J Nonverbal Behav (2009) 33:73–88
DOI 10.1007/s10919-008-0063-9

ORIGINAL PAPER

Do Iconic Hand Gestures Really Contribute to the Communication of Semantic Information in a Face-to-Face Context?

Judith Holler · Heather Shovelton · Geoffrey Beattie

Published online: 20 January 2009

Abstract Previous research has shown that iconic gestures are effective at communicating semantic information, particularly about the size and relative position of objects. However, the conclusions of these experiments have been somewhat limited by the fact that the methodology has typically involved presenting gesture–speech samples on video rather than in an actual face-to-face context. Because these different viewing conditions can impact on addressees' behavior and perception, and therefore potentially impact on the amount of information they receive from gestures, the present study compares the communicative effectiveness of iconic gestures when viewed in a face-to-face context compared to when viewed on video. The results are quite striking in that gestures seemed at least as effective, and in some cases even more effective at communicating position and size information when they occurred in the face-to-face condition compared to video.

Keywords Iconic gesture · Face-to-face communication · Semantic features · Size information · Relative position information

Introduction

There is now considerable evidence that iconic hand gestures play a crucial role in human communication as they have been shown to convey core semantic information over and above that contained in the accompanying speech in a variety of experimental and quasi-experimental situations (e.g., Beattie and Shovelton 1999a, b, 2001, 2002, 2005, in press; Riseborough 1981; Rogers 1978). However, all of this evidence derives from studies which involve playing *video-recorded* gesture–speech samples to decoders and therefore does not allow us to reach strong conclusions about the role such gestures might play in face-to-face communication. This study aims to close this gap by comparing the amount of semantic information recipients receive in contexts in which gesture and speech information is

J. Holler (✉) · H. Shovelton · G. Beattie
School of Psychological Sciences, University of Manchester, Coupland Building 1, Manchester M13 9PL, UK
e-mail: judith.holler@manchester.ac.uk

presented on video and when they occur in a more natural face-to-face context. If we fail to find that significant amounts of information are transmitted by gesture in a face-to-face context this would have a devastating effect on the growing consensus in this area that gestures are highly communicative, based upon considerable experimental evidence (but almost all, unfortunately, deriving from video-presented stimulus material).

In face-to-face conversation, people use not only words but also a multitude of nonverbal behaviors to communicate. A considerable amount of research has focussed on facial communication, particularly with respect to the communication of emotion (e.g., Darwin 1872; Ekman 1993, 1999; Ekman et al. 1980). Other types of nonverbal behavior, such as posture, the spatial orientation of interactants, and gaze patterns, have also been thoroughly studied, mainly in the context of the communication of affect and interpersonal relationships (e.g., Argyle 1975; Beattie 2003), or, alternatively, in the sequential organization of interaction (e.g., Goodwin 1981). Another aspect of nonverbal behavior which researchers have paid increasing attention to in recent years is a certain type of gesture—namely those iconic hand movements that frequently accompany everyday talk (McNeill 1985, 1992). In contrast to other forms of nonverbal communication involved in the communication of interpersonal attitudes and emotional states, iconic gestures are very closely linked to the *semantic* contents of the speech that they accompany, in both form and timing. Speech and gesture *together* comprise an utterance and both together externalize thought; they are believed to emerge from the same underlying cognitive representation and to be governed, at least in part, by some of the same cognitive processes (Kendon 1980, 2000; McNeill 1985, 1992).

McNeill (1992, p. 13), one of the pioneers in the research on iconic gestures, provides a well-known example which illustrates how iconic gestures can represent semantic information relating to an underlying idea which is not contained in the accompanying speech. In this case, a speaker describes an event from a cartoon story in which one character is pursuing another, using an umbrella as a weapon. The speaker describes this event by saying 'she chases him out again'. However, only the accompanying iconic gesture shows that the character which the speaker is referring to is gripping an umbrella, with one hand, and swinging it through the air. According to McNeill, this example demonstrates that iconic gestures are a core part of the speaker's utterance. Thus, a recipient of the message needs to understand the speech and simultaneously interpret the gesture to understand the full representation that the speaker was trying to communicate.

A number of experimental studies have demonstrated that participants who see these accompanying gestures as well as hearing the speech do indeed pick up significantly more relevant information (e.g., Graham and Argyle 1975; Riseborough 1981; Rogers 1978), and, more recently, a number of studies have tried to determine precisely what kinds of information are communicated effectively by these gestures (Beattie and Shovelton 1999a, b, 2001, 2002, 2005; see also Beattie 2003). In their studies, Beattie and Shovelton split complex messages into a range of semantic categories (for example, *identity, description of action, shape, size, movement, direction, speed, relative position*; see Beattie and Shovelton 2002). This research has demonstrated that gestures communicate additional information overall, but they appear to be particularly effective at communicating information about the *size* of objects and the *relative position* of objects with respect to one another. These two semantic categories seemed to have the most robust effects in the various statistical analyses used in their experimental studies.

