

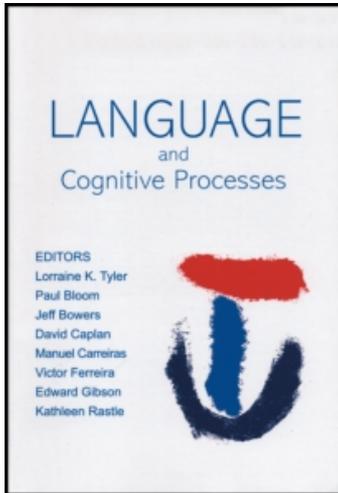
This article was downloaded by: [The University of Manchester]

On: 10 February 2009

Access details: Access Details: [subscription number 788787424]

Publisher Psychology Press

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Language and Cognitive Processes

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title-content=t713683153>

### Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task

Judith Holler<sup>a</sup>; Katie Wilkin<sup>a</sup>

<sup>a</sup> School of Psychological Sciences, University of Manchester, Manchester, UK

First Published: February 2009

**To cite this Article** Holler, Judith and Wilkin, Katie (2009) 'Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task', *Language and Cognitive Processes*, 24:2, 267 — 289

**To link to this Article:** DOI: 10.1080/01690960802095545

**URL:** <http://dx.doi.org/10.1080/01690960802095545>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task

Judith Holler and Katie Wilkin

*School of Psychological Sciences, University of Manchester, Manchester, UK*

Much research has been carried out into the effects of mutually shared knowledge (or common ground) on verbal language use. This present study investigates how common ground affects human communication when regarding language as consisting of both speech and gesture. A semantic feature approach was used to capture the range of information represented in speech and gesture. Overall, utterances were found to contain less semantic information when interlocutors had mutually shared knowledge, even when the information represented in both modalities, speech *and* gesture, was considered. However, when considering the gestures on their own, it was found that they represented only marginally less information. The findings also show that speakers gesture at a higher rate when common ground exists. It appears therefore that gestures play an important communicational function, even when speakers convey information which is already known to their addressee.

### INTRODUCTION

A question that has sparked much interest and debate among gesture researchers over many years is why we gesture when we speak. Various theories explaining their functions and usage have been proposed. For example, some have argued that gestures facilitate lexical retrieval during the speech production process and thus fulfil mainly speaker-beneficial functions

---

Correspondence should be addressed to Judith Holler, School of Psychological Sciences, Coupland Building 1, University of Manchester, Manchester M13 9PL, UK. E-mail: judith.holler@manchester.ac.uk

We would like to thank Sotaro Kita as well as two anonymous reviewers for their invaluable comments on an earlier version of this manuscript. We would also like to thank the Economic and Social Research Council, UK, for a research grant awarded to the first author (RES-061-23-0135) which allowed us to carry out this research, as well as the participants who took part in our study.

---

© 2008 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

<http://www.psypress.com/lcp>

DOI: 10.1080/01690960802095545

(e.g., Butterworth & Hadar, 1989; Krauss, Chen, & Gottesman, 2000; Krauss, Morrel-Samuels, & Colasante, 1991; Rimé & Shiaratura, 1991). Others have argued that gestures are intended for the addressee and thus are primarily socially and interactionally motivated (e.g., Bavelas & Chovil, 2000; Kendon, 1983, 1985, 2004) (see also De Ruiter, 2000; Freedman, 1977; Kita & Özyürek, 2003; McNeill, 1992, 2005, for further examples of gesture production models). However, these theoretical positions are not necessarily exclusive of each other. A number of researchers, including many of the above, acknowledge that gestures may indeed be multifunctional, fulfilling various speaker-beneficial as well as addressee-related functions. It is therefore ever more important to investigate under which circumstances gestures fulfil which kinds of functions, and how different processes integral to human communication influence gesture use.

Several earlier investigations have shown how the social context in which people communicate can influence gestural communication. For example, researchers have manipulated visibility between interactants and found that, in face-to-face communication, speakers use a higher rate of representational gestures (Alibali, Heath, & Meyers, 2001), produce spatially more differentiated and defined anaphoric gestures (Gullberg, 2006), and they use more gestures which contribute information over and above the information contained in speech (Bavelas, Kenwood, Johnson, & Phillips, 2002). Other studies have shown that speakers orient some gestures specifically towards their addressee (Furuyama, 2000) and that they represent certain types of spatial information differently depending upon the location of their addressee(s) (Özyürek, 2000, 2002). These studies provide important evidence that the social context in which face-to-face communication is embedded can influence speakers' gestural communication, and indeed how this influence is manifest. Further insights can be gained from studying the influence of variables which tap more 'covert' aspects of communication than, for example, visibility and addressee location. One such aspect is the speaker's consideration of the addressee's knowledge, thinking and understanding.

The present study focuses on knowledge when it is, or is not, mutually shared between the speaker and the addressee (and relevant with respect to the topic of conversation). A considerable number of experimental studies have manipulated common ground, i.e., the knowledge, assumptions, and beliefs that interactants mutually share (Clark, 1996; Clark & Schäfer, 1989) to investigate its influence on communication. For example, Isaacs and Clark (1987) asked participants to carry out a referential communication task in which postcards showing New York landmarks had to be arranged in a certain order by both experts (i.e., participants with knowledge about New York City) and novices (i.e., participants without any knowledge about New York City). The amount of knowledge participants shared about the

New York landmarks significantly affected their communication. For example, when experts communicated with other experts they used fewer words to refer to the individual landmarks than when novices communicated with each other, and more of the references consisted of proper names (e.g., *Rockefeller Centre*) rather than descriptions of the landmarks. Fussell and Krauss (1989) conducted a referential communication task in which participants created references to abstract line drawings either for themselves (non-social condition, involving maximal common ground) or for other participants (social condition, involving relatively little common ground). They found, amongst others, that references participants had created for themselves consisted of considerably fewer words, showed more diversity in terms of the vocabulary used and contained more figurative language.

Apart from the knowledge that interlocutors share from the outset of a conversation, common ground also accumulates over the course of it. Experimentally, equivalent conditions have been created by using successive trials in which participants refer to the same stimuli; here, participants have to ascertain that a certain form of reference or description unequivocally identifies a particular referent and thus forms part of their mutual knowledge (e.g., Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Isaacs & Clark, 1987). With increasing common ground, speakers are able to use fewer words and fewer turns (e.g., Isaacs & Clark, 1987; Fussell & Krauss, 1992) and to provide less informational content (e.g., Fussell & Krauss, 1992). In short, interactants draw upon common ground for conversational coordination, thus allowing them to communicate effectively and efficiently by constructing syntactically more simplified, shortened, and informationally less complex utterances (see also Garfinkel, 1967, and Vygotsky, 1986/1934, for examples of the same kind of phenomenon manifesting itself in everyday discourse).

