

Gesture and the communicative intention of the speaker*

Alissa Melinger and Willem J. M. Levelt

Max-Planck-Institute for Psycholinguistics, Nijmegen

This paper aims to determine whether iconic tracing gestures produced while speaking constitute part of the speaker's communicative intention. We used a picture description task in which speakers must communicate the spatial and color information of each picture to an interlocutor. By establishing the *necessary minimal content* of an intended message, we determined whether speech produced with concurrent gestures is less explicit than speech without gestures. We argue that a gesture must be communicatively intended if it expresses necessary information that was nevertheless omitted from speech. We found that speakers who produced iconic gestures representing spatial relations omitted more required spatial information from their descriptions than speakers who did not gesture. These results provide evidence that speakers intend these gestures to communicate. The results have implications for the cognitive architectures that underlie the production of gesture and speech.

Keywords: gesture, interaction, speech production

In the course of speaking, people regularly produce gestures that are temporally synchronized and semantically tied to their speech. Intuitively, one might assume that speakers produce these co-speech gestures to create a fuller, more expressive message than can be created with speech alone. However, research on gesture and communication has yet to clearly demonstrate that speakers use gestures to convey information or that speakers view their own gestures as communicatively informative (see Kendon, 1994 for a review).

The question of whether and how gestures contribute to communication is complex and should be broken into separate questions focusing on the addressee and the speaker independently. While some research has shown that

listeners' comprehension of speech was not influenced or supplemented by the presence of gestures (Krauss, Morrels-Samuels, & Colasante, 1991; Krauss, Dushay, Chen, & Rauscher, 1995; Feyereisen, van de Wiele, & Dubois, 1988), other studies show that listeners do incorporate gesturally expressed information into their broader understanding of a narrative (Beattie & Shovelton, 1999a, 1999b, 2002; Cassell, McNeill, & McCullough, 1999; Graham & Argyle, 1975; Gullberg, 2003). It is often implicitly assumed that demonstrating that listeners use gestured information is tantamount to demonstrating that speakers *intended* that the information they provided in gesture be understood. However, as argued by de Ruiter (2000) and Bavelas, Kenwood, Johnson and Phillips (2003), whether listeners do or do not extract information from gestures is an independent empirical issue from whether speakers use gestures to communicate. The present paper will address the latter issue.

The intensity of the debate about the communicative functions of gestures varies greatly for different types of gestures. Most researchers agree that *deictic* or *pointing* gestures, which identify real or abstract entities or locations in space, are often intended to communicate. Deictic gestures produced in lieu of speech or with deictic referring expressions such as "here" or "there" are especially uncontroversial. Instead, the debate centers on *iconic* gestures, as defined in McNeill (1992). Iconic gestures, also known as representational or lexical gestures, crucially share a transparent relationship with some semantic aspect of the concurrent speech, often representing concrete or abstract entities, traits, or activities. The type of iconic gesture addressed in the current paper is tracing gestures, which highlight spatial characteristics of to-be-described images (cf. Müller, 1999).

Demonstrating that speakers use iconic gestures to communicate is a difficult task. Consider an example from McNeill (1985) in which the speech and gesture express different aspects of an event:

Speech: 'She chases him out again'

Gesture: Hand, gripping an object, swings from left to right

McNeill interprets this gesture as expressing the manner in which the chasing event (described in speech) occurs, namely by swinging an umbrella. However, it is not clear whether the speaker intended to convey this manner information to his interlocutor. Just because the speaker produces the gesture is not sufficient evidence. Assuming so leads down a slippery slope. The gesture encodes many characteristics of the event, such as the speed of the back and forth motion, the handgrip on the umbrella, and the angle at which the umbrella is held.

Any of these gestured elements may have been intended as part of the speaker's message, but it cannot be taken for granted that they are. Likewise, one also cannot take for granted that such gestures are *not* intended to communicate. This is in fact the position explicitly taken by Krauss, Chen, and Gottesman (2000; p. 266). These authors explicitly assume that gestures of the type described above are not communicatively intended.

Clearly, gestures can convey rich semantic information, some of which is redundant with speech and some of which is supplementary. Unfortunately, it is very difficult for an analyst to determine which of these meaning components are meaningfully intended and which not. Classic measures like speech rate or number of words do not provide adequate insight into the speaker's communicative intention. Investigating the listener's understanding of the discourse is likewise insufficient. Rather, what is needed is a method that grants insight into the speaker's intended message and examines whether components of that intended message are expressed with gesture.

In this paper, we investigate whether *speakers* consider the iconic tracing gestures that they produce while speaking to be communicative. Our aim is to determine whether speakers use these gestures intentionally to convey part of their message. We address this issue with a picture description task in which we evaluate whether gesture use influences the content of concurrent and subsequent spoken utterances.

Few studies have explicitly addressed whether iconic gestures form part of the speaker's communicative intention. One study that points towards an intentional use of gesture investigated the rate of gesturing in different interaction situations. Cohen and Harrison (1973; see also Cohen, 1977) showed that gesture frequency is greater in face-to-face interactions than in non-visible (intercom) interactions, suggesting that speakers use gestures strategically when appropriate. This difference persists even when the listener is actually non-present. While describing pictures, speakers produce more gestures when they think a future listener will see a video of their description compared to when they think a future listener will only hear the audio portion of their description (Bavelas et al., 2002).

