Journal of Memory and Language **44**, 169–188 (2001)
doi:10.1006/jmla.2000.2752, available online at http://www.academicpress.com on IDEAL®

Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen

Martha W. Alibali and Dana C. Heath

Carnegie Mellon University

and

Heather J. Myers

Stanford University

Do speakers gesture to benefit their listeners? This study examined whether speakers use gestures differently when those gestures have the potential to communicate and when they do not. Participants watched an animated cartoon and narrated the cartoon story to a listener in two parts: one part in normal face-to-face interaction and one part with visibility between speaker and listener blocked by a screen. The session was videotaped with a hidden camera. Gestures were identified and classified into two categories: *representational* gestures, which are gestures that depict semantic content related to speech by virtue of handshape, placement, or motion, and *beat* gestures, which are simple, rhythmic gestures that do not convey semantic content. Speakers produced representational gestures at a higher rate in the face-to-face condition; however, they continued to produce some representational gestures in the screen condition, when their listeners could not see the gestures. Speakers produced beat gestures at comparable rates under both conditions. The findings suggest that gestures serve both speaker-internal and communicative functions.

Key Words: gesture; communication; narrative; context.

Do speakers gesture to benefit their listeners? Many studies have demonstrated that speakers' spontaneous hand gestures have communicative effects (e.g., Alibali, Flevares, & Goldin-Meadow, 1997; Beattie & Shovelton, 1999; Graham & Argyle, 1975; Kelly & Church, 1998). However, few studies have examined whether speakers produce gestures *in order to* communicate. In the present study, we examine whether speakers use gestures differently when those gestures have the potential to communicate and when they do not. If speakers produce gestures in order to aid listeners' comprehension, they should produce fewer gestures when their listeners are unable to see those gestures.

Several investigators have claimed that gestures are produced, not for communicative purposes, but to facilitate speech production (e.g., Rauscher, Krauss, & Chen, 1996; Rimé & Shiaratura, 1991). These investigators argue that any communicative effects of gestures are minimal and at best epiphenomenal (Krauss, Morrel-Samuels, & Colasante, 1991). If gestures are produced primarily to facilitate speech production, then gesture production should not be affected by the visible presence of the interlocutor. Indeed, it is a common observation that people gesture when speaking on the telephone (de Ruiter, 1995). This fact is consistent with the view that speakers gesture for themselves rather than for their listeners.

This research was supported by the Undergraduate Research Initiative at Carnegie Mellon University and by the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. We thank Janet Bavelas, Herb Clark, Susan Goldin-Meadow, Lisa Haverty, Sotaro Kita, Michael Matessa, David McNeill, Asli Özyürek, Jan-Peter de Ruiter, Mandana Seyfeddinipur, and the anonymous reviewers for thoughtful comments on previous versions of the article. We are also grateful to Lindsey Clark, Julia Lin, Nicole McNeil, and Leslie Petasis for assistance in coding and establishing reliability. Portions of these data were presented at the 1997 meeting of the Midwestern Psychological Association, Chicago, Illinois.

Address correspondence and reprint requests to Martha W. Alibali, Department of Psychology, University of Wisconsin—Madison, 1202 W. Johnson St., Madison, WI 53706. E-mail: mwalibali@facstaff.wisc.edu.

0749-596X/01 $35.00

PREVIOUS RESEARCH ABOUT VISIBILITY AND GESTURE PRODUCTION

The literature contains several conflicting reports about whether visibility between speaker and listener influences gesture production. As described below, and as summarized in Table 1, some investigators have argued that speakers gesture more when they can see their listeners, and others have reported marginal or nonsignificant differences. The goal of the present study was to explore a possible explanation for these conflicting reports: specifically, that visibility affects speakers' production of some types of gestures but not others.

Four studies have reported that speakers gesture more when interacting face-to-face than when they are unable to see their listeners. In two of these studies (Cohen, 1977; Cohen & Harrison, 1973), participants were asked to give directions from one location to another, either face-to-face or over an intercom. In both studies, speakers used more *illustrator* gestures, operationally defined as "hand movements that were performed in conjunction with the verbal encoding" (Cohen, 1977, p. 58) in the face-to-face condition than in the intercom condition. The *illustrator* category was originally defined by Ekman and Friesen (1969) and includes several subtypes, among them gestures that depict semantic content and gestures that rhythmically accompany speech. In another study, participants instructed the experimenter where to place puzzle pieces on a grid, either face-to-face or with a screen between participant and experimenter (Emmorey & Casey, in press). Participants produced more gestures in the face-to-face condition than in the screen condition; however, no information was provided about how the gestures were identified and coded. In another study, speakers described abstract figures or complex synthesized sounds, either face-to-face or over an intercom (Krauss, Dushay, Chen, & Rauscher, 1995). Again, speakers produced more gestures per minute when speaking face-to-face; however, no information was provided about how the gestures were identified and coded.

Two studies have failed to find reliable differences in gesture production when visibility between speaker and listener was blocked. In one study, participants described a set of four photographic portraits to other participants, who were then asked to identify those portraits from a larger set (Lickiss & Wellens, 1978). There were no differences in gesture production when speakers interacted over a two-way videotelephone (audiovisual access) or over a voice intercom (audio-only access). Again, no information was provided about how the gestures were identified and coded. In another study, Rimé (1982) asked pairs of participants to "explain to each other their opinions on movies and to express what they liked to find in the cinema" (p. 117), either face-to-face or with visibility blocked by an opaque screen. Speakers' use of *communicative* gestures, which Rimé defined as "any gesture accompanying or paralleling the content or rhythm of the verbal flow" (p. 119), was assessed in both conditions. This category of gestures appears comparable to illustrators (as defined by Cohen, 1977) in that it presumably includes both gestures that depict semantic content and gestures that rhythmically accompany speech. Rimé found that speakers produced slightly but not significantly more communicative gestures in the face-to-face condition. There was also no statistically reliable difference across conditions in the overall duration of the gestures, considered as a percentage of time spent speaking.

One additional study compared effects of listener visibility on gesture production in 6-year-old and 11-year-old participants and reported reliable differences for older but not for younger children (Doherty-Sneddon & Kent, 1996). In this study, children were asked to describe a route marked on a map to a partner who had the same map without the route marked on it, either face-to-face or with a screen erected between them. The experiment was set up with the children's maps on either side of a two-way easel so that "it was impossible for the children to see each other's hands unless they raised them in a deliberate attempt" to show a gesture to the partner (p. 951). Only such deliberately raised gestures were scored as gestures; therefore, it appears that even the gestures produced in the screen condition were used in an explicitly communicative fashion. Hence, it is difficult to compare these findings to those reported in other studies.

TABLE 1

Overview of Previous Research

Study	Manipulation	Task	Camera hidden?	Type of gestures examined, and operational definitions	Summary of results
Cohen and Harrison (1973)	Face-to-face vs intercom	Giving directions	Yes	*Illustrators*: "movements which are directly tied to speech, serving to illustrate what is being said verbally" (Ekman & Friesen, 1969)	More illustrators in face-to-face condition
Cohen (1977)	Face-to-face vs intercom	Giving directions	Yes	*Illustrators*: "hand movements that were performed in conjunction with the verbal encoding" (p. 58)	More illustrators in face-to-face condition
Lickiss and Wellens (1978)	Videotelephone vs intercom	Describing photographic portraits	No [a]	No information provided	No significant differences across conditions
Rimé (1982)	Face-to-face vs partition	Talking about movies	Yes	*Communicative* gestures: any gestures "accompanying or paralleling the content or rhythm of the verbal flow" (p. 119)	No significant differences across conditions in % time producing gestures Slightly but not significantly more gestures in face-to-face condition
Bavelas, Chovil, Lawrie, and Wade (1992, Expt. 2)	Face-to-face vs partition	Talking about "close call" events	No	*Topic* gestures: any gestures that "depict semantic information directly related to the topic of discourse" (p. 473). *Interactive* gestures: any gestures that "refer . . . to some aspect of the process of conversing with another person," e.g., citing the listener's previous contribution, seeking agreement, forestalling the turn (p. 473)	*Topic gestures*: No significant differences across conditions *Interactive gestures*: More in face-to-face condition
Krauss, Dushay, Chen, and Rauscher (1995)	Face-to-face vs intercom	Describing abstract figures or complex, synthesized sounds	No [a]	No information provided	More gestures in face-to-face condition
Doherty-Sneddon and Kent (1996, Expt. 1)	Face-to-face vs partition	Describing route on map to partner	No [a]	*Communicative* gestures: any gestures in which "hands were raised in order to 'show' [the] gesture to a partner" (p. 951)	6-year-old participants: no significant differences across conditions 11-year-old participants: more communicative gestures in face-to-face condition
Emmorey and Casey (in press)	Face-to-face vs partition	Telling experimenter where to place puzzle pieces on a grid	No [a]	No information provided	More gestures in face-to-face condition

[a] It was not stated explicitly in the article that the camera was hidden, so we infer that it was not.