This previous research has, however, focused almost exclusively on verbal communication, and has not taken into consideration the gestural side of utterances. Speech-accompanying hand gestures are conceived of as being an integral part of language, emerging from the same underlying mental representation, and are governed – at least in part – by the same cognitive processes (e.g., Kendon, 1980, 2000; McNeill, 1985, 1992). Thus, the question arises as to how mutually shared knowledge affects communication when we consider both speech *and* gesture.

Some studies have already provided first insights into this issue. For instance, Gerwing and Bavelas (2004) asked participants to describe the play actions they performed with a certain set of toys to either someone who played with the same or with different kinds of toys. Their findings revealed that speakers use less precise, less complex, or less informative gestures when talking to participants with whom they mutually share knowledge, as compared with when communicating with participants with whom they do not share this knowledge. Similarly, Holler and Beattie (2007) found that

gestures appear less elaborate when common ground exists than when it does not. They also compared gesture rate but found no significant effect of common ground on the number of iconic gestures in relation to the number of words produced. Furthermore, Holler and Stevens (2007) used a referential communication task which involved participants locating a target entity in an array of other entities of different sizes. Their analysis focused on the representation of the size of particularly large entities in the array. They found that common ground influences the semantic interplay of gesture and speech in the representation of such size information. Entities were represented gesturally as significantly smaller when their actual size was already known to the addressee, as compared with when addressees had no pre-existing shared knowledge regarding the respective entities' size. Furthermore, Parrill (in press) also found an effect of common ground on the encoding of semantic information in gesture. Her analysis focused on a cartoon event in which a cat 'melted down some stairs' and the number of times speakers included information about this 'ground' component in their description of the event. Amongst other things, the results showed a main effect of common ground in that speakers encoded less ground information in gesture when talking to participants who had watched the cartoon event with them than when they had not.

Overall, these studies provide a first glimpse of how common ground can affect the qualitative nature of gestures, the encoding of semantic information in the gestural modality, and the semantic interaction of gesture and speech. But in everyday talk we, of course, communicate more than just size and ground information. As pointed out by Gerwing and Allison (2008) as well as Holler and Stevens (2007), the effect of common ground on the representation of a wider range of semantic features in speech and gesture is still needed. The present study examines the information contained in speech and gesture relating to individual semantic categories (cf. Beattie & Shovelton, 1999a, 1999b, 2001; Bergmann, 2006; Bergmann & Kopp, 2006; Gerwing & Allison (2008); Holler & Beattie, 2002, 2003, 2004). The 10 semantic features included in this analysis were designed to capture the range of information contained in the respective parts of the stimulus material (see Method section), including information about entities and their identity, movement, any specific actions, path/direction information, the manner of action, the appearance of entities, position, and the size and shape of entities. The semantic analysis is applied to specific parts of speakers' narratives based on a story involving everyday events. The main aim of this analysis is to apply a broader view of language which comprises both speech and gesture when testing the effect of common ground on the communication of semantic messages. As seen above, common ground leads to reduced verbal messages (both in terms of word length and semantic content), and to gestures that are less complex and which encode less semantic information, at

least with respect to size and ground information. What we do not yet know is how common ground affects the communication of meaning when we consider utterances as consisting of both speech and gesture (subsequently referred to as speech-gesture utterances). Our analysis will tackle this question. Additionally, we will consider the representation of semantic information in the two modalities, gesture and speech, separately. In line with previous findings, we predict that speakers will verbally communicate less semantic content when common ground exists as compared to when it does not. However, we make no directional predictions about how speech is affected regarding individual semantic features. Concerning gesture, we make no directional predictions regarding the effect of common ground on the amount of semantic information represented overall, due to a lack of previous research in this respect. Regarding the effect of common ground on the gestural representation of individual semantic features, we make one directional prediction, namely for the semantic feature 'size'. Because previous research has shown that speakers tend to represent less size information in gesture when common ground exists (Holler & Stevens, 2007), we predict to also find this difference in the present study. Preceding the semantic analyses, we will also compare the number of words and gestures used when knowledge is mutually shared and when it is not. Based on previous findings, we expect speakers to use significantly fewer words when common ground exists. But we make no specific predictions regarding the number of gestures or gesture rate as previous findings have yielded no conclusive evidence in this respect.

## METHOD

### Experimental design

The present study used a between-subjects design with two conditions: the 'common ground' (CG) condition, in which some shared knowledge about the stimulus material was experimentally induced, and the 'no common ground' condition (N-CG), in which participants did not share any experimentally induced common ground.

### Participants

Eighty students (forty female and forty male) from the University of Manchester took part in this experiment, and all received either payment or experimental credits for their participation. All individuals were right handed and native English speakers. Each participant was allocated to a same-sex pairing, which was then randomly assigned to one of the two experimental conditions (allowing 10 female-female, and 10 male-male pairs

in each condition). None of the participants taking part as a pair knew each other before the experiment.

## Materials

The stimulus material consisted of video footage taken from a German children's television series, 'Neues aus Uhlenbusch' (permission was sought and granted from the copyright holders, Zweites Deutsches Fernsehen (ZDF)). Two episodes were edited into two briefer stories (each of approximately six minutes) (but note that only one story formed the basis of the present analysis, see below). The two stories depicted children and adults interacting in a rural setting, e.g., carrying out typical everyday activities such as grocery shopping or playing in a field, and both followed a coherent storyline (a clear narrative structure with a beginning, middle, and an end). Six individual scenes from each story were then selected. The scenes were between two and five seconds in length, which was judged as ideal such that they did not include too much detail that had to be remembered but at the same time contained enough information for participants to talk about. In order to test whether participants were able to remember which parts of the entire video they had seen as individual scenes, a pre-test was carried out with 10 additional participants, who were first shown the respective six individual scenes, and then the entire video (with the same procedure then being repeated for the second video). The test showed that they were able to identify the individual scenes from the entire video with great accuracy (112 out of the 120 scenes were correctly identified, i.e., 93.3%). The stories and the scenes were played to all participants without sound (the content was still easily understandable; all participants produced coherent narratives of accurate contents). A Panasonic video player and a 15" television set were used to do so. The experiment was filmed split-screen using two wall-mounted cameras providing frontal views of the participants.

## Experimental procedure

Participants were asked to sit on chairs facing each other at a comfortable conversational distance. They were informed that the experiment was looking into how people communicate information and how well that information is understood (however, nonverbal, or indeed gestural communication was not mentioned). Within each pair participants were randomly assigned to either the role of 'speaker' or 'addressee'. It was explained that, firstly, the speaker would watch a story while the addressee was out of the room, and that he/she would then be asked to describe in detail what happened in that story to the addressee. This procedure was then repeated for the second story. The addressee was asked to attend as closely as possible to each description because they would be asked questions about the

individual stories at the end. Both participants were told that the addressee's performance on this test would be taken as a reflection of their *joint achievement* in the task (the reason for this was to motivate participants to do well in the task and to consider it as a collaborative activity).

