Do Conversational Hand Gestures Communicate?

Robert M. Krauss, Palmer Morrel-Samuels, and Christina Colasante
Columbia University

In 5 experiments, male and female undergraduates viewed gestures and tried to select the words that originally accompanied them; read interpretations of gestures' meanings and tried to select the words that originally had accompanied them; tried to recognize gestures they previously had seen, presented either with or without the accompanying speech; and assigned gestures and the accompanying speech to semantic categories. On all 4 tasks, performance was better than chance but markedly inferior to performance when words were used as stimuli. Judgments of a gesture's semantic category were determined principally by the accompanying speech rather than gestural form. It is concluded that although gestures can convey some information, they are not richly informative, and the information they convey is largely redundant with speech

Although spontaneous hand gestures[1] are ubiquitous accompaniments to conversational speech, neither the process by which they are generated nor the functions they serve for speaker and addressee are well understood. Such gestures are usually assumed to be communicative. However, among researchers and writers who have addressed the subject, there is no clear consensus about the amount and kind of information they convey. Historically, gestures have been endowed with extraordinary communicative powers. Some of the claims seem fanciful at best. Consider the contention of the 16th century French philosopher and essayist Michel de Montaigne:

> There is no motion, nor jesture, that doth not speake, and speakes in a language, very easie, and without any teaching to be understood: nay, which is more, is a language common and publike to all: whereby it followeth (seeing the varietie, and severall vse it hath from others) that this must be deemed the proper and peculier speech of humane nature (Montaigne, cited in Bevington, 1984).

Two centuries later, the German philosopher Artur Schopenhauer observed that

> Natural *gesticulation*, such as commonly accompanies any lively talk, is a language of its own, more widespread even, than the language of words—as far, I mean, as it is independent of words

and alike in all nations (Schopenhauer, 1903, p. 109. Italics in original).

More recent assessments have been considerably less extravagant, but typically they have concluded that one of the important functions of gestures accompanying speech is to convey information to a listener (Argyle, 1974; Birdwhistell, 1970; Cohen, 1977; Graham & Argyle, 1975; Kendon, 1983; Slama-Cazacu, 1976; for dissenting views see Feyereisen, van de Wiele, & Dubois, 1988; Moscovici, 1967; Rimé & Schiaratura, 1991). For example, Birdwhistell (1970) argues that gestures visibly replicate information conveyed in the speech channel, thereby contributing to the redundancy of the accompanying speech. He also suggests that some gestures may be the structural equivalents of certain linguistic forms and may be used as substitutes for them. Kendon states that

> Gesticulation arises as an integral part of an individual's *communicative* effort and that, furthermore, it has a direct role to play in this process. Gesticulation . . . is important principally because it is employed, along with speech, in fashioning an effective utterance unit (Kendon, 1983, p. 27. Italics in original).

Despite the pervasiveness of the assumption that gestures play a significant role in communication, the supporting empirical evidence is surprisingly tenuous. Two kinds of studies are cited in support of the assumption (cf. Bull, 1983; Kendon 1983):

1. *Studies of the effects of visual accessibility on gesturing.* A number of studies have found that speakers gesture more in

[1] Throughout this article, we use the term *gesture* to refer to spontaneous hand gestures that accompany speech—what have been termed "illustrators" (Ekman & Friesen, 1972), "gesticulations" (Kendon, 1983) or "speech-focused movements" (Butterworth & Beattie, 1978). These should be distinguished from gestural signs (e.g., the "thumbs-up" sign) that have conventionalized meanings and can be used in the absence of speech. The latter, variously referred to as "emblems" (Ekman & Friesen, 1972) or "autonomous gestures" (Kendon, 1983), clearly serve a communicative function.

face-to-face conversation than they do when their addressee cannot see them (Cohen, 1977; Cohen & Harrison, 1973). For example, Cohen and Harrison found that speakers giving directions used about twice as many gestures face-to-face as they did when communicating over an intercom. Data collected in our laboratory has shown that speakers describing abstract figures or complex synthesized sounds use 12% to 15% more gestures per unit of time in the face-to-face condition. Rimé (1982) found only a marginally significant difference between the two conditions.

Cohen and Harrison (1973) and others interpret the greater frequency of gesturing when speaker and addressee can see each other as evidence for the proposition that such gestures are used communicatively. However, all of the relevant studies report that a substantial amount of gesturing occurs when the listener cannot see the speaker. Such gestures may simply reflect long-standing habits, or they may be different in some fundamental way from the kind of gestures used in face-to-face interaction. Still, it is difficult to see how the relatively small difference in gesturing between face-to-face and voice-only situations provides strong support for the proposition that the primary function of such gestures is communicative.

2. *Studies of the effectiveness of communication with and without gesturing.* A number of investigators have reported that addressees' comprehension is greater when they can see their speakers' gestures (Berger & Popelka, 1971; Graham & Argyle, 1975; Riseborough, 1981; Rogers, 1978). In the most frequently cited of these, Graham and Argyle had speakers describe abstract line drawings to a small audience of listeners who then tried to reproduce the drawings. For half of the descriptions, speakers were allowed to gesture; for the remainder, they were required to keep their arms folded. Graham and Argyle found that audiences of the nongesturing speakers reproduced the figures somewhat less accurately. Although this finding is consistent with the proposition that gestures are communicative, it is also possible that prohibiting speakers from gesturing affected their speech production. Speakers who were allowed to gesture may have produced better verbal descriptions of the stimuli, which, in turn, would have enabled their audiences to reproduce the figures more accurately.[2] Other studies attempting to test the effectiveness of communication with and without gestures were designed in ways that make it difficult to draw unequivocal conclusions.[3]

In two of our own studies, audiotapes of people describing nonsense figures in a referential communication task communicated as well as videotape presentations of the same people talking and gesturing (Dushay, 1991; Krauss, Dushay, Chen, & Bilous, 1991). Moreover, an extensive series of studies by the British Communication Studies Group concluded that people convey information just about as effectively over the telephone as they do when they are face-to-face with their coparticipants (Short, Williams, & Christie, 1976; Williams, 1977).[4] Although someone speaking to a person he or she cannot see may compensate verbally for information that would ordinarily be conveyed by gestures (and other visible displays), it may also be the case that the contribution such information typically makes to communication is of little consequence. In any event, the communicative value of these visible displays has yet to be demonstrated conclusively.

Despite the absence of compelling evidence, to many, the communicativeness of gestures is self-evident. Gestures communicate, it is contended, because, like the words they accompany, they convey semantic information (cf. Kendon, 1983; McNeill, 1985, 1987). And indeed, from the perspective of a naive observer, gesture and speech do seem to convey closely related meanings. When both are available, it is not difficult to see how a particular gesture might be meaningfully related to a particular word or phrase. However, it may be that much of the gesture's meaning is illusory. In the absence of speech, the very same gesture's meaning can be quite opaque, communicating little, if anything. It is at least possible that the apparent relationship of gestural meaning to the meaning of the accompanying speech is a post-hoc construction deriving primarily from the listener–viewer's comprehension of the speech and bears no systematic relation to the movements observed.

Surprisingly, there have been very few attempts to determine how accurately or consistently naive observers can interpret gestures. In a recent experiment, Feyereisen et al. (1988) showed to subjects some videotaped gestures excerpted from classroom lectures along with the following three possible interpretations of each gesture: the word(s) in the accompanying speech that had been associated with the gesture (the *correct response*); the meaning most frequently attributed to the gesture by an independent group of judges (the *plausible response*); and a meaning that had been attributed to the gesture by only one judge (the *implausible response*). Subjects tried to select the response that most closely corresponded to the gesture's meaning. Not surprisingly, the plausible response (the meaning most often spontaneously attributed to the gesture) was the one most often chosen; what is surprising is that on average the correct response was not chosen more often than the implausible response.

