Journal of Memory and Language 79-80 (2015) 1–17

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Reduction in gesture during the production of repeated references

Marieke Hoetjes *, Ruud Koolen, Martijn Goudbeek, Emiel Krahmer, Marc Swerts

Tilburg Center for Cognition and Communication (TiCC), Faculty of Humanities, Tilburg University, PO Box 90153, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 6 September 2012
revision received 14 October 2014

Keywords:
Gesture
Repeated references
Reduction
Visibility

ABSTRACT

In dialogue, repeated references contain fewer words (which are also acoustically reduced) and fewer gestures than initial ones. In this paper, we describe three experiments studying to what extent gesture reduction is comparable to other forms of linguistic reduction. Since previous studies showed conflicting findings for gesture rate, we systematically compare two measures of gesture rate: gesture rate per word and per semantic attribute (Experiment I). In addition, we ask whether repetition impacts the form of gestures, by manual annotation of a number of features (Experiment I), by studying gradient differences using a judgment test (Experiment II), and by investigating how effective initial and repeated gestures are at communicating information (Experiment III). The results revealed no reduction in terms of gesture rate per word, but a U-shaped reduction pattern for gesture rate per attribute. Gesture annotation showed no reliable effects of repetition on gesture form, yet participants judged gestures from repeated references as less precise than those from initial ones. Despite this gradient reduction, gestures from initial and repeated references were equally successful in communicating information. Besides effects of repetition, we found systematic effects of visibility on gesture production, with more, longer, larger and more communicative gestures when participants could see each other. We discuss the implications of our findings for gesture research and for models of speech and gesture production.

© 2014 Elsevier Inc. All rights reserved.

Introduction

When we communicate, we continuously refer to objects and persons in our vicinity. Typically, the same target is referred to multiple times during an exchange, and speakers may use both speech and gesture when doing this. For example, a speaker who wants to point out a particular building for her[1] addressee can produce an initial description such as "the brown building at the back of the university campus shaped like this", accompanied by two hand gestures indicating first the location and then depicting the shape of the building. Later in the interaction, when she refers to the same building again, a typical description might be "the building shaped like this", produced in tandem with only the shape gesture.

A substantial body of literature has shown that, as the preceding example suggests, repeated references consist of fewer words (e.g., Brennan & Clark, 1996; Clark &

* Corresponding author at: Room D 404, PO Box 90153, 5000 LE Tilburg, The Netherlands.

E-mail addresses: m.w.hoetjes@tilburguniversity.edu, m.w.hoetjes@uvt.nl (M. Hoetjes), r.m.f.koolen@tilburguniversity.edu (R. Koolen), m.b.goudbeek@tilburguniversity.edu (M. Goudbeek), e.j.krahmer@tilburguniversity.edu (E. Krahmer), m.g.j.swerts@tilburguniversity.edu (M. Swerts).

[1] Throughout this paper, 'she' will be used to refer to the speaker, and 'he' will be used to refer to the addressee.

http://dx.doi.org/10.1016/j.jml.2014.10.004
0749-596X/© 2014 Elsevier Inc. All rights reserved.

Wilkes-Gibbs, 1986). In addition, we know from various studies that repeated references can be reduced acoustically as well, in such a way that, for example, the second realisation of the word "building" in our example may be less intelligible (when heard in isolation) than the initial one (e.g., Aylett & Turk, 2004; Bard et al., 2000; Fowler, 1988). Finally, and most importantly for the current study, a number of studies have shown that repeated references are also accompanied by fewer gestures (e.g., de Ruiter, Bangerter, & Dings, 2012; Holler & Stevens, 2007; Holler, Tutton, & Wilkin, 2011; Levy & McNeill, 1992).

Most of the earlier studies on gesture reduction focused on numeric, quantitative reduction, and while they agree that repeated references contain fewer gestures per description than initial ones, a closer look reveals a mixed pattern of results. To study the relative contribution of gesture and speech to repeated references, researchers generally focus on *gesture rate*. Reconsider our example: the initial description combines 13 words with 2 gestures, and thus has a gesture rate per word of .15 (=2/13). The repeated reference consists of 5 words and 1 gesture, suggesting that in this case the gesture rate has actually increased to .2. Indeed, some studies (e.g., Holler & Wilkin, 2009; Holler et al., 2011) found a general increase in gesture rate per word, while others did not (de Ruiter et al., 2012), or found a reduction in gesture rate (Galati & Brennan, 2014; Jacobs & Garnham, 2007).

An alternative is to look at gesture rate as a function of the semantic attributes in a referring expression. In our initial example, four attributes of the target were included (colour, type, location, shape), and combined with two gestures, yielding a gesture rate per attribute of .5. The repeated example with one gesture mentions two attributes (type, shape), and thus has a gesture rate per attribute of .5 as well. This highlights the importance of how gesture rates can be conceptualised, indicating that different metrics may yield different results.

In view of the mixed results of earlier studies, and given the importance of comparing different metrics for gesture rate, we will systematically compare these two in the current study, asking (1) whether repeated references lead to reduction in gesture per words, (2) whether repeated references lead to reduction in gesture per attribute, and (3) whether we can observe any differences in how these gesture rates develop with repetition.

In addition, we investigate whether the gestures produced in repeated references themselves are different in form from comparable initial gestures. It could be, for instance, that the initial shape gesture in our running example is produced with two hands, depicting the shape precisely and multiple times, while the repeated reference is accompanied by a single one-handed gesture only vaguely suggesting the shape of the target building. Alternatively, it might be that repeated gestures are similar to initial ones in general form, but differ in more gradient ways, much like repeated articulations of the same word ("building") tend to be less clearly articulated. This question has received little attention in the literature, although some studies have looked at some qualitative aspects and generally find evidence for reduction in form (e.g., Galati & Brennan, 2014; Gerwing & Bavelas, 2004; Holler & Stevens, 2007). However, these studies tend to vary with respect to the measures that are used, resulting in an incomplete understanding of how repetition influences how speakers realize their gestures qualitatively. We systematically compare the gestures produced during initial and repeated references, asking (1) whether repeated gestures differ in general form from initial ones and (2) whether there are perceivable gradient differences between initial and repeated gestures. In addition, (3) taking the analogy with repeated realisation of words seriously, we predict that repeated gestures are less "intelligible" when presented without context than initial ones; a prediction which has not been tested before.

By combining quantitative and qualitative analyses, as we do in this paper, we hope to reconcile the conflicting earlier results on gesture rate in repeated references, and further our understanding of the relative contribution of gesture and speech in repeated references, which also has implications for psycholinguistic models of speech and gesture production.

Background

Reduction in speech

Roughly speaking, we can divide previous research on reduction in spoken repeated references into studies that look at reduction at the acoustic level, and studies that look at reduction at the lexical level.

The idea that certain predictable words are reduced acoustically has a long history. Lieberman (1963) compared productions of the word "nine" in a context where it was not predictable ("The word you are about to hear is nine.") with those in a context where it was fully predictable, at least for a native speaker of English ("A stitch in time saves nine", meaning that it is better to do something now than wait until later). Lieberman (1963) found that in the unpredictable context, the word "nine" was longer, had a higher pitch peak (F0) and was rated as more intelligible when taken out of context.

One way in which words can become more predictable is by producing them repeatedly. In particular, realisations of words that represent new information in a discourse tend to be articulated differently (e.g. longer duration, higher pitch) than realisations of the same words occurring later in the discourse, where they express given information (Aylett & Turk, 2004; Bard et al., 2000; Brown, 1983; Fowler & Housum, 1987; Kaland, Krahmer, & Swerts, in press; Lam & Watson, 2010). As in the "nine" example of Lieberman (1963), the references to given information are generally less intelligible when presented in isolation than the references to new information (e.g., Bard et al., 2000; Fowler, 1988; Fowler & Housum, 1987).