Addressees were told that they should not ask any questions or interrupt the speaker, but that they could provide feedback in a more subtle manner, for example, by nodding their head, or providing 'backchannel responses' (such as 'uh-hu', 'yeah', 'right', 'ok') to signal their understanding during the descriptions. The reason for this was that dialogic, verbal interaction has been shown to influence gesture frequency (Beattie & Aboudan, 1994). We thus aimed to reduce the possibility that differences in the degree of verbal interaction, rather than our manipulation of common ground, may be responsible for any effects found on gesture (both in terms of frequency/gesture rate as well as semantic content).

Participants' common ground was manipulated using the six scenes from each story. These were played before each corresponding video in a random, non-chronological order. In the CG condition, before the addressee was asked to leave the room, both participants watched the six scenes from that story together, and they were told that the scenes were extracts from a longer story. In the N-CG condition, the speaker also watched the six scenes from that story in the same random order, but this time in the absence of the addressee (the rationale behind this was to ensure that the salience of the six scenes with respect to the remaining content of the story would be equivalent in both conditions). In both conditions, once the speaker had watched the full video, the addressee was asked to re-enter the room and the speaker was asked to tell the respective story as a whole. In the CG condition, it was also emphasised that the participants should bear in mind that they already had some shared knowledge regarding parts of the story. At the end of the experiment, participants were fully debriefed, they were given the option to view and have their data removed from the tape (all were happy for their data to be used).

## Analysis

For the present analysis, only one story was chosen (at random) owing to the time-consuming nature of the analysis. Participants' gestural and verbal behaviour relating to the selected scenes was included in the analysis. During the analysis it was noticed that some speakers may have confused or 'merged' one of the six events with a different event occurring in the video. The respective scene shows three boys jumping down from a ledge into some hay. However, the video also contains a brief scene of children jumping up and down on some hay bales. Utterances such as 'and the kids were jumping around in the hay' could therefore not be unequivocally categorised as a

vaguer, more elliptical, description of the target jump event. To avoid this potential confound in the data the respective scene was excluded from the analysis.

### Speech segmentation

All descriptions relating to the target scenes were transcribed in detail. To identify the respective parts in the narratives, each event was defined in terms of what it comprised semantically. The segments of speech included in the analysis therefore constituted semantic, or ideational, rather than sentential units (cf. Holler & Beattie, 2002; see also Butterworth, 1975) which varied in length between a single clause and several sentences. Not included in the analysis were meta- and paranarrative statements (see McNeill, 1992), such as, 'this is the scene that we saw', or, 'that's the first scene'. The reason for excluding these statements was that the analysis focused on differences in the *semantic* content of the event descriptions only. Also, these kinds of statements appeared to occur more frequently in the common ground condition as part of speakers' direct reference to their shared knowledge. (While this may be another potentially interesting difference worthy of examination, it would go beyond the scope of the present study.) Also excluded were speakers' further interpretations of the target semantic events which involved drawing upon knowledge or information from other parts of the story. This constituted, for instance, interpretations and inferences about why a character behaved as they did in the respective event. For example, in one scene a little boy picks up a bit of litter and puts it in a bin; not included in the analysis was additional information about the fact that a particular person told him to do so, as it was not part of the experimentally manipulated shared knowledge. The following examples were taken from two different speakers in our data (speech in parentheses was excluded from the analysis): One speaker referred to this scene by saying, '*the little boy picks up a piece of paper that was on the floor (I don't remember if he dropped it – I'm not sure how it got on the floor) and put it in a, a bin on a lamppost thing, like a wire bin*'. Another speaker described it by saying, '(he pointed to something, the old man, with his walking stick, and) *the little boy picked up a wrapper off the floor and put it in the bin*'.

Two independent coders identified the units of speech referring to the respective semantic events for eight speakers selected at random. The speech data from these speakers comprised 17% of the overall speech data included in the analysis. The percentage agreement was calculated by dividing the number of words that had been identified as being part of the event description by both judges by the total number of words that had been identified as being part of the scene descriptions by the two judges (i.e., agreements + disagreements). This resulted in 87.6% agreement (all discrepancies were discussed and resolved).

## Gesture coding

Based on McNeill (1992), hand gestures representing semantic information referring to concrete concepts were coded as iconic gestures, whereas gestures that referred to abstract concepts were coded as metaphoric gestures. Any pointing gestures were classed as deictic gestures (which included both concrete and abstract deictic gestures). In addition, we also coded the data for interactive gestures (see Bavelas, Chovil, Lawrie & Wade, 1992). The inter-observer reliability for the coding of gestures (performed on 25% of the data) was 79.9%, indicating high agreement. Again, all discrepancies were resolved through discussion.

The semantic analysis itself focused on iconic and deictic gestures only. Interactive gestures occurred primarily with speakers' direct references to their shared knowledge which were not included in the present analysis. Another reason was that, by definition, they do not represent any semantic information. Metaphoric gestures occurred only on three occasions.

## Semantic analysis of gesture and speech

Both the speech segments and the gestures were analysed with respect to the semantic information they represented, based on an established micro-analytic coding method using a range of individual semantic features (cf. Beattie & Shovelton, 1999a, 1999b, 2001; Bergmann, 2006; Bergmann & Kopp, 2006; Gerwing & Allison, 2008; Holler & Beattie, 2002, 2003, 2004). As in previous studies, the range of semantic features included in this analysis was considered to capture most of the semantic information contained in the respective parts of the story the participants narrated. Each verbal utterance, as well as each iconic and deictic gesture describing the target scenes, was analysed for the semantic information it contained, based on the semantic categories applied (see Example 1). Hence, our analysis focused on the amount of information represented, regardless of whether the information was complementary or redundant with regard to the information in the respective other modality. The reason for this was that we aimed here to analyse whether common ground affects the encoding of semantic information in terms of the overall amount and type; to evaluate the redundancy of the systems requires an essentially different perspective, namely one which assumes that representing a certain semantic element twice (once in gesture, once in speech) is just as communicative as representing this information once. While this may be another interesting question to answer, we decided to analyse our data here without making this assumption. Overall, 10 generic semantic categories were used, some including sub-categories which were collapsed owing to the small number of data points, as well as partial information (such as one aspect of the direction of movement, e.g., upwards,

when the trajectory was upwards *and* arched; size information when only the size of a part of an entity was represented, or when it was imprecise as in the case of quantifiers (more than one/some), etc.):