However, one issue with the experimental paradigm employed by previous studies trying to measure the communicative effectiveness of iconic gestures is that it has only presented *video-recorded* gesture–speech segments to experimental participants. This means that we cannot necessarily generalize these findings to face-to-face talk. After all, the behavior of individuals when watching and listening in face-to-face communication may be different from their behavior when watching and listening to a video-recording. For example, face-to-face communication places specific constraints on the gaze behavior of both speakers *and* listeners (see Beattie 1983). Participants in dyadic interaction gaze at their interlocutor around 75% of the time when listening (Argyle and Ingham 1972; but see also Kendon 1967, and Argyle and Cook 1976, for findings which show considerable variability in this respect), and they gaze particularly at the speaker's face (e.g., Gullberg and Holmqvist 1998) and even more specifically at the speaker's eye-region (see Argyle and Cook 1976) during actual talk. Gaze at other parts of the body during face-to-face conversation can not only be discriminated but may on occasion be socially negatively sanctioned (see Argyle and Cook 1976). So, one prediction might be that when viewing *video-recordings* of speakers talking, participants may spend more time gazing at parts of the body other than the face, including the hands.

However, Gullberg and Holmqvist (2006) recently compared the actual patterns of the gaze behavior of participants in a face-to-face condition to that of participants watching speakers in two types of video screen conditions, one on a life-sized video screen and one on a small-sized screen (28″) (the latter is most similar to the size of the screen used in Beattie and Shovelton's experiments, see above). This research found that, overall, the pattern of gaze behavior was actually very similar in the three conditions; the life-sized video screen condition differed hardly at all from the face-to-face condition, which led the authors to conclude that using this type of video screen yields ecologically valid findings which can be compared to gaze behavior in a face-to-face condition. Only some very small differences emerged when comparing the smaller video screen condition to the face-to-face condition; in this video screen condition, participants spent slightly less time focusing on the face (91% of the viewing time, as compared to 96% of the viewing time in the face-to-face condition) and—contrary to what one might have assumed—they spent slightly *less* time fixating gestures (0.2% of the overall viewing time in the small-screen video condition compared to 0.5% in the face-to-face condition) and they visually fixated a significantly smaller proportion of gestures (3% in the small-screen condition compared to 7% in the face-to-face condition). Instead, participants in the small-screen video condition spent a bit more time looking at body-parts other than the face and gestures than in the life-sized video screen condition (1.3% of viewing time in the face-to-face condition, 1.4% in the life-sized video condition, and 5.6% in the 28″ screen video condition).

However, whether any of these differences would actually affect the amount of information that recipients would glean from the gestural representations is currently not known. A recent eye-tracking study carried out by Gullberg and Kita (forthcoming; cited in Gullberg 2003) tested the association between overt fixations of gestures and up-take of information about motion direction. This study showed, firstly, that addressees sometimes overtly fixated gestures but did not process the information from these gestures, and, secondly, that addressees often picked up information (at a level above chance) from gestures that they themselves did not overtly fixate but which were looked at by the speaker-gesturer him or herself. This suggests that overt visual fixation of gestures is not a necessary prerequisite for processing semantic information from gestures, at least with respect to directional information. Moreover, the authors' conclude that speaker-fixation of gestures may provide an important interactional cue influencing addressees' information up-take from gesture.

Thus, while recipients' eye movement patterns may not differ to a great extent when comparing face-to-face and video presentation contexts, what might differ is recipients' use of potential social cues received from the speaker-gesturer, such as the speaker's fixations of his or her own gestures (for example, because recipients in face-to-face communication may feel more socially engaged and involved in the communication process, which may potentially alter their attention and perceptiveness of social information cues). One assumption therefore is that recipients may nevertheless pick up different amounts of gesturally represented semantic information in video and face-to-face contexts, despite only small differences in *overt* visual fixations.

In addition, previous studies using video-recorded segments have introduced other experimental features which could potentially have considerably *inflated* possible communicative effects of gesture. For example, only clause length segments were used in the majority of studies (Beattie and Shovelton 1999b, 2001, 2002; see also Beattie 2003) and the segments were often played twice (Beattie and Shovelton 1999b, 2001, 2002). Clearly, in everyday talk gesture–speech segments are embedded in a broader communicational context, in the case of storytelling within a complete and often complex narrative. And in everyday talk recipients usually do not get a second chance to view a gesture, unless they ask for clarification in conversation (or the speaker initiates a self-repair) and the rephrased utterance is accompanied by the same or a similar gesture (for examples see Kendon 2004). These two experimental features could easily have resulted in an over-estimation of the communicative effects of gesture (cf. Holler 2003; Holler and Beattie 2003).