Finally, one study has suggested that the effects of visibility differ for different types of gestures. Bavelas and colleagues asked participants to talk about “a close-call incident or near-miss incident that you have had,” either face-to-face or with visibility blocked by a partition (Bavelas, Chovil, Lawrie, & Wade, 1992). Bavelas et al. classified all gestures into two categories based on their function in the communicative situation: (1) *topic* gestures, which “depict semantic information directly related to the topic of discourse,” and (2) *interactive* gestures, which “refer instead to some aspect of the process of conversing with another person,” such as citing the listener’s previous contribution, seeking agreement, or forestalling the turn (p. 473). Bavelas et al. found that interactive gestures, which were comparatively rare overall (fewer than 15% of all gestures), were used at a higher rate in the face-to-face condition. However, they found no difference across conditions in the rate of topic gestures, which depict semantic information.

Because the existing studies have used widely different tasks and different schemes for identifying and classifying gestures, as summarized in Table 1, at present it is difficult to draw any firm conclusions about the effects of reciprocal visibility on gesture production. One possible explanation for the conflicting results is that the categories used to classify gestures are too broad. In this regard, Bavelas et al.’s (1992) finding that visibility influenced interactive gestures but not topic gestures is suggestive. However, even within the broad category of *topic* gestures (or *communicative* gestures, or *illustrators*), visibility might influence some types of gestures but not others. The tasks that have revealed effects of visibility (e.g., Cohen and Harrison’s directions task and Emmorey & Casey’s puzzle task) may have elicited different types of gestures than the tasks that have shown no effects of visibility (e.g., Rimé’s movie task and Bavelas’s close-call task).

PURPOSE OF THE PRESENT STUDY: RECONCILING CONFLICTING RESULTS

We hypothesize that the effects of visibility on gesture production differ for different types of gestures. To investigate this hypothesis, we used a monologue cartoon narration task, which has been shown to reliably elicit gestures in many prior studies (e.g., Beattie & Shovelton, 1999; McNeill, 1992; Özyürek & Kita, 1999). In evaluating the data, we distinguish between gestures that represent some aspect of the content of speech, which we term *representational* gestures, and motorically simple gestures that do not represent speech content, which we term *beat* gestures. This distinction is accepted and used by many investigators, although different investigators sometimes use different labels (e.g., Feyereisen & Harvard, 1999; Krauss, Chen, & Chawla, 1996; McNeill, 1992).

The distinction between representational and beat gestures is supported by both theoretical analyses (e.g., Butterworth & Hadar, 1989; Hadar, 1989) and empirical research, including studies showing (a) differential responses to experimental manipulations (e.g., spontaneous vs rehearsed speech; Chawla & Krauss, 1994), (b) different distributions across different types of tasks (e.g., Feyereisen & Harvard, 1999), and (c) different patterns of breakdown in different types of aphasia (e.g., Cicone, Wapner, Foldi, Zurif, & Gardner, 1979; Hadar, Wenkert-Olenik, Krauss, & Soroker, 1998; McNeill, Levy, & Pedelty, 1990). However, not all investigators accept the representational/beat distinction. In particular, Bavelas and colleagues have proposed an alternative approach that focuses on the function of gestures (as inferred based on both form and meaning) rather than on typological classification (Bavelas, 1994; Bavelas, Chovil, Coates, & Roe, 1995; Bavelas et al., 1992). As noted above, they have identified gestures that are specialized for dialogic interaction, called *interactive* gestures. According to Bavelas and colleagues (1992), *interactive* gestures subsume the category of beats and, in addition, include some representational gestures. Interactive gestures were expected to be rare in the narrative monologue task that we used in this study because listeners did not engage in dialogic give-and-take with the speakers (see Bavelas et al., 1995). However, we acknowledge that many of the gestures that we classified as beats in this study, as well as a subset of the gestures that we classified as representational, could be construed as interactive gestures in functional terms.

We suggest that the conflicting reports in the literature about visibility and gesture production may have arisen because visibility affects some types of gestures but not others. Specifically, we hypothesize that the effects of visibility on gesture production may depend on whether the gestures convey semantic information. As noted above, *representational* gestures depict semantic information by virtue of handshape, placement, or motion. However, *beat* gestures "do not present a discernible meaning" (McNeill, 1992, p. 80), even when the visual channel is available. Thus, we hypothesize that the effects of visibility may differ for representational gestures and for beats.

Further, we propose two specific alternatives about how the effects of visibility may differ for these two types of gestures. One possibility is that only gestures that convey semantic information will be affected by the visibility manipulation. Under this *semantic information hypothesis*, speakers produce representational gestures in order to convey semantic information, so they should produce such gestures less often when their listeners are unable to see those gestures. Beat gestures, in contrast, do not convey semantic information. Hence, production of beat gestures should not depend on whether the listener can see them. Thus, according to this hypothesis, representational gestures will be produced less often when speakers cannot see their listeners, but beat gestures will be unaffected.

An alternative possibility is suggested by the work of Tuite (1993). Tuite argues that all gestures are built upon a kinesic base, or "rhythmical pulse," that is associated with speaking. Features of semantic content can be "overlaid" upon this rhythmical pulse. According to Tuite, beat gestures are the simplest type of gestures—a simple "kinetic realization of the underlying pulse" (p. 99). Representational gestures are beat gestures that have features of the speaker's internal representation overlaid upon them. Tuite's hypothesized "rhythmical pulse" should not depend on listener visibility. Thus, under the *rhythmical pulse hypothesis*, overall levels of gesture production should not differ when speakers can and cannot see their listeners. However, speakers might produce beat gestures *instead of* representational gestures when visibility is blocked. Instead of gesturing less overall, speakers might simply decline to overlay semantic features on the rhythmical pulse and therefore produce beat gestures rather than representational gestures. According to this view, representational gestures will be produced less often when speakers cannot see their listeners, but beat gestures will be produced *more* often.

In brief, the goal of the present study was to examine whether speakers modify their production of beat and representational gestures when they can see their listeners and when they cannot. To address this issue, we compared speakers' gesture production as they narrated a cartoon story to a listener in two parts: one part in normal face-to-face interaction and one part with visibility between speaker and listener blocked by a screen. According to the *semantic information hypothesis*, representational gestures should be produced less often in the screen condition and beat gestures should be unaffected. According to the *rhythmical pulse hypothesis*, representational gestures should be produced less often in the screen condition and beat gestures should be produced more often.

If changes in gesture are observed in the screen condition, then we must consider the possibility that such changes derive from changes in speech. That is, the visibility manipulation may influence speech, and changes in gesture may be parasitic on changes in speech. To foreshadow the results, we do observe changes in gesture, so we also consider whether the visibility manipulation leads to changes in speech. Speakers might modify their speech content in one of two ways when they cannot see their listeners. First, speakers might simply *avoid* information that is best conveyed with gesture (or with speech and gesture together) when gesture cannot be used communicatively. For example, speakers might avoid talking about spatial relationships when they cannot use gestures communicatively. Such a shift might lead to a decrease in the amount of speech. Alternatively, speakers might *compensate* with speech when they cannot use gestures communicatively. That is, information that would ordinarily be conveyed in gesture might be shifted into speech. For example, spatial relationships that might ordinarily be conveyed solely through gesture might be expressed in speech instead. Such a shift might lead to an increase in the amount of speech.

If speakers modify their speech content when they cannot see their listeners, then formulating speech might be more effortful in the screen condition than in the face-to-face condition. This additional effort might be manifested in reduced fluency of speech. Indeed, Rimé (1982) reported that speakers produced more filled pauses (e.g., "um" and "uh") when they could not see their listeners. However, Lickiss and Wellens (1978) reported that speakers produced fewer speech errors when they could not see their listeners. As noted above, neither of these two studies reported significant effects of visibility on gesture production.