Graham and Heywood (1975) focused on the speaker's use of gesture by examining the effect that gesture prohibition has on the content of speech. Graham and Heywood compared the rate and content of speech produced when gestures were allowed to when they were prohibited. When gestures were discouraged, Graham and Heywood found an increase in the number of words used to describe spatial relations and a decrease in the use of deictic expressions

compared to when gestures were allowed. Their results suggest that gestures are used to convey (spatial) information that is not expressed in the verbal message. However, gesture prohibition has been shown to lead to changes in speech rate and fluency for reasons associated with speech production processes, e.g., the speaker-directed functions of gesture (Rauscher, Krauss, & Chen, 1996). Thus gesture prohibition in a task aimed at evaluating the communicative functions of gestures is not ideal. A stronger criticism concerns Graham and Heywood's analysis. Since their analysis focused on the number of words used to describe spatial information rather than on the presence or absence of certain spatial characteristics, it is not possible to determine whether the gestures were actually intended as part of the communicative message. Clearly, it is possible to express the same information with more or fewer words; thus, the observed decrease in the number of words does not inform us about the content of the speech.

This criticism brings to light a critical methodological criterion for determining whether a gesture forms part of a speaker's communicative intention, namely that gesture use or prohibition must influence the *content* of the concurrent speech, not only its rate or other surface characteristics. Of course, this measure is difficult because it requires that one first determine what the speaker's communicative intention was. In natural discourse, the communicative intention extends far beyond the sum total of the words expressed (consider, for example, sarcasm). As a result, identifying the intended meaning of a naturally occurring utterance often depends greatly on hermeneutic tactics that do not provide a high degree of certainty.

In experimental settings, however, speakers are assigned specific tasks which are designed to restrict the set of possible communicative intentions. With an experimental setting one can determine the *necessary minimal content* of a message. For example, if one were instructed to name all the colors present in an image, then the necessary minimal content of the resulting message would be the color information. Speakers are always free to include additional information, such as the relative positions of the colors to each other, but this would be considered unnecessary content.

For purposes of the present study, we used a picture description task in which participants describe networks of colored circles arrayed along a path. We identified the necessary minimal content of each description as the order and color of the circles and the spatial relationships between circles, usually described with directional terms such as *left*, *right* and *straight*.

Although directional information is necessary to the message in this task, it can still be omitted from picture descriptions under certain circumstances.

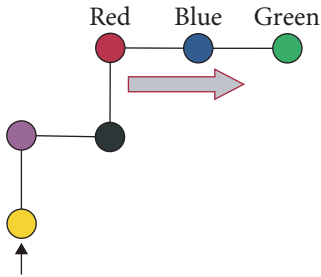


Figure 1. An example picture in which the direction stays the same.

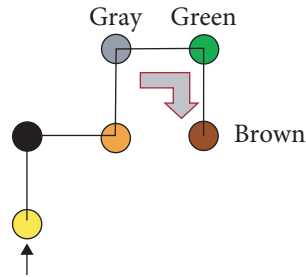


Figure 2. An example picture in which the direction changes.

In Levelt (1996), the criteria for directional omission (called directional ellipsis) were investigated. Levelt reasoned that the decision to elide directional terms could be either conceptually or lexically motivated; the decision could be based on the *image* containing a repeated direction or on the *verbal description* using a repeated lexical item. To discriminate between these two possibilities, Levelt compared ellipsis production in descriptions of two types of pictures, those with repeated directions, as in Figure 1, and those with changes in direction, as in Figure 2. Additionally, he contrasted *deictic perspective* descriptions to *intrinsic perspective* descriptions (Miller & Johnson-Laird, 1976). Crucially, different linguistic perspectives result in the use of different linguistic terms to express the same spatial relations, as seen in examples (1a) and (1b).

Deictic description for Figure 1:

- (1) a. You begin with a yellow circle. *Above* that you see a purple circle. To the *right* you see a black circle. Above the black circle you see a red circle. To the *right* of the red circle is a blue circle and then again to the *right* is a green circle.

Intrinsic description for Figure 1:

- b. You begin with a yellow circle. Then you go *straight* to a purple circle. Then you go to the *right* to a black circle and then *left* to a red circle. At the red circle you go *right* to a blue circle and then *straight ahead* to a green circle.

For each of these figures, the last two transitions can be described with lexical repetition, *right and then right again*, in only one of the two perspectives, namely in the deictic perspective for Figure 1 and in the intrinsic perspective for Figure 2. In the intrinsic perspective, the final two transitions of Figure 1

are described as *to the right and then straight*; in the deictic perspective, the final two transitions of Figure 2 are described as *to the right and then down*. Thus by comparing same and different direction transitions described in the two linguistic perspectives, Levelt could distinguish whether decisions to elide directional information were based on lexical or conceptual factors.

Levelt found that directional information was only excluded when the direction of movement was repeated, as in Figure 1, and never when it changed, as in Figure 2, suggesting that the decision to elide was based on the conceptual representation (i.e., the image) used by the speaker to construct their description, rather than on the desire to avoid the repetition of a lexical direction term.

Levelt's (1996) study provides an ideal backdrop for the current investigation. Since speakers in that study spoke to a tape recorder, not to an interlocutor, gestures could not be used to convey information; thus, Levelt's data provide a measure of the frequency with which speakers omit directional information independent of gesture use. In the current investigation, we examine the frequency and distribution of directional omissions with and without concomitant gestures.