1. *Entity (E)*: representation of the involvement/existence of an entity (speech examples: the kid, he, a car, it; gesture examples: a hand movement showing someone picking up an imaginary object from the ground; the hand representing the body of a car; a deictic gesture locating an entity in the gesture space).
2. *Identity (I)*: further information about an entity, such as whether the entity is animate or inanimate, as well as general category knowledge (e.g., that the entity is a human, or a type of animal; what category of inanimate entity, e.g., car, crops; speech examples: car, boy, man, owl, wheat, cane (but not: he, she, it, they, etc.); gesture examples: representation of specific features of an entity, such as wings, or particular clothing worn by a character in the story).
3. *Movement (MO)*: representation of movement, but not necessarily its (accurate) path nor type of action; speech examples: coming past (vs. driving past), to put (vs. to throw), to leave (vs. to run away); gesture examples: representation of general movement of an entity, such as in a lax, brief hand movement towards the right as if something is travelling out of the middle of the gesture space).
4. *Action (A)*: representation of the type of action taking place (e.g. grabbing, holding, walking, driving, poking, throwing, running, seeing, sitting). N.B. This was scored in addition to *Movement* when the type of action, by definition, also involved movement (speech) or this movement was gesturally represented. Actions such as 'seeing' were therefore not scored for *Movement* but for *Action* only.
5. *Direction/path (D)*: representation of the direction/trajectory/path of a movement or an activity (e.g., up, down, straight, diagonal, a curvy line). This category also contained information about actions that were directed at other entities (D-O) (e.g., shouting *at* someone, saying something *to* the driver, shaking a stick *at* someone).
6. *Manner of action (MA)*: representation of the manner of action, i.e., the way in which an action is being carried out (such as, speech: walking *quickly*, waving the stick *angrily*, sitting *in a row*, coming out *one after the other*; gesture: representation of a speedy movement, representation of individual entities as being positioned in a line/row).
7. *Appearance (AP)*: representation of information about the appearance of entities (often adjectival, for example, looking *cute*, a *fluffy* owl), including information about their specific category type (e.g., *barn owl*, *wheat field*).

8. *Position (P)*: representation of the position of one entity relative to another, or relative to its surrounding space (speech examples: the man got hit by the car; the kids hide behind the grass; gesture examples: representing a car with one hand, a man with another as well as the distance/contact between them). This semantic category also includes location information (P-Loc) (e.g., over there, in the distance, nearby).
9. *Size (S)*: representation of the size of entities (speech examples: a *little* bin, *tall* grass, an *average-sized* house; gesture examples: showing the width of a field, size of a window, body of an owl). This category also includes quantity information (S-Q) (e.g., some, little, all; quantity information expressed through singular and plural markers of nouns and pronouns [except uncountable nouns without a quantifier, such as 'the litter', which could be a small or a big amount, and the pronoun 'it', which could refer to both a single person/thing as well as a group of individuals/undefined quantity of an entity], and quantity information expressed in gesture [e.g. representing an entity as being a big group/more than one individual]).
10. *Shape (SH)*: representation of the shape of entities (e.g., outline of a square window frame, a round bin, reference to the shape of a car).

Example 1, and the following description, illustrates the semantic coding of one speech segment and its accompanying iconic and deictic hand gestures. Codes for the individual semantic features are included as superscript, following the corresponding speech elements. The square brackets mark individual gestures, indicated as subscript preceding the respective gesture and numbered consecutively.

(1)  
 $G_1$ [the boy<sup>E, I, S-Q</sup>]  $G_2$ [picks up<sup>MO, A, D</sup> the piece<sup>S-Q</sup>] of litter<sup>E, I</sup>  
 $G_3$ [and puts<sup>MO</sup> it<sup>E</sup> in<sup>P</sup> the bin<sup>E, I, S-Q</sup>]

G1: abstract deictic gesture pointing towards the right hand side of the gesture space, referring to the boy.<sup>(E)</sup>

G2: iconic gesture showing someone grabbing an imaginary small object which is enclosed by the grip of the hand and moved upwards (palm pointing downwards).<sup>(E, E, MO, A, D, P, S)</sup>

G3: iconic gesture showing someone holding a small object enclosed in the hand which moves down and forwards, stopping at about chest height in front of the speaker's body.<sup>(E, E, MO, A, D, P, P, S)</sup>

Each gesture and each speech segment was therefore given a score from 0 (when no information relating to a particular semantic category was represented), up to any number (representing the number of times informa-

tion relating to a particular category was represented). Hence, each gesture and each speech segment could receive multiple codes, and none of the codes were mutually exclusive. With respect to Example (1), the verbal utterance was given a total score of 15 (4 for *Entity*, 3 for *Identity*, 2 for *Movement*, 1 for *Action*, 1 for *Direction*, 0 for *Manner*, 0 for *Appearance*, 1 for *Position*, 3 for *Size* and 0 for *Shape*). The three gestures accompanying the verbal utterance received a total score of 16 (5 scores for *Entity* (1 for the deictic gesture indicating the existence of the entity referred to, 1 for the iconic representation of the agent involved in picking something up, and 1 for the iconic representation of the agent involved in putting something down, 2 for the representations of the object being picked up and put down), 0 for the representation of *Identity* of the agents or objects involved, 2 for the representation of *Movement*, 2 for the representation of *Action* (grabbing/holding while moving upwards, holding while moving downwards), 2 for the representation of *Direction* (up and down, in an arch-type motion), 0 for *Manner*, 0 for *Appearance*, 3 for *Position* (2 for the representation of an entity being positioned within the agent's hand, 1 for representing this entity as being positioned at about chest height with respect to the agent; 0 for representing that the entity was picked up from the ground), 2 for *Size* (an entity so small that it, or at least part of it, can be enclosed with one hand), and 0 for *Shape* (as the hand could be holding something round, as well as square or of any other shape).

Two analysts independently coded 16% of the gesture and speech data included in the analysis to test inter-observer reliability. Their overall percentage agreement was 86.6% (79.8% for coding gesture, and 93.3% for coding speech). All discrepancies were discussed and the analysis was based on the scores that were revised on the basis of these decisions.

## RESULTS

The first part of the results section includes an analysis of the number of words and gestures (and gesture rate), as well as some examples of our experimental data. The second part includes an analysis of the semantic information contained in gesture and speech. The comparisons consider all semantic categories combined, as well as each of the semantic categories on an individual basis. For both of these steps, we analyse the overall amount of semantic information speakers provided in their speech-gesture utterances, which includes all gestural and all verbal utterances (in other words, we collapsed the data across modalities). Then we consider the amount of information represented in the two modalities, speech and gesture separately, i.e., their individual contributions to the utterance. For the statistical analyses, parametric and non-parametric tests are used, depending on

whether the data are normally distributed or not (as shown by Kolmogorov–Smirnov tests), and we adopt an alpha level of .05 throughout the manuscript.