Taken at face value, these results appear to indicate that although judges exhibit a fair degree of consensus in their interpretations of gestural meanings, the meanings the gestures convey are not the same as the meanings of the words they accompany.[5] However, several limitations of the Feyereisen et

---

[2] Graham and Argyle (1975) raised this possibility but rejected it on the basis of the results of a study by Graham and Heywood (1975). We discuss this study later in this article.

[3] For example, in the Berger and Popelka (1971) study and in Experiments 2 and 3 of Riseborough (1981), the gestures were preplanned and enacted; the relation of such pantomimic enactments to spontaneously generated gestures is unclear. In Rogers (1978) and in Experiment 3 of Riseborough, communicative effects for gestures were found only when noise made the audio track difficult to understand.

[4] More recent research has found some effects attributable to the lack of visual access (Rutter, Stephenson, & Dewey, 1981; Rutter, 1987), but generally speaking these effects involve the perceived social distance between communicators, not their ability to convey information. There is no reason to believe that the presence or absence of gesture per se is an important mediator of these differences.

[5] Feyereisen, Van de Wiele, & Dubois (1988) do not discuss the possibility that meanings of "correct" and "plausible" responses are related in some systematic way. No relationship is evident from an examination of the responses themselves, but since the lectures originally were in French, nonFrancophone judgments of translations may be insensitive to subtle semantic nuances.

al. (1988) experiment suggest caution in generalizing the results: There were only two speakers, the gestures and the correct responses seem to have been selected on an ad hoc basis by the experimenters, and no attempt was made to assess the similarity of the correct and plausible responses and so forth.

In this article, we report five experiments we conducted to examine the information that conversational hand gestures convey to naive observers. In the first two experiments, we assessed the extent to which gestures and the speech they accompany convey the same meanings. In the second pair of experiments, we examined the hypothesis that gestures convey information that is complimentary to (i.e., related to but not redundant with) the information conveyed by speech. The final experiment examines the relative contributions of verbal and visual information to judgments of a gesture's semantic category.

Experiment 1

In Experiment 1, we showed subjects a series of brief videotaped segments, each of which contained a gesture, along with two potential meanings. Their task was to decide which of the two was closer to the gesture's meaning.

Method

Stimuli. In this and the remaining experiments, we used as stimuli two sets of 30 video clips that had been excerpted from longer narratives. The clips ranged from 0.54 to 7.71 s ($M = 2.49$ s; $SD = 1.35$), and each showed a person talking and gesturing. The clips had been selected from a set of videotapes of nine men and eight women who were describing a set of 13 photographic slides depicting a variety of subjects (landscapes, abstractions, buildings, machines, people, etc.) to a female confederate seated nearby who was positioned so that she could see the describer but not the picture. This yielded a total of 221 narrative descriptions and a corpus of 2,328 hand movements whose displacement had been determined by a computer–video interface to have exceeded 1.5 in. (3.8 cm) per 0.8 s (see Morrel-Samuels & Krauss, 1990, for details). On average, describers made about five such hand movements during each 70-s narration.

Naive subjects provided with transcripts viewed the videotapes in groups of 10. The tape was stopped at major constituent boundaries, and subjects indicated whether they had seen a gesture and, if they had, the word or phrase in the accompanying speech they perceived to be related to it. No definition of *related to* was provided, and subjects who asked were instructed to use whatever criterion seemed appropriate to them. We will refer to these words and phrases as the gestures' *lexical affiliates* (Schegloff, 1984). Movements whose lexical affiliates were agreed on by 8 or more of the 10 viewers were considered to be speech-related gestures. From a total of 193 such gesture–lexical affiliate pairs, 60 were selected for use in this (and other) experiments.[6] (See Morrel-Samuels, 1989; Morrel-Samuels & Krauss, 1991, for a fuller account of the process by which the stimuli used in this experiment were selected.) The 60 segments were randomly partitioned into two sets of 30 and edited in random order onto separate videotapes. Each segment showed a full-body front view of a seated speaker who was videotaped at a distance of 18 ft (5.5 m) and displayed half-screen on a 19-in. (48.3 cm) monochrome video monitor; the other half of the screen (which showed a side view of the same speaker) was masked off.

Subjects. Twenty-four college undergraduates (12 men and 12 women) served as subjects. They were tested in groups of 4 in sessions that lasted about ½ hr. Subjects' participation fulfilled a course re-

quirement.[7] As were subjects in the other experiments reported below, all were native English speakers.

Procedure. Seated about 2.5 m from the monitor, subjects viewed the two stimulus tapes played at normal speed but without sound. The order of presentation was randomly determined. After each segment, we stopped the tape and showed the subject two lexical affiliates and asked him or her to choose the one that he or she deemed to be closer in meaning to the meaning of the gesture. One of the two lexical affiliates was the one that had originally accompanied the gesture; the other was the lexical affiliate of a randomly chosen gesture.

Results and Discussion

The mean proportion of raters selecting the correct lexical affiliate (we call this the *identification accuracy* score) was .765 ($SD = .208$), which was reliably higher than the chance level of 50%, $t(59) = 9.86$, $p < .0001$. Although identification accuracy was far from perfect, on average it was quite good: For 93% of the gestures, a majority of subjects selected the correct lexical affiliate; on nearly half of them, at least 90% of the judges made the correct choice.

It is clear that subjects who have seen a gesture can distinguish its lexical affiliate from a randomly selected lexical affiliate with reasonable accuracy. What is less clear is how they accomplished this. One possibility, consistent with the view that gestures are communicative, is that subjects imputed a meaning to the gesture and then chose the lexical affiliate that was closer in meaning. If that is how our subjects performed the task, one would expect their success to be inversely related to the similarity in meaning of the paired lexical affiliates—that is, the task should be more difficult when the lexical affiliates were similar in meaning. To investigate this, we had an independent sample of 12 subjects rate the similarity in meaning of each of the 30 pairs of lexical affiliates on a 7-point scale that ranged from *not at all similar* (1) to *extremely similar* (7). Overall, raters perceived the 30 pairs to be relatively dissimilar in meaning ($M = 2.6$; $SD = 1.07$); no pair received an average rating of 6 or 7, and only 10% were rated at or above the midpoint of the scale, resulting in a distribution with a marked negative skew.

To determine whether accuracy was related to the similarity in meaning of the paired lexical affiliates, we computed the correlation between the mean similarity rating for each pair and their combined mean identification accuracy score[8] and found it to be positive rather than negative, albeit of marginal statistical reliability, $r(59) = .223$, $p < .087$; that is, the likeli-

---

[6] A number of typological schemes differing in relatively minor ways have been proposed for classifying conversational gestures. We have not attempted to code our stimulus gestures into such categories because we believe that the use of all such coding schemes begs important questions about the relation of gesture and speech. However, to give readers familiar with such typologies some sense of what the stimulus gestures were like, it may be helpful to note that all of them fell within the general category of illustrators (Ekman & Friesen, 1972); in our judgment none would have been coded a baton, an ideograph, or an emblematic movement.

[7] This was the case in all of the experiments reported below except where specifically noted otherwise.

[8] We combined the accuracy scores of the two lexical affiliates in the pair.

hood of selecting the correct lexical affiliate was not inversely related to the similarity in meaning of the pair. However, because the distributions of both similarity ratings and identification accuracy were so badly skewed (positively in the first case and negatively in the second), a product–moment correlation probably does not provide a good test of the hypothesis.

To investigate the relationship between the meaning a gesture conveys and the semantic content of its lexical affiliate, we coded each of the 60 lexical affiliates into one of four semantic categories: *locations* (e.g., "There's another young girl to the woman's right" and "passing it *horizontally* to the picture"[9]), *actions* (e.g., "rockets or bullets *flying* out" and "seems like it's going to *swallow them up*"), *objects* (e.g., "scarf or *kerchief* around her head" and "actual *frame* of the window and the Venetian blind"), and *descriptions* (e.g., "one of those *pointillist* paintings" and "which is *covered* with paper and books"). The 60 lexical affiliates were distributed fairly equally among the four coding categories (approximately 33%, 22%, 22%, and 23%, respectively), and two coders working independently agreed on 85% of the categorizations ($\kappa = .798$).