Bard et al. (2000), for example, tested whether speakers adjust the reduction in their references to what the listener does or does not know. Bard and colleagues studied this using the Map Task paradigm (Anderson et al., 1991), in which pairs of speakers communicated about a route on a schematic map with labelled landmarks (like a rope bridge or a banana tree). By manipulating the maps, the knowledge of speakers and listeners was manipulated

independently. Words introducing landmarks to two successive listeners were less intelligible when they were repeated, whether they were new for the second listener or not (Experiment 1). In addition, repeated references became less intelligible, also when the listener expressed that he could not see the landmark (Experiment 2). This suggests that speakers reduce repeated references, irrespective of the needs of the listener. Bard et al. (2000) suggest that this pattern of results can be explained by assuming a two-component language production model, consisting of a fast component, which depends on the speaker's knowledge, and a slow, optional component drawing inferences about what the listener knows (but see e.g., Galati & Brennan, 2010; Galati & Brennan, 2014 for a different take on this issue).

Lexical reduction in repeated references has been documented in a seminal paper by Clark and Wilkes-Gibbs (1986), in which pairs of participants engaged in a director–matcher task. In this task, one participant (the director) is instructed to describe an array of humanoid tangram figures, in such a way that another participant (the matcher) can rearrange the figures in front of him such that they match the described ordering. Crucially, this task is repeated six times, so that each tangram figure is discussed multiple times, during different trials. In a typical example, a director might describe a figure in trial 1 as "a person who's ice skating, except they're sticking two arms out in front," while in trial 6 the same figure is referred to simply as "the ice skater" (Clark & Wilkes-Gibbs, 1986, p. 12). This general finding has been replicated many times, and is often explained in terms of an emerging common ground between interlocutors (Clark & Brennan, 1991), where common ground can informally be understood as the information that is shared by interlocutors (or which they assume to share). In this view, common ground makes it possible to reduce repeated references, because speakers can rely on common ground in subsequent references. By repeatedly referring to a target, interlocutors quickly agree on how to refer to an object, and in doing so establish these as common ground. The emergence of a "conceptual pact" (Brennan & Clark, 1996) such as "the ice skater" is a good illustration of this; over time, interlocutors form a shared conceptualization of a particular target, which allows them to refer to it in a more efficient way (using fewer words).

This short overview illustrates that reduction – both acoustically and lexically- in speech has been well established. In recent years, reduction in gesture has been studied as well, and we turn to these studies next.

Reduction in gesture

Speech-accompanying, or co-speech gestures (henceforth called gestures) can be defined as the (usually manual) symbolic movements that people make while they speak (Kendon, 2004; McNeill, 1992). As the phrase co-speech gestures suggests, these movements are closely related to the speech they accompany. Indeed, it has long been suggested that gesture and speech are tightly connected at the semantic level (Kendon, 1972; Kendon, 1980; Kendon, 2000; Kendon, 2004; McNeill, 1985; McNeill, 1992; McNeill & Duncan, 2000), and many studies found quantitative support for this claim (e.g., Kita & Özyürek, 2003; Krahmer & Swerts, 2007; So, Kita, & Goldin-Meadow, 2009). For example, So et al. (2009) found, in a scene description experiment, that speakers could use gesture locations to identify referents in discourse, but that they tended to do this only when the referent was also identified in the accompanying speech. The authors interpret this as an illustration of gesture going "hand-in-hand" (So et al., 2009, p. 123) with speech. Similar ideas have been expressed by, among others, Bavelas, Gerwing, Sutton, and Prevost (2008) and Clark (1996). Clark, for instance, argued that gestures, much like intonation, are an integral part of the communicative signal, suggesting that it would be "difficult to produce the speech without the gesture" (Clark, 1996, p. 179).

Based on considerations such as these, a reduction in speech might be accompanied by a reduction in gesture, and this is indeed what has been claimed. Levy and McNeill (1992), for instance, conducted an analysis of four narratives describing a commercial film and noted that speakers were more likely to gesture in their initial references to people than in later references to the same people in the same scenes. In addition, the authors suggested that new information should not only be accompanied by more gestures, but also by more complex ones than given information.

Various studies have followed up on these initial observations, looking at both quantitative and qualitative analyses of gesture, but the pattern of results is "complex" (Holler et al., 2011, p. 3), with various "conflicting findings" (de Ruiter et al., 2012, p. 235), partly because studies rely on different methods, ranging from collecting narrations to referential communication tasks, and consider a range of differing dependent variables.

Gerwing and Bavelas (2004) was the first test of the idea that gestures referring to given information are "sloppier" (p. 176) than those referring to new information, just like words referring to given information are produced with a sloppier articulation. The authors tested this by having participants play with a number of toys, including a finger cuff (also known as a Chinese finger trap, which 'traps' ones' index fingers at both ends of a small cylinder), and afterwards asked them to explain, without the toys being present, to two other participants what they did with these toys. One of the listeners in this triad had played with the same toys, the other one with different ones, and the speaker was aware of this. Gerwing and Bavelas (2004) concentrated on the gestures that speakers used in their initial identification of the finger cuff, and found that when speakers described it to the participant who had also played with this toy, their gestures were more "elliptical" (Gerwing & Bavelas, 2004, p. 170), compared to the gestures made when describing the toy to a person who had not played with it before (i.e., no common ground), in which case the associated gestures were more elaborate and complex. This was established by having two independent analysts judge which of the two dialogues in each triad contained gestures that conveyed "more information, were more complex, or were more precise" (p. 168) and revealing that the two judges reliably selected the no common ground dialogues as the ones having more

informative gestures. A qualitative analysis of a number of gestures confirmed that gesture parts depicting new information were larger and more precise (Gerwing & Bavelas, 2004, p. 182).

Holler and Stevens (2007) obtained similar results in a referential communication task. They asked participants to locate targets in *Where's Wally?* pictures, and observed that when speakers referred to the size of an object in one of these pictures to an addressee for whom this information was new (unknowing recipients), they generally represented it only in gesture or in gesture and speech. By contrast, when the size information was shared knowledge, speakers mainly realised this information in speech only. In addition, Holler and Stevens (2007) had two independent judges score the perceived size of gestures on a 7-point Likert-scale, and found that size scores for gestures produced to knowing recipients were lower than those for unknowing ones.

Similarly, Jacobs and Garnham (2007) asked speakers to retell a comic strip story multiple times, either to the same listener or a different one. They found that repeated narration to the same listener resulted in a decreased gesture rate, but this did not occur when retelling to different addressees, for whom the story was new. Galati and Brennan (2014), using a similar design, found that speakers who retold a story to an old addressee (i.e., one who had heard the story before) gestured less frequently than when they retold it to a new addressee. In addition, Galati and Brennan (2014) showed that the gestures in retellings to old addressees were smaller and less precise than in those retold to new addressees.

However, other studies have yielded results that are only partly compatible with this. Holler and Wilkin (2009), for example, had speakers narrate stories to an addressee, where some narrative scenes were part of the common ground, because speaker and addressee had watched them together. Using a semantic feature account, the authors found that utterances, taking into account information from speech and gesture, expressed less semantic content when there was common ground between speaker and listener. However, they also reported the "paradoxical result" (Galati & Brennan, 2014, p. 449) that speakers gestured at a higher rate (per 100 words) in the common ground condition, suggesting that gestures are relatively more communicatively important when there is common ground. Holler et al. (2011) similarly found that gesture rate increased with accumulating common ground, when objects were repeatedly referred to.