### Number of words and gesture rate

In line with previous findings, speakers in the CG condition used significantly fewer words ( $M = 73.15$ ,  $SE = 4.30$ ) to narrate the scenes than speakers in the N-CG condition ( $M = 105.40$ ,  $SE = 8.61$ ),  $t = 3.35$ ,  $df = 27.92$ ,  $p = .001$ , one-tailed.

The number of iconic and deictic gestures was then counted for each participant and divided by the number of words (to control for differences in the length of verbal utterances). A  $t$ -test revealed that speakers in the CG condition used gestures at a higher rate, i.e., more gestures per one hundred words ( $M = 0.17$ ,  $SE = 0.01$ ), than speakers in the N-CG condition ( $M = 0.14$ ,  $SE = 0.01$ ),  $t = 2.24$ ,  $df = 38$ ,  $p = .031$ , two-tailed.

Closer examination of the speech and gesture data revealed that speakers in both conditions appeared to accompany the core components of the scene descriptions with gestures; that is, those components that were retained despite speech being more elliptical when common ground existed. The additional information provided by speakers in the N-CG condition was not always represented by individual gestures, and these participants used more words overall to narrate the events (including the core components of them). Speakers in the CG condition used only about 8% fewer gestures overall: namely, 245 gestures (212 iconic and 33 deictic gestures) as compared with 290 gestures in the N-CG condition (263 iconic and 27 deictic gestures); neither the number of iconic gestures, CG:  $M = 10.60$ ,  $SE = 0.83$ ; N-CG:  $M = 13.15$ ,  $SE = 1.41$ , nor the number of deictic gestures, CG: *median* = 1.50, *range* = 4.00; N-CG: *median* = 1.00, *range* = 3.00, differed significantly between the two groups,  $t = 1.56$ ,  $df = 30.61$ ,  $p = .13$  and  $U = 173.00$ ,  $n_1 = 20$ ,  $n_2 = 20$ ,  $p = .48$ , respectively. This explains why speakers in the CG condition actually used *more* gestures per 100 words.

For example, for one of the scenes, speakers in both conditions described a boy picking up some litter and then throwing it into a bin; however, in the N-CG condition speakers tended to add information such as what kind of litter it was, the fact that the litter was on the ground, what type of bin it was and the relative position of the bin (Examples 1–6; square brackets indicate the approximate start and end points of individual gestures;  $IC$  = iconic gesture,  $DE$  = deictic gesture,  $INT$  = interactive gesture; the latter gesture type has been marked up for completeness, despite the analyses focusing on iconic and deictic gestures only, see Method section.)

*CG condition – scene 1*

1. ‘<sub>DE</sub>[the boy] <sub>IC</sub>[picks up the piece] of litter <sub>IC</sub>[and puts it in the bin]’
2. ‘(and then you see) the little boy <sub>IC</sub>[pick up some rubbish] <sub>IC</sub>[and put it in the bin]’
3. ‘he <sub>IC</sub>[picks something up] and <sub>IC</sub>[puts it in the bin]’

*N-CG condition – scene 1*

4. ‘<sub>DE</sub>[there’s like a wrapper on the floor] and <sub>IC</sub>[the little boy] <sub>IC</sub>[picks up the wrapper] and like <sub>IC</sub>[puts it in the bin]’
5. ‘the kid <sub>IC</sub>[picks up a bit of litter off the] floor and <sub>IC</sub>[puts it in a], <sub>IC</sub>[in a, in a litter bin] <sub>IC</sub>[which is a little basket] <sub>IC</sub>[attached to a] lamppost’
6. ‘<sub>IC</sub>[the little boy] <sub>IC</sub>[picks up a piece of paper] <sub>DE</sub>[that was on the floor] <sub>IC</sub>[and put it in a] <sub>IC</sub>[bin on a lamppost] thing, <sub>IC</sub>[like a wire bin]’

Regarding another scene, speakers in both conditions described a car driving past and almost hitting a man crossing the road. For this scene, speakers in the N-CG condition tended to comment in more detail on the type of car involved, the fact that the car comes very close to the man, how it manages to avoid bumping into him, and the man’s reaction to almost being hit (Examples 7–12):

*CG condition – scene 2*

7. ‘<sub>IC</sub>[he nearly got run over by the car], like <sub>IC</sub>[jumped out of] the way of the car’
8. ‘he almost <sub>IC</sub>[gets run over by a car]’
9. ‘the man, <sub>DE</sub>[the old man] <sub>IC</sub>[crosses a road] and <sub>IC</sub>[the car comes] and nearly <sub>IC</sub>[hits him], <sub>IC</sub>[and he waves his stick at it]’

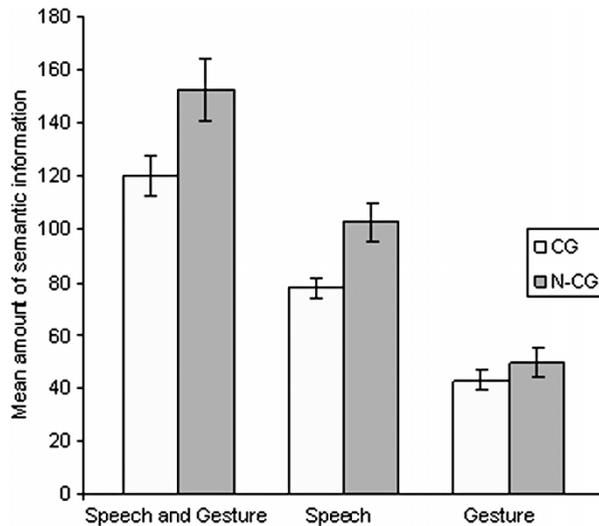
*N-CG condition – scene 2*

10. ‘as <sub>IC</sub>[he’s crossing the road] there’s an old Volkswagen Beetle <sub>ur</sub>, that <sub>IC</sub>[comes around a corner], <sub>um</sub>, looks as though <sub>IC</sub>[it breaks hard] <sub>IC</sub>[and then carries on and goes round] him. <sub>INT</sub>[It looks like the old man sort of, says something...to the driver or whatever]’
11. ‘<sub>IC</sub>[the man walks] <sub>IC</sub>[up the road], <sub>IC</sub>[where the car like drives past him] and like, <sub>IC</sub>[almost hits him and stuff], and like <sub>IC</sub>[drives straight past him] – <sub>IC</sub>[he like waves his stick in the air and like] gets a bit <sub>DE</sub>[angry with him]’
12. ‘<sub>IC</sub>[an old-school v-er, VW Beetle] <sub>IC</sub>[coming down the road], <sub>ur</sub>, <sub>IC</sub>[the Beetle sort of swerves] and misses him and the guy <sub>RE</sub>[turns round] <sub>IC</sub>[just as it’s about to come] towards him, <sub>IC</sub>[manages to get out of] the way just in time, and as it <sub>IC</sub>[drives past] he <sub>IC</sub>[waves his cane at] it like <sub>INT</sub>[disapprovingly]’

The above examples show that, although there is some overlap in how speakers in each condition narrate the events, there are also some differences, both in gesture and in speech. One question, therefore, is how the difference in the number of words used to narrate the scenes affects the semantic contents of the verbal utterances. Furthermore, we have seen that the gestures used by speakers in both conditions refer, to some extent, to similar aspects of the scenes. Another question therefore is how similar, or different, the gestures are in each condition in terms of their semantic contents. The subsequent semantic analysis addresses these questions.