A multiple regression with the semantic category of the two lexical affiliates as independent variables accounts for one third of the variation in similarity ratings of the paired lexical affiliates, $F(2, 27) = 6.53$, $p < .005$; in making their similarity judgments, subjects apparently did take the semantic category of the lexical affiliates into account. Accuracy was greater when the two lexical affiliates came from different semantic categories (78%) than when they came from the same semantic category (65%). Unfortunately, in only 4 of the 30 pairs were both lexical affiliates in the same semantic category; nevertheless, despite this lopsided distribution, a t test of the difference yielded a probability value that was close to the conventional level for statistical significance, $t(28) = 1.96$, $p = .0646$. Identification accuracy did not vary as a function of the correct lexical affiliate's semantic category.

These results provide some evidence that gestures convey meanings that are related to the meanings of the words they accompany. If subjects did not spontaneously impute meanings to the gestures when they performed the identification task, how could they have chosen the correct lexical affiliate as frequently as they did? One possibility is that subjects were able to read the speakers' lips. But the small size of the videotaped image and its low resolution quality makes this unlikely. Another possibility is that subjects used features of the lexical affiliate (e.g., its length) that were unrelated to its meaning but correlated with some property of the gesture (e.g., its duration). We have no direct way of testing whether such strategies were used, but it seems reasonable to suppose that their use would increase over trials, as subjects discovered the difficulty of ascertaining gestural meanings. This would have resulted in a serial order effect for identification accuracy, and we found no such effect.

Experiment 2

Even assuming that our subjects' performance in Experiment 1 derived from the communicativeness of the gestures and not from some artifactual strategy, we find that the results suggest that the communicative value of the gestures is not very great. The task required only that subjects chose between two lexical

affiliates whose meanings, according to our similarity ratings, tended to be quite dissimilar. Even a vague hint as to meaning might have been sufficient to allow the binary choice; yet accuracy, although substantially better than chance, was far from perfect.

To determine whether meanings subjects attribute to the conversational gestures correspond to the meanings of the gestures' lexical affiliates, we had subjects view each gesture and write down their impressions of its meaning (we call these *interpretations*). A separate sample of subjects read each interpretation and rated its similarity to the meanings of two lexical affiliates, one of which had originally accompanied the gesture.

Method

Subjects. Six undergraduates (3 men and 3 women) provided interpretations of the gestures; we refer to these subjects as *interpreters*. Thirty-six subjects (20 men and 16 women) judged how similar the interpretations were to the two lexical affiliates; we refer to these subjects as *judges*.

Procedure. Interpreters were told that they would see a series of videotaped clips of people who were talking and gesturing but that they would be unable to hear what was being said. Their task was to observe the gesture carefully and to try to interpret what the speaker intended it to mean. No definitions were offered for such terms as *intended meaning* or *interpretation*, and none of the subjects seemed to have questions about them. Interpreters viewed the clips individually and the tape was paused after each clip to permit them to write their interpretations. This produced a set of 360 interpretations. Each judge was given the 60 interpretations produced by 1 interpreter, printed one to a page. Also on the page were two lexical affiliates, labeled A and B, and a 6-point scale with poles labeled *very similar to A* and *very similar to B*. One of the two lexical affiliates had originally accompanied the gesture that served as stimulus for the interpretation and the other had accompanied a randomly chosen gesture. Judges were asked to use the 6-point scale to indicate which of the two lexical affiliates was closer in meaning to the interpretation and the degree of the similarity of meaning. Because the scale had no middle category, subjects were forced to select one of the lexical affiliates as more similar. The position on the page of the correct lexical affiliate was randomly varied, as was the order of the pages. Each set of interpretations was rated by six judges, and each judge rated all and only the interpretations of 1 interpreter.

Results and Discussion

We derived two measures from the ratings: (a) a 6-point scale measure for which we transformed the scales so that a rating of 6 indicated a judgment that the interpretation was very similar to the incorrect alternative and a rating of 1 indicated a judgment that the interpretation was very similar to the correct alternative and for which we then calculated the mean rating for each gesture; and (b) a binary accuracy measure for which we dichotomized the ratings into correct (ratings 1, 2, or 3) or incorrect (ratings 4, 5, or 6) values, producing a measure that reflected the choice of the correct alternative (irrespective of the degree of similarity seen between the interpretation and the correct lexical affiliate) and for which we calculated the mean value for each gesture. Because the two measures produced

---

[9] The italicized words are those judged by subjects to be related in meaning to the meaning of the gesture.

identical results, we report results only for the binary accuracy measure.

A 6 (interpreters) × 60 (interpretations) repeated measures ANOVA revealed that interpreters differed marginally in their ability to produce interpretations that yielded the correct lexical affiliate, $F(5, 30) = 2.34$, $p < .07$, but that the 60 gestures differed markedly in the quality of the interpretations they elicited, $F(59, 1770) = 6.53$, $p < .0001$. The two factors also interacted significantly, $F(295, 1770) = 2.97$, $p < .0001$. The mean binary accuracy score was 0.62 ($SD = 0.173$), which is reliably greater than the chance value of 0.50, $t(59) = 12.336$, $p < .0001$.

Accuracy also varied as a function of the lexical affiliate's semantic category. A one-way analysis of variance (ANOVA) of mean binary accuracy scores for the 60 gestures with semantic category as the independent variable yielded $F(3, 56) = 4.72$, $p < .005$. Accuracy was greatest when the lexical affiliates were actions (73%), somewhat lower for locations (66%), and considerably lower for object names and descriptions (57% and 52%, respectively). The first of these values is reliably larger than the last two (Dunnett's $d = 2.61$ and 3.432, respectively, both $ps < .01$); mean accuracy for locations is higher than the mean for descriptions ($d = 2.44$, $p < .01$) and marginally higher than the mean for object names ($d = 1.56$, $.10 > p > .05$). The first two means differ reliably from 50%, $t(56) = 5.29$ and 4.51, respectively, both $ps < .0001$); the latter two do not ($ts < 1$).

The results indicate that gestures elicit interpretations that are more similar in meaning to their own lexical affiliates than to the lexical affiliates of other gestures. However, neither our data nor those reported by Feyereisen et al. (1988) suggest that such gestures are capable of conveying the broad-ranged and richly nuanced meanings commonly expressed in speech. In the present experiment, accuracy averaged only 12% above chance. The interpretations of the best of the 6 interpreters yielded the correct lexical affiliate only 66% of the time, and the highest accuracy score by any judge for any interpreter was only 72%. Thus, although gestures may be a guide to what is being conveyed verbally, it would be difficult to claim on the basis of these data that they are a particularly discriminating guide.

In our situation, gestures were most informative when they referred to actions and locations; they were markedly less effective when they referred to object names and descriptions. In Experiment 1, recognition accuracy did not vary as a function of the correct lexical affiliate's semantic category. We have no clear explanation for this apparent inconsistency in the results of the two experiments. It may be that presenting both interpretations and lexical affiliates in verbal form caused judges to focus more closely on semantic factors. Again, it should be stressed that our test of communicativeness—that is, whether the interpretation enabled a judge to discriminate the correct lexical affiliate from a randomly selected affiliate that, on average, was relatively dissimilar in meaning—was relatively undemanding. With so lenient a criterion, the fact that performance was barely better than chance undermines the plausibility of the claim that gestures play an important role in communication when speech is fully accessible.