In sum, based on past research, different predictions could be made with respect to how the amount of semantic information communicated by iconic gestures compares in video and face-to-face contexts. Because in video presentation conditions recipients are not constrained in their gaze behavior through social and conversational norms, they may pay more attention to the speaker's gestures. This could mean that in a face-to-face context they pick up *less* information from gestures than when viewing these same gestures on video. At the same time, eye-tracking studies show that only small differences exist in terms of *overt* visual fixation patterns when comparing video and face-to-face contexts, but we do not yet know whether, and if so how, these differences affect information up-take. Additional findings have shown that one aspect potentially influencing addressees' covert attention to gestures is speakers' overt attention to their own gestures. It is possible that this and other interactional influences affect how recipients process gestural information when they feel more socially engaged. This could mean that recipients glean *more* information from gestures in a face-to-face context. In addition, we have mapped out several methodological aspects which may have artificially enhanced information-uptake from gestures in previous studies employing video presentations. This, again, would provide grounds for assuming that gestures are actually *less* communicative in a face-to-face context. And, of course, face-to-face and video presentation contexts may differ in other respects too. For example, the quality of the image resolution may affect how much information recipients glean from individual gestures. Furthermore, research has shown that memory for events is significantly enhanced when event information (e.g., a science demonstration, magic tricks) is presented live instead of on video, at least in the case of children (Roebers et al. 2004; Thierry and Spence 2004). Both of these factors may mean that gestures could be *more* communicative when observed in a face-to-face context rather than on video.

Thus, the present study aims to answer the question of how communicative iconic gestures, which have been shown to convey a significant amount of information in a video presentation context, are in a face-to-face context. We concentrate here on gestures designed to carry size and positional information because we know that iconic gestures communicate information relating to these semantic categories particularly well, based on previous work (and we use a systematic approach to measuring communicative effects of gestures as used by Beattie and Shovelton 1999a, b, 2001, 2002, 2005). Four experimental conditions were used. In one condition, participants watched and listened to cartoon narrations in a face-to-face context (gesture + speech (face-to-face) condition). In a second condition, they watched and listened to video-recorded cartoon narrations presented on a TV screen (gesture + speech (video) condition). In the third condition, participants were presented with video-recorded narrations but the monitor was turned off, hence providing them with only the speech (speech only (video) condition). In the fourth condition, participants were presented with video-recordings of the narratives but this time the sound was turned off, preventing them from hearing the speech but allowing them to see the speech-accompanying hand gestures (gesture only (video) condition). The last two experimental conditions were used in addition to the face-to-face and the video condition in which participants were presented with both gesture and speech in order to test how much information each of the two channels would communicate on their own about the respective events our analysis focuses on. Because the present study differs methodologically in a number of respects from previous work (with the presentation of entire narratives, and no repeated presentation), the aim is to compare the results we obtain for these two experimental conditions with those from previous studies (Beattie and Shovelton 1999b, 2001, 2002). Also, because the iconic gestures selected as stimuli had been shown to be highly communicative in these previous studies but were here delivered by an actor, we aimed to obtain evidence that the gestures themselves were still communicative even though they were not produced spontaneously.

Method

Overview and Definition of Terms

The experimental paradigm used involved one actor telling cartoon stories, face-to-face, to one interlocutor at a time (each communication with a different interlocutor is henceforth referred to as an 'experimental trial'). These narrations were video-recorded and later played to additional listener–viewers, under different video presentation conditions (i.e., gesture + speech, gesture only, speech only). Thus, each experimental trial recorded in the face-to-face condition was played to three additional recipients (i.e., each video-recording was played three times, once in each of the three different presentation conditions). The cartoon stories told by the actor were scripted based upon spontaneous cartoon narratives obtained in a different experimental study (Beattie and Shovelton 1999a). In this previous study, participants were asked to look at images from a cartoon story and to tell the story to the experimenter to show how well they could understand and tell cartoon stories. Following the same procedure, each participant told two further cartoon stories. These cartoon narrations were video-recorded (and then used by Beattie and Shovelton 1999a, b, 2001, as stimulus material to carry out various analyses). These video-recordings of cartoon narratives served as the basis for creating the scripted cartoon stories used in the present study (delivered by the actor), as described below. The participants who told the stories in this previous experiment (Beattie and Shovelton 1999a) will therefore be referred to as the '*original* speakers/narrators', and their narratives as the '*original* cartoon narrations/narratives'. The narratives delivered by the actor in the present study will be referred to as the '*scripted* cartoon narratives/stimulus stories'.

Participants

Overall, 150 participants took part in the experiment as recipients, or 'decoders' (i.e., as participants who were presented with the actor's narratives, under the four presentation conditions). Of these 150 participants, 39 participants took part in the gesture + speech (face-to-face) condition. Due to technical problems, the first and the 33rd experimental trials (i.e., the narratives delivered to participant one and participant 33) did not record, hence reducing it to 37 valid recordings of the actor's cartoon narrations that could be used in all four conditions. Therefore, the data from 148 participants (37 in each condition) was used in the analyses for this study. In addition, one participant who took part in the 'speech only (video) condition' did not follow the instructions—this case was handled as missing data for this condition in the statistical analyses. All of the participants were recruited from the University of Manchester population using opportunity sampling and were compensated financially for their participation.