The picture description task constrains the necessary minimal content of the message. Our analysis will determine whether the use of gesture influences the expression of that necessary content. For example, does the restriction against change of direction omissions persist when speakers can additionally convey directional information via gesture? If necessary information is omitted from speech more often and under different circumstances in the presence of a compensatory gesture, this would suggest that the gesture was communicatively intended by the speaker. In other words, to conclude that gestures are communicatively intended, we must show that the *use* of gesture influences the *content* of the concurrent or subsequent speech.

The approach taken for the current study is different from prior studies which investigated similar questions in that, (a) speakers were not prohibited from gesturing but rather allowed to choose whether they gestured or not, (b) the speaker's minimal necessary content was identified by setting a specific task that speakers must complete, and (c) rather than focusing on measures such as number of words or gestures per minute, we assess what information is included or excluded given the presence of a co-speech gesture. Thus, we can reveal a relationship between gesturing and speakers' decisions of what to express verbally.

To test our hypothesis, we conduct two analyses. In analysis I, we compare the number and type of directional omissions produced by gesturing and

non-gesturing speakers. If gestures form part of the speaker's communicative intention, then their use may be correlated with a reduction in explicit directional terms. Thus, gesturers may be more likely to omit directional information from their speech than non-gesturers and spatial information may be omitted in a broader range of circumstances, including *change of direction* situations, when gestures can compensate for the omitted information. In contrast, if gestures are not used to communicate, then the content of speech, measured by the rate of directional omissions, should be unaffected by gesture use. Furthermore, the type of directional omissions produced should not differ between gesturers and non-gesturers.

Another indication that speakers intend their gestures to communicate information would be if a gesture produced early in the discourse had an effect on our second analysis thus focuses on the content of the subsequent text. In the picture description task, participants often include overview information, such as in examples given in (2), that provided the addressee with a general idea of what the end product should look like. Participants can provide overview information at the onset of their description, functioning as a preview of the image, and in the middle or end of a description, functioning as a summary. The use of overviews may influence a speaker's decision to omit information later in the linguistic description if the speaker feels that the information has already been presented.

- (2) a. this is some sort of T with an additional roof at one side
 b. this figure you can almost view as some sort of staircase
 c. and the figure then is an F
 d. so that is the staircase

Overviews can include (implicitly or explicitly) the direction of the transitions between circles. If speakers intend overviews to be useful and informative to the addressee, then the inclusion of an overview may influence the speaker's decisions about what needs to be expressed in subsequent utterances. Specifically, if an overview is intended to convey crucial information about the arrangement of circles in the image, then redundant subsequent directional information may be omitted. One prediction, then, is that more directional omissions will be observed in descriptions with overviews than in descriptions without overviews. Furthermore, since only overviews that precede the description should influence the content of the speech, this difference should be confined to the overviews at the beginnings of descriptions; summary overviews should not be related to the frequency of directional omissions in speech. If overviews at the beginning, but not the end, of a description are related to the omission of

direction information from speech, then this would suggest that the direction of influence is from gesture to speech and not the reverse.

In face-to-face interactions, speakers can provide overview information verbally, or bimodally. To clearly attribute changes in content to the information conveyed in gesture, gestured overviews must be distinguished from spoken overviews. If the gestured overview is as effective at imparting shape information as speech, then more directional omissions should follow gestured overviews than when no picture overview is provided in either modality. However, if the speaker does not intend the gestured overviews to convey shape information, then there should be no change in the rate of directional omissions in speech following gestured overviews.

Experiment

Speakers described networks of colored circles connected by lines, creating a path, to a visible interlocutor. Participants were explicitly instructed to identify the color of each circle and the spatial relationship between each circle (or the direction of the transition from one circle to the next) in their descriptions. Given this specific task, the minimal intended message for each image can be objectively determined before any description is elicited. Our analyses are designed to determine whether the inclusion of gestures is related to the frequency and type of directional omissions. If speakers use gestures to convey part of the intended message, the use of gestures should be related to the omission of necessary, and therefore intended, information. If speakers do not intend their gestures to be informative, then there should be no relationship between the occurrence of gestures and the omission of information.

Method

Participants. Thirty university-aged native speakers of Dutch were paid for their participation.

Materials. Sixteen images depicting networks of colored circles were constructed. Each image included an explicit start point and five or six colored circles. No color was repeated within an image. Half of the pictures had branching paths. Both *repeated direction transitions*, as in Figure 1 above, and *change of direction transitions*, as in Figure 2 above, were included. In all, 480 total picture descriptions were collected.

Procedure. Participants were instructed to describe the images to their interlocutor, being sure to mention the direction from one circle to the next as well as the color of each circle. They were also instructed that their descriptions should follow the path, not jump from one circle to adjacent but unconnected circles. Sample descriptions were provided to set a standard minimal level of detail. The instructions stressed that speakers could use any means to convey the picture information, including using their hands if they so desired. In this way, we allowed speakers to choose whether they wanted to gesture or not rather than explicitly prohibiting them.

The addressees in the study were experimental confederates. We chose to use confederates rather than real interlocutors in order to ensure that speakers' interactions were maximally equivalent. However, if the speaker had been aware that the interlocutor already had experience with the pictures, then the completeness of their descriptions might have been affected. Therefore, interlocutors behaved as first time participants and speakers were instructed that the interlocutor's task was to answer, in writing, questions about the image after each description. Participants were asked to make their descriptions specific enough to meet this demand.