It might be argued that conversational gestures are richly communicative but that the information they convey is different from the semantic information conveyed by speech. Although this position has been advanced (cf. Deutsch, 1952;

Feldman, 1959), we are aware of no credible empirical support for it, and the meager evidence that does exist opposes it (cf. DeMeijer, 1989; Graham, Ricci-Bitti, & Argyle, 1975). However, regardless of whether one accepts or rejects the idea that gestures can serve as a second channel of communication, one should remember that the gestures we used as stimuli in these experiments were chosen from a large number of gestures specifically because they were perceived by naive subjects to be meaningfully related to their lexical affiliates. So although it may be that gestures and speech can convey different messages, that is not the case for the ones we used insofar as the perceptions of naive subjects are concerned.

Alternatively, one might contend that although gesture and speech both convey semantic information, they are not completely redundant, that is, that the information conveyed gesturally is related to, but not necessarily the same as, the information conveyed verbally. According to Kendon, "When verbal language is available and is used, gesture is employed in cooperation with it. It serves to complement it in various ways" (Kendon, 1983, p. 38). McNeill (1987, p. 20) asserts "that gestures and grammatical sentence structures compensate for one another" and suggests that a speaker may shift semantic information from speech to gestural form when the grammatical structure of the sentence is particularly complex. Slama-Cazacu (1976) has advanced a roughly similar proposal.

Although this position is not implausible, it is a difficult one to evaluate empirically because there is no objective criterion against which a gesture's informational content can be assessed. Investigators can (and frequently do) attribute complex meanings to gestures on the basis of careful observation of them and their intuitions about the relation of gestural form to meaning, yet we know of no case in which they have tested the validity of their attributions. It is not clear how one would go about doing this. And although we regard descriptions of carefully observed examples (e.g., Kendon, 1972; McNeill, 1987, Chap. 7; Slama-Cazacu, 1976; Sherzer, 1973) as fruitful sources of hypotheses, we find them less than satisfactory as evidence in support of those hypotheses.

Experiment 3

One approach to examining the kinds of information that gestures and speech convey is through the use of a recognition memory paradigm. It is possible to compare recognition accuracy for the lexical affiliates, for the gestures that accompanied those lexical affiliates, and for the speech and gestures combined. If gestures conveyed information that is different from the information conveyed by speech, one might expect that speech and gestures combined would be better recognized than either speech or gestures separately. On the other hand, if gestures simply conveyed a less rich version of the information conveyed by speech, one might expect that adding gestural information to speech would have little effect on recognition memory, compared with memory for the speech alone.

In Experiment 3, we had subjects view, hear, or view and hear taped segments of speech, gestures, or speech and gestures combined. Then, in a forced-choice recognition test, we determined whether they could distinguish previously presented segments from segments to which they had not been exposed.

Method

Stimuli. The two videotapes, each containing 30 speech–gesture segments that were used in Experiments 1 and 2, were used in this experiment. The experiment was performed in two phases: a presentation phase and a recognition phase. For the recognition phase, a tape consisting of 30 pairs of randomly selected gestures was constructed. Within a given pair, the 2 gestures came from different tapes used in the presentation phase; hence, for each pair of segments that was seen or heard in the recognition phase, only 1 segment had previously been presented.

Procedure. Stimuli were presented in one of three modes: Subjects either viewed and heard one of the two presentation videotapes (audio–video condition), viewed the picture with the sound turned off (video-only condition), or heard the sound without seeing the picture (audio-only condition). In the presentation phase, subjects viewed or heard one of the two 30-segment presentation tapes. They were instructed to try to remember each segment and told that they would later be tested to see how well they could recognize them. The presentation phase followed the recognition phase by about 5 min. Subjects viewed or heard the 30 paired recognition segments and tried to select the segment they had seen or heard before.

Subjects. A total of 72 undergraduates (24 in each presentation-recognition condition) served as subjects. Men and women were roughly equally distributed among conditions.

Results and Discussion

The experiment constitutes a 3 (modes: audio, video, and audio–video) × 2 (presentation tape: A vs. B) mixed ANOVA. ANOVA of the proportion of correct recognitions revealed large effects for mode, $F(2, 33) = 7.58$, $p < .002$. Recognition accuracy did not differ for the two presentation tapes either as a main effect or in interaction with mode (both $Fs < 1$). The mean proportions correct in the video-only, audio-only, and audio–video conditions were .733, .826, and .840, respectively. All three values are reliably greater than .50, the expected value if subjects had guessed, $t(59) = 10.46, 15.22$, and 18.49, respectively, all $ps < .0001$). However, subjects who saw (but did not hear) the stimulus segments recognized them less well than subjects who heard (but did not see) them, or who both saw and heard them (Sheffé $F = 5.68$ and 7.50, respectively, both $ps < .001$) Adding gestural information to the speech does not improve recognition; the difference between the audio–video and audio-only conditions (.014) is small and unreliable $(F = 0.126)$.

The pattern of correlations among recognition rates for stimulus segments in the three experimental conditions presents a complex and somewhat surprising picture. Although mean recognition rates in the audio-only and audio–video conditions are nearly identical, the correlation between the two conditions across stimulus segments is relatively small, $r(59) = 0.285$, $p < .03$). Recognition in the video-only condition is uncorrelated with recognition in the other two conditions, with audio-only, $r(59) = -.049$, ns; with audio–video, $r(59) = .153$, ns. The multiple correlation of audio-only and video-only with audio–video is .331, only marginally higher than the simple correlation between audio-only and audio–video. It seems reasonable to infer that the information subjects used to recognize segments in the video-only condition did not play an important role in recognition of audio–video segments, and that this visual information was not redundant with the information that entered into the

recognition rates of subjects in the audio-only condition. Such results would be consistent with the hypothesis that subjects encoded the gestures as inchoate movements rather than as meaningful units.

Accuracy for video-only segments was low relative to the rate for the audio-only segments, and combining speech and gesture in the audio-video condition did not improve recognition. These results are consistent with the notion that, compared with words, the informative value of gestures is relatively low and largely redundant with the information conveyed by speech. However, it would be premature to draw such a conclusion on the basis of these data.

To begin with, despite the fact that all of the information in the audio–video presentations was contained in either the audio-only or video-only presentations, recognition rates in the single-channel (i.e., audio-only and video-only) conditions account for less than 10% of the variance in recognition rates in the audio–video condition, raising questions about how subjects in the latter condition performed the recognition task. One possibility is that they combined the vocal and visual information in a form that is not captured by a simple linear regression model. Or, to put it another way, perhaps the informativeness of any element of vocal or visual information is not independent of the total communicative context. If this were so, our experiment would not be a fair test of the informative value of gestures, because by abstracting gestures from the context of their accompanying speech, we made it more difficult for subjects to ascertain their distinctive meanings and, hence, to recognize them subsequently.

Experiment 4

In Experiment 4, instead of presenting gestures out of their communicative context, we first presented gestures together with the speech they accompanied and then asked subjects to discriminate gestures they had seen before from new gestures.

Method

Stimuli. The presentation and recognition tapes used in Experiment 1 were used as stimuli in this experiment.

Procedure. Like Experiment 3, Experiment 4 was performed in two phases: a presentation phase and a recognition phase. In the presentation phase, all subjects viewed and heard 30 brief excerpts of people talking and gesturing. They were informed of the recognition condition to which they had been assigned and told they would later be asked to distinguish segments to which they had been exposed from new segments on the basis of the video portion only, the audio portion only, or the combined audio–video segment. The instructions strongly stressed the importance of attending to the aspect of the display they would later be asked to recognize. About 5 min after completing the presentation phase, subjects performed a forced-choice recognition test with 30 pairs of segments seen or heard in the mode to which the subject had been instructed to attend. In each pair, one segment had previously been exposed and the other was new.