### Amount of semantic information encoded in speech-gesture utterances

Collapsing across the individual semantic categories we initially considered the amount of semantic information represented in each condition (see Figure 1). Firstly we considered only the information represented verbally. This showed, as expected, that speakers in the CG condition represented less semantic information than speakers in the N-CG condition,  $t = 3.06$ ,  $df = 38$ ,  $p = .002$ , one-tailed; (namely, on average, 25% less or, when considering the total number of semantic elements represented, 1550 compared to 2057, respectively). Next we considered not only the informational content of the



**Figure 1.** Mean amount of semantic information in the CG (Common Ground) and N-CG (No Common Ground) conditions, in speech and gesture combined, in speech alone and in gesture alone (error bars represent the standard error of means).

verbal utterances, but also their gestural content. In total, the speech-gesture utterances in the CG condition contained fewer semantic elements than in the N-CG condition (in terms of the total number of semantic elements, 2407 as compared to 3047 elements, respectively), and a *t*-test showed that those in the CG condition represented significantly less information than those in the N-CG condition (namely, on average, 32% less),  $t = 2.33$ ,  $df = 38$ ,  $p = .026$ , two-tailed. Finally, when considering only the amount of semantic information within gestures, we found that speakers in the CG condition represented less information than speakers in the N-CG condition (namely a total amount of 857, compared to 990 semantic elements, respectively); however the difference in the mean amount of information represented in the two conditions (namely 7% on average) failed to reach statistical significance,  $t = 0.96$ ,  $df = 38$ ,  $p = .341$ , two-tailed.

### Semantic feature analysis of gesture and speech

We then analysed the amount of information in participants' speech-gesture utterances for each of the ten semantic categories individually (see Table 1). The findings show, firstly, that the speech-gesture utterances in the CG condition contained less semantic information for each of the semantic

TABLE 1  
Amount of semantic information contained in utterances in the N-CG (No Common Ground) and CG (Common Ground) conditions, for speech (S) and gesture (G) combined

<i>Semantic category</i>	<i>Modality</i>	<i>N-CG Mean (SE)</i>	<i>CG Mean (SE)</i>	<i>t-value</i>	<i>df</i>	<i>p-value (two tailed)</i>
Entity	S+G	37.90 (2.98)	29.80 (2.07)	2.23	38	.032*
Identity	S+G	14.90 (1.14)	11.35 (0.82)	2.52	38	.016*
Movement	S+G	18.25 (1.21)	16.70 (1.15)	0.93	38	.359
Action	S+G	9.95 (0.73)	9.00 (0.60)	1.01	38	.319
Direction	S+G	9.95 (0.91)	8.55 (0.75)	1.19	38	.242
Manner	S+G	5.55 (0.92)	4.40 (0.61)	1.04	38	.304
Appearance <sup>a</sup>	S+G	4.00 (11.00)	2.00 (6.00)			.001* <sup>a</sup>
Position	S+G	18.10 (1.63)	14.80 (1.21)	1.63	38	.112
Size	S+G	30.10 (2.26)	22.40 (1.53)	2.82	38	.008*
Shape	S+G	2.70 (0.64)	1.30 (0.26)	2.02	25.20	.054

\* indicates statistical significance.

<sup>a</sup> Owing to non-normality of the data, a Mann-Whitney test was used and the figures therefore represent Median and (*Range*):

Appearance,  $U (n_1 = 20, n_2 = 20) = 84.50$ .

TABLE 2  
Amount of semantic information contained in utterances in the N-CG (No Common Ground) and CG (Common Ground) conditions, for speech (S) alone.

<i>Semantic category</i>	<i>Modality</i>	<i>N-CG Mean (SE)</i>	<i>CG Mean (SE)</i>	<i>t-value</i>	<i>df</i>	<i>p-value (two-tailed)</i>
Entity	S	24.00 (1.73)	18.30 (1.16)	2.74	38	.009*
Identity	S	14.85 (1.16)	11.35 (0.82)	2.47	38	.018*
Movement	S	10.05 (0.60)	8.85 (0.58)	1.44	38	.158
Action	S	7.40 (0.53)	6.50 (0.41)	1.35	38	.186
Direction	S	2.50 (0.32)	1.60 (0.30)	2.04	38	.048*
Manner <sup>a</sup>	S	2.50 (10.00)	2.00 (4.00)			.221 <sup>a</sup>
Appearance <sup>a</sup>	S	4.00 (11.00)	2.00 (5.00)			.001 <sup>a</sup>
Position	S	11.95 (1.05)	9.05 (0.58)	2.42	29.68	.022*
Size	S	23.15 (1.66)	17.40 (1.09)	2.90	38	.006*
Shape <sup>a</sup>	S	0.50 (3.00)	0.00 (2.00)			.086 <sup>a</sup>

\* indicates statistical significance.

<sup>a</sup> Owing to non-normality of the data, a Mann-Whitney test was used and the figures therefore represent Median and (*Range*):

Manner, U ( $n_1 = 20, n_2 = 20$ ) = 154.50

Appearance, U ( $n_1 = 20, n_2 = 20$ ) = 80.50

Shape, U ( $n_1 = 20, n_2 = 20$ ) = 136.50.

features, with no exception. Statistical comparisons show that the speech-gesture utterances in the CG condition contained significantly less information with regard to *Entity*, *Identity*, *Appearance*, and *Size* (the result for *Shape* information approached significance).

Next, we considered the amount of semantic information encoded in speech only, and this showed that speakers in the CG condition represented less semantic information than speakers in the N-CG condition, for all semantic features. The statistical results (Table 2) show that speakers in the CG condition encoded significantly less semantic information than speakers in the N-CG condition about *Entity*, *Identity*, *Direction*, *Appearance*, *Position*, and *Size*.

Finally, regarding the semantic information encoded in gesture, we found that speakers represented less information in the CG condition concerning all semantic categories except the category *Shape*, where the median value was the same in both conditions; regarding *Appearance* and *Identity* the median was 0.00 in both conditions (see Table 3). However, none of these differences were statistically significant; the result for *Size* information was marginally significant.

TABLE 3  
Amount of semantic information contained in utterances in the N-CG (No Common Ground) and CG (Common Ground) conditions, for gesture (G) alone.