Based on these considerations, we examined the effects of visibility on the amount, content, and fluency of narrators' speech. To assess fluency, we assessed the rate of speech, the rate of filled pauses, and the rate of speech errors. To assess content, we examined three aspects of speech content that have been linked to gesture production by other investigators: (1) nonnarrative (i.e., meta- or extranarrative) content, which has been linked to the production of beat gestures (McNeill, 1992); (2) spatial content, which has been linked to the production of representational gestures (Krauss, 1998); and (3) the use of verbs that convey manner information (e.g., "climb" as opposed to "go," see Talmy, 1985), which has been linked to the production of representational gestures (McNeill, in press; Özyürek & Kita, 1999). Finally, it is possible that the manipulation would lead to changes in the relationship between gestures and speech content. Thus, we also consider the effects of visibility on the strength of the relation between beat gestures and nonnarrative content and on the strength of the relation between representational gestures and spatial content.

METHOD

Participants

Sixteen undergraduate students (8 males and 8 females) served as narrators in the study. On a postexperiment questionnaire, 14 of the narrators reported that English was their native language. The remaining 2 were fluent speakers of English who had used English as a primary language, 1 for 6 years and 1 for 20 years. An additional 16 students (8 males and 8 females) served as listeners. Participants were scheduled to come to the laboratory in same-gender pairs and were assigned to experimental roles by the toss of a coin.

Participants were told that the focus of the study was story narration and that the study involved either (a) retelling a cartoon story after watching a cartoon or (b) listening to another person retell the cartoon story. They were informed that the session would be audiotaped. In addition to being audiotaped, the experimental session was videotaped with a hidden camera. At the conclusion of the session, participants were fully debriefed and were offered the option of having their videotape erased immediately if they did not wish their data to be used in the study. All participants consented to have their data included.

Procedure

After experimental roles were assigned, the narrator watched a "Tweety and Sylvester" cartoon, "Canary Row," in two 4-min segments, while the listener waited out of earshot in a room adjoining the laboratory. Each of the two segments consisted of four episodes. In each episode, Sylvester (a cat) attempted to catch Tweety Bird (a canary) in a different way. For further details about episode content, see the Appendix, and for a scene-by-scene description of the cartoon, see McNeill (1992). Following each segment, the narrator described to the listener what happened in the cartoon. Narrators were instructed simply to tell the listener what happened in the cartoon. Listeners were instructed to listen carefully and were told that they would later be asked to retell the story to the experimenter. Listeners were also instructed not to ask any questions. As noted above, the retellings were audiotaped as well as covertly videotaped. After each of the narrator's two retellings, the listener then retold the story to the experimenter, while the narrator waited in the adjoining room.

For the narrator's retelling of one of the cartoon segments (i.e., four episodes), the narrator spoke to the listener face-to-face. For the retelling of the other segment, an opaque wooden screen was placed between narrator and listener so that visibility between them was completely blocked. The order of conditions was counterbalanced across pairs. Eight narrators retold the cartoon in the face-to-face condition first and in the screen condition second. The remaining eight narrators retold the cartoon in the screen condition first and in the face-to-face condition second.[1]

All of the listeners sat quietly while listening to the narrator retell the story, and some occasionally provided back-channel feedback (e.g., nodding or saying “uh-huh”). In one case, a listener provided a word for which a narrator was searching. Listeners also sometimes smiled or laughed aloud, since the cartoon story can be quite funny, depending on the narrator’s skill at retelling. Of course, in the screen condition, narrators could hear audible back-channel responses and laughter, but could not see nods or smiles.

Transcribing Speech

Each narration was transcribed from the videotape, and all filled pauses (e.g., “um”), word fragments, and repeated words were included in the transcripts. As the speech was transcribed, the transcripts were segmented into units in preparation for gesture transcription (see below). Each transcript unit consisted of a verb and its associated arguments and modifiers, with the exception that prepositional phrases were treated as separate units if they were set off from the main clause by a pause. To illustrate, the following two examples each consist of two units, with the break between units marked by a slash:

(a) “the cat gets thrown out the window / and falls down to the street”
(b) “he’s back in his room (pause) / right across the way from Tweety”

Identifying and Coding Gestures

All of the hand gestures that each speaker produced with each unit of the verbal transcript were identified from the videotape. Each unit was viewed repeatedly in both regular and slow motion in order to identify the gestures. In most cases, the hand(s) returned to rest position after each individual gesture. When multiple gestures were produced in succession without the hand(s) returning to rest position, the boundaries between gestures were determined based on changes in the handshape, motion, or placement of the hands. Each individual gesture was then classified as either representational or beat.

Representational gestures ($N = 1280$, or 64.7% of the 1977 gestures observed) were defined as gestures that depict semantic content via the shape, placement, and/or motion trajectory of the hands. Most representational gestures have three movement phases (preparation, stroke, and retraction; McNeill, 1992). Such gestures are sometimes referred to in the literature as “lexical movements” (Krauss et al., 1996). Representational gestures were further classified into four subtypes, based on those described by McNeill (1992): (1) *iconics*, which are gestures that depict concrete referents (e.g., making climbing motions with the hands to convey “climb”); (2) *metaphorics*, which are gestures that depict abstract referents metaphorically (e.g., gently waving the hand back and forth to represent “music”) or that indicate spatial locations to metaphorically refer to characters, locations, or parts of the story (e.g., pointing to the left to indicate Tweety and to the right to indicate Granny); (3) *spatial deictics*, which are gestures that convey direction of movement (e.g., pointing upward to convey upward movement); and (4) *literal deictics*, which are gestures that indicate concrete objects in order to refer to those objects or to similar ones (e.g., pointing to the wooden screen in order to indicate a piece of wood). We also observed a few isolated instances ($N = 8$, or 0.4% of the 1977 gestures observed) of *emblems*, which are gestures that have a conventional form and meaning (e.g., holding up the index and middle fingers to mean “two”). These gestures were omitted from our analyses.

Beat gestures ($N = 689$, or 34.9% of the 1977 gestures observed) were defined as motorically simple, rhythmic gestures that do not depict semantic content related to speech. Beat gestures have only two movement phases (e.g., up/down), and most are produced using one hand in a loose, untensed handshape. Such gestures are sometimes referred to in the literature as “motor move-

[1] One of the two nonnative English speakers retold the cartoon in the face-to-face condition first and the other retold the cartoon in the screen condition first.

ments" (Hadar, 1989; Krauss et al., 1996) or "batons" (Efron, 1941/1972; Ekman & Friesen, 1969). There were also some instances in which beat gestures were *superimposed* on representational gestures (*N* = 67, or 9.7% of the 689 beat gestures observed). In such cases, a representational gesture was held briefly, and the entire gesture was then moved in a rhythmic, beat-like motion. In such cases, the initial gesture was classified as representational, and the additional movements were classified as beats.

A single coder initially coded all of the data, and each transcript (both speech and gesture) was then reviewed and checked by a second coder. These checked codes were used in the data analysis. To establish reliability, a third trained coder independently assessed 25% of the data (half of the narrative from each of eight participants). Agreement between this third coder and the checked codes was 92% for identifying individual gestures and 85% (*N* = 406) for categorizing gestures as beat or representational. For gestures that both agreed were representational, agreement was 87% (*N* = 247) for categorizing the type of representational gesture (iconic, metaphoric, spatial deictic, or literal deictic). Finally, as an additional reliability check, a fourth coder, who was unaware of the purpose of the study and who was blind to the experimental hypotheses, independently assessed another 25% of the data (half of the narrative from each of eight participants). Agreement between this fourth coder and the checked codes was 89% for identifying individual gestures and 86% (*N* = 359) for categorizing gestures as beat or representational.

Coding Speech

We considered the effects of visibility on the amount, fluency, and content of speech. Amount of speech was measured in number of words and number of transcript units. Fluency of speech was assessed using three different measures: (1) speech rate, measured in words per second; (2) the rate of filled pauses, defined as the number of filled pauses per 100 words; and (3) the rate of speech errors, defined as the number of speech errors per 100 words. Four types of speech errors were identified (see Levelt, 1983): (a) *repetitions*, in which one or more words are simply repeated (e.g., "and *they, they* see each other in their binoculars"); (b) *repairs*, in which one or more words within a given syntactic frame are altered (e.g., "Tweety gets all flustered by *it, Sylvester*"); (c) *fresh starts*, in which the speaker shifts to a new syntactic frame (e.g., "*Sylvester—there's* a gutter that goes up the side of the apartment building"); and (d) *uncorrected syntactic errors*, in which the speaker produces a syntactic error such as an agreement error or a word deletion, but does not overtly repair the error (e.g., "now obviously Sylvester just *go* down the pipe").