Subjects. A total of 72 undergraduates (24 in each condition) served as subjects. Men and women were about equally distributed across conditions.
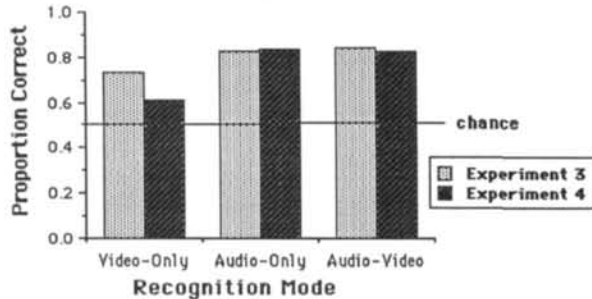
Figure 1. Recognition accuracy in the three recognition conditions for Experiments 3 and 4.

Results and Discussion

A 3 (modes: audio, video, and audio–video) × 2 (presentation tape: A vs. B) × 2 (experiments: 3 vs. 4) mixed ANOVA was performed, combining the recognition data of the two experiments. Large effects were found for presentation–recognition mode, $F(2, 33) = 40.23$, $p < .0001$; experiment, $F(1, 33) = 5.69$, $p < .02$; and the Mode × Experiment interaction, $F(2, 33) = 4.75$, $p < .02$. No other reliable effects were found (all $Fs < 1$). The means for the six conditions are plotted in Figure 1. For the audio-only and audio–video conditions, recognition rates are virtually identical to those found in the previous experiment; hearing speech in its gestural context did not improve subsequent recognition and, again, speech accompanied by gesture was no better recognized than speech alone (both $Fs < 1$). However, in the video-only condition, there were substantial differences in performance across the two experiments. Compared with subjects who initially saw only the gestures, subjects who had viewed gestures and simultaneously heard the accompanying speech were subsequently less likely to recognize them. Indeed, their mean recognition rate was only about 10% better than the chance level of 50%. The difference in video-only recognition accuracy between the two experiments (.733 vs. .610) is highly reliable, $F(1, 33) = 14.97$, $p < .0001$. Apart from this difference, Experiment 4 replicates the results of Experiment 3. Recognition rates in the audio-only and audio–video conditions are virtually identical, and both differ significantly from the video-only condition.

To examine the effect of semantic category on recognition in the two experiments, we calculated accuracy rates for the 30 gesture–lexical affiliate pairs (summing across subjects) in each of the two recognition series within each of the three presentation modes and subjected them to a 4 (semantic categories) × 3 (recognition conditions) × 2 (experiments) mixed ANOVA. Recognition varied reliably as a function of semantic category, $F(3, 56) = 3.14$, $p < .04$. Overall, descriptions were recognized reliably less well (72%) than object names (82%), actions, or locations (both 78%). A statistically significant semantic Category × Presentation Modality × Experiment interaction, $F(6, 112) = 2.26$, $p < .04$, reflected the fact that in Experiment 3 (but not Experiment 4), locations were recognized about as well in the video mode as in the other two modes; for the other three semantic categories, video recognition accuracy was reliably lower than accuracy in the audio or audio–video conditions.

Thus, presenting gestural information in context negated the one instance in which video recognition accuracy was not inferior to recognition accuracy in the audio or audio–video conditions.

Rather than improving recognition accuracy for gestures, putting visual information in its communicative context appears to affect it adversely. Recognition accuracy in the video-only condition in Experiment 4 was better than would be expected by chance, but it was not a great deal better and it was reliably less good than in Experiment 3. It is not clear why this happened. One possibility is that the speech served as a distractor—that is, subjects may have been unable to ignore the accompanying speech and, as a result, attended less well to the visual information. However, if this accounts for the lower accuracy of subjects who heard the sound track while trying to remember the gestures, it is interesting that the distraction effect is asymmetrical. Adding visual information to the audio segments does not improve later recognition of the audio segments, but neither does it harm it.

An alternative possibility is that subjects used the verbal context to impute meanings to the gestures and used these meanings to encode the gestures in memory. If the meanings were largely a product of the lexical affiliate—that is, if there was only a weak relationship between a gesture's forms and its imputed meaning—they would be of little use in the subsequent recognition task. As we speculated earlier, the transparent meanings that gestures appear to have when seen in the context of the speech they accompany may be illusory, deriving mainly from the perceiver's understanding of what the speech means.

Experiment 5

The hypothesized explanation for the low recognition accuracy of gestures initially seen in the context of the accompanying speech is highly speculative, because we have no direct way of ascertaining the strategies subjects used in trying to remember and recognize the gestures. However, the explanation rests on an assumption that is testable, namely, that perceived gestural meanings derive mainly from the meanings of their lexical affiliates, rather than from the form of the gestures themselves. It is possible to estimate the relative weight people implicitly give to gestural and speech information when they judge one component of gesture's meaning: its semantic category. It seems reasonable to suppose that if the gesture's form makes only a minor contribution to its perceived meaning, remembering the meaning will be of limited value in trying to recognize the gesture.

In Experiment 5, subjects assigned gestures to semantic categories based either on seeing the gesture alone or on seeing the gesture and hearing the accompanying speech. Two additional groups categorized the gestures' lexical affiliates, one group from the audio track and the other from verbatim transcriptions. From these four sets of judgments, we were able to estimate the relative contribution of speech and gestural information to this component of a gesture's perceived meaning.

Method

Videotapes of the 60 gestures and their lexical affiliates were used as stimuli. Forty undergraduates, approximately evenly divided between

men and women, served as subjects. Ten saw the videotape but did not hear the soundtrack; we will refer to this as the video-only condition. Ten both saw and heard the videotape; we will refer to this as the audio–video condition. Ten heard the soundtrack but did not see the video; we will refer to this as the audio-only condition. Subjects in this latter condition were members of a seminar who volunteered to participate after class; they were run in a group. The remaining 10 read transcriptions of the 60 lexical affiliates along with a few additional words for context as in the examples above; we refer to this as the transcript condition. Subjects in this group were paid for participating.

Subjects in the video-only and audio–video conditions judged whether the meaning of the gesture displayed in each segment referred to an action, location, object name, or description. No further elaboration of the semantic categories was given, and subjects who asked were instructed to use their own criteria in making the assignments. Instructions in the audio–video condition stressed that it was the meaning of the gestures that was to be categorized. In the audio-only condition, subjects were instructed to categorize the meaning of the phrase on the soundtrack, and in the transcript condition, the written phrase.

Results and Discussion

Although subjects' categorizations of the gestures' and lexical affiliates' intended meanings displayed considerable variability, it was not random. In Table 1 the distribution of semantic categories attributed to gestures or lexical affiliates in the four conditions is shown as a function of the semantic category of the lexical affiliate. All three departed reliably from chance, audio–video $\chi^2(9) = 379.74$; video-only $\chi^2(9) = 36.42$; audio-only $\chi^2(9) = 312.4$; transcript $\chi^2(9) = 750.06$, all $ps < .0001$. As inspection of Table 1 reveals, the association between the semantic category assigned to the gesture and the semantic category of the lexical affiliate is greater when the coder can hear the sound than when the coder cannot hear the sound (audio–video $\kappa = .427$, video-only $\kappa = .134$).

The primary question of interest here is what is the relative influence of speech and gestural form on judgments of a gesture's semantic category? Unfortunately, with categorical data of this kind, there is no obvious best way to pose such a question statistically. A log-linear analysis is problematic because many cells have very small expected values. If our data consisted of ratings on a continuous variable (i.e., if subjects had assigned a numerical value rather than a category to each of the stimulus segments), we could simply use multiple regression to

calculate the proportion of variance in the audio–video ratings that is accounted for by ratings in each of the three other conditions. Unfortunately, our four unordered semantic categories do not readily lend themselves to this approach. Nevertheless, we performed such a calculation, using the 16 frequencies in the corresponding cells of video-only, audio-only, and transcript tables as the independent variables and the values in the cells of the audio–video table as the dependent variable. Overall, the model accounted for 92% of the variance in the cell frequencies of the audio–video matrix, $F(3, 12) = 46.10$; $p < .0001$, but the contribution of the video-only matrix was negligible. The β coefficient for the video-only matrix is $-.026$, $t(16) = .124$, $p < .90$; for the audio-only condition, $\beta = .511$, $t(16) = 3.062$, $p < .01$; and for the transcript condition, $\beta = .42$, $t(16) = 3.764$, $p < .003$. Suggestive as these results are, the analysis does not take between-subjects variance into account and should be interpreted with some caution.