Stimulus Material and Experimental Design

Three scripted cartoon narratives were created based on a set of 14 original cartoon narrations of three different cartoon stories (Ivy the Terrible, Korky the Cat and Billy Whizz; see Beattie and Shovelton 1999a, b, 2001), i.e., 42 original cartoon narratives in total. Although the three cartoon stories used as stimuli had quite different storylines, the 14 original cartoon narrations relating to the same respective cartoon story were all found to contain the same basic idea units (Butterworth 1975) constituting the storyline (i.e., they contained the same basic semantic information, resulting in highly similar storylines). The scripted stories used as stimuli here incorporated these core idea units and were thus representative of the original cartoon narratives. We identified those idea units of each cartoon which the different original narrators had frequently accompanied with gestures (varying between 1 and 4 narrators, with a mean of 2.9 narrators gesturing for each of the respective idea units). Of those original narrators who had gestured with the idea units, we picked those gestures which previous analyses (Beattie and Shovelton 1999b, 2001) had shown to be particularly communicative regarding position and size information (however, *all* gestures produced with the same idea unit by different speakers represented the same core concept). We included these gestures in the scripted stimulus stories (three gestures per story) at the appropriate points and accompanied by the original speech. The scripted cartoon stories thus contained both speech and gesture, and they were composites deriving from a number of individual speakers (up to four per story). However, we were careful to check for both coherence and style and amended the stories accordingly, such as replacing 'he' with 'Billy' when the new preceding context would not allow for the use of a pronoun (this resulted in only minimal amendments). Since the original narrators had all included the same basic idea units in their narratives merging the individual parts of these narratives into a coherent whole was easily possible. Some meta-narrative comments (such as 'now this is really funny actually') were excluded from the scripted versions of the cartoon narratives, because such comments were considered as being more dependent on the particular social dynamics of the original interaction (for example, a narrator announcing that something funny is coming up in response to an interlocutor who seemed to be particularly amused by the story told). Apart from this, the speech constituting the scripted narratives consisted of the original words produced by the respective original speakers and was accompanied by the gestures which the speakers originally produced (for examples of the kind of gestures included, an example of the verbal scripts used and the selected contributions from individual speakers, please see Appendix 1).

This means that with regard to each idea unit, in terms of semantic content, the participants in our study received the same information in both gesture and speech as those participants who originally listened to the cartoon narratives in the Beattie and Shovelton (1999b, 2001) studies. Because these cartoon narratives were *spontaneously* produced by the original narrators, the semantic information our analysis focuses on (i.e., position and size information) was sometimes explicitly represented in gesture only (nine instances) and sometimes explicitly in gesture and speech (three instances); in addition, recipients may of course make inferences based on information they glean from the context of the story and their knowledge of the world. Thus, how the semantic information was packaged in our scripted stimulus cartoon narratives was similar to natural communication where the distribution of semantic information across the gesture and the speech modalities varies from one gesture–speech utterance to the next (cf. Holler and Beattie 2002, 2003).

A professional actor was asked to deliver the scripted cartoon narratives. We chose this option rather than using spontaneous narratives produced by different speakers to be able to compare across trials, and, importantly, to assure that the gestures of interest would be communicative in a video context in terms of position and size information (which represents the basis for our comparison). If a number of spontaneous narrators would not have produced communicative gestures, this would have reduced our data points considerably, as for these gestures we would not have been able to compare the claimed significant communicative contribution of iconic gestures in video-contexts with that of iconic gestures in a face-to-face context.

The actor rehearsed the cartoon stories and practiced performing the original gestures in a natural and consistent manner based on the original video-recordings of the respective gestures. The actor was given no direction about his gaze behavior. Two of the experimenters (HS and JH) checked the actor's gestural performance for consistency and congruency with respect to the original gestures (in terms of the representation of the relevant size and relative position information represented in the gestures). A third independent coder checked 14 of the 37 recordings for the representation of size and relative position information (i.e., 37.8% of the stimulus data); her scores showed 95.8% agreement with the experimenters' coding.

In total, there were four sets of 37 participants each, all of whom watched the narrator telling the three scripted (and carefully rehearsed) cartoon stories. In the gesture + speech (face-to-face) condition, each participant watched and listened to the actor narrating live the three stories. The actor's performances (i.e., the 37 individual trials of the face-to-face condition) were video-recorded at the same time (showing only the actor) and then played three more times each (to three different recipients), once in each condition (the gesture + speech (video) condition, the speech only (video) condition (where the stimuli were played with the picture turned off), and the gesture only (video) condition (where the clips were played without any sound)). Thus, each of the 37 narrations (which each consisted of the actor telling the three different stories in a row) was seen/heard by four corresponding interlocutors, but under different presentation conditions.

Questionnaires were used which contained nine questions relating to each story, consisting of a mixture of target and filler questions (the target questions always tapped information relating to the two semantic features, *size* or *relative position*, such as 'what is the size of the table?'; 'where is the table in relation to the ceiling?'). The questionnaire for story 1 contained six filler questions and three target questions, the one for story 2 contained four filler questions and five target questions, and the questionnaire for story 3 contained five filler questions and four target questions. Out of the 27 questions participants had to answer overall across the three stories, six related to size and six to relative position. The order of the questions was randomized. All questions were open ended and participants were encouraged to write down as much information as they thought they had relating to the respective questions. They were asked to not only restrict their answers to information which was provided to them in an explicit manner. The reason for this was that gesturing is often considered as a more unconscious or implicit form of representing semantic information (e.g., Broaders et al. 2007). And, of course, understanding speech requires inferential processes which form part of the mental representations we build up while listening. The participants' open ended answers were then scored for size and position information (see 'Analysis' section).