To further complicate the picture, de Ruiter et al. (2012) found no evidence for an increase in gesture rate in repeated references, but also little or no evidence for a decrease in gesture rate. de Ruiter et al. (2012) explicitly contrasted the aforementioned hand-in-hand hypothesis (So et al., 2009) with an alternative, which they call the trade-off hypothesis (based on observations in, among others, Bangerter, 2004; de Ruiter, 2006; Melinger & Levelt, 2004; Van der Sluis & Krahmer, 2007). This hypothesis suggests that when speaking gets harder, speakers will rely more on gestures (and vice versa, although this second part was not tested by de Ruiter et al., 2012). This leads to the prediction that during the production of repeated references (which, as argued above, are easier to produce than initial ones), speakers will rely less on gestures, which should lead to a decrease in gesture rate. de Ruiter and colleagues studied this using an adaptation of the tangram matching task, inspired by Bangerter (2004), in which directors could identify targets to matchers from a mutually visible array of tangram figures on a wall poster. Since the trade-off between speech and gesture may depend on the type of gesture, the authors coded deictic (pointing) gestures as well as iconic gestures, illustrating a feature of the target (for instance its shape). The authors studied the gesture rate per 100 words, and, in general, found little support for the trade-off hypothesis (with one exception: the gesture rate for pointing gestures decreased when speakers produced repeated references, de Ruiter et al., 2012, p. 244).

To sum up: some studies find evidence that gesture rate decreases when information is shared or repeated (e.g., Galati & Brennan, 2014; Jacobs & Garnham, 2007), some find that it increases (Holler & Wilkin, 2009; Holler et al., 2011), and others find that it stays the same (de Ruiter et al., 2012). However, as illustrated in our opening example, and also noted by others (Galati & Brennan, 2014; Holler & Wilkin, 2009), it is not only the number of words speakers use, but also the semantics of their utterances that are relevant. Galati and Brennan (2014, p. 444) even suggest that gesture rates per words can be misleading and that rates per "unit of semantic content" should be considered as well.

Considering the qualitative aspects of gestures referring to given information: there is indeed some evidence that these are reduced in comparison to gestures referring to initial information, but so far only a limited number of studies have looked into this, all using different measures, ranging from, for instance, an analysis of which dialogue contains more informative and precise gestures (Gerwing & Bavelas, 2004), to coding of size information in gesture as judged on a 7-point scale (Holler & Stevens, 2007), the location of the gesture in gesture space (Holler et al., 2011), and the distance between hands in two-handed gestures or displacement of the hand in one-handed gestures, both on a 7-point scale (Galati & Brennan, 2014).

This paper aims to further our understanding of gesture production when referring to new or given information, by systematically comparing gesture rates per word and per semantic attribute, and by looking in detail at the qualitative aspects of the produced gestures, by manual annotation, but also using judgment studies of gestural precision and intelligibility.

Visibility and gesture

Following many previous studies (see Bavelas & Healing, 2013, for discussion), we include visibility as an additional variable in our design, in such a way that one group of participants will be able to see each other (mutual visibility), while the other group is prevented from doing so using a screen (no visibility).

Traditionally, gesture researchers have used visibility-designs to get a better understanding of the extent to

which speakers produce gestures for their addressees[2]. For example, Alibali, Heath, and Myers (2001, p. 169) write "if speakers produce gestures in order to aid listeners' comprehension, they should produce fewer gestures when their listeners are unable to see those gestures." Indeed, various studies have found that the gesture rate (per word) decreases when participants are not able to see each other, although speakers do still produce gestures when the listener cannot see them (e.g., Alibali et al., 2001; Bavelas et al., 2008). It has also been found that the decrease in gesture rate in part depends on the *kind* of gestures under consideration; the rate with which speakers produce beat gestures, for example, is roughly the same with and without visibility, while deictics and (obligatory) iconics (i.e., iconic gestures needed for understanding) are more frequent with mutual visibility (e.g., Alibali et al., 2001; Bavelas et al., 2008; de Ruiter et al., 2012).

These results raise an obvious question: why do speakers still produce some gestures in the no-visibility condition? This is unexpected when one assumes that speakers produce gestures for the benefit of their addressees. Various explanations have been offered, including the suggestion that these gestures may serve cognitive needs of the speaker (e.g., Alibali et al., 2001; Kita, 2000; Krauss, 1998; Melinger & Kita, 2007). But alternative interpretations have also been defended: speakers may produce gestures that are not visible for the addressee out of habit (e.g., Cohen & Harrison, 1973) or for an imagined audience (e.g., Fridlund, 1994).

Clearly, these are complicated issues, but one consensus that seems to be emerging is that different gestures can have multiple functions (e.g., Alibali et al., 2001), with some gestures being more speaker- and others more addressee-oriented. Perhaps more important for the current study is that, besides gesture rate, visibility may also influence the qualitative form of the gesture (e.g., Bavelas et al., 2008; Gullberg, 2006). Bavelas and colleagues, for example, found that speakers describing an 18th century dress with a distinctive shape used larger gestures in a mutual visibility condition (as if placing the dress around their own body, Bavelas et al., 2008, pp. 509–510) as opposed to speakers describing the same dress via telephone (in which case the gestures were more likely to be on the same scale as the dress on the picture). We include visibility in our design to study whether and when gesture reduction, both in terms of gesture rate and in terms of gesture form, is more speaker- or more addressee-oriented.

The present studies

To further our understanding of gesture production when speakers refer to new or given information, we conduct a series of production and judgment experiments. In Experiment I we collect data from speakers who refer repeatedly to the same target. For this, we rely on a director-matching referential communication task. Referential communication tasks do not require speakers to tell a narrative, and hence references need not be embedded in a larger structure where different factors (such as relative importance to the overall narrative) may conceivably influence the realisation of referring expressions (see de Ruiter et al., 2012; Holler & Stevens, 2007 for similar arguments). We opt for abstract, hard to describe figures with different shapes ("Greebles", Gauthier & Tarr, 1997), which are expected to result in spontaneous descriptions containing both verbal and gestural references to these shapes, both in initial and in repeated descriptions. Both initial and repeated references to the same target are fully transcribed and analysed in terms of the semantic attributes used by speakers. All gestures produced by speakers during these references are analysed as well, allowing us to study both the number of gestures per 100 words and the number of gestures per semantic attribute. Since we are primarily interested in how speakers attenuate their descriptions as a function of repetition, we focus on the individual speaker and not on interactive aspects in our analyses (cf. Bavelas & Healing, 2013).

Besides the quantitative analyses, in which we compare gesture rate per 100 words and per semantic attribute as a function of repetition, we also study how the gestures themselves differ between initial and repeated references. When speakers repeatedly express the same shape in gesture, can we observe qualitative differences between these gestures? Based on the literature, we approach this question from two perspectives, and using two different methods. On the one hand, relevant gestures of initial and repeated descriptions are manually annotated and compared. Based on earlier work (Galati & Brennan, 2014; Holler & Stevens, 2007), we expect gestures produced during repeated references to a target to be smaller. In addition, we ask whether other systematic "discrete" differences can be observed, where we expect gestures produced during repeated references to be shorter in duration, more often produced with one hand, and containing less repetitive movements during the stroke. On the other hand, a conceivable alternative is that the gestures do not change in this discrete manner, but instead differ in a more gradient way, in line with, for instance, Gerwing and Bavelas (2004). This possibility is tested using a judgment test (Experiment II), in which naïve participants are asked to say which of two gestures, one taken from an initial and one from a repeated reference, contains "more information, is more complex, or more precise" (as in Gerwing & Bavelas, 2004, but then applied at the level of gesture rather than dialogue). Finally, if gestures from repeated descriptions are indeed sloppier, analogously to the way in which repeated words are articulated in a sloppier way (as suggested by Gerwing & Bavelas, 2004), we would expect that these are less intelligible/communicative as well. We test this in Experiment III, where participants are shown video clips with either a gesture from an initial or from a repeated reference to a Greeble object, and are asked to indicate which from a pair of Greebles is the one the speaker is gesturing about. Our findings have

[2] It is worth noting, incidentally, that visibility designs have also been used in studies where gesture is not the main focus of attention, such as Clark and Krych (2004) and the aforementioned study by Bard et al. (2000, p. 6), who had participants separated by a "flimsy barrier" in one of their experiments (see also Anderson, Bard, Sotillo, Newlands, & Doherty-Sneddon, 1997).