An alternative approach to the analysis of these data is through the use of multivariate ANOVA (MANOVA). Each of the 4 matrixes in Table 1 is the mean of 10 matrixes, 1 for each of the 10 subjects in that condition. We can treat the values in the cells of each subject's 4 × 4 matrix as 16 dependent variables and compute a MANOVA using the four presentation conditions as a between-subjects variable. Given a significant overall test, we could then determine which of the six between-subjects conditions contrasts (i.e., audio–video vs. audio-only, audio–video vs. transcript, audio–video vs. video-only, audio-only vs. transcript, audio-only vs. video-only, and transcript vs. video-only) differ reliably. The overall test (Wilks's λ) indicates the presence of reliable differences among the four conditions, $F(36, 74.59) = 6.72$, $p < .0001$. The F ratios for the six between-conditions contrasts are shown in Table 2. As that table indicates, the video-only condition differs reliably from the audio–video condition and from the audio-only and transcript conditions as well. The latter two conditions differ reliably from each other, but not from the audio–video condition. These results are consistent with our multiple regression analysis in their indication that judgments of a gesture's semantic category made solely from visual information are quite different from the same judgments made when the accompanying speech is accessible. Judgments of a gesture's semantic category made in the presence of speech are not reliably different from judgments made of the

Table 1

Subjects' Assignments of Gestures or Speech to Semantic Category as a Function of the Semantic Category of the Lexical Affiliate

| Semantic category of lexical affiliate | Semantic category judged from | | | | | | | | | | | | | | | |
| | Audio + video | | | | Video only | | | | Audio only | | | | Transcript | | | | |
	A	L	O	D	A	L	O	D	A	L	O	D	A	L	O	D	%
A	91	7	5	27	54	13	20	43	69	9	6	46	91	1	2	36	21.7
L	32	110	6	52	55	44	28	73	18	100	16	66	28	107	11	54	33.3
O	4	28	52	46	28	22	32	48	3	5	47	75	0	0	109	21	21.7
D	12	31	12	85	26	21	20	73	9	21	30	80	5	20	6	109	23.3
Sum	139	176	75	210	163	100	100	237	99	135	99	267	124	128	128	220	
%	23.2	29.3	12.5	35.0	27.2	16.7	16.7	39.5	16.5	22.5	16.5	44.0	20.7	21.3	21.3	36.7	

Note. Data are for 600 judgments per condition (10 subjects × 60 judgments). A = actions; L = locations; O = object names; D = descriptions.

Table 2

Value of Multivariate F Ratios (Wilks's Lambda) for Between-Conditions Contrasts

Condition	1	2	3	4
1. Audio + video	—	1.61	3.14	6.52**
2. Audio only		—	5.33*	17.6***
3. Transcript			—	12.85***
4. Video only				—

Note. For all contrasts, $df = 12, 7$.
* $p < .05$.   ** $p < .01$.   *** $p < .0001$.

speech alone. Unlike the regression analysis, the MANOVA takes the within-cell variances into account, but it does not readily yield an index of the proportion of variance accounted for by each of the independent variables.

Taken together, the multiple regression and MANOVA analyses lead to a relatively straightforward conclusion: When people can hear the words that accompany a gesture, their understanding of the gesture's semantic category is largely a product of what they hear rather than what they see. In addition, both analyses suggest that the audio-only and the transcript condition make unique contributions to judgments made in the audio–video condition. Although judgments made in the audio-only and transcript conditions are highly correlated, $r(15) = .815$, $p < .0001$, the MANOVA indicates that they also differ reliably. In the regression analysis, the two account for independent shares of the audio–video variance. Because the speech and transcript contain the same semantic information, these results suggest that subjects use paralinguistic information in the speech when they formulate rudimentary notions of the gesture's meaning.

On average, matching the lexical affiliate's semantic category requires 1.9 bits of information.[10] From the data in Table 1, one can calculate that the gestures transmit an average of 0.92 bits, reducing uncertainty about a lexical affiliate's semantic category by nearly half (see Wickens, 1989). Knowing a lexical affiliate's semantic category (e.g., that it refers to an action) constrains its possible meanings, and the fact that conversational gestures convey some measure of information about semantic category establishes that they are at least to that degree communicative. Of course, a lexical affiliate's semantic category is only a small part of the information conveyed. *Flying out* in the phrase "rockets or bullets flying out" and *swallow them up* in "seems like it's going to swallow them up" both refer to actions, but the two actions are different in a variety of respects. Even when one knows a word or phrase's semantic category, considerable uncertainty about its meaning remains. Unfortunately, there is no practical way to estimate the additional uncertainty, and, as a result, we cannot calculate in any general way the amount of information a gesture communicates about the meaning of its lexical affiliate.

General Discussion

In summary, in our five experiments, we have found the following:

1. After seeing a gesture, subjects can discriminate its lexical affiliate from the lexical affiliate of another gesture at a better-than-chance level.

2. Interpretations of a gesture's intended meaning are judged more similar to the meaning of its lexical affiliates than to the meaning of another gesture's lexical affiliate. However, the semantic category of its lexical affiliate is an important determinant of how accurately a gesture will be interpreted: On average, lexical affiliates characterized as actions or locations are correctly selected about 69% of the time; the corresponding value for object names and descriptions is 55%.

3. Recognition accuracy for a gesture in a forced-choice recognition test is reliably better than chance, but considerably less good than for the speech the gesture originally accompanied. Recognition for speech and gesture combined is not better than for speech alone.

4. When a gesture initially is presented together with speech, subsequent recognition accuracy for the gesture is depressed, relative to accuracy when the gesture initially is presented in isolation; accuracy for speech and for speech and gesture combined is unaffected.

5. Judgments of a gesture's semantic category, made when the accompanying speech is accessible, derive almost entirely from the semantic and paralinguistic content of the speech. However, a gesture seen in the absence of speech conveys some information about the semantic category of its lexical affiliates.

We have used a variety of methods to assess the communicativeness of hand gestures, each tapping a somewhat different dimension of the information they convey. Taken together, the results are consistent with the notion that conversational gestures convey semantic information related to the semantic content of the speech they accompany. However, it also is clear that the relationship of gesture to speech is relatively imprecise and unreliable. In a sense, the question is whether the cup is partly empty or partly full. It is clear that gestures do not communicate as articulately as speech (something no contemporary theorist has claimed), but they do convey *some* information. Under ordinary circumstances, do they add anything to the information conveyed by speech when the speech is intelligible? Our guess is that they do not, but we lack a method for quantifying the information in a meaningful way.

We do not contend that gestures are incapable of serving communicative purposes or that they are never used in that way. Deictic gestures are often used to convey locational information, and Bavelas, Chovil, Lawrie, and Wade (1990) have identified a class of "interactive gestures" that they hypothesize serve to maintain the addressee's involvement in the ongoing interaction. In addition, speakers may use gestures (as well as other nonvocal displays) communicatively when they find speech difficult to produce or difficult for their addressee to comprehend. Travelers to foreign countries are often impressed by how much can be accomplished by some energetic pantomime and a few words of vocabulary. Nevertheless, granting these and other exceptions, the gestures in our corpus in this

---

[10] If the four semantic categories occurred with equal frequency, selection of the correct alternative would require two bits; any deviation from equiprobability reduces the uncertainty of the alternatives.

study seem to convey relatively little information, and it is difficult to see how they could play an important role in communication. We have no reason to believe that they are atypical of the conversational gestures people ordinarily use.