<i>Semantic category</i>	<i>Modality</i>	<i>N-CG Mean (SE)</i>	<i>CG Mean (SE)</i>	<i>t-value</i>	<i>df</i>	<i>p-value (two-tailed)</i>
Entity	G	13.90 (1.67)	11.50 (1.14)	1.19	38	.242
Identity <sup>a</sup>	G	0.00 (1.00)	0.00 (0.00)			.799 <sup>a</sup>
Movement	G	8.20 (0.76)	7.85 (0.64)	0.35	38	.727
Action	G	2.55 (0.34)	2.50 (0.34)	0.10	38	.919
Direction	G	7.45 (0.79)	6.95 (0.57)	0.51	38	.610
Manner	G	2.25 (0.44)	2.15 (0.42)	0.62	38	.871
Appearance <sup>a</sup>	G	0.00 (1.00)	0.00 (1.00)			.799 <sup>a</sup>
Position	G	6.15 (0.78)	5.75 (0.88)	0.34	38	.736
Size	G	6.95 (0.93)	5.00 (0.72)	1.66	38	.053
Shape <sup>a</sup>	G	1.00 (10.00)	1.00 (3.00)			(one-tailed) .369 <sup>a</sup>

\* indicates statistical significance.

<sup>a</sup> Owing to non-normality of the data, a Mann-Whitney test was used and the figures therefore represent Median and (*Range*):

Identity, U ( $n_1 = 20$ ,  $n_2 = 20$ ) = 190.00

Appearance, U ( $n_1 = 20$ ,  $n_2 = 20$ ) = 190.00

Shape, U ( $n_1 = 20$ ,  $n_2 = 20$ ) = 166.50.

## DISCUSSION

The aim of the present study was to add to our understanding of the effect of mutually shared knowledge on the communication of meaning. Firstly, we analysed the number of words and gestures speakers used to narrate semantic events when they did, or did not, share knowledge with their addressee. In line with previous findings (e.g., Fussell & Krauss, 1992, Isaacs & Clark, 1987), speakers used significantly fewer words in their narrations when common ground existed. However, analyses revealed that they in fact produced proportionally *more* gestures (iconic + deictic gestures) in relation to the number of words used. Thus, speakers tend to use representational gestures at a higher rate when common ground exists (at least in the context of narratives). So, rather than speakers depending on gesture less when they convey information which the addressee already knows (as one might have intuitively assumed), representational gestures seem to actually carry a greater communicational weight with regard to the overall utterance.

Previous work has shown that verbal utterances tend to convey less information when common ground exists (e.g., Fussell & Krauss, 1992), and the present study replicated these earlier findings. The question emerges whether this claim still holds if we apply a broader notion of language, which views utterances as consisting of both, speech *and* gesture (e.g., Kendon,

1980, 2004; McNeill, 1985, 1992). Based on our results, we can maintain that, when we consider the semantic information represented in both of these modalities, speakers represent significantly less information when they mutually share knowledge with their addressee about the content of talk. Further, although speakers represented less semantic information when considering the gestural modality on its own, this difference was not statistically significant.

We also examined the information represented within the utterances by splitting the information up according to semantic category. When collapsing across modality, we found that the speech-gesture utterances produced when common ground existed contained less information about *Entity*, *Identity*, *Appearance*, and *Size* (with *Shape* approaching significance). When considering each modality in isolation, the verbal components of speakers' utterances were found to contain less information about *Entity*, *Identity*, *Direction*, *Appearance*, *Position*, and *Size* when common ground existed. Speakers also represented slightly less information gesturally, though none of these differences reached significance. One potential exception is *Size* information: based on Holler and Stevens (2007) we predicted that size information would be represented less frequently in gesture, and this difference was marginally significant.

To sum up, the findings show that speech-gesture utterances contain less semantic information when common ground exists, while, at the same time, gestures appear to carry a greater communicational weight due to a higher gesture rate (the gestures do not, however, appear to take up additional information which would have been encoded verbally when no common ground existed). The findings can be interpreted as supporting the notion of speakers using gesture and speech as a flexible, adaptable resource which allows them to best meet the communicational demands of a situation (cf. Kendon, 1983, 1985, 2004; Holler & Beattie, 2003). Retaining relatively high information content in gesture may facilitate speakers' creation of more elliptical verbal utterances without compromising effective, successful communication. According to such a view, gesture and speech may be working hand in hand to represent a semantically reduced but nevertheless sufficiently informative message. (This may be particularly beneficial when speakers may be in slight doubt as to whether the information they believe to be mutually shared knowledge is indeed known to, or remembered, by their addressee.) Also, being overly explicit and prolix in speech is seen as uncooperative conversational behaviour (Grice, 1975). Representing slightly more information than necessary in gesture, however, possibly may not be considered as violating conversational rules to quite the same extent (because, for example, we may be used to gesture responding to competing demands in conversation, aiding both the speaker and the addressee in the communication process (see below); what is deemed conversationally

appropriate in terms of gestural behaviour may therefore also depend considerably upon the communicational task and the message to be encoded). The observation of a stronger effect of common ground on speech than on gesture may therefore be evidence of cooperative behaviour (in the Gricean sense) with speakers using gesture and speech jointly to communicate meaning most efficiently (Clark & Brennan, 1991; Clark & Krych, 2004; Isaacs & Clark, 1987). We therefore do not interpret the lack of a strong effect on the semantic information encoded in gesture as evidence that speakers do not take into account mutually shared knowledge in 'designing' their gestures. Rather, we argue that speakers use speech and gesture together in a way that allows them to most appropriately and effectively communicate with a particular recipient. Thus, speech-gesture utterances, if considered as wholes, can indeed be seen as 'recipient designed' (Bavelas et al., 2002).

Enfield, Kita, and de Ruiter (2007) have discussed a similar function of pointing gestures in a corpus of conversations with Lao speakers. Their analysis distinguishes two different kinds of pointing gestures, namely those which are rather large gestural movements, involving the whole arm, outstretched, and the elbow being raised (B-points), and those which are comparatively smaller and faster, more casually performed gestures, involving mainly a movement of the hand while the arm is not fully outstretched (S-points). They found that whereas B-points are mainly used with demonstratives, thus carrying much of the utterance's information, S-points appear to play a more secondary function, signalling in the manner in which they are performed that the information they convey may be of lesser importance to the addressee. Also, Enfield et al. (2007) argue that these S-points may supplement speech 'in a situation where the referent seems likely but not certain to be recognizable for [...] interlocutors' (p. 1730). They argue that this more subtle and 'off-the-record' manner of representing information provides the speaker with a compromise solution, or what they call a 'safety net for supposition', allowing the speaker to construct verbal utterances which do not underestimate common ground while at the same time not risking to provide less information than is needed. This is very much in line with the conclusions we are drawing here based on data comprising both iconic and deictic gestures produced in quite a different context.