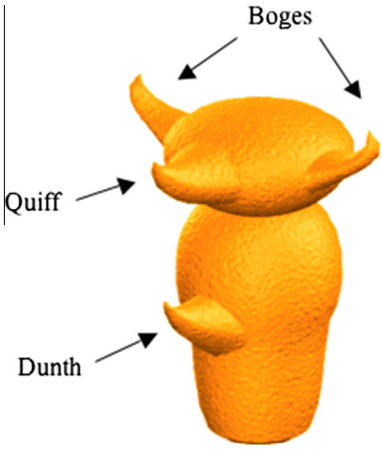

Fig. 1. Example Greeble, in this case with the main body shape "Tasio" and of the gender "Glip" (names in figure refer to specific types of protrusions).

implications for current psycholinguistic models of speech and gesture production, which we describe in the General Conclusion and Discussion section of this paper.

Experiment I: Production of repeated references

Participants

In total, 162 speakers of Dutch took part in the experiment. In the visibility condition there were 106 participants, all undergraduate students (31 male, 75 female, age range 18–29 years old, $M = 21$ years and 7 months), who took part in pairs as partial fulfilment of course credits. From these pairs, data from 5 pairs was left out because there were technical problems, leading to a data set consisting of data from 48 pairs of participants (48 directors and 48 matchers). In the no-visibility condition there were 56 participants, all undergraduate students (21 male, 35 female, age range 17–30 years old, $M = 20$ years and 7 months). From these pairs, data from one pair was left out because the participants had not understood the procedure of the experiment, leading to a data set consisting of data from 27 pairs of participants (27 directors and 27 matchers). In both conditions, participants were randomly assigned the role of director or matcher.

Stimuli

The stimulus materials consisted of pictures of Greebles[3], which are hard to describe, small yellow objects, initially designed so as to share abstract characteristics with human faces. These Greebles vary in terms of their main body shapes ("Samar", "Galli", "Radok", "Tasio"), their gender ("Plok", "Glip"), the different types of protrusions that they have ("Boges", "Quiff", "Dunth") and in terms of the shapes and sizes of these protrusions (see Fig. 1 for an example Greeble, and see Gauthier & Tarr, 1997, for a more detailed description of the Greebles and their properties).

Since directors would naturally be unfamiliar with the specialized vocabulary developed to describe Greebles ("Tasio", "Glip", etc.), they were expected to describe in detail the shapes and protrusions in both their initial and repeated descriptions, for which both speech and gesture would be helpful. In this way, we could collect sequences of shape descriptions, both in word and gesture, for initial and repeated descriptions. In order to make the Greebles look less like animate figures (which might possibly cause participants to rely less on the shape information in their descriptions), they were turned upside down compared to the way in which they were presented in Gauthier and Tarr (1997).

Two picture grids containing 16 Greebles were created. Each picture grid was used for 15 trials, which made a total of 30 trials. The order in which the directors were presented with the two picture grids was counterbalanced over participants. In each trial, there was one target object (marked by a red square), which was surrounded by 15 distractor objects. An example of a picture grid can be seen in Fig. 2.

The crucial manipulation in the task was that several Greebles had to be described repeatedly. In each of the picture grids, two Greebles had to be described twice, and two Greebles had to be described three times; five Greebles were referred to only once. Repeated references to the same object always had a reference to another object in between and were never the first or the last trial of the picture grid. We analysed all descriptions of the Greebles that had to be described three times (i.e. a total of twelve trials per participant; 2 grids $\times$ 2 target Greebles $\times$ 3 descriptions).

Procedure

The experiment was performed in a lab, where the director and the matcher were seated at a table opposite each other (see Fig. 3 for an example of the setup).

The procedure for both visibility conditions was identical, apart from the fact that in the no-visibility condition, there was a large opaque screen between participants, obscuring the view of their entire body (Fig. 3 shows the visibility condition). Both participants were filmed during the experiment, with slightly different camera positions, depending on the visibility condition: in the visibility condition, one camera was positioned behind the matcher (filming the director) and another camera was positioned to the side of the director (filming the entire setup, as in Fig. 3). In the no-visibility condition, both cameras were situated at the side of the screen, one filming the director and one filming the matcher.

The participants were given written instructions and had the opportunity to first ask questions, after which the experiment started. The director was presented with the trials on a computer screen (which was positioned to her side, as in Fig. 3), and was asked to provide a description of the target in such a way that it could be distinguished from the 15 distractor objects. The matcher had a box filled with 16 stacks of cards (one small stack for

[3] Images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University. URL: http://www.tarrlab.org/.

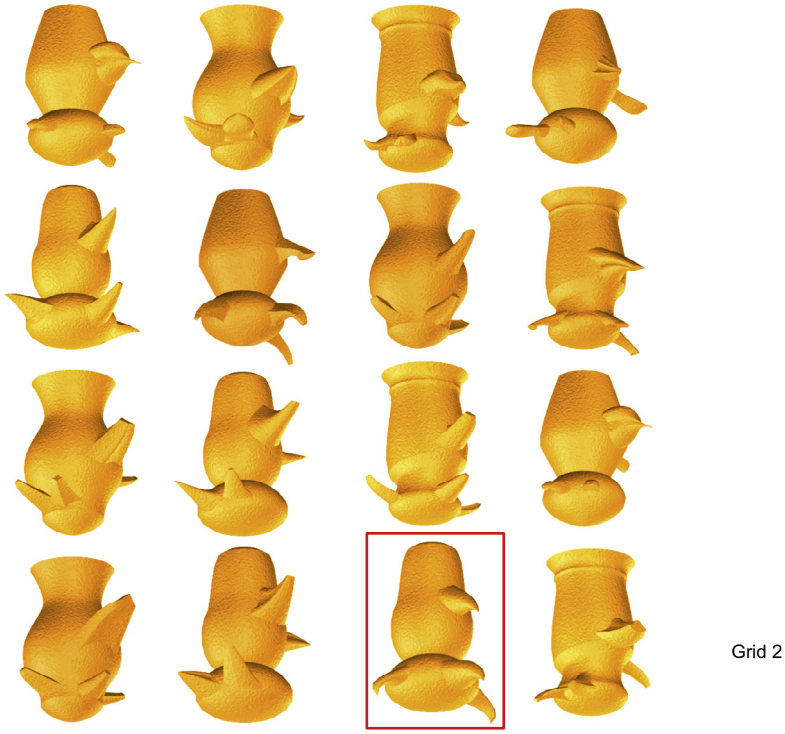

Fig. 2. Example of one of the picture grids presented to the director, in Experiment I. The object with the square surrounding it is the target object of that particular trial.

Fig. 3. Setup of experiment I, in visibility condition, matcher sits on the left and director sits on the right.

each Greeble) in front of him, which were not visible to the director (regardless of the visibility condition the participants were in). The cards in the matcher's box showed the same objects as on the director's screen, but these objects were ordered differently for the director and the matcher. Directors were made aware of this, and it was explained during the instruction phase that visual location on the screen could thus not be used, since the matcher saw the figures in a different order in front of them. The instructions stressed that directors were free to describe their target in any way they wanted, but the use of gesture was not explicitly mentioned. The instructions did mention that it was possible that some targets occurred multiple times.

Based on the director's target description, the matcher had to pick the corresponding card from the box in front of him. Once the matcher had found the card that he thought was being described, the experimenter advanced the director to the next trial. Matchers were instructed not to interrupt the director or ask any questions, but for each new object first wait for the director to finish their description, after which they could indicate that they had found the described object. This instruction was inspired by similar instructions in, among others, Alibali et al. (2001) and Mol, Krahmer, Maes, and Swerts (2009). By instructing our participants in this way, we could collect initial and repeated descriptions in situations that are as comparable as possible, to ensure that any effects could be attributed to our manipulations, and not to possible differences in verbal interaction (see Holler & Wilkin, 2009, p. 273 for a similar argument). After 15 trials, the director was shown the second picture grid containing 16 new objects, and the matcher was presented with a new box filled with stacks of cards of these objects.