Speakers were seated across from their interlocutor. A 12-inch divider separated them, allowing face-to-face interactions while restricting the addressee's view of the image. The barrier was low enough to allow gestures to be seen by the addressee (in a slightly raised gesture space) while also providing a space where unseen gestures could be produced. Gestures that were visible to the interlocutor, produced in the normal or slightly raised gesture space, were viewed as addressee-directed while gestures produced in the speaker's lap or under the table were viewed as speaker-directed (cf. Anderson, Robertson, Kilborn, Beeke, & Dean, 1997). The addressee-directed gestures are the focus of the analyses.¹

The experimenter sat behind and to the left of the participant. Each picture was handed to the participant individually and placed on the table. The interlocutor was instructed not to speak and to provide minimal feedback so as not to influence the speaker's level of linguistic detail.² Following each trial there was a pause of approximately 30 sec. to allow the addressee to answer questions about the image (thus maintaining the pretense that the task was to describe the images in a way that allowed the addressee to answer questions about them).

Sessions were video recorded from two viewpoints. One camera was placed directly above the speaker aiming downwards. A second camera was to the speaker's left and captured the gestures from a side view.

Coding system

The videotapes were used to create a transcription of the speech as well as a record of all gestures. Transcripts were produced by a native Dutch speaker blind to the hypotheses under investigation but familiar with the gesture transcription system proposed by McNeill (1992). Given the nature of our gesture data, we chose to score whether a stretch of speech occurred with concurrent gestures or not rather than attempting to quantify the number of individual gestures. Picture descriptions were coded as including co-expressive iconic gestures, (a) throughout the verbal description, (b) for portions of the verbal description or (c) not at all. When gestures were only produced during parts of the descriptions, the transcriber noted which portions of the verbal description were produced with co-expressive gestures. Since the majority of the gestures expressed spatial relations, directions or shapes, it was fairly straightforward to determine whether gestures were semantically tied to the concurrent speech or not.

In a second step of coding, the transcriber identified all instances of directional omissions and noted, for each omission, whether the speaker produced a compensatory iconic gesture. Each directional omission was coded either as a *same direction* or a *change of direction* omission, depending on the characteristics of the image. Cases where speakers underspecified the direction of a transition (e.g., the red ball is *next to* the green one) were also coded as omissions. Underspecified directions reduce the world of possible directions for a transition (in this case to *left* or *right*) but they do not uniquely identify the direction, leaving it ambiguous for the listener. Each picture was also coded as including or not including a picture overview. Overviews occurring at the beginning, middle or end of the descriptions were distinguished. They were also divided by modality of presentation: gesture, speech or both.

Results and Discussion

Gesture use

Thirteen of the thirty speakers used gestures constantly in combination with their verbal descriptions. An additional four speakers produced few manual gestures but regularly used head movements to convey left/right information. Thirteen speakers produced few or no gestures in the visible gesture space.

The gestures observed in this elicitation procedure predominantly expressed spatial notions, such as direction of transition (e.g., *left, right, up, down,*

etc.) or spatial relations (e.g., *above*, *below*, *next to*, etc.). Other gestures accompanied words or phrases that highlighted the overall shape of the image (e.g., *this one is shaped like a backwards F*). Many gestures also had a deictic component in that the end-point of the gesture indicated the abstract location of the circle within the diagram being created.

Figure 3–5 show some typical gestures elicited in this study. In Figure 3, the speaker uses a gesture to illustrate the circle's location relative to prior and subsequent circles (*Purple is in the middle*). Additionally, her hand shape illustrates the shape of the circle. In Figure 4, the speaker provides general information about the shape of the image (*This is a cross with, on the right hand side, a longer piece*). In the first image she uses both hands to set up the middle point of the cross and in the second she moves one hand to the right, illustrating which side is longer. In Figure 5, the speaker produces a two-handed gesture expressing the spatial relationship between circles (*To the right of purple is first blue*). Speakers often used two hands when expressing spatial relations, one hand to mark the location of the given circle and one to indicate the relative position of the new circle. Two hands are used especially often when one circle will be mentioned again in the description, one hand serving to anchor the location of the recurring circle.

Speech patterns

Prior studies in which speakers describe similar images to a tape recorder or to non-visible interlocutors found a 3-to-1 preference for deictic descriptions over intrinsic descriptions (Levelt, 1996; Melinger & Kita, in press). In the current study, only five speakers consistently used the intrinsic perspective, one used a combination of the two perspectives, two used neither perspective³ and the remaining 22 produced consistently deictic descriptions.

Of the thirteen speakers who gestured consistently, ten produced seemingly complete verbal descriptions (the validity of this impression is further investigated in analyses I and II) while two divided the information load between the two modalities (one speaker alternated between these two strategies). Examples (3a) and (3b) are taken from the translated transcriptions of gesturing speakers who either produced full verbal descriptions or divided descriptions, respectively. Note that in (3a) each transition from one circle to the next is explicitly labeled with a directional term. This verbal description contains all the necessary information to reproduce the shape and colors of the image. In contrast, (3b) contains no directional terms at all. Thus, for the addressee to comprehend what the image looks like, he must extract the spatial information



Figure 3. An example of a speaker producing a gesture that shows a circle's shape and location.



Figure 4. An example of a speaker producing a gesture that shows a general characteristic of the image.



Figure 5. An example of a speaker producing a gesture that shows the relationship between two circles.

from the gestures. It is important to note that the gestural behaviors of these two types of speakers do not obviously differ; both speakers gesture throughout the entire description of the image, indicating circle locations and transition directions with gesture. What is different between the speakers is whether the direction information is additionally encoded in speech.