We assessed three aspects of the content of narrators' speech: (1) use of verbs that convey manner, (2) spatial content, and (3) nonnarrative content. To assess use of manner verbs, we identified 12 motion events in the cartoon (6 in each half). We then examined whether speakers described the events using verbs that conveyed manner information or verbs that were bleached of manner information (e.g., tiptoed vs went along the trolley wires and crawled vs went up the pipe; see Talmy, 1985). To assess spatial content, we calculated the rate of spatial prepositions (e.g., up, across, over, etc., used either as verb particles or in prepositional phrases) per 100 words in each episode.

To assess nonnarrative content, we calculated the proportion of words that conveyed nonnarrative content in each condition. Words were identified as having nonnarrative content if they were used to convey meta- or extranarrative information rather than information about the story line itself. Meta- or extranarrative information included information about the structure of the cartoon (e.g., "that was the end"), about watching or retelling the cartoon (e.g., "I'm mixing this up"), or about the camera movement (e.g., "then the camera pans down"). To calculate the proportion of words that conveyed nonnarrative content, we divided the number of words used to express nonnarrative content by the total number of words for each condition.

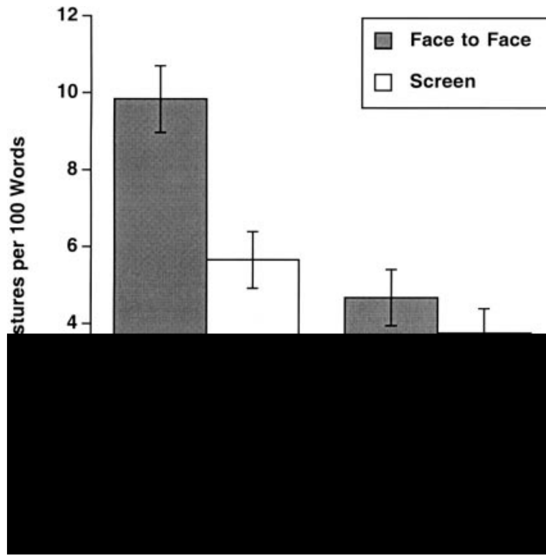
RESULTS

Most studies of the effect of visibility on gesture production have examined the rate of gestures per minute of speech. However, it is possible that people speak at different rates when they can see their listeners and when they cannot. If this were the case, then any difference in gestures per unit of time could reflect differences in speech rate rather than gesture rate per se. To avoid this interpretive difficulty, we chose to use the rate of gestures per 100 words as our primary dependent measure. We also report findings for gestures per minute in order to facilitate comparison with the work of other investigators. No effects of gender were found, so all analyses collapse across male and female pairs. Except where noted, the effects of experimental condition did not depend on the order of conditions. In addition, except where noted, all results are unchanged if the two nonnative speakers are omitted from the analysis. Unless otherwise noted, all statistical tests are significant at $p < .01$.

The results are organized into three sections. In the first section, we consider the effects of the visibility manipulation on narrators' production of gestures. In the second section, we consider the effects of visibility on narrators' speech, focusing on the amount, content, and fluency of speech. Finally, in the third section, we consider the effects of the visibility manipulation on the relationship between speech content and gesture production.

Did Visibility Influence Gesture Production?

Rate of gestures per 100 words. Our main goal was to establish whether visibility between speaker and listener influenced speakers' production of representational and beat gestures. To address this question, we used a repeated-measures ANOVA with condition (face-to-face and screen) and gesture type (representational and beat) as within-participants factors and order (face-to-face first and screen first) as a between-participants factor. The dependent measure was the rate of gestures per 100 words, calculated for each speaker by averaging across the cartoon episodes in each condition. As seen in Fig. 1, the visibility manipulation influenced representational gestures more strongly than beat gestures, yielding a significant interaction between condition and gesture type, $F(1, 14) = 10.47$. Speakers produced representational gestures at a higher rate when they could see their listeners than when they could not (left set of bars), $F(1, 14) = 34.49$. Speakers produced beat gestures at a slightly higher rate when they could see their listeners; however, this effect was not reliable (right set of bars), $F(1, 14) = 1.68$, n.s. fifteen of the sixteen speakers produced representational gestures at a higher rate in the face-to-face condition than in the screen condition, and the remaining speaker produced representational gestures at comparable rates in both conditions. For beat gestures, patterns were inconsistent across participants. Nine speakers produced beats at a higher rate in the face-to-face condition, five produced beats at a higher rate in the screen condition, and two produced beats at comparable rates in both conditions.



[redacted]

FIG. 1. Mean rates of representational and beat gestures per 100 words in each condition.

We also examined whether the effects of visibility were consistent across the eight cartoon episodes. Speakers produced more representational gestures in the face-to-face condition than in the screen condition in every one of the eight episodes. Indeed, univariate ANOVA with episode and condition as factors revealed a main effect of condition, $F(1, 108) = 39.23$. For beat gestures, the effect of condition was not significant, $F(1, 108) = 2.11$, $p = .15$.

The results thus far do not support Tuite's theory, which predicts that beat gestures should increase when visibility is blocked because beat gestures would be produced instead of representational gestures. However, one might argue that the preceding analyses do not adequately test Tuite's theory because superimposed beats (i.e., beats performed "on top of" representational gestures) were included the analysis. A decrease in superimposed beats could have counteracted an increase in "pure" beats, leading to an apparent lack of change in beat rate, even though the rate of "pure" beats might have increased. To test this possibility, and to provide a fairer test of Tuite's theory, we eliminated all superimposed beats from the data set and reanalyzed the data. The results were unchanged: speakers produced representational gestures at lower rates when they were unable to see their listeners, but produced beat gestures at comparable rates under both conditions, again resulting in a significant interaction between condition and gesture type, $F(1, 14) = 11.18$. Importantly, speakers produced beat gestures at comparable rates when they could see their listeners and when they could not (face-to-face $M = 4.24$, $SE = 0.73$, vs screen $M = 3.47$, $SE = 0.57$, per 100 words), $F(1, 14) = 1.12$, n.s.

Rate of gestures per minute. The same patterns were observed when the data were analyzed in terms of gestures per minute. The dependent measure was the rate of gestures per minute, calculated for each speaker by averaging across the cartoon episodes in each condition. Again, the interaction between condition and gesture type was significant, $F(1, 14) = 12.57$. Speakers produced more representational gestures per minute when they could see their listeners than when they could not (face-to-face $M = 14.82$, $SE = 1.72$, vs screen $M = 8.37$, $SE = 1.18$), $F(1, 14) = 43.03$. However, speakers produced beat gestures at similar rates per minute in both conditions (face-to-face $M = 7.42$, $SE = 1.27$, vs screen $M = 5.91$, $SE = 1.08$), $F(1, 14) = 2.39$, p = .14.[2] The pattern of results was unchanged when superimposed beats were eliminated from the data: the interaction between condition and gesture type remained significant, $F(1, 14) = 14.05$, and speakers produced beat gestures at comparable rates per minute when they could see their listeners and when they could not (face-to-face $M = 6.69$, $SE = 1.23$ vs screen $M = 5.48$, $SE = 1.00$), $F(1, 14) = 1.48$, n.s.

Thus, for both gestures per word and gestures per minute, the rate of representational gestures was higher when speakers could see their listeners and lower when they could not. However, even though speakers produced more representational gestures when they could see their listeners, they produced such gestures at surprisingly high rates when the screen was in place. Speakers produced an average of 5.65 ($SE = 0.73$) representational gestures per 100 words and 8.37 ($SE = 1.18$) representational gestures per minute in the screen condition. Thus, even without reciprocal visibility, speakers often produced representational gestures.

Subtype analyses. We next considered whether the visibility manipulation influenced all types of representational gestures. We focus on the three subtypes that were most frequent in our data: iconic gestures, metaphoric gestures, and spatial deictic gestures.