Data analysis

Speech annotation

For the speech analysis we analysed the duration and the number of words for each reference (this served as a manipulation check, and to compute the number of gestures per 100 words). The duration was based on the moment at which the matcher indicated that the correct object card had been found. This moment was the end point of one reference, and the beginning of another reference (a new trial was shown to the director as soon

as the matcher had found the correct object). To analyse the number of words, all speech within a reference was transcribed orthographically. Repetitions, hesitations, false starts and corrections were all transcribed and counted as words.[4]

From the transcribed speech data we annotated the number of attributes per reference, so that we could compute the number of gestures per attribute. The number of attributes is a measure of the references' semantic content. When constructing the trials, we made sure that all targets could be distinguished by means of 4 attributes. We designed an annotation scheme containing 45 attributes that speakers could potentially use when describing a Greeble. This scheme was based on the basic characteristics of Greebles (main body shape, gender, protrusions) and was expanded with attributes describing all other properties that they can possibly have (mainly concerning the protrusions' shapes, locations and sizes). An example of a participant's description of a Greeble and its annotated attributes can be seen below. The annotation shows the ID of each attribute, the name of each attribute, followed by the value of this attribute and the part of the reference (in Dutch) that the attribute consists of. A combination of an attribute and a value is referred to as a property of the target.

Example of a participant's description of a Greeble (in Dutch and English literal translation), followed by the accompanying, systematic, attribute annotations.

> "Eh dit is weer die klassieke vaasvorm met die taille, eh er zit aan de rechterkant echt een hele brede eh uitsteeksel"
> *"Uh, this is again that classic vase shape with that waist, uh, there is on the right side really a very wide uh protrusion"*

> <ATTRIBUTE ID = "a1" NAME = "family" VALUE = "galli" > die klassieke vaasvorm met die taille</ATTRIBUTE>
> <ATTRIBUTE ID = "a2" NAME = "DunthLocation" VALUE ="right" > aan de rechterkant</ATTRIBUTE>.
> <ATTRIBUTE ID = "a3" NAME = "DunthWidth" VALUE = "wide" > hele brede</ATTRIBUTE>.
> <ATTRIBUTE ID = "a4" NAME = "Protrusion"VALUE = "dunth" > uitsteeksel </ATTRIBUTE>.

Gesture annotation

For the gesture analysis we used the multimodal annotation programme ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). To analyse the quantity of gestures, all gestures occurring during the critical trials (12 per director) were identified and selected. For the qualitative analyses we annotated a subset of these gestures in detail. To make the analyses for first, second and third references as comparable as possible, we selected for each reference the first gesture that a speaker produced when describing the shape of the target object. For these gestures, only the stroke (i.e. the most effortful and meaningful part of the gesture, see Kendon, 1980; Kendon, 2004; McNeill, 1992) was analysed in detail, without sound. The onset of the stroke was determined by the first video frame in which the most effortful movement started, and the offset of the stroke was determined by the first video frame in which the stroke phase turned into a post-stroke hold phase, or a retraction phase. When a director produced a reference without a gesture, this was treated as a missing value in our analyses on gesture form.

Table 1
Distribution of iconic, deictic and beat gestures, over initial, second and third references, in Experiment I. For each director, only the first gesture that was produced when describing the shape of the target object was included in this analysis.

Repetition	Iconic	Deictic	Beat
1	214	6	3
2	194	6	4
3	178	5	4

For the gestures that were annotated in detail, we determined the type of gesture, differentiating between iconic, deictic and beat gestures (following McNeill, 1992). Iconic gestures were considered as such when a gesture depicted a particular feature of the target object, such as its main shape or the shape of one of the protrusions. Deictic gestures were pointing gestures, generally used to indicate a specific location of one of the object's protrusions. Beat gestures consisted of a simple rhythmic movement without a semantic relation to the speech it accompanied. We found that overwhelmingly, iconic gestures were used, see Table 1. Therefore, the different types of gestures were taken together in all qualitative gesture analyses, as described below.

We took the following aspects of gesture form into account:

- Gesture duration: the duration of the stroke (as defined above), in seconds.
- Gesture size: indicating whether the gesture was produced with a finger (1), the hand (2), the forearm (3) or the entire arm (4). If a gesture involved movement of, say, hands and forearm, we noted down the highest score (3).
- Number of hands: indicating whether the gesture was produced with one or with two hands.
- Number of repeated strokes: a stroke was considered repeated when (near) identical strokes followed each other without a retraction phase in between.

The assumption was that gestures associated with initial references would have a longer duration, a larger size, were more likely to be produced with two hands and to repeat the stroke. To assess the reliability of the coding, a subset of 23 gestures (produced by 23 participants) was coded by a second independent annotator, who was blind to the experimental conditions. There was agreement on 83% of cases for gesture size, on all cases for the number of hands, and on 91% of cases for the number of repeated strokes.

Statistical analyses

The experiment consisted of a 3×2 design, with factors Repetition (levels: initial, second, third) and Visibility

[4] Contractions were counted as one word, however, there was only one type of contraction in the data (namely, the Dutch 'zo'n', '*such a*').

(levels: no screen, screen). The statistical procedure consisted of two repeated measures ANOVAs, one by participants (F_1) and one by items (F_2). On the basis of these, *minF'* was computed (Clark, 1973), to see whether the results could be generalised over participants and items simultaneously, while keeping the experiment wise error rate low (Barr, Levy, Scheepers, & Tily, 2013, p. 268). We used Mauchly's test for sphericity to test for homogeneity of variance. When this test was significant we applied a Greenhouse–Geisser correction on the degrees of freedom, but for the purpose of readability we report the uncorrected degrees of freedom for these cases. Bonferroni corrections were used for post hoc multiple comparisons. We only report when analyses show significant results.

Results

Manipulation check

As expected based on previous literature, reference duration and the number of words used were lower in repeated references and were unaffected by (a lack of) mutual visibility, while the number of gestures decreased in repeated references and when there was no mutual visibility.

Fig. 4 provides an overview of the mean reference duration across all conditions. The reference duration decreased in repeated references, $F_1(2,144) = 53.160$, $p < .001$, $\eta_p^2 = .425$; $F_2(2,9) = 9.992$, $p = .005$, $\eta_p^2 = .689$; $minF'(2,13) = 8.411$, $p = .005$. Post-hoc tests showed that all three references differed significantly from each other (all $p < .05$). Fig. 5 provides an overview of the mean number of words across all conditions. The number of words decreased in repeated references, $F_1(2,144) = 46.497$, $p < .001$, $\eta_p^2 = .392$; $F_2(2,9) = 20.348$, $p < .001$, $\eta_p^2 = .819$; $minF'(2,18) = 14.153$, $p < .001$. Post-hoc tests showed that all three references differed significantly from each other (all $p < .05$). Fig. 6 provides an overview of the mean number of gestures across all conditions. The number of gestures decreased in repeated references, $F_1(2,144) = 13.102$, $p < .001$, $\eta_p^2 = .154$; $F_2(2,9) = 7.089$, $p = .014$, $\eta_p^2 = .612$; $minF'(2,21) = 4.600$, $p = .022$. Post-hoc tests showed that initial references differed from both second and third references (both $p < .05$), whereas second and third references did not differ ($p = .51$). There was also an effect of visibility, with fewer gestures being produced when participants could not see each other, $F_1(1,72) = 10.361$, $p = .002$, $\eta_p^2 = .126$; $F_2(1,9) = 176.878$, $p < .001$, $\eta_p^2 = .952$; $minF'(1,79) = 9.787$, $p = .002$.