If conversational gestures are intimately related to speech but are insufficiently informative to serve a communicative function, what function might they serve? One possibility is that they serve no function at all. It has been proposed that gestures are simply an atavistic vestige of the evolutionary process that resulted in speech (Hewes, 1973). Of course, arguments of this sort are notoriously difficult either to support or to falsify empirically, and we are inclined to agree with Nottebaum's assessment of Hewes's proposal as "a novel idea unsupported by any compelling evidence" (Nottebaum, 1973, p. 15). In a similar vein, Dittmann and Llewelyn (1969) have suggested that gestures are more-or-less random movements whose function is to dissipate tension during lexical search. The results of these studies (especially Experiment 2) reduce the plausibility of this hypothesis.

An alternative view takes as its point of departure the fact that gestures and lexical affiliates are coordinated temporally, a finding that has been interpreted to indicate that the two behaviors have a common origin (Kendon, 1980, 1983; Morrel-Samuels & Krauss, 1991; Levelt, Richardson, & La Heij, 1985; McNeill, 1985, 1987; Schegloff, 1984). However, there is little consensus as to the locus of that common origin or the process by which gestures are transformed into articulate movements. It is our assumption that the common origin of gesture and speech is at the level of communicative intention. According to this view, the speaker's communicative intention activates both an abstract propositional representation that, after suitable processing, results in an utterance (Levelt, 1989) and a motoric representation that may be reflected in a gestural movement.

Typically, but not inevitably, the communicative intention will activate a motoric representation of a concept that is also expressed in the utterance. The words a speaker utters are the product of an attempt to represent lexically the concepts contained in an underlying abstract proposition. For a number of reasons (among them the real-time processing demands of speech production), the individual may be unable to access a lexical entry that precisely expresses the sense represented in the proposition,[11] and on such occasions the motoric representation could convey a sense that for some purposes might be more informative than the one expressed in speech. More typically, however, concepts represented gesturally will also be represented in speech, and the sense the gesture conveys will tend to be shallower and less differentiated.

What functions might such motoric representations serve?[12] We do not think their functions are primarily communicative. On some occasions, gestures may serve to convey information that can less readily be expressed verbally, but the low quality of the information conveyed by the gestures in our corpus leads us to doubt that this happens very often.

Several theorists have suggested that gestures might help the speaker formulate utterances, particularly when there is difficulty retrieving elusive words from lexical memory (DeLaguna, 1927; Ekman & Friesen, 1972; Freedman, 1972; Mead, 1934; Moscovici, 1967; Werner & Kaplan, 1972). If this were the case, one might expect restricting gestures to adversely affect speech production, and there is some evidence that it does, although it

is far from clear. According to Schlauch (1936), Dobrogaev (1929) found that preventing speakers from gesturing resulted in decreased fluency, impaired articulation, and reduced vocabulary size. However, examination of the original publication provided no details of procedure or findings, and it was impossible to evaluate the claim. Rimé, Schiaratura, Hupet, & Ghysselinckx (1984) reported that the vividness of imagery in speech decreases when a speaker's hands, arms, legs, feet, and head are immobilized. On the other hand, Graham and Heywood (1975) concluded that preventing speakers from gesturing has no consequential effects on their speech; however, given their small sample size (the speech of only 6 speakers was examined) and the fact some reliable differences in fluency were found, their conclusion strikes us as premature.

We believe that gestures can aid in the process of lexical access, especially when the word being sought is difficult to retrieve. Consistent with this view is the finding that gestures tend to precede or to be uttered simultaneously with their lexical affiliates, but not to follow them (Butterworth & Beattie, 1978; Morrel-Samuels, 1989; Schegloff, 1984). For example, all of the 60 gestures in our corpus were initiated before or simultaneously with the onset of their lexical affiliate; none were initiated after the onset of the lexical affiliate. The asynchrony ranges from 0 s to 3.8 s, with a mean of 0.75 s, and there is a reliable inverse relation between the magnitude of the asynchrony and the familiarity of the lexical affiliate: Gestures tend to be displayed more-or-less simultaneously with familiar lexical affiliates and to precede unfamiliar ones (Morrel-Samuels, 1989; Morrel-Samuels & Krauss, 1991).

If gestures are not primarily communicative devices, why do they occur more frequently when speaker and addressee can see each other? One possibility is that arousal deriving from the physical presence of another person in the face-to-face situation (Zajonc, 1965) potentiates gestures along with other motor responses. Another intriguing, but highly speculative, possibility is that the increased frequency of gesturing reflects the greater difficulty of formulating speech in the presence of others. It is well established that in face-to-face situations, a speaker will alternately gaze toward and away from the addressee (Kendon, 1967). In spontaneous speech, gaze is most likely to be averted during periods of hesitant articulation when speech planning is assumed to be taking place; compelling the speaker to gaze continuously at the addressee results in less fluent speech, presumably because it interferes with planning (Beattie, 1983). Consistent with this conjecture is the finding that filled pauses, considered an indicator of semantic planning (Beattie, 1983), occur more frequently when discussants are face-to-face than when they cannot see each other (Ballard, 1989; but see also Rutter & Stephenson, 1977). Rather than serving a communica-

---

[11] This may occur when the speaker is temporarily unable to retrieve a term or when the appropriate term is not represented in the speaker's lexicon. It may also be the case that the speaker's language fails to provide such a term. It is often contended that any idea can be expressed in any language, but certainly the ready availability of monolexemic forms facilitates expression.

[12] See Zajonc and Markus (1984) for a lucid discussion of motoric representations and the role they can play in the encoding and retrieval of a variety of types of information.

tive function, the higher rate of gesturing in the face-to-face situation may reflect the greater information processing demands such situations impose on speakers, which make it more difficult for them to formulate coherent speech.

Our results may also help explain why the impression that gestures convey complex meanings is so strong and persistent and why the apparent meanings of gestures seen in the context of speech seem so self-evident. In Experiment 5, despite instructions to focus on the gesture's meaning, subjects' perceptions of a gesture's semantic category derived overwhelmingly from the speech that they heard, rather than from the movements that they saw. If one has a clear idea of what a gesture is intended to mean, it probably is not too difficult to think of a way in which its movements can be related to that meaning. Some of these "derivations" may be correct and others may be illusory, but even when they are correct, the relation of gestural form to meaning usually is too tenuous to be interpretable in the absence of speech. Nevertheless, the process may yield a clear sense that form and meaning are intimately related.

Our data indicate that, in addition to semantic information, conversational gestures convey some of the information communicated paralinguistically in speech. This possibility has been suggested by several writers, without clear empirical support (but see Morrel-Samuels, 1989). Gestures perceived as emphatic tend to fall on stressed syllables (Bull & Connelly, 1985) but so too do head nods and shakes and other nongestural movements (Bull & Connelly, 1985; Hadar, Steiner, & Rose, 1984). The functions of these movements both for speech production and communication are not well understood.

What about the possibility that the presentation conditions used in our experiments caused us to underestimate the extent to which gestures communicate? Specifically, it might be argued that by presenting videotapes of speakers who were unfamiliar to the subjects and by excerpting the gestures out of the ongoing stream of communication, we created an unnatural situation that adversely affected subjects' ability to ascertain gestural meaning. In response, we first note that in the Feyereisen et al. (1988) study, no effects were found for subjects' familiarity with the speaker. But more generally, it seems to us that criticisms of this sort rest on a set of unstated assumptions (e.g., that the gestures of a familiar person communicate better than those of someone who is unfamiliar, that gestures require a discourse context to be comprehensible, and that a gesture's "true meaning" can be established by examining it intensively). To the best of our knowledge, none of these assumptions have been subjected to an empirical test.