Iconic gestures depict concrete referents (e.g., making a swinging motion with the hand to convey "swing"). Given that the lion's share of representational gestures (75%) were iconic, it is not surprising that speakers produced iconic gestures at a higher rate when they could see their listeners than when they could not (face-to-face $M = 7.19$, $SE = 0.74$, vs screen $M = 4.13$, $SE = 0.60$, per 100 words), $F(1, 14) = 21.98$. The same pattern held for *metaphoric* gestures, which comprised 16% of all representational gestures. Metaphoric gestures depict abstract referents metaphorically (e.g., making a circular hand movement to represent "continuing") or indicate spatial locations to metaphorically refer to characters, locations, or parts of the story (e.g., pointing to the right to indicate Sylvester's building and to the left to indicate Tweety's building). Speakers produced metaphoric gestures at a higher rate when they could see their listeners than when they could not (face-to-face $M = 1.68$, $SE = 0.30$, vs screen $M = 0.95$, $SE = 0.19$, per 100 words), $F(1, 14) = 7.32$, $p < .02$.

A slightly different pattern was observed for *spatial deictic* gestures, which comprised 8% of

[2] With the two nonnative English speakers excluded, $F(1, 12) = 1.12$, $p = .31$.

all representational gestures. Spatial deictic gestures convey direction of movement (e.g., pointing upward to convey upward movement). Speakers who narrated in the face-to-face condition before the screen condition produced spatial deictics more often when they could see their listeners than when they could not (face-to-face M = 1.13, SE = 0.27, vs screen M = 0.24, SE = 0.07, per 100 words). However, speakers who narrated in the screen condition first produced spatial deictics about equally often under both conditions (face-to-face M = 0.62, SE = 0.22, vs screen M = 0.80, SE = 0.33, per 100 words). This pattern yielded a significant interaction between condition and order, $F(1, 14) = 5.37$, $p < .05$, and the main effect of condition did not reach significance, $F(1, 14) = 2.37$, $p = .15$.

Did the Visibility Manipulation Influence Narrators' Speech?

The results thus far show that speakers produce more representational gestures when they can see their listeners than when they cannot. We next consider whether this change in gesture production was accompanied by changes in the amount, fluency, or content of speech.

Amount of speech. If speakers compensate with speech when they cannot use gestures communicatively, they might use more words when speaking to listeners they cannot see than to listeners they can see. Alternatively, if speakers avoid content that is best expressed using gesture, they might use fewer words when speaking to listeners they cannot see. To address these possibilities, we compared the number of words per episode in the face-to-face and screen conditions. Although participants produced slightly more words per episode in the screen condition, the effect was not reliable (face-to-face M = 127.1, SE = 10.0, vs screen M = 141.2, SE = 10.5, words per episode), $F(1, 108) = 1.60$, n.s. The same pattern held for the number of transcript units per episode (face-to-face M = 20.3, SE = 1.6, vs screen M = 21.5, SE = 1.6, transcript units per episode), $F(1, 108) < 1$, n.s. Thus, there was no evidence that the amount of speech was affected by the visibility manipulation.

Fluency of speech. If speakers modify their speech when they cannot see their listeners, either to compensate for the absence of gesture or to avoid content that is best expressed in gesture, then formulating speech should be more effortful in the screen condition than in the face-to-face condition. This additional effort might be manifested in reduced fluency of speech. We assessed fluency using three dependent measures: (1) the rate of words per second, (2) the rate of filled pauses, and (3) the rate of speech errors.

We first compared speech rate across conditions. Overall, participants spoke more slowly in the screen condition (face-to-face M = 2.51, SE = 0.13, vs screen M = 2.39, SE = 0.14 words per second), $F(1, 14) = 5.73$, $p < .05$. This pattern was strong in participants who narrated in the screen condition first, but was absent in participants who narrated in the face-to-face condition first, yielding a marginally significant interaction between order and condition, $F(1, 14) = 3.50$, $p = .08$. The same pattern held when the nonnative speakers were excluded from the analysis; however, the interaction of order and condition reached significance, $F(1, 12) = 4.78$, $p < .05$, and the main effect of condition declined to marginal significance, $F(1, 12) = 3.56$, $p = .08$. One possible interpretation of the interaction is that it is due to a warm-up effect. Overall, participants spoke more slowly during the first half of the experiment than during the second half. They also spoke more slowly in the screen condition than in the face-to-face condition. In the screen first group, these effects added up to yield a substantial difference across conditions, whereas in the face-to-face first group these effects cancelled one another out.

We next compared the rate of filled pauses (e.g., pauses that contained filler words such as "um" and "uh") across conditions. Filled pauses were produced at a slightly higher rate in the screen condition than in the face-to-face condition (face to face M = 2.44, SE = 0.37 vs M = 3.23, SE = 0.53 per 100 words), $F(1, 14) = 4.85$, $p < .05$; this effect declined to marginal significance when the nonnative speakers were excluded from the analysis, $F(1, 12) = 4.24$, $p = .06$. This finding suggests that narrators spoke less fluently when they could not see their listeners. However, it could also be the case that speakers used filled pauses to signal dysfluencies (e.g., delays in word retrieval) in the screen condition and gestures to signal dysfluencies in the face-to-face condition. Speakers may have been equally fluent in both conditions, but may have signaled their dysfluencies to the listener in different ways.

Finally, we compared the rate of speech errors across conditions. We found no systematic differences in the rate of speech errors per 100 words when speakers could and could not see their listeners (face-to-face $M = 2.98$, $SE = 0.36$, vs screen $M = 3.16$, $SE = 0.35$, per 100 words), $F(1, 14) < 1$, n.s.

Content of speech. We assessed content using three dependent measures: (1) the types of verbs used to describe motion events, (2) the proportion of words that conveyed nonnarrative content, and (3) the rate of spatial prepositions.

We first examined whether the visibility manipulation influenced the verbs speakers used to describe motion events. In face-to-face interaction, English speakers frequently convey the manner associated with a given motion both in words (such as roll, swing, crawl; Talmy, 1985) and in gestures (McNeill, in press). For example, a speaker might express that Sylvester went up the pipe with a climbing motion by saying, “he *climbed* up the pipe,” while producing a gesture that depicts climbing. In this example, the speaker expressed the manner of motion (climbing) in both speech and gesture. In other cases, speakers express manner information only in speech (e.g., “he *climbed* up the pipe” with a gesture that indicates direction), only in gesture (e.g., “he *went* up the pipe” with a gesture that depicts climbing), or in neither speech nor gesture (e.g., “he *went* up the pipe” with a gesture that indicates direction).

If speakers compensate with speech when they cannot use gestures communicatively, they might express manner information in speech rather than in gesture, so they might use more verbs that convey manner when they are unable to see their listeners. Alternatively, if speakers avoid manner information altogether, they might use fewer verbs that convey manner when they are unable to see their listeners. To test these possibilities, we identified 12 motion events in the cartoon and examined whether speakers described the events using verbs that conveyed manner information or verbs that did not (e.g., *tiptoed* vs *went* along the trolley wires). We found no systematic differences in the mean proportion of verbs that conveyed manner when speakers could and could not see their listeners [for the first half of the cartoon, 65% vs 70% of the target verbs, $t(14) = 0.39$, n.s.; for the second half, 81% vs 77% of the target verbs, $t(14) = 0.41$, n.s.]. Thus, there was no evidence that speakers’ choice of verbs of motion was affected by visibility.

We next examined whether the visibility manipulation influenced speakers’ use of nonnarrative speech. McNeill (1992) has argued that, in narrative discourse, representational gestures tend to accompany narrative content (i.e., the story itself), whereas beat gestures tend to accompany nonnarrative content (i.e., meta- or extranarrative content, such as information about the structure of the cartoon, about watching or retelling the cartoon, or about the camera movement). Since speakers typically produce representational gestures with narrative content, it is possible that speakers might focus less on such content (to the extent possible in a narrative task) when they are unable to use gestures communicatively and might instead produce more nonnarrative speech. To address this possibility, we compared the proportion of words that each speaker used to express nonnarrative content in each condition. We found no systematic difference in the mean proportion of nonnarrative words across conditions (face-to-face $M = 0.16$, $SE = 0.02$, vs screen $M = 0.15$, $SE = 0.02$), $F(1, 14) < 1$, n.s. Thus, there was no evidence that the amount of narrative speech was affected by the visibility manipulation.

Finally, we considered whether visibility influenced speakers’ expression of spatial content. If speakers compensate with speech when they cannot use gestures communicatively, then speakers might use more words that convey spatial content when they are unable to see their listeners. Alternatively, speakers might choose to avoid spatial content altogether, so they might use fewer words that convey spatial content when they are unable to see their listeners. To test these possibilities, we assessed the rate of spatial prepositions (e.g., up, across, or over, used either as verb particles or in prepositional phrases) per 100 words in each episode. We found no systematic differences in the rate of spatial prepositions when speakers could and could not see their listeners (face-to-face M = 8.06, SE = 0.27, vs screen M = 8.23, SE = 0.33, per 100 words), $F(1, 14) < 1$, n.s.