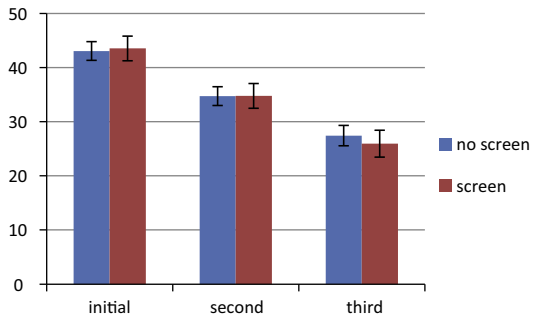

Fig. 4. Mean duration (in seconds) for each reference in Experiment I, in both visibility conditions. Error bars represent standard errors.

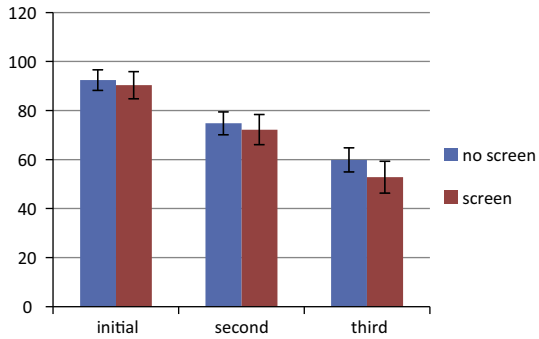

Fig. 5. Mean number of words for each reference in Experiment I, in both visibility conditions. Error bars represent standard errors.

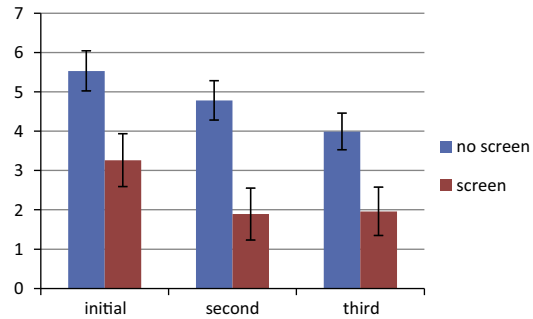

Fig. 6. Mean number of gestures for each reference in Experiment I, in both visibility conditions. Error bars represent standard errors.

Gesture rate

Turning to the two measures of gesture *rate*, we firstly found that there was no significant effect of repetition on the number of gestures per 100 words (see Table 2), indicating that the decrease in the number of words and the number of gestures, as reported in the manipulation check, is proportionally the same, i.e., number of words and number of gestures decrease to the same extent (as in de Ruiter et al., 2012). However, for the number of gestures per attribute, we did find an effect of repetition[5] (see Table 2). The number of gestures per attribute was lower in second references as compared to initial references, and higher in third references as compared to second references, $F_1(2,144) = 21.577$, $p < .001$, $\eta_p^2 = .231$; $F_2(2,9) = 16.346$, $p = .001$, $\eta_p^2 = .784$; $minF'(2,27) = 9.300$, $p < .001$. Post-hoc tests showed that showed that second references differed from both initial and third references (both $p < .05$), whereas initial

[5] We also conducted analyses on the number of attributes per reference, and found that initial references ($M = 11.09$, $SE = .31$) contained fewer attributes than second references ($M = 15.37$, $SE = .63$), which in turn contained more attributes than third references ($M = 8.82$, $SE = .34$), $F_1(2,144) = 93.467$, $p < .001$, $\eta_p^2 = .565$; $F_2(2,9) = 15.084$, $p = .001$, $\eta_p^2 = .770$, $minF'(2,12) = 12.98$, $p = .001$. Post-hoc tests (with Bonferroni correction) showed that all three references differed significantly from each other (all $p < .05$).

Table 2
Mean values, standard errors and confidence intervals of the two types of gesture rate in Experiment I: number of gestures per 100 words, and number of gestures per attribute, in initial, second and third references.

Gesture rate	Repetition	Mean (SE)	95% Confidence interval	
			Lower bound	Upper bound
Gestures/100 words	1	4.928 (.472)	3.986	5.870
Gestures/100 words	2	4.421 (.467)	3.491	5.351
Gestures/100 words	3	6.046 (1.102)	3.849	8.242
Gestures/attribute	1	.430 (.040)	.350	.510
Gestures/attribute	2	.227 (.025)	.178	.276
Gestures/attribute	3	.385 (.047)	.292	.479

Table 3
Mean values, standard errors and confidence intervals of the two types of gesture rate in Experiment I: number of gestures per 100 words, and number of gestures per attribute, in conditions of visibility (no screen) and no-visibility (screen).

Gesture rate	Visibility	Mean (SE)	95% Confidence interval	
			Lower bound	Upper bound
Gestures/100 words	No screen	7.587 (.703)	6.185	8.990
Gestures/100 words	Screen	2.676 (.928)	.825	4.526
Gestures/attribute	No screen	.515 (.041)	.434	.596
Gestures/attribute	Screen	.180 (.053)	.073	.286

and third references did not differ significantly from each other ($p = .67$).

For both measures of gesture rate we found an effect of visibility (see Table 3). When there was no mutual visibility, fewer gestures per 100 words were produced than when there was mutual visibility, $F_1(1,72) = 17.787$, $p < .001$, $\eta_p^2 = .198$; $F_2(1,9) = 36.065$, $p < .001$, $\eta_p^2 = .800$; $minF'(1, 54) = 11.912$, $p = .001$, and likewise fewer gestures per attribute were produced, $F_1(1,72) = 24.974$, $p < .001$, $\eta_p^2 = .258$; $F_2(1,9) = 133.359$, $p < .001$, $\eta_p^2 = .937$; $minF'(1,79) = 21.03$, $p < .001$.

Finally, as is illustrated in Fig. 7, for the number of gestures per attribute there was a significant interaction between repetition and visibility, $F_1(2, 144) = 8.348$, $p = .001$, $\eta_p^2 = .104$; $F_2(2,9) = 6.951$, $p = .015$, $\eta_p^2 = .607$; $minF'(2, 29) = 3.7928$, $p = .034$, which shows that the effect of repetition, with fewer gestures per attribute in second references, followed by more gestures per attribute in third references, is especially prevalent in the visibility condition.

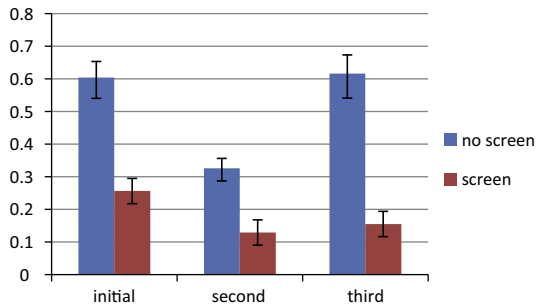

Fig. 7. Mean number of gestures per attribute for each reference in Experiment I, in both visibility conditions. Error bars represent standard errors.

Gesture form

In addition to the gesture rate measures, we analysed several qualitative aspects of the gestures. Table 4 shows the mean values and standard errors for these variables in all three references.

The statistical analyses showed that, although the decrease in gesture duration for repeated references was significant in F_1 and F_2, it was not significant in $minF'$, $F_1(2,166) = 3.781$, $p = .026$, $\eta_p^2 = .061$; $F_2(2,9) = 4.577$, $p = .043$, $\eta_p^2 = .504$; $minF'(2, 41) = 2.070$, $p = .139$. For gesture size, number of hands and number of repeated strokes, there was a comparable numerical effect, with second references obtaining somewhat lower scores than initial ones, and third references lower still, but these differences were not statistically reliable. There was no interaction between repetition and visibility for any of these variables, but there was an effect of visibility on gesture duration and gesture size (see Table 5). Gestures were shorter in duration when there was no mutual visibility, $F_1(1,58) = 6.084$, $p = .017$, $\eta_p^2 = .085$; $F_2(1,9) = 36.161$, $p < .001$, $\eta_p^2 = .801$; $minF'(1, 67) = 5.208$, $p = .026$. Gestures produced without mutual

Table 4
Overview of mean results (M and SE) of Experiment I, for gesture duration (in seconds), gesture size (range 1–4), number of hands (range 1–2, with e.g. 1.70 indicating 70% two-handed gestures) and number of repeated strokes, in initial, second and third references.