Procedure

Participants took part in the experiment in individual trials. In the gesture + speech (face-to-face) condition, participants were told that they and another participant (the actor) would be taking part in the communication experiment together, which tested how well people understand cartoon stories, told by another person, which they do not know anything about. The actor played the role of a confederate who pretended to be, initially, as unknowing about the cartoons as the addressee–participant. To create this impression, the actor pretended to meet the experimenter for the first time when arriving for the experiment, and he was given time to examine each cartoon story for a few minutes before starting to narrate the story. The order in which the actor narrated the three stories was randomized across trials. The participants were informed that after each story they would be given a questionnaire and that they had to answer the questions, in writing, by providing as much detail as they could, even if the amount of detail seemed trivial to them (they were allowed to take as much time as they needed to complete each questionnaire). Participants were told that they were not allowed to ask questions. This sort of passive listening was considered comparable to the conversational situation in which one person listens to another taking an extended turn, such as when telling a more complex story of some kind. Of course, much of everyday talk is of a more interactive nature, but opting for this degree of experimental control and thus limiting it to this particular kind of communicational situation was here necessary to allow us to compare the amount of information communicated in the different presentation conditions.

In the video conditions, participants were given the same basic instructions but now the actor's performances captured on video were played to the participants on a small-screen (28″) television. Again, the participants were not given the opportunity to ask any questions of clarification about what they had witnessed.

Analysis

Answers about size and position were scored as correct when it was explicitly stated by participants that, for example, a certain ball was 'big' (size of ball), or 'two baskets were placed together' (relative position of the two baskets). However, statements which contained nothing but a comparative judgement, such as 'the table is of the size that a family can fit around it' were deemed as not precise enough to judge the listener's size impression—after all, family-sized tables can vary quite considerably in terms of their size. Also, different participants used the same comparative (e.g., 'the size of a tennis ball'), but in some cases preceded by the words ‘it is *small*—the size of a tennis ball’ and in other cases (but relating to the same question) by the words ‘it’s *big*—the size of a tennis ball). Thus, only when accurate information was explicitly given was it scored as correct. Regarding information about relative position, some answers were also scored as not precise enough. For example, in response to the question ‘where is the table in relation to the ceiling?’, the answer ‘under the ceiling’ does not unequivocally show that information about the table’s exact position has been communicated. The answers ‘close to the ceiling’ or ‘just below the ceiling’, however, were considered as precise enough.

Twenty-five percent of the questionnaires (i.e., 40 in total, 10 questionnaires randomly chosen from each condition) were additionally scored by a second judge (blind to the experimental conditions). The judges’ agreement was measured using Cohen’s Kappa, which resulted in the following values, per condition: ‘gesture only (video)’ $K = .95$ (size) and .84 (position); ‘speech only (video)’ $K = .90$ (size) and .82 (position); ‘gesture + speech (video)’ $K = .87$ (size) and .86 (position); ‘gesture + speech (face-to-face)’ $K = .81$ (size) and .79 (position). The few discrepancies that did emerge were resolved through discussion and the questionnaires were re-scored accordingly.

For each answer that was scored as including relevant information about relative position or size one data point was given. For each participant, these data points were summed across cartoons because each cartoon related to only a small number of target questions (namely between three and five, as seen above), and because we were not primarily interested in the effect of different stories on the communicativeness of iconic gestures. We could have used one long story instead of the three shorter ones, but the aim was for the stimuli to be comparable to those used in previous studies. In addition, longer stories might have resulted in participants not being able to remember everything. However, to assure that the patterns revealed in our Results section holds across the three stories (and are thus stimulus independent) an overview of the data for each cartoon separately has been included in the Appendix (see Appendix 2). The graph shows a very similar pattern for all three cartoon stories.

Results

All statistical tests reported in this section are two-tailed, in line with our experimental hypotheses, and an alpha level of .05 is used throughout. First, we analyzed how participants in the four conditions differed in terms of the information they received about the cartoons, considering both semantic categories together. Figure 1 provides a summary of the data, including the means and standard error for each condition. A one-way ANOVA was carried out which revealed that, overall, participants in the four experimental conditions received significantly different amounts of information, F (3, 143) = 15.69, $p < .0001$.

Tukey post-hoc comparisons showed that the following conditions differed significantly: speakers received more information in the gesture + speech (face-to-face) condition than in the gesture + speech (video) condition. Furthermore, in line with previous findings, speakers also received more information in the gesture + speech (video) condition than in the speech only (video) condition. Further differences existed between the gesture + speech (face-to-face) and speech only (video) conditions, the gesture + speech (face-to-face) condition and the gesture only (video) condition, as well as the gesture + speech (video) and gesture only (video) conditions. Results were not

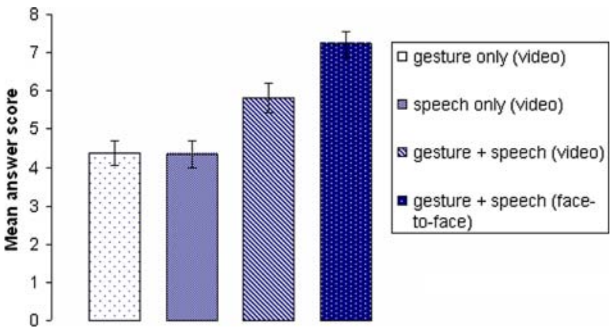


Fig. 1 Mean number of questions answered correctly in the four experimental conditions, collapsing the position and size data (error bars represent standard error of means)

significant for the comparison of the gesture only (video) and the speech only (video) conditions.