- (3) a. Uhm, you start with a black circle. And then upwards is a red circle. To the left, that is a green circle. Then upwards again, there is a yellow circle. And then from the yellow to the right there is a blue circle.
- b. Yes, it is again such a figure. Yes, let's see...It starts over here with a yellow one. Then it goes to a black one, to an orange one, to a blue one, to a green one and then again to a brown one.

The fact that some speakers choose to divide the necessary information between two modalities already provides a hint that these tracing gestures were intended to communicate; if information that must be communicated is only expressed in gesture, then the gesture must be viewed as the source of that information. To address the issue of the intentional use of gestures *redundant* with speech, which are at the heart of the communicative debate, we examine more closely the content of the descriptions produced by gesturing speakers compared to the content of the descriptions produced by non-gesturing speakers.

Analysis I: Directional omissions in the concurrent speech

For this analysis, two types of omissions were excluded from consideration. First, omissions from speakers who divided their information load between the two modalities were considered different in nature from the omissions of speakers who otherwise produced full and free-standing linguistic descriptions. Thus, no more than two consecutive omissions were allowed within a picture; if more than two consecutive transitions were omitted in speech, then the description was excluded from this analysis. Second, omissions of the first transition within a picture, from the start point circle to the second circle, were also excluded from the analysis. Speakers often failed to mention this direction, perhaps because the first movement in all pictures was upwards, or straight. Prior studies without a face-to-face interaction have revealed the same tendency to omit the (fully predictable) direction of the first movement (Melinger & Kita, in press).

In total, there were 97 directional omissions produced from 78 picture descriptions. Twenty of the 30 participants produced at least one directional omission. We compared the overall omission rate for consistent gesturers ($N = 10$) to consistent non-gesturers ($N = 13$).⁴ We also looked at the type of directional omissions (Same vs. Different) produced by these two groups. Average omission rates for gesturing and non-gesturing speakers are provided in Table 1.

As predicted, gesturers omitted directional information more often than non-gesturers, although this difference was only marginally significant, Mann-Whitney U: $z = 1.74$, $p = .08$.⁵ Furthermore, *both* same and change of direction transitions were omitted, but only by gesturing speakers. Gesturers produced a comparable number of same and different directional omissions, Wilcoxon Sign: $z < 1$, while the non-gesturers produced only same direction omissions, Wilcoxon Sign: $W = 21$, $n_{s/r} = 6$, $p < .05$. Finally, the distribution of same and different direction omissions produced by gesturers and non-gesturers is significantly different, Fisher's Exact Probability Test: $p < .01$. The fact that gesturers omitted directional information more often than non-gesturers supports the claim that the decision to gesture is related to what information is explicitly included in speech. The pattern of omissions produced by non-gesturers, namely failing to omit change of direction transitions, replicates the original finding from Levelt (1996). The fact that gesturers omitted change-of-direction transitions as often as they omitted same-direction transitions reinforces the relationship between gesture production and the content of speech.

While on the surface the linguistic content of gesturers and non-gesturers appeared equivalent, closer inspection revealed that gesturers omitted necessary directional information from their speech more often than non-gesturers. This difference suggests that speakers who produce gestures expressing spatial information are free to be less linguistically explicit about spatial relations than speakers who do not. This freedom may come from the knowledge that the necessary information is being conveyed by the gesture as well. Thus, we observe a relationship between gesture production and the frequency and type of directional omissions in speech; gesture production seems to be tied to decisions about what speakers express verbally.

Table 1. Average number (and standard error) of same and different directional omissions produced by gesturing and non-gesturing speakers.

Omissions	Gesturer (N = 10)	Non-Gesturer (N = 13)
Same direction	2.8 (0.9)	0.9 (0.5)
Different direction	2.0 (0.7)	0.0 (0.0)
OVERALL	4.8 (1.5)	0.9 (0.5)

Furthermore, we also observed that gesturers produced a type of directional omission in face-to-face interactions that were not observed in non-face-to-face interactions (Levelt, 1996) or by our non-gesturers. One possible explanation for the difference is that change of direction omissions are only interpretable with a concomitant gesture. Speakers, aware of the interpretation problem, only omit non-retrievable information when they produce a gesture that compensates for the missing information. Gesturers therefore have a broader range of situations in which directional omissions are felicitous.

The analyses presented thus far can only show a correlation between gesture production and directional omissions. An alternative interpretation of the results is that the direction of influence is reversed; namely, what speakers express verbally influences what is expressed in gesture. This alternative possibility is addressed in analysis II.

Analysis II: Previewing with gesture

For this analysis, we included omissions from speakers who divided their information load between the two modalities because it is possible that the use of an overview is related to the decision to adopt this description style.

Across the 30 participants, 95 picture descriptions included overview information, some with multiple overviews at various points in the description. Across modalities, the corpus included 79 initial overviews, 13 medial overviews and nine final overviews. Sixteen of the 30 participants produced at least one overview; 15 participants produced at least one initial overview, three participants produced at least one medial overview, and five produced at least one final overview.

Table 2 shows the total number of picture descriptions that did or did not include an overview as well as the number with initial and non-initial overviews. The proportion of these picture descriptions that omitted directional information is also presented. The difference in the likelihood of omitting directional information in a description that included an overview at any point compared to descriptions that did not was significant, $t(42) = 2.6$, $p < .02$.⁶ This difference was slightly greater when descriptions with initial overviews are compared to descriptions without overviews, $t(42) = 2.7$, $p < .01$. Due to the small number of tokens, a Chi-squared test was conducted to determine if the proportion of directional omissions was greater in descriptions including non-initial overviews compared to descriptions without an overview; no significant difference was found.