	Initial (SE)	Second (SE)	Third (SE)
Gesture duration	1.11 (.07)	0.92 (.06)	0.99 (.07)
Gesture size	3.27 (.06)	3.17 (.06)	3.14 (.07)
Number of hands	1.70 (.04)	1.67 (.05)	1.58 (.06)
Number of repeated strokes	0.16 (.05)	0.18 (.04)	0.15 (.04)

Table 5
Mean values, standard errors and confidence intervals in Experiment I, for gesture duration (in seconds) and gesture size (range 1–4), in conditions of visibility (no screen) and no-visibility (screen).

	Visibility	Mean (SE)	95% Confidence interval	
			Lower bound	Upper bound
Duration	No screen	1.147 (.055)	1.036	1.258
Duration	Screen	.873 (.096)	0.681	1.066
Size	No screen	3.654 (.052)	3.549	3.758
Size	Screen	2.731 (.090)	2.551	2.912

visibility were also smaller than with mutual visibility (see Table 5), $F_1(1,58) = 78.052$, $p < .001$, $\eta_p^2 = .574$; $F_2(1,9) = 154.267$, $p < .001$, $\eta_p^2 = .945$; $minF'(1, 50) = 51.828$, $p < .001$.

Summarising the main findings of Experiment I, we found that for gesture rate there was no effect of repetition on the number of gestures per 100 words, but that there was an effect of repetition on the number of gestures per attribute: these were lower in second references than in initial ones, and then increased in third references back to the level of initial references. Lack of visibility caused both gesture rates to be lowered. For gesture form we found no significant effects of repetition, although we did find effects of visibility on gesture duration and gesture size.

Experiment II: Precision judgment of repeated references

In this judgment test participants judged gesture precision, looking at pairs of gestures taken from initial and repeated (third) references, as produced in Experiment I, to see whether there might be more gradient differences in gesture.

Participants

In total, 39 Dutch undergraduates (14 male, 25 female, age range 18–29 years old, $M = 20$ years 8 months) took part. Twenty participants took part in the visibility condition, and 19 participants in the no-visibility condition, all as partial fulfilment of course credits. The participants had no previous knowledge of and had not taken part in Experiment I.

Stimuli

For the visibility condition, 66 pairs of video clips were selected from the visibility condition of Experiment I. For the no-visibility condition, 31 pairs of video clips were selected from the no-visibility condition of Experiment I. The pairs of video clips contained minimal pairs of gestures with one gesture in each video clip, produced by the same director, illustrating the main shape of the same object. One video clip showed a gesture produced in an initial description of an object, the other video clip showed a gesture produced during a third description of the same object. The order in which the initial and third gestures were presented in the pairs of video clips was counterbalanced over trials. In each trial, a picture of the target object that was described during gesture production was positioned above the video clips, and the participants were told that the gestures were produced when describing this particular picture.

Procedure

The participants were presented with the pairs of video clips. For each pair of video clips, they had to decide in which video clip they thought the gesture was "the most precise". It was explained to participants that a gesture "is more precise, for example when it provides more information about the shape of the object or when it is more complex" (English translation of Dutch instruction). Experiment II was a forced choice test, and although repeated viewing of the video clips was possible, participants were asked to go with their first intuition, and repeated viewing hardly occurred. The judgment test took about 20 min and was administered without sound.

Data analysis

In each trial, one point was given when an initial gesture was chosen to be the most precise and no points (0) were given when a repeated gesture was chosen to be the most precise. We conducted a binomial test to check for significance (i.e. whether the distribution between 0 and 1 was equal, or not). We looked at the overall number of times that an initial gesture was chosen to be the most precise, as well as at both visibility conditions separately.

Results

A binomial test showed that, overall, initial gestures were chosen significantly more often (in 1085, or 57%, of 1909 cases) than repeated gestures, $p < .001$. This was the case for both visibility conditions; in 765, or 58%, of cases in the visibility condition ($p < .001$), and in 320, or 54%, of cases in the no visibility condition ($p = .039$), the initial gesture was chosen to be the most precise[6]. These results show that participants consider gestures from initial references to be the most precise, regardless of whether these gestures were produced in contexts of mutual visibility or not.

Experiment III: Gesture interpretation

Finally, in Experiment III, we ask whether repeated gestures, when presented without context, are less 'intelligible' than initial gestures. Previous studies on speech (e.g., Bard et al., 2000) found that words taken from repeated references, when presented without context, were less intelligible. The question is whether a similar process occurs for gesture. To answer this question a final experiment was set up where participants had to watch a

[6] In a previous version of the precision judgment experiment the participants were not shown a picture of the target object and were not given additional information about what they should consider to be precise; the effect we found was essentially the same.

selection of gestures taken from Experiment II, and choose which Greeble object was the target associated with the gesture they were shown.

The hypotheses are, firstly, that it is more difficult to choose the correct object when the gesture was produced in a repeated reference (and hence participants will make more incorrect choices), compared to when the gesture was produced in an initial reference, and, secondly, that it is more difficult to choose the correct object when the gesture is produced in a context without mutual visibility.

Participants

Participants were 35 Dutch university students (6 male, 29 female, age range 18–30 years old, $M = 21$ years old,) who took part in the experiment as partial fulfilment of course credits. The participants had not taken part in either Experiment I or Experiment II.

Stimuli

The experiment was set up in a 2×2 design, with the within subject factors visibility (levels: no screen, screen) and repetition (levels: initial, third). Eighty gestures were semi-randomly selected from the precision judgment experiment, so that they were evenly distributed over the two factors; gestures from contexts with (40 gestures) and without (40 gestures) mutual visibility between the director and the matcher, half of which in turn were taken from contexts of initial and half from third references. To control for individual variation between the directors' gestures, sets of gestures of the same director producing a gesture about the same object (as in the minimal pairs of video clips in Experiment II) were selected. The video clips were ordered semi-randomly, in such a way that video clips showing the same director gesturing about the same object were never presented one after the other. To control for possible learning effects, two reverse stimulus orders were used.

Procedure

The experiment consisted of 80 slides, with one video clip of one gesture on each slide. For each slide, there was a separate piece of paper with two Greeble objects on it, picture A and picture B. The task for the participants was to choose for each video clip whether the gesture in the video clip was produced in a description of object A or in a description of object B. The participants noted down their answers on an answer sheet. One of the two objects that the participants could choose from was always the object that was being described (i.e. the correct answer), and the alternative object was always a Greeble object with a main body shape different from the correct answer. The order of the correct answers (A or B) was counterbalanced over the trials in the experiment. The experiment was preceded by two practice trials to get the participants used to the short video clips.

Participants were given written instructions and the possibility to ask questions. The slide presentation was opened and participants were allowed to go through the slides and the booklet of Greeble object pictures by themselves. The video clips started playing as soon as a new slide was opened and participants were allowed to watch each video clip only once. For each video clip the participants had to choose A or B from the accompanying page in the booklet of object pictures. Participants were encouraged to go with their first intuition, also in cases where they found the task difficult. The experiment took about 20 min and was administered without sound.

Table 6
Scores for number of correct trials, across conditions, in Experiment III.

	No visibility	Mutual visibility	Total
Initial gesture	376	578	954
Third gesture	364	533	897
Total	740	1111	1851

Data analysis

Each correct answer given by each participant received one point. To test whether participants were better able to pick the correct object, depending on whether a gesture was produced in an initial or repeated gesture which was produced with or without mutual visibility, we conducted chi-square analyses.