In sum, we found that narrators spoke slightly more slowly and produced more filled pauses (i.e., "ums" and "uhs") in the screen condition than in the face-to-face condition. However, we found no systematic differences across conditions in the amount or content of speech. Given the observed differences in fluency, it seems likely that there may be other, subtle differences in speech across conditions that we have not detected. Nevertheless, the present findings suggest that the visibility manipulation had a much more dramatic effect on the rate of representational gestures than on the amount and content of speech.

Did the Visibility Manipulation Influence the Integration of Gestures with Speech Content?

Finally, we assessed whether the manipulation influenced the relationship between speech content and gesture production. We explored two kinds of relationships that have been reported in the literature. First, Krauss and colleagues have claimed that representational gestures are associated with spatial content (Krauss, 1998; Rauscher et al., 1996). Second, as noted above, McNeill (1992) has argued that beat gestures are associated with nonnarrative content. We examined whether the strength of these relationships varied as a function of the visibility manipulation. To address these issues, we used the transcript unit as the unit of analysis (see "Method").

To examine the relation between spatial content and representational gestures, we scored whether each transcript unit included a spatial preposition and whether it was accompanied by a representational gesture. For each participant, we then calculated the proportion of units that were accompanied by representational gestures for units with spatial content (e.g., "Tweety drops a bowling ball *down* the drain pipe") and for units without spatial content (e.g., "Tweety and Granny are driving it"). Consistent with Krauss's claim, participants produced representational gestures with more of the units that included spatial content, (M = 0.59, SE = 0.04, vs M = 0.32, SE = 0.03), $F(1, 14) = 130.44$. This pattern held both in the face-to-face condition, $F(1, 14) = 137.40$, and the screen condition, $F(1, 14) = 60.84$. The three-way interaction of condition, condition order (screen first and face-to-face first), and type of unit (with vs without spatial content) was also significant, $F(1, 14) = 11.77$, as was the two-way interaction of condition and type of unit, $F(1, 14) = 7.69$, $p < .02$. For participants who narrated in the face-to-face condition first, the association between spatial content and representational gestures was stronger in the face-to-face condition than in the screen condition. For participants who narrated in the screen condition first, the association between spatial content and representational gestures was comparable in both conditions.

To examine the relation between nonnarrative content and beat gestures, we scored whether each unit included nonnarrative content and whether it was accompanied by a beat gesture. For each participant, we then calculated the proportion of units that were accompanied by beat gestures for units with nonnarrative content (e.g., "the camera's panning across") and for units without nonnarrative content (e.g., "he tries to go up the drainpipe").[3] Across subjects, units with narrative content were much more common than units with nonnarrative content (M = 86% narrative units, SE = 2%). Overall, consistent with McNeill's claim, participants produced beat gestures with proportionately more of the units that included nonnarrative content (M = 0.33, SE = 0.04, vs M = 0.23, SE = 0.02), $F(1, 14) = 5.78$, $p < .05$. However, this pattern held only in the face-to-face condition, $F(1, 14) = 11.39$, and not in the screen condition, $F(1, 14) < 1$, n.s. (see Fig. 2). The interaction of condition and type of unit (nonnarrative vs not) was marginally significant, $F(1, 14) = 3.90$, $p = .07$. As seen in Fig. 2, when speakers could see their listeners,

[3] It should be noted that our category of nonnarrative units does not include all units that, according to McNeill's (1992) theory, should be accompanied by beat gestures.

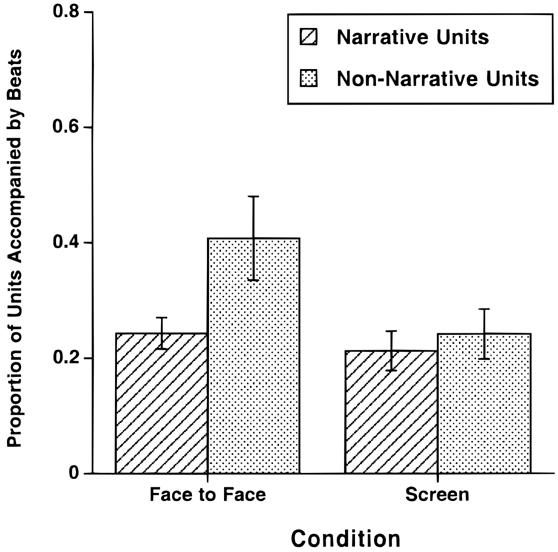


FIG. 2. Proportion of narrative and nonnarrative units accompanied by beat gestures in each condition.

beat gestures were associated with nonnarrative content. However, this was not the case when speakers could not see their listeners. This finding suggests that speakers may produce beat gestures with nonnarrative content for listeners' benefit.

DISCUSSION

The findings do not support the *rhythmical pulse hypothesis*, which holds that visibility between speaker and listener should not influence overall levels of gesture production. Instead, the findings support the *semantic information hypothesis,* which holds that visibility between speaker and listener should influence speakers' production of gestures that convey semantic information. There were three main findings. First, representational gestures were produced at higher rates when speakers could see their listeners than when they could not. Second, representational gestures did not disappear in the screen condition. Speakers continued to produce representational gestures at a fairly high rate even when they could not see their listeners. Third, beat gestures were produced at comparable rates overall when speakers could see their listeners and when they could not. We discuss each of these findings and its implications in turn.

Representational Gestures Showed Effects of the Visibility Manipulation

In this experiment, the rate of representational gestures depended on visibility. Speakers produced representational gestures more frequently when they could see their listeners than when they could not. This finding suggests that speakers produce representational gestures in order to communicate to their listeners. It is tempting to interpret this finding as evidence that speakers intend to use gestures communicatively (see Cohen & Harrison, 1973, for discussion). However, in our view, these data do not directly address the issue of intention, so no firm conclusions on this issue can be drawn. Regardless of intention, the present findings are consistent with the view that speakers produce representational gestures for communicative purposes.

However, some alternative interpretations of this finding are possible. First, the observed differences in representational gestures may derive from differences in listener behavior across conditions. Perhaps speakers were "rewarded" with smiles and nods for producing representational gestures in the face-to-face condition, so they produced these gestures at a higher rate when they could see their listeners. In the screen condition, when speakers could not see their listeners' smiles and nods, they produced representational gestures at a lower rate. In the present experiment, listener visibility and availability of feedback cannot be separated, so we cannot definitively rule out this alternative; future work with a confederate listener would be required. However, we find this alternative account unlikely because most of the listeners were quite impassive and smiled or nodded only rarely. Further, the effects of listener visibility were consistent across narrators, despite modest variations in listener behavior. Finally, some backchannel feedback (e.g., laughter) was also available in the screen condition.

Another possibility is that, when the screen was set up, narrators may have "guessed" that gesture production was the focus of the experiment and they may have attempted to suppress their gesture production in the screen condition. Representational gestures may have been more easily suppressed than beat gestures, and hence representational gestures, but not beat gestures, showed effects of the visibility manipulation. According to this view, the present findings reveal only how easily representational gestures are suppressed. Again, we cannot rule out this alternative; however, we find it unlikely because speakers continued to produce both representational gestures and beat gestures at fairly high rates in the screen condition. If speakers were intentionally trying to suppress their gestures, we expect that they would have done a much better job of doing so.

Our findings add to the growing body of literature that shows that representational gestures are produced more often when a listener is watching. However, this does not necessarily imply that such gestures have communicative effects. Many studies have presented evidence that gestures do have communicative effects (see Kendon, 1994, for a review); however, several other studies have failed to find such effects (e.g., Krauss et al., 1995). It is possible that gestures have communicative effects primarily when they supplement or "mismatch" speech, not when they are redundant with speech (Goldin-Meadow, Alibali, & Church, 1993; Goldin-Meadow, Wein, & Chang, 1992; McNeill, Cassell, & McCullough, 1994). Alternatively, gestures may have communicative effects primarily when speakers' verbal messages are complex relative to their listeners' skills (McNeil, Alibali, & Evans, 2000).