Table 2. Total number of picture descriptions with or without overviews (initial or non-initial) and the proportion of these picture descriptions with an omission of directional information.

Description type	Total number of descriptions	Proportion of descriptions with directional omission
Initial Overview	79	0.64
Non-initial Overview	22	0.14
All Overviews	95	0.43
No Overview	385	0.17

Table 3. Total number of initial overviews produced manually, verbally, and in both modalities and the proportion of descriptions with subsequent directional omissions.

Description type	Total number of descriptions	Proportion of descriptions with directional omission
Overview in speech	32	0.34
Overview in gesture	20	0.95
Overview in both modalities	27	0.70
No overview	385	0.17

Table 3 shows the total number of initial overviews produced manually, verbally, and in both modalities and the proportion of these descriptions with subsequent directional omissions. Again, because the number of tokens in this sample is small, tests that allow generalizations across speakers are not possible.⁷ Instead, we conducted Chi-squared tests to evaluate differences in the distribution of directional omissions given the presence or absence of overviews. Compared to picture descriptions without an overview, the proportion of directional omissions was greater following gestured overviews, $\chi^2 = 114.75$, $df = 1$, $p < .001$, spoken overviews, $\chi^2 = 5.3$, $df = 1$, $p < .05$, and bimodal overviews, $\chi^2 = 50.61$, $df = 1$, $p < .001$. Furthermore, overviews presented manually or bimodally had a significantly stronger influence on the inclusion of subsequent directional information than spoken overviews, $\chi^2 = 78.61$, $df = 1$, $p < .001$; $\chi^2 = 24.54$, $df = 1$, $p < .001$, respectively. Thus, the inclusion of an initial overview in all modalities increased the likelihood that speakers omitted directional content from subsequent utterances; interestingly overviews with a gestural component exhibited a stronger effect than purely verbal overviews.

The fact that speakers left out information more often when they had expressed that same information in a picture overview suggests that the speaker intended the overview to convey part of the communicative message. Furthermore, overviews led to the omission of directional information even when

the overview was only expressed manually, highlighting the role of gesture. Gestured and bimodal overviews also preceded more directional omissions than spoken overviews, possibly due to the different types of information typically conveyed in the two modalities. Overviews presented gesturally generally traced the whole shape of the image, providing maximal directional information. Speech overviews tended to provide general shape information by comparing the image's shape with that of some familiar entity (e.g., staircases, crosses, the letter F).

All of the predictions regarding overviews were born out by this analysis. Initial overviews lead to an increased number of directional omissions while medial and final overviews did not. The temporal relationship between initial overviews and omitted directional information also suggests that the decision to gesture influences the content of speech rather than the reverse relationship where the content of speech influences decisions of whether or not to gesture.

General discussion

Using a picture description elicitation task in a face-to-face interaction, we collected descriptions in which speakers freely chose whether or not to produce gestures that complement their verbal descriptions. With two different analyses, we demonstrate that the decision to gesture influences decisions about what is explicitly mentioned in speech. Since we did not prohibit speakers from gesturing, the differences in the content of speech cannot be attributed to external factors such as distractions due to the prohibition device or the detrimental effect gesture prohibition has been shown to have on speech (e.g., Rauscher et al., 1996).

The results of our analyses converge to support claims that gestures are intended by speakers to supplement and complement their speech. We found that some speakers took full advantage of the face-to-face interaction by dividing the information load of their descriptions between the two modalities; an obvious sign that the gestures were intended to be informative. We also found subtle differences between the linguistic descriptions of gesturers and non-gesturers. First, gesturers produced more directional omissions than non-gesturers. Second, gesturers produced different types of directional omissions than non-gesturers, namely change of direction omissions were only observed with a concomitant visible gesture. Finally, when speakers provided a picture overview in gesture at the beginning of their description, they were more likely to leave out subsequent directional terms from their speech. The difference

between overviews that function as previews and those that function as summaries suggests that the effects reported in this paper are driven by gesture's influence on what is encoded in speech and not by linguistic content's influence on gesture. Taken together, these results show that, for the type of iconic tracing gestures produced in this task, speakers do express part of their message via the manual modality. Thus, this study presents some of the first direct evidence for the speaker's intentional use of gestures for communication.

One possible concern however is that our conclusions are based on a small subset of the elicited gestures. Only some of the gestures in our sample are non-redundant with speech and not all the gestured overviews led to the subsequent omission of directional information. Were the fully redundant gestures also intended to convey information? This is impossible to conclude for all the reasons outlined in the introduction. Some of them may have been communicatively intended and some may not have been. Since there is no logical necessity that intended gestures must lead to a change in speech, our diagnostic is insufficient in many cases. However, in thinking about this issue, it is interesting to consider the difficulty of the addressee's task. It is very difficult to re-construct spatial images from speech alone; this is perhaps why spatial gestures are so common in description tasks and why gestures that convey spatial information have been found to be more effective at communicating information compared to other gestures (Beattie & Shovelton, 1999a, 1999b). Gestures that are fully redundant with the speech can still reinforce the information expressed verbally, making a description maximally comprehensible. Thus, while it is possible that gestures that did not co-occur with directional omissions were also communicatively intended, we cannot be sure this is the case.