Secondly, we considered information received about relative position separately from size information. This analysis showed that for information about relative position (see Fig. 2) the four conditions again differed significantly, F (3, 143) = 19.40, $p < .0001$.

Tukey post-hoc tests were conducted which revealed that, considering the position scores alone, participants in the gesture + speech (face-to-face) condition still received more information than those in the gesture + speech (video) condition, albeit not to a statistically significant extent. However, we were again able to replicate past findings in that participants received more information in the gesture + speech (video) condition than in the speech only (video) condition. The difference between the gesture + speech (face-to-face) and the speech only (video) condition was also significant. Further significant differences existed between the gesture + speech (face-to-face) condition and the gesture only (video) condition, as well as the gesture + speech (video) and gesture only (video) conditions. Again, no significant differences existed between the gesture only (video) and the speech only (video) conditions.

The same analysis was conducted for information about size (see Fig. 3), which again revealed an overall significant difference between the four conditions, F (3, 143) = 4.37, $p < .006$.

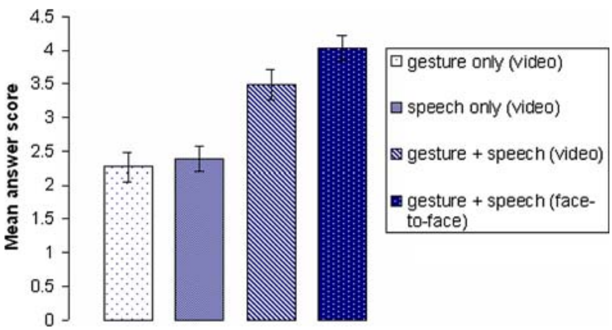


Fig. 2 Mean number of questions answered correctly, for the semantic category 'relative position' only (error bars represent standard error of means)

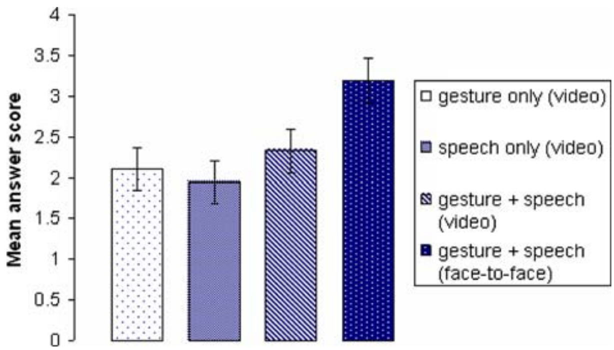


Fig. 3 Mean number of questions answered correctly, for the semantic category 'size' only (error bars represent standard error of means)

Tukey post-hoc comparisons revealed the following results: Again, participants in the gesture + speech (face-to-face) condition received more information than participants in the gesture + speech (video) condition, but this difference failed to reach statistical significance. Further, although participants in the gesture + speech (video) condition received more information than participants in the speech only (video) condition, this difference did not reach statistical significance. However, comparing the gesture + speech (face-to-face) and the speech only (video) conditions did yield a significant difference, as did the comparison of the gesture + speech (face-to-face) condition and the gesture only (video) condition. The gesture + speech (video) and the gesture only (video) conditions did not differ significantly, and neither did the gesture only (video) and speech only (video) conditions.

Discussion

The aim of the present investigation was to test the communicative effectiveness of iconic hand gestures in various contexts. The results were quite clear. Iconic gestures which communicate size and position information successfully when seen on a small-screen television do also communicate effectively in a face-to-face context. The results on 'position' and 'size' together found that iconic gestures in fact communicated significantly *more* information in a face-to-face condition than in the video mediated condition. Although the differences between the gesture + speech (face-to-face) condition and the gesture + speech (video) condition were not statistically significant when considering the scores for 'size' information and 'relative position' information *separately*, the scores were still higher in the face-to-face condition than in the video condition in the cases of both semantic features. This means that, despite the changes in methodology to that employed by Beattie and Shovelton (e.g., 1999a, b, 2001, 2002), and potential differences in recipients' gaze behavior, which we assumed might make it harder for gestures to emerge as effective in the communication of semantic information, the gestures did not diminish in their communicative power; quite the contrary in fact. We can thus conclude that such iconic gestures really do communicate information about size and relative position, even under more natural face-to-face circumstances.