Representational Gestures did not Disappear in the Screen Condition

Although speakers produced more representational gestures when they could see their listeners, they produced many representational gestures even when visibility was blocked. This finding is consistent with the view that representational gestures play a role in speech production (Kita, 2000; Krauss, 1998; Rimé & Shiaratura, 1991) as well as in communication. Indeed, a recent study showed that children blind from birth also produce representational gestures (Iverson & Goldin-Meadow, 1997). Thus, there is mounting evidence that representational gestures serve a function for the speaker that is independent of their function for the listener.

However, there are some alternative interpretations of the finding that speakers continued to produce representational gestures when visibility was blocked. First, speakers may have produced representational gestures in the screen condition simply out of habit. Second, speakers may have produced representational gestures in the screen condition because they imagined their communication partner and not because such gestures are involved in speech production (Fridlund, 1994). Indeed, Fridlund argues that "imaginal interactants" can never be excluded from consideration (p. 166). Of course, the communicative setting in the screen condition in this study, although not face-to-face, was certainly not nonsocial. Further research will be needed to establish the extent to which gesture production depends on the social context (see Özyürek, 2000).

Finally, speakers may have produced representational gestures in the screen condition because gestures and speech together make up "composite signals" that depend on speakers' communicative intentions and that cannot be separated (see Clark, 1996; Engle, 1998, for discussion). According to Engle (1998), spoken and gestured composite signals are *integrated* units of communication rather than combinations of independently interpretable signals from each channel. The integrity of such composite signals requires both components to be produced together, and hence they are produced together even when listeners cannot see them. According to this view, the fact that speakers continued to produce representational gestures in the screen condition does not necessarily imply that such gestures are involved in speech production.

The present study was not designed to test the function of representational gestures for speakers, and no definitive conclusions can be drawn based on the finding that representational gestures did not disappear in the screen condition. However, other aspects of the present findings lend support to the idea that representational gestures play a role in speech production. The decrease in representational gestures in the screen condition coincided with a reduction in verbal fluency, manifested both in a slight decrease in speech rate and in an increase in the rate of filled pauses. Thus, speech was more dysfluent when representational gesture production decreased. This combination of findings is consistent with the claim that representational gestures do play a role in the process of speech production.

Nevertheless, the present findings cannot specify the particular point in the process of speech production in which gesture is involved. The pattern of results is consistent with either of two views: (1) that gesture plays a role in accessing lexical items (e.g., Butterworth & Hadar, 1989; Krauss, 1998) or (2) that gesture plays a role in conceptualizing the message to be verbalized (e.g., Kita, 2000). Further research will be needed to tease apart these possibilities (see Alibali, Kita, & Young, 2000).

Beat Gestures Did Not Show Effects of Visibility

In this study, speakers produced beat gestures at comparable rates overall when they could see their listeners and when they could not. However, the association of beat gestures with nonnarrative content varied depending on whether speakers could see their listeners. In the face-to-face condition, beat gestures patterned with nonnarrative content; however, in the screen condition, they did not. This finding suggests that beat gestures may be used to mark nonnarrative speech for the benefit of the listeners. Speakers may use beat gestures to highlight nonnarrative speech so that listeners can better apprehend the narrative structure.

In both conditions, many beats were produced with units that did not contain nonnarrative content. In many cases, these beats appeared to be bound to the rhythm of speech. If one accepts that such beats are tightly linked to the prosodic structure of speech, it is not surprising that they were not affected by the visibility manipulation. In general, the findings suggest that some beats (i.e., those that are linked to prosody) may be used for speaker-internal purposes, whereas other beats (i.e., those that signal shifts away from the story line) may be used for communicative purposes. More generally, these findings suggest that beat gestures are a functionally heterogeneous category. Indeed, Bavelas and her colleagues have recently questioned the functional significance of the gestures classified as beats in the current study, and they have argued that such gestures serve a diversity of interactive functions (Bavelas et al., 1992; 1995). The present findings highlight the need for further study of beat gestures and their role in speech production and communication.

Reconciling Conflicting Findings in the Literature

The present results suggest two possible explanations for the conflicting findings in the literature about whether visibility influences gesture production. First, most investigators who have addressed this issue have examined broad categories of gestures (e.g., *illustrators* or *communicative* gestures). The present study is the first visibility study in which representational and beat gestures have been examined separately. The conflicting findings in the literature may be due to the fact that some studies used tasks that elicited primarily beat gestures, which do not show effects of visibility, whereas other studies used tasks that elicited primarily representational gestures, which do show effects of visibility. Indeed, a recent study confirmed that different types of tasks elicit different distributions of gesture types. Feyereisen and Harvard (1999) found that speakers produced more beat gestures when speaking about abstract topics and more representational gestures when speaking about topics that involve visual or motor imagery. Among visibility studies, those that have demonstrated effects of visibility on gesture production (e.g., Cohen, 1977; Cohen & Harrison, 1973; Krauss et al., 1995) used tasks with high spatial content (giving directions, describing abstract figures), which may have elicited primarily representational gestures. In contrast, studies that revealed nonsignificant or null effects of visibility (e.g., Rimé, 1982) used expository or recollection tasks (e.g., discussing "what one likes to find in the cinema"), which may have elicited primarily beat gestures. In this regard, the present findings support the utility of distinguishing representational and beat gestures.

Second, prior studies that have demonstrated effects of visibility used tasks in which speech content was well controlled across participants. For example, in Cohen's work (Cohen, 1977;

Cohen & Harrison, 1973), participants all provided route directions for the same set of routes. In contrast, the studies that revealed nonsignificant or null effects of visibility used open-ended tasks, which presumably led to high variability in task performance across participants. Such heterogeneity in participants' performance could lead to inadequate statistical power and Type II errors. The narrative task used in the present study proved to be a well-chosen test bed for hypotheses about visibility, not only because it routinely elicits both representational and beat gestures but also because it controls speech content across participants. All participants described the same scenes in the same order; hence, extraneous task-related variability in speech content was minimized.

Implications for Theories about How Gesture Is Produced

The present findings have implications for theories about the internal processes that underlie gesture production. The experiment was not designed to provide conclusive evidence for a specific model of gesture production; however, our findings are not consistent with Tuite's suggestion that semantic content is simply overlaid on a "rhythmical pulse" in order to generate representational gestures (Tuite, 1993). The rhythmical pulse that Tuite posits should not vary across conditions, so according to his model, the overall frequency of gesture should not change across conditions. Instead, his theory predicts a decrease in representational gestures and a corresponding increase in beat gestures. We found an overall decrease in gesture rate when visibility was blocked, due to a dramatic decrease in representational gestures, and a minimal, nonsignificant decrease in beat gestures. This pattern is inconsistent with the rhythmical pulse hypothesis.

Our findings are consistent with the view that different types of gestures are generated by different processes (Krauss et al., 1996). At a more general level, our findings indicate that models of gesture production must incorporate a mechanism by which characteristics of the communicative setting can influence the production of representational gestures. Furthermore, our data show that such a mechanism must be able to influence whether or not a gesture is produced and not simply what kind of gesture is produced.

Practical Implications and Conclusions

The present findings have important practical implications for investigators who use gestures as a source of information about mental processes (e.g., Alibali, Bassok, Solomon, Syc, & Goldin-Meadow, 1999; Schwartz & Black, 1996). Specifically, our results suggest that participants in such studies may produce more content-laden representational gestures if they talk to the experimenter or to another participant face-to-face. Because representational gestures are used in order to communicate, experimental tasks that are designed to elicit gestures should involve social communication. However, it should be noted that verbal and gestured protocols provided in social communicative settings may differ in important ways from protocols provided in nonsocial settings. In particular, verbal protocols provided in nonsocial settings are thought to accurately reflect the contents of working memory (Ericsson & Simon, 1993), but this may not be the case for protocols provided in social settings. Hence, methodological decisions about how to elicit gestures in experimental settings must depend crucially on the nature of the research questions being addressed.

The present study showed that visibility between speaker and listener had different effects on representational and beat gestures. Taken together, the results indicate that both representational and beat gestures serve speaker-internal and communicative functions. Let us first consider representational gestures. In this study, speakers produced representational gestures at higher rates when they could see their listeners than when they could not, suggesting that representational gestures are produced in order to communicate. However, speakers continued to produce representational gestures even when their listeners could not see those gestures, and the decrease in representational gestures coincided with a decrease in fluency. These findings suggest that representational gestures may also play a role in speech production. Next, consider beat gestures. In this study, speakers produced beat gestures at similar rates when they could see their listeners and when they could not, suggesting that such gestures may serve a speaker-internal function. However, beat gestures were more closely linked with nonnarrative speech content when speakers could see their listeners than when they could not, suggesting that at least some beat gestures have a communicative function.