An additional concern comes from drawing conclusions about the communicative function of gesture based on gestures produced in silence. In analysis II, we found that overviews presented only in the manual modality influenced the subsequent content of speech. These overviews, however, are in some ways different from gestures accompanying speech; namely, in the former case speech highlights the gesture as the sole source of information. Most of the gesture-only overviews were preceded by deictic phrases that select the gesture as referent, e.g., *The figure looks like this...* In these instances, the gesture is given special status in the discourse. Furthermore, when gesture is the only source of information, it is somehow odd to ask the question of whether or not it is communicative. Rather, the specific question at issue is whether speech-accompanying gestures such as those discussed in Analysis I are intended to communicate. Perhaps the gesture-only overviews fall outside the scope of this question.

However, both manual-only and bimodal overviews influenced the content of the subsequent speech more so than speech alone. This suggests that the manual component of bimodal overviews contributed to the observed increase in omission rates. Thus, although the non-speech-accompanying overviews may hold a special status outside the purview of the communicative debate, the speech-accompanying overviews are clearly relevant to the issue at hand.

The results from analysis II suggest that speakers may view information conveyed in gesture as shared knowledge between speaker and hearer, forming part of their common ground (Clark & Brennan, 1991). If the previewed information were not viewed by the speaker as available to the listener, then subsequent utterances should not be dependent on the prior information. It appears that once something has been expressed with a gesture, the speaker considers it common ground and thus it can influence the manner in which she expresses subsequent information. This interpretation is speculatively based on the interpretation of the present data; further experiments need to be conducted to definitively address the issue of gestures and common ground.

By comparing the behavior of the participants in the current study to participants from similar studies (e.g., Levelt, 1996; Melinger & Kita, in press), one can begin to understand why many prior studies failed to find supportive evidence of the communicativeness of gestures (e.g., Krauss et al., 1991, 1995). Many of the participants in this study took advantage of the face-to-face interaction, omitting from speech some components that were expressed in gesture. These omissions are not observed in the absence of a visible interlocutor, although speakers do produce (speaker-directed) gestures (cf. Melinger & Kita, in press). Prior studies that found that listeners do not integrate gestured information into their understanding of an utterance used stimuli produced by speakers addressing non-present listeners. These speakers therefore had no reason to use gestures informatively (Krauss et al., 1991, 1995). It makes sense, then, that listeners who did not see the gestures were not hindered in their understanding. Following this argument, if the descriptions elicited in this study were presented to new participants without the video, listeners should have lower accuracy when recreating the images from descriptions that included gestures than from descriptions without gestures. This investigation, however, must remain a topic for future research.

The present results also have implications for the cognitive architectures responsible for coordinating speech and gesture production. They demonstrate that the speaker's communicative intention underlies both speech and gesture production. The speaker's intention declares what information is to be expressed. The conceptualizer (Levelt, 1989) then distributes the information

between modalities, sending information to be expressed propositionally to the message generator and information to be expressed manually to the sketch generator (de Ruiter, 2000), motor planner (Krauss et al., 2000) or other gesture generating component. The data support the idea that information that forms part of the intended message can optionally be sent to both generator, producing gestures that are redundant with speech, or to one or the other generators, producing speech without a corresponding gesture or a gesture that adds information not encoded in speech. Crucially, however, the data show that intended information must be expressed by one of the two modalities, as speakers did not omit change of direction information without a compensatory gesture.

To account for the observed relationship between speech and gesture, one of two structural characteristics must be included in the model's architecture. Either speech and gesture have a common origin in some component akin to Levelt's (1989) conceptualizer or the gesture generator must feed back into the conceptualizer. Models such as the one proposed by de Ruiter (2000), which include both of these structural characteristics, can easily account for the observed relationship between speech and gesture content. In contrast, models such as the one proposed by Krauss et al. (2000), which adopt neither of these structural characteristics, cannot account for the observed relationships.

To conclude, we have provided evidence that iconic co-speech gestures, specifically those expressing spatial relations, can be used communicatively. When speakers express information in gesture, that same information can be excluded from the concurrent and subsequent utterances in a way that does not or cannot occur in the absence of gestures. The findings are drawn primarily from iconic tracing gestures. The extent to which these findings generalize to other classes of iconic gestures is an open question. Furthermore, these results say nothing of how addressees interpret and use these gestures. It is possible, although unlikely, that addressees do not incorporate the meaning expressed by these gestures into their understanding of the picture description. The results also do not exclude the possibility that gestures may additionally or alternatively serve a speaker-directed function (Krauss et al., 2000; Kita, 2000; de Ruiter, 2000). Allowing for the multifunctional role of gestures complicates the situation for the gesture researcher. Not only must we identify the full range of gesture types that can be used communicatively but we must also determine a method for discriminating the primary purpose of any given gesture. In addition to presenting evidence for the communicative function of some iconic gestures, the present study provides a method for beginning such a research program, by utilizing the notion of a necessary minimal content.

Notes

* This work benefited greatly from comments from Marianne Gullberg, Mandana Seyfedinipur, Sotaro Kita, and Adam Kendon.

1. Unfortunately, there were too few non-visible gestures produced by our speakers to conduct any reliable analyses.
2. While preventing the interlocutor from speaking may have had the effect of reducing the number of gestures produced by speakers generally, we opted for this approach to ensure that each speaker's interaction with the interlocutor was maximally equated, thereby reducing variance in gesture and speech behavior due to differences in the interlocutor's behavior.
3. Some speakers did not produce directional terms in speech and therefore it was impossible to identify their linguistic perspective.
4. For the analysis, we focused exclusively on speakers who fall within one of these two categories, excluding speakers who gestured inconsistently.
5. For these analyses, it was not possible to transform the data into proportions. Therefore, we conducted non-parametric tests on the total number of omissions produced by each speaker.
6. Since half of the speakers only contributed data to one of the two conditions, analyses were conducted on independent groups to allow all speakers to be considered.
7. Sixteen participants produced at least one picture overview, nine produced at least one in speech only, five in gesture only, and ten in both modalities.