What we currently cannot draw any firm conclusions about is why the iconic gestures did in fact communicate more semantic information alongside speech in the face-to-face context than in the video presentation context. Because this study did not set out to test exactly which factors may influence the gestural communication of semantic information and to what extent, this question will need to be answered in future studies. One possible explanation is that the gestures were more communicative when presented face-to-face because the gestural representations were of a larger scale (compared to gestures shown on a 28″ television screen image), and they may have appeared clearer due to potential differences in the quality of the television image compared to the live presentation. Another possible explanation is that recipients presented with gestures in the face-to-face context felt more socially engaged in the communication which may make them more perceptive of social cues potentially directing recipients' covert attention (such as speakers fixating their own gestures); this could result in increased information up-take compared to recipients who watched the speaker in a socially isolated context (i.e., on a video screen).

Furthermore, this study aimed to determine whether participants who are presented with gesture and speech together (on video) receive more information than those who are presented with only the speech. In this study, we found this to be the case both when we considered the data by collapsing the position and size scores, as well as when considering only the scores for relative position. With regard to size, however, the basic information was effectively communicated as well by speech alone as by speech and gesture (but the actual scores were still higher in the gesture + speech (video) condition than in the speech only condition). This is not entirely surprising as size information has been shown to be somewhat less reliably communicated by gestures than relative position in some previous studies (e.g., Beattie and Shovelton 2002 Study 2). Size information tends to be more easily represented in speech than relative position, and it is perhaps also more easily inferable from speech (including potential intonation cues) than information about relative position. This may explain why the benefit of perceiving the gestural information in addition to the speech was less pronounced for size information.

Finally, the data also replicate previous findings by showing that gestures do communicate a considerable amount of semantic information even in the absence of speech. More precisely, information about these two semantic features was as well communicated through gesture alone as through speech alone in all three comparisons (that is position and size, position only and size only). This is in line with previous research showing that although iconic gestures are designed to accompany speech, they convey semantic information to recipients even when speech is absent or when it is very difficult to hear (e.g., Beattie 2003; Beattie and Shovelton 1999a, b; Riseborough 1981; Rogers 1978). In these previous studies, the amount of information communicated by gestures on their own was, overall, considerably less than the information received from speech. The fact that in the present study we obtained values showing that information was equally well communicated by the two modalities is of course due to us using gesture stimuli which we knew were highly communicative when presented on video, and because we only considered those two semantic categories which iconic gestures have been shown to communicate reliably in studies using video presentation. We therefore have to view our results in the light of this, and it would not be appropriate to conclude that iconic gestures *generally* communicate as much information as speech does.

One potential limitation of the present study is that it does not allow us to partial out the effects of facial cues. It could be argued that recipients may have gleaned information from the speaker's gaze, at least in some cases. However, we have checked for this possibility and although no specific instructions were given to the speaker, his gaze was either directed at the recipient or was averted as when speakers engage in the cognitive planning of utterances (see Beattie 1983). Thus, in instances where the gestures represented information about the relative position of an object, for example about a table being raised towards the ceiling (represented by the hands rising upwards), the speaker continued to look at the recipient (or averted his gaze), but he did *not* look at his hands or the ceiling/upward. Also, in this study we did not manipulate the availability of information from the speaker's lips, and although it is unlikely that they did convey any relevant *semantic* information (for a more detailed discussion, see Rogers 1978), we cannot determine yet with absolute certainty whether recipients gleaned information from these potential cues. However, future investigations may provide further insight into whether any potential interaction effects exist between the perception of lip and eye cues and the semantic communication of iconic gestures.

In summary, the present study lends strong support to previous experimental research which found that iconic gestures can effectively communicate semantic information alongside speech. Moreover, it transcends this earlier work by demonstrating that the conclusions about the power of these gestures are still valid, even when interlocutors communicate face-to-face. However, we cannot necessarily generalize our findings to iconic gestures that appear not very communicative in video-contexts, or to the communication of semantic features which iconic gestures seem to communicate in certain instances but much less reliably so (cf. Beattie and Shovelton 1999a, b, 2001). These aspects will need to be addressed by future research.

In 2003, Goldin-Meadow posed the question: 'will someone who gleans a great deal of information from gesture when watching and listening to a video clip also glean a great deal of information from gesture when participating in a conversation, listening to a story, and so on? Very little work has been done on this question—it's a wide-open issue' (p. 82). The present study has provided a first tentative glimpse of what the answer to this question may be.

Acknowledgments We would like to thank two anonymous reviewers, as well as Prof. Judith Hall, for their invaluable comments on an earlier version of this manuscript. We would also like to thank Lucy Treanor for her assistance in collecting the data, William Newell for his invaluable acting performances, as well as the participants who took part in this investigation. The research was funded through a grant from the Faculty of Medical and Human Sciences at the University of Manchester awarded to Judith Holler. During the duration of this study, Geoffrey Beattie and Heather Shovelton were supported by a research grant from the Economic and Social Research Council (RES-000-22-1917).