At this stage, however, we must remain tentative regarding the exact reasons for the observed influence of common ground on gesture and speech in this narrative task. Experimental evidence to date suggests that gestures may indeed be multifunctional, and thus may play a role in social interaction when communicating information with an addressee (e.g., Alibali et al, 2001; Gullberg, 2006; Özyürek, 2002), as well as facilitating lexical retrieval (e.g., Hadar, Dar, & Teitelman, 2001; Pine, Bird, & Kirk, 2007), or the conceptual planning of speech (e.g., Hostetter, Alibali, & Kita, 2007; Kita & Özyürek, 2003), to name but a few. Regarding the present data, we do not consider

lexical retrieval to be a likely explanation for our findings because the speakers in our data did not appear to experience greater difficulty retrieving lexical items when common ground existed. Indeed, the lexical items used by these speakers appeared, if anything, less carefully chosen and to be more high-frequency items (e.g., 'car' rather than 'VW Beetle'). However, we do accept that common ground may potentially affect conceptual planning processes, for example in that speakers have to consider the knowledge they mutually share with their addressee (however, note that Horton and Keysar, 1996, provide evidence suggesting that speakers do not take common ground into account during the initial stages of utterance planning). In conclusion, the present study has provided additional insight into how common ground affects verbal and gestural behaviour. Future studies are needed to shed more light on the exact social, interactional, and cognitive processes involved in the observed effects.

## REFERENCES

- Alibali, M. W., Heath, D. C., & Meyers, H. J. (2001). Effects of visibility between speakers and listeners on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188.
- Bavelas, J. B., & Chovil, N. (2000). Visible acts of meaning: An integrated message model of language in face-to-face dialogue. *Journal of Language and Social Psychology*, 19, 163–194.
- Bavelas, J., Chovil, N., Lawrie, D., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, 15, 469–489.
- Bavelas, J. B., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture*, 2, 1–17.
- Beattie, G., & Aboudan, R. (1994). Gestures, pauses and speech: an experimental investigation of the effects of changing social context on their precise temporal relationships. *Semiotica*, 99, 239–272.
- 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.
- Bergmann, K. (2006). *Koexpressivität von Sprache und Gestik in Wegbeschreibungen*. Unpublished Diploma Thesis, Bielefeld University, Faculty of Technology.
- Bergmann, K., & Kopp, S. (2006). Verbal or visual? How information is distributed across speech and gesture in spatial dialog. *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (SemDial-10)*, Potsdam, Germany, pp. 90–97.
- Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, 4, 75–87.
- Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review*, 96, 168–174.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

- Clark, H. H., & Brennan, S. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition*. Washington, DC: American Psychological Association.
- Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62–81.
- Clark, H. H., & Schäfer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*, 259–294.
- Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1–39.
- De Ruiter, J.-P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge: Cambridge University Press.
- Enfield, N. J., Kita, S., & de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*, 1722–1741.
- Freedman, N. (1977). Hands, words, and mind: On the structuralization of body movement during discourse and the capacity for verbal representation. In N. Freedman & S. Grand (Eds.), *Communicative structures and psychic structures: A psychoanalytic interpretation of communication* (pp. 109–132). New York: Plenum Press.
- Furuyama, N. (2000). Gestural interaction between the instructor and the learner in origami instruction. In D. McNeill (Ed.), *Language and gesture* (pp. 99–117). Cambridge: Cambridge University Press.
- Fussell, S. R., & Krauss, R. M. (1989). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, *25*, 203–219.
- Fussell, S. R., & Krauss, R. M. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, *62*, 378–391.
- Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Gerwing, J., & Allison, M. (2008). *Coordinated semantic integration of verbal and gestural contributions in conversation: an experimental investigation*. Manuscript submitted for publication.
- Gerwing, J., & Bavelas, J. B. (2004). Linguistic influences on gesture's form. *Gesture*, *4*, 157–195.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, Vol. 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Gullberg, M. (2006). Handling discourse: gestures, reference tracking and communication strategies in early L2. *Language Learning*, *56*, 155–196.
- Hadar, U., Dar, R., & Teitelman, A. (2001). Gesture during speech in first and second language: Implications for word retrieval. *Gesture*, *1*, 151–165.
- 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). How iconic gestures and speech interact in the representation of meaning: are both aspects really integral to the process? *Semiotica*, *114*, 86–116.
- Holler, J., & Beattie, G. (2004). The interaction of iconic gesture and speech. In A. Camurri & G. Volpe (Eds.), *Lecture notes in computer science, 5<sup>th</sup> International Gesture Workshop, Genova, Italy, April 15–17, 2003, Selected revised papers* (pp. 63–69). Heidelberg, Germany: Springer Verlag.
- Holler, J., & Beattie, G. (2007). Gesture use in social interaction: How speakers' gestures can reflect listeners' thinking. Lorenza Mondada (Ed.), *Proceedings of the 2<sup>nd</sup> Conference of the International Society of Gesture Studies*, Lyon, France 15–18 June 2005.
- Holler, J., & Stevens, R. (2007). An experimental investigation into the effect of common ground on how speakers use gesture and speech to represent size information in referential communication. *Journal of Language and Social Psychology*, *26*, 4–27.
- Hostetter, A., Alibali, M., & Kita, S. (2007). I see it in my hands' eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, *22*, 313–336.

- Horton, W., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, 59, 91–117.
- Isaacs, E. A., & Clark, H. H. (1987). References in conversations between experts and novices. *Journal of Experimental Psychology: General*, 116, 26–37.
- 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. (1983). Gesture and speech: How they interact. In J. M. Wiemann & R. P. Harrison (Eds.), *Nonverbal interaction* (pp. 13–45). Beverly Hills, CA: Sage.
- Kendon, A. (1985). Some uses of gesture. In D. Tannen & M. Saville-Troike (Eds.), *Perspectives on silence* (pp. 215–234). Norwood, MA: Ablex.
- 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.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48, 16–32.
- Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical retrieval: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.
- Krauss, R. M., Morrel-Samuels, P., & Colasante, C. (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, 61, 743–754.
- 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.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (Ed.), *Language and gesture* (pp. 64–83). Cambridge: Cambridge University Press.
- Özyürek, A. (2002). Do speakers design their co-speech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, 46, 688–704.
- Parrill, F. (in press). The hands are part of the package: Gesture, common ground and information packaging. In J. Newman & S. Rice (Eds.), *Empirical and experimental methods in cognitive functional research*. Stanford, CA: CSLI.
- Pine, K., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science*, 10, 747–757.
- Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman & B. Rimé, *Fundamentals of nonverbal behavior* (pp. 239–281). Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1986). *Thought and language*. A. Kozulin (Ed.). London: MIT Press. (Original work published 1934)