In sum, we suggest that both beat and representational gestures are multifunctional and that there is not just a single answer to the question of why speakers gesture. Research on gesture production should move beyond asking whether gestures are produced to aid communication *or* to aid speech production. Instead, research should examine how different speakers use gestures in different types of contexts for both speaker-internal and communicative purposes.

APPENDIX

Description of Cartoon Episodes

Episode 1: Garbage

Using binoculars, Sylvester spies Tweety in a window at the "Broken Arms" apartment building, across the street from his building. Sylvester goes into the main entrance of Tweety's building, but he is kicked out and lands in a pile of garbage.

Episode 2: Drainpipe

Sylvester climbs up the drainpipe next to Tweety's window and climbs in the window, trying to catch Tweety. Granny hits Sylvester with an umbrella and throws him out the window.

Episode 3: Bowling Ball

Sylvester starts to climb up the *inside* of the drainpipe next to Tweety's window. Tweety drops a bowling ball down the drainpipe. Sylvester swallows the bowling ball, falls down out of the drainpipe with the bowling ball in his belly, and rolls into a bowling alley.

Episode 4: Monkey

Sylvester knocks out an organ grinder's monkey and steals his outfit. Disguised as a monkey, Sylvester climbs up the drainpipe next to Tweety's window and climbs in the window. Granny notices the "monkey" and offers him a penny for his cup. Sylvester tips his cap, and Granny hits him on the head with the umbrella, saying, "I was hep to ya all the time!"

Episode 5: Bellhop

Sylvester is hiding in a mailbox at the front desk of Tweety's building. Granny calls the front desk to say that she is checking out, and she asks the clerk to "send up a boy to get my bags and bird." Sylvester, dressed as a bellhop, appears at Granny's door, and collects her bags and the covered birdcage. He discards the suitcase and tiptoes downstairs with the birdcage. In the alley he removes the cover and finds Granny hiding in the cage. She hits him on the head with her umbrella.

Episode 6: Catapult

Sylvester sets up a seesaw with a crate and board under Tweety's window, and then he stands on one end of the seesaw holding a 500-pound weight. He throws the weight onto the other end of the board and is propelled into the air. As he flies past Tweety's window he grabs Tweety. He then lands on the board again, holding Tweety, and the weight is propelled into the air again by his landing on the board. Sylvester runs off, but as he does so, the weight falls on his head, flattening him. Tweety escapes from his grasp.

Episode 7: Swing

Sylvester sets up a rope between his building and Tweety's building, which he plans to use to swing into Tweety's window, Tarzan-style. He leaps off his window ledge holding the rope, slams into the side of Tweety's building, and falls to the ground.

Episode 8: Trolley

Sylvester is walking on the overhead trolley wires. The trolley car approaches him from behind, bell ringing. Sylvester runs with the trolley car following him. When the car connector reaches him, Sylvester receives a shock and jumps up from the wire as if exploding. He lands on the wire and runs a few more steps before the connector reaches him again and he receives another shock. After several shocks, the camera pans to a view of the trolley driver—Tweety Bird—and the bell ringer—Granny.

REFERENCES

Alibali, M. W., Bassok, M., Solomon, K. O., Syc, S. E., & Goldin-Meadow, S. (1999). Illuminating mental representations through speech and gesture. *Psychological Science*, **10,** 327–333.

Alibali, M. W., Flevares, L., & Goldin-Meadow, S. (1997). Assessing knowledge conveyed in gesture: Do teachers have the upper hand? *Journal of Educational Psychology,* **89,** 183–193.

Alibali, M. W., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes*, **15,** 593–613.

Bavelas, J. B. (1994). Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction,* **27,** 201–221.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin,* **21,** 394–405.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes,* **15,** 469–489.

Beattie, G., & Shovelton, H. (1999). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, **123,** 1–30.

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, **96,** 168–174.

Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology*, **30,** 580–601.

Cicone, M., Wapner, W., Foldi, N., Zurif, E., & Gardner, H. (1979). The relation between gesture and language in aphasic communication. *Brain and Language*, **8,** 324–349.

Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge Univ. Press.

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

Cohen, A., & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, **6,** 341–349.

de Ruiter, J.-P. (1995). Why do people gesture at the telephone? In M. Biemans & M. Woutersen (Eds.), *Proceedings of the Center for Language Studies Opening, Academic Year 95–96* (pp. 49–55). Nijmegen, The Netherlands: University of Nijmegen.

Doherty-Sneddon, G., & Kent, G. (1996). Visual signals and the communication abilities of children. *Journal of Child Psychology and Psychiatry*, **37,** 949–959.

Efron, D. (1941/1972). *Gesture, race and culture*. The Hague: Mouton.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, **1,** 49–98.

Emmorey, K., & Casey, S. (in press). Gesture, thought, and spatial language. *Gesture*.

Engle, R. A. (1998). Not channels but composite signals: Speech, gesture, diagrams, and object demonstrations are integrated in multi-modal explanations. In M. A. Gernsbacher & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 321–326). Mahwah, NJ: Erlbaum.

Ericsson, K. A., & Simon, H. (1993). *Protocol analysis* (rev. ed.). Cambridge, MA: MIT Press.

Feyereisen, P., & Harvard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior*, **23,** 153–171.

Fridlund, A. J. (1994). *Human facial expression: An evolutionary view*. San Diego: Academic Press.

Goldin-Meadow, S., Alibali, M. W., & Church, R. B. (1993). Transitions in concept acquisition: Using the hand to read the mind. *Psychological Review*, **100,** 279–297.

Goldin-Meadow, S., Wein, D., & Chang, C. (1992). Assessing knowledge through gesture: Using children's hands to read their minds. *Cognition and Instruction*, **9,** 201–219.

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.

Hadar, U. (1989). Two types of gesture and their role in speech production. *Journal of Language and Social Psychology*, **8,** 221–228.

Hadar, U., Wenkert-Olenik, D., Krauss, R., & Soroker, N. (1998). Gesture and the processing of speech: Neuropsychological evidence. *Brain and Language*, **62,** 107–126.

Iverson, J., & Goldin-Meadow, S. (1997). What's communication got to do with it? Gesture in children blind from birth. *Developmental Psychology*, **33,** 453–467.

Kelly, S. D., & Church, R. B. (1998). A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development*, **69,** 85–93.

Kendon, A. (1994). Do gestures communicate? A review. *Research on Language and Social Interaction*, **27,** 175–200.

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture: Window into thought and action* (pp. 162–185). Cambridge, UK: Cambridge Univ. Press.

Krauss, R. M. Dushay, R., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, **31,** 533–552.

Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science*, **7,** 54–60.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology*, **28,** 389–450.

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

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, **14,** 41–104.

Lickiss, K. P., & Wellens, A. R. (1978). Effects of visual accessibility and hand restraint on fluency of gesticulator and effectiveness of message. *Perceptual and Motor Skills*, **46,** 925–926.

McNeil, N. M., Alibali, M. W., & Evans, J. L. (2000). The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior*, **24,** 131–150.

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

McNeill, D. (in press). An ontogenetic universal and how to explain it. In C. Mueller & R. Posner (Eds.), *The semantics and pragmatics of everyday gestures*. Berlin: Verlag Arno Spitz.

McNeill, D., Cassell, J., & McCullough, K.-E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language in Social Interaction*, **27,** 223–237.

McNeill, D., Levy, E. T., & Pedelty, L. L. (1990). Gesture and speech. In G. R. Hammond (Ed.), *Cerebral control of speech and limb movements* (pp. 203–256). Amsterdam: Elsevier/North Holland.

Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (Ed.), *Language and gesture: Window into thought and action*. (pp. 64–83). Cambridge, UK: Cambridge Univ. Press.

Özyürek, A., & Kita, S. (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture and conceptualization. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 507–512). Mahwah, NJ: Erlbaum.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, **7,** 226–231.

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

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

Schwartz, D., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, **20,** 457–497.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description III: Grammatical categories and the lexicon* (pp. 57–149). Cambridge, UK: Cambridge Univ. Press.

Tuite, K. (1993). The production of gesture. *Semiotica*, **93**, 83–105.

(Received September 21, 1998)
(Revision received July 11, 2000)