Appendix 1

Example cartoon script (for 'Billy Whizz') including speech and gestures (the gesture description is included in square brackets and transcribed in italics underneath and the part of speech that it accompanies. Images of the gestures are also included; the speaker's face has been blurred in these images to provide anonymity). Insertions mark the parts of the narrative originally contributed by different speakers (here, speakers 1–4):

[SPEAKER 1:] It's a family, with a woman, a husband and a little child. And the father's saying 'We'd like to buy a kitchen table and chair set.' And the salesman in the furniture shop says 'Certainly.' And so then the salesman says 'How about a bench set instead of a table and chair set' and the father says that that's not very safe, because he imagines them having breakfast and Billy is throwing the food all over the chairs. Then he says 'It's no use for our Billy'. So the salesman says 'How about a table with single swivel chairs?' And the Mother says 'Oh no, we've had chairs like those before.' And she remembers when Billy jumped onto the table instead of onto the chairs and the table went [spinning and all the food went all over the floor].

Right arm moves in two large clockwise circles, while the left hand moves away from the right arm

And she says ‘Yikes, he’s like a mad spinning top.’ So the salesman comes up with a solution, ‘I have the very table to suit you. It will be very safe.’ So the Mum says ‘Great.’ [SPEAKER 2:] And the guy has come up with a particular kitchen table set which is, it’s on a [pole which] moves up and down.

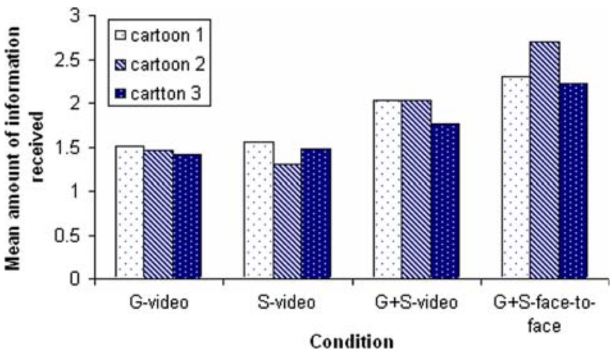
Hands apart, palms pointing towards each other, fingers curved, hands make two large rapid movements up and down

[SPEAKER 3:] So when Billy comes in he says ‘Where are the family?’ as the table can be [raised up towards the ceiling], so he can’t jump onto the table.

Hands are shoulder width apart, palms facing down, hands move upwards

And the Father says ‘Ha ha, the table lifts us out of harms way.’ [SPEAKER 4:] But then Billy’s laughing his head off ‘cos the table then moves to the very top of the ceiling and squashes the entire family.

Appendix 2



References

Argyle, M. (1975). *Bodily communication*. London: Routledge.

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Cambridge: Cambridge University Press.

Argyle, M., & Ingham, R. (1972). Gaze, mutual gaze, and proximity. *Semiotica, 1*, 32–49.

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

Beattie, G. (2003). *Visible thought: The new psychology of body language*. London: Routledge.

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

Beattie, G., & Shovelton, H. (1999b). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of Language and Social Psychology, 18*, 438–462.

Beattie, G., & Shovelton, H. (2001). An experimental investigation of the role of different types of iconic gesture in communication: A semantic feature approach. *Gesture, 1*, 129–149.

Beattie, G., & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology, 93*, 179–192.

Beattie, G., & Shovelton, H. (2005). Why the spontaneous images created by the hands during talk can help make TV advertisements more effective. *British Journal of Psychology, 96*, 21–37.

Beattie, G., & Shovelton, H. (in press). An exploration of the other side of semantic communication. How the spontaneous movements of the human hand add crucial meaning to narrative. *Semiotica*.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General, 136*, 539–550.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research, 4*, 75–87.

Darwin, C. (1872). *The expression of the emotions in man and animals*. London: John Murray.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist, 48*, 384–392.

Ekman, P. (1999). Facial expressions. In T. Dalgleish & T. Power (Eds.), *The handbook of cognition and emotion* (pp. 301–320). Sussex, UK: Wiley.

Ekman, P., Friesen, W. V., & Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39*, 1125–1134.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. London: Belknap Press of Harvard University Press.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *Journal of Human Movement Studies, 1*, 33–39.

Gullberg, M. (2003). Eye movements and gestures in human interaction. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eyes: Cognitive and applied aspects of eye movements* (pp. 685–703). Oxford: Elsevier.

Gullberg, M., & Holmqvist, K. (1998). *Focus on gesture: Visual attention towards gestures in conversation*. LUCS, Lund: Department of Cognitive Science.

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition, 14*, 53–82.

Holler, J. (2003). *Semantic and pragmatic aspects of representational gestures: Towards a more unified model of communication*. Unpublished doctoral thesis, University of Manchester, UK.

Holler, J., & Beattie, G. (2002). A micro-analytic investigation of how iconic gesture and speech represent core semantic features in talk. *Semiotica, 142*, 31–69.

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture, 3*, 127–154.

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

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. Ritchie Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). The Hague: Mouton.

Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). Cambridge: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

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

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

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

Roebers, C. M., Gelhaar, T., & Schneider, W. (2004). 'It's magic!' The effects of presentation modality on children's event memory, suggestibility, and confidence judgments. *Journal of Experimental Child Psychology, 87*, 320–335.

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

Thierry, K. L., & Spence, M. J. (2004). A real-life event enhances the accuracy of preschoolers' recall. *Applied Cognitive Psychology, 18*, 297–309.