972) and Ekman and Friesen (1969b) have called emblems, such as the hitchhiking sign, which are stereotypic actions used in non-speaking settings, although these are clearly visible acts of meaning.

3. Roberts and Bavelas (1996) have explicitly added the addressee’s understanding as an essential criterion as well and have explicitly included gestures and facial displays.

4. In our framework, the terms communicative and noncommunicative are more appropriate than posed and spontaneous.

REFERENCES