References

- Anderson, Anne H., Alasdair Robertson, Kerry Kilborn, Suzanne Beeke, & Elizabeth Dean (1997). Dialogue despite difficulties: A study of communication between aphasic and unimpaired speakers. In T. Givon (Ed.), *Conversation. Cognitive, communicative and social perspectives* (pp. 1–39). Amsterdam & Philadelphia: John Benjamins.
- Bavelas, Janet, Christine Kenwood, Trudy Johnson, & Bruce Phillips (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture*, 2, 1–17.
- Beattie, Geoffrey & Heather Shovelton (1999a). Do iconic hand gestures really contribute anything to the semantic information conveyed by speech? An experimental investigation. *Semiotica*, 123, 1–30.
- Beattie, Geoffrey & Heather Shovelton (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, Geoffrey & Heather Shovelton (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *Journal of British Psychology*, 93, 179–192.
- Clark, Herbert & Susan Brennan (1991). Grounding in communication. In L. B. Resnick & J. M. Levine (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington DC: American Psychological Association.
- Cassell, Justine, David McNeill, & Karl-Erik McCullough (1999). Speech-gesture mismatch: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition*, 7, 1–33.
- Cohen, Akiba (1977). The communicative functions of hand illustrators. *Journal of Communication*, 27, 54–63.
- Cohen, Akiba & Randall P. Harrison (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology*, 28, 276–279.
- de Ruiter, Jan-Peter (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge: Cambridge University Press.
- Feyereisen, Pierre, Michèle v. d. Wiele, & Fabienne Dubois (1988). The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive/ European Bulletin of Cognitive Psychology*, 8, 3–25.
- Graham, Jean A. & Michael Argyle (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, 10, 57–67.
- Graham, Jean A. & Simon Heywood (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, 5, 189–195.
- Gullberg, Marianne (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*. Oxford: Elsevier.
- Kendon, Adam (1994). Do gestures communicate? A review. *Research on Language and Social Interactions*, 27, 175–200.
- Kita, Sotaro (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.
- Krauss, Robert M., Robert A. Dushay, Yihsiu Chen, & Frances Rauscher (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology*, 31, 533–552.
- Krauss, Robert M., Palmer Morrel-Samuels, & Christine Colasante (1991). Do conversational hand gestures communicate? *Journal of Personality and Social Psychology*, 61, 743–754.
- Krauss, Robert M., Yihsiu Chen, & Rebecca Gottesman (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). Cambridge: Cambridge University Press.
- Levelt, Willem J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, Willem J. M. (1996). Perspective taking and ellipsis in spatial descriptions. In P. Bloom, M. A. Peterson, M. F. Garrett, & L. Nadel (Eds.), *Language and space* (pp. 77–107). Cambridge, MA: MIT Press.

- McNeill, David (1985). So you think gestures are non-verbal? *Psychological review*, 92, 350–371.
- McNeill, David (1992). *Hand and mind. What the hands reveal about thought*. Chicago: Chicago University Press.
- Melinger, Alissa & Sotaro Kita (In press). Does gesture help processes of speech production? Evidence for conceptual level facilitation. *Proceedings of the Berkeley Linguistics Society*, USA, 27.
- Miller, George A. & Phillip Johnson-Laird (1976). *Language and perception*. Cambridge, MA: Harvard University Press.
- Müller, Cornelia (1999). Lectures on gesture. University of Chicago.
- Rauscher, Frances, Robert M. Krauss, & Yihsiu Chen (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science*, 7, 4, 226–231.

Author's address

Alissa Melinger
 FR 4.7 Psycholinguistik
 Geb. 17.1, Room 1.16
 Saarland University
 66041 Saarbrücken
 Germany
 E-mail: melinger@coli.uni-sb.de

Willem J. M. Levelt
 Max Planck Institute for Psycholinguistics
 Postbus 310
 Nijmegen, 6500 AH
 The Netherlands
 E-mail: PIM@mpi.nl

About the authors

Alissa Melinger received her Ph.D. in linguistics and cognitive science from the University of Buffalo, SUNY in 2001. She then joined the Speech Production group and Gesture Project at the Max Planck Institute for Psycholinguistics, Nijmegen. She is currently a research scientist in Psycholinguistics at Saarland University, Saarbrücken, Germany. Her research interests include semantic and syntactic processing in speech production, the interface between gesture and speech production and the functions of gesture for speaking.

Willem J. M. Levelt is founding director of the Max Planck Institute for Psycholinguistics, Nijmegen, Professor of Psycholinguistics at Nijmegen University, and President of the Royal Netherlands Academy of Arts and Sciences. He was an active member of the MPI Gesture Project since its inception and has been involved in several gesture related research projects and doctoral theses. Relevant publications include, *Speaking: From intention to articulation*; Pointing and voicing in deictic expressions with Richardson and La Heij in *Journal of Memory and Language*, *SLOT*; an experimental platform for investigating multi-modal communication, with de Ruiter, Rossignol, Vuurpijl, & Cunningham in *Behavior, Research, Methods Instruments, & Computers*.