Attempts to generalize from our results should take into account the fact that our corpus of gestures and lexical affiliates was not a random sample of those in our speakers' narratives, much less a representative sample of the gestures found in conversations generally. The gestures we used were taken from a particular kind of communication situation—narrative descriptions of pictures to an attentive, but relatively passive, listener—and we would expect this to affect the types of gestures speakers used. Although four semantic categories were sufficient to categorize all 60 lexical affiliates, it is unlikely that they would be adequate for a more representative sample of lexical affiliate-gesture pairs. In addition, the criterion by which gestures were selected for use in our experiments was that naive

subjects agreed on the relation of gesture and word. In this process, a very large number of hand movements was excluded. Some of these, no doubt, were the sort of meaningless self- and object manipulations that Ekman and Friesen (1972) have called adapters, but many others appeared to be related to the accompanying speech, despite the fact that our subjects did not agree on precisely how they were related. Thus, our data may overestimate the degree to which speech content is also conveyed gesturally.

To summarize our findings, we began with the question: "Do conversational hand gestures communicate?" The answer, drawn from the results of our five experiments, is: "They do to a limited extent, but certainly not as much as some theories of gesture would lead us to expect."

References

Argyle, M. (1974). *Bodily communication*. London: Methuen.

Ballard, G. (1989). *Filled pause and gesture in conversation as a function of visual access*. Senior honors essay, Columbia University.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1990). *Interactive gestures*. Unpublished manuscript, University of Victoria, British Columbia.

Beattie, G. W. (1983). *Talk: An analysis of speech and non-verbal behaviour in conversation*. Milton Keyes, England: Open University Press.

Berger, K. W., & Popelka, G. R. (1971). Extra-facial gestures in relation to speech-reading. *Journal of Communication Disorders, 3,* 302–308.

Bevington, D. (1984). *Action is eloquence: Shakespeare's language of gesture*. Cambridge, MA: Harvard University Press.

Birdwhistell, R. L. (1970). *Kinetics and context*. Philadelphia: University of Pennsylvania Press.

Bull, P. (1983). *Body movement and interpersonal communication*. London: Wiley.

Bull, P., & Connelly, G. (1985). Body movement and emphasis in speech. *Journal of Nonverbal Behavior, 9,* 169–187.

Butterworth, B., & Beattie, G. (1978). Gesture and silence as indicators of planning in speech. In R. N. Campbell, & P. T. Smith (Eds.), *Recent advances in the psychology of language: Formal and experimental approaches*. New York: Plenum Press.

Cohen, A. A. (1977). The communicative functions of hand illustrators. *Journal of Communication, 27,* 54–63.

Cohen, A. A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology, 28,* 276–279.

DeLaguna, G. (1927). *Speech: Its function and development*. New Haven, CT: Yale University Press.

DeMeijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior, 13,* 247–267.

Deutsch, F. (1952). Analytic posturology. *Psychoanalytic Quarterly, 21,* 196–214.

Dittmann, A. T., & Llewelyn, L. G. (1969). Body movement and speech rhythm in social conversation. *Journal of Personality and Social Psychology, 23,* 283–292.

Dobrogaev, S. M. (1929). Ucnenie o reflekse v problemakh iazykovedeniia [Observations on reflexes and issues in language study]. *Iazykovedenie i Materializm,* 105–173.

Dushay, R. D. (1991). *The association of gesture with speech: A reassessment*. Unpublished doctoral dissertation, Columbia University.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *Journal of Communication, 22,* 353–374.

Feldman, S. S. (1959). *Mannerisms of speech and gesture in everyday life.* New York: International Universities Press.

Feyereisen, P., van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive, 8,* 3–25.

Freedman, N. (1972). The analysis of movement behavior during the clinical interview. In A. W. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 153–175). Elmsford, NY: Pergamon Press.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology, 10,* 57–67.

Graham, J. A., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology, 5,* 185–189.

Graham, J. A., Ricci-Bitti, P., & Argyle, M. (1975). A cross-cultural study of the communication of emotion by facial and gestural cues. *Journal of Human Movement Studies, 1,* 68–77.

Hadar, U., Steiner, T., & Rose, C. (1984). The relationship between head movements and speech dysfluencies. *Language and Speech, 27,* 429–444.

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

Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica, 26,* 22–63.

Kendon, A. (1972). Some relationships between body motion and speech: An analysis of an example. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–210). Elmsford, NY: Pergamon Press.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *Relationship of verbal and nonverbal communication.* Amsterdam: Mouton de Gruyter.

Kendon, A. (1983). Gesture and speech: How they interact. In J. M. Weimann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 13–45). Beverly Hills, CA: Sage.

Krauss, R. M., Dushay, R. D., Chen, Y., & Bilous, F. (1991). [The communicative value of gestures in referential communication]. Unpublished raw data.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Levelt, W., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24,* 133–164.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological Review, 92,* 350–371.

McNeill, D. (1987). *Psycholinguistics: A new approach.* New York: Harper & Row.

Mead, G. H. (1934). *Mind, self and society.* Chicago: University of Chicago Press.

Morrel-Samuels, P. (1989). *Gesture, word and meaning: The role of gesture in speech production and comprehension.* Unpublished doctoral dissertation, Columbia University.

Morrel-Samuels, P., & Krauss, R. M. (1990). Cartesian analysis: A computer-video interface for measuring motion without physical contact. *Behavioral Research Methods, Instrumentation and Computers, 22,* 466–470.

Morrel-Samuels, P., & Krauss, R. M. (1991). *Word familiarity predicts the temporal asynchrony of hand gestures and speech.* Unpublished manuscript, Columbia University.

Moscovici, S. (1967). Communication processes and the properties of language. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 225–270). San Diego, CA: Academic Press.

Nottebaum, F. (1973). Comment on Hewes's paper. *Current Anthropology, 14,* 15–16.

Rimé, B. (1982). The elimination of visible behaviour from social interactions: Effects on verbal, nonverbal and interpersonal behavior. *European Journal of Social Psychology, 12,* 113–129.

Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 239–281). Cambridge, England: Cambridge University Press.

Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion, 8,* 311–325.

Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5,* 172–183.

Rogers, W. T. (1978). The contribution of kinesic illustrators toward the comprehension of verbal behaviors within utterances. *Human Communication Research, 5,* 54–62.

Rutter, D. (1987). *Communicating by telephone.* Elmsford, NY: Pergamon Press.

Rutter, D. R., & Stephenson, G. M. (1977). The role of visual information in synchronising conversation. *European Journal of Social Psychology, 7,* 29–37.

Rutter, D. R., Stephenson, G. M., & Dewey, M. E. (1981). Visual communication and the content and style of communication. *British Journal of Social Psychology, 20,* 41–52.

Schegloff, E. (1984). On some gestures' relation to speech. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action* (pp. 226–296). Cambridge, England: Cambridge University Press.

Schlauch, M. (1936). Recent Soviet studies in linguistics. *Science and Society, 1,* 152–167.

Schopenhauer, A. (1903). *Studies in pessimism.* New York: Modern Library.

Sherzer, J. (1973). Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language and Society, 2,* 117–131.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* Chichester, England: Wiley.

Slama-Cazacu, T. (1976). Nonverbal components in message sequence: "Mixed syntax." In W. C. McCormack & S. A. Wurm (Eds.), *Language and man: Anthropological issues.* Amsterdam: Mouton de Gruyter.

Werner, H., & Kaplan, B. (1972). *Symbol formation.* New York: Wiley.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Erlbaum.

Williams, E. (1977). Experimental comparisons of face-to-face and mediated communication: A review. *Psychological Bulletin, 84,* 963–976.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269–284.

Zajonc, R. B., & Markus, H. (1984). Affect and cognition: The hard interface. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.), *Emotions, cognition and behavior* (pp. 73–102). Cambridge, England: Cambridge University Press.