Results

In Table 6 the total scores for all four conditions are shown. Results from the chi-square test of goodness-of-fit showed that there was an equal distribution for initial and repeated gestures, $\chi^2(1) = 1.755$, $p = .185$. There was, however, not an equal distribution for mutual visibility, $\chi^2(1) = 74.360$, $p < .001$. Thus participants were better at selecting the correct object based on a gesture taken from a description in which the director and the matcher could see each other than when the gesture was taken from a description in which the director and the matcher could not see each other, but whether the gesture was taken from an initial or a repeated description had no effect. A chi-square test of independence was conducted to examine the relation between repetition and visibility, and we found no significant relation between the two, $\chi^2(1) = .262$, $p = .609$.

General conclusion and discussion

In this paper, we studied how speakers gesture during initial and repeated references to hard to describe objects, i.e., Greebles. To this end, we used an adaptation of the director-matcher, referential communication paradigm (e.g., Clark & Wilkes-Gibbs, 1986; de Ruiter et al., 2012; Holler & Stevens, 2007; Krauss & Weinheimer, 1966), combined with a visibility manipulation such that some participant pairs could see each other (mutual visibility), while others could not. Our findings extend earlier research by providing arguably the largest (in terms of participants) and most comprehensive (in terms of different analyses) study on gesture in repeated references to date.

Earlier research has shown that repeated references in successful communication are different from initial ones, in the sense that they contain fewer words (e.g., Clark & Wilkes-Gibbs, 1986), that these words can be reduced acoustically (e.g., Bard et al., 2000), and that repetition causes speakers to gesture less (e.g., Levy & McNeill, 1992). Our findings in Experiment I were in line with this, showing that our paradigm worked as intended. Our main foci of attention in the present study were the influence of repetition on two different types of gesture rate (with respect to words and semantic content) and on gesture form.

Repetition and gesture rate

In view of earlier, inconsistent findings, we systematically compared reduction in gesture rate per word with reduction in gesture rate per attribute. We found a small numeric increase comparing the first and the last reference for the gesture rate per word (consistent with the pattern observed by Holler et al., 2011). However, in our data this difference was not statistically reliable, similar to the findings of de Ruiter et al. (2012). The similar reduction in repeated references in words and in gestures (causing gesture rate per word to stay the same) thus offers evidence for the "hand-in-hand" hypothesis (So et al., 2009).

When looking at the gesture rate per attribute, a more nuanced picture emerges. Comparing the first and third reference to a target revealed no differences in gesture rate per attribute, which again appears to be in line with the hand-in-hand hypothesis. However, the second reference is associated with a reduced gesture rate, as compared to the preceding and following one. This drop in gesture rate per attribute is caused by an increase of the number of attributes that are included in the second description, which is not mirrored by an increase in the number of gestures (nor the number of words for that matter). We conjecture that this U-shaped pattern is related to the nature of the task. Describing Greebles is hard – speakers have not been confronted with these objects before, and they do not have a vocabulary ready when they start the director–matching task. This might explain the relatively high gesture per attribute rate during the initial descriptions, and could be interpreted as evidence for the trade-off hypothesis (when speaking gets harder, speakers gesture more, de Ruiter et al., 2012). However, during the experiment speakers gradually learn which attributes are useful when describing a particular Greeble, and how to convey these efficiently in words and gesture (cf. the reduction in the numbers of words and gestures, which is fully consistent with earlier studies). During the third and final description, speakers use fewer attributes, presumably because they have learned which set of attributes is most helpful in distinguishing the target Greeble from the others, causing a relative increase of gesture per semantic attributes. Interestingly, this pattern is most clearly observed in the mutual visibility condition, which we discuss in more detail below.

Taken together, these results show that it is important to look at both the gesture rate per word and the gesture rate per attribute, since these can reveal subtly different effects. However, it also raises an important question: when should researchers rely on gesture rate per word and when on gesture rate per semantic attribute?

Gesture rate: per word or per attribute?

If there were a one-to-one correspondence between words and attributes, it should not matter how gesture rates are computed. However, although words and attributes are obviously related, it is easily seen that they do not necessarily stand in a one-to-one relationship. On the one hand, some attributes require more words to be realized in a referring expression than others. In general, it can be assumed, for instance, that premodifiers (i.e., adjectives occurring before the head noun) consist of fewer words than postmodifiers (such as preposition phrases or relative clauses), and whether an attribute is expressed as a pre- or a postmodifier is more or less coincidental and may differ from one language to another (see e.g., Goudbeek & Krahmer, 2012, for discussion). In addition, utterances may include hedges ("I think") and fillers ("uh"), which do not have a direct counterpart in the semantic representation of the description; it is conceivable that such non-attribute related words occur more often in initial than in repeated references, which might complicate reduction patterns. In a somewhat similar vein, it is often assumed that gestures encode meanings in a globally and non-compositional fashion, with one gesture expressing various meanings (e.g., Galati & Brennan, 2014; Hostetter & Alibali, 2008; McNeill, 1992). Hostetter and Alibali (2008, p. 501), for example, discuss the English example "She climbed up the ladder" produced with a single gesture consisting of wiggling fingers moving upwards horizontally, thereby combining various meaning components. It is interesting to observe that the possibilities of gesture to express multiple meanings simultaneously may differ with task and domain; in the Greeble dataset we tend to find that a single gesture expresses a single semantic attribute. For all of these reasons, the relation between meanings on the one hand, and words and gestures on the other, is not straightforward. By only computing gesture rate per word, one risks missing important information (such as the U-shaped pattern in gestures per attribute that we observed).

As we have seen, with some notable exceptions, most gesture researchers only compute gesture rate per word, presumably, at least to some extent, because it is easier and less time consuming. Defining a semantic representation for a task can be complicated, in particular when the task is relatively open ended. An advantage of Greebles, and one of the reasons why we opted for using them in this study, is that their body shapes and protrusions differ in predictable ways, which facilitated the development of a semantic representation. Our data collection is thus "semantically transparant" (in the terminology of Van Deemter, Gatt, van der Sluis, & Power, 2012) in the sense that we know the semantic attribute–values of the target Greebles as well as of all distractors, thus enabling semantic annotation of speech and the subsequent computation of gesture rates per attribute.

In general, if time and resources allow, if a clear semantic representation for the task can be defined, and if in said task the relation between attributes and words is not one-to-one (which might especially be the case in complex domains), researchers are advised to report both gesture rates per word and per attribute. In addition, as we shall discuss below, this distinction also has implications for models of speech and gesture production. Finally, it may be worth noticing that observations such as the above (the nature of the task; how meanings are expressed in words; how much information can be conveyed by a single gesture) may also partly explain why earlier research revealed conflicting results when looking at gesture rates per word, as described in the Introduction.

Repetition and gesture form

Besides gesture rate, we also studied whether the gestures produced during repeated references are different in their realization from comparable gestures produced during initial ones, asking whether there are discrete differences in form and/or whether differences are more gradient in nature, with repeated gestures appearing less "precise" than initial ones.

For this purpose, in Experiment I we compared gestures expressing the same property of a Greeble (its general form or body shape). When looking at gesture form, we found that gestures during initial references numerically lasted somewhat longer than gestures produced during repeated references. However, these findings, while significant in F_1 and F_2, were not significant in the *minF'* analyses, and hence cannot be considered statistically reliable. We did find clear effects of visibility, with gestures that could be seen by the addressee lasting significantly longer and being bigger than ones that were not visible.

We also asked, in two different ways, whether there were gradient differences between initial and repeated gestures. One judgment study (Experiment II) presented participants with minimal pairs of gestures, taken from an initial and a repeated reference, and asked which of the two was more "precise" for a particular Greeble object. The results of this judgment study revealed that initial gestures were indeed perceived as being more precise than repeated ones. These findings are consistent with the observations of Gerwing and Bavelas (2004), although it is important to note that their findings were obtained by two annotators comparing larger stretches of dialogue. Another study (Experiment III) presented participants with a video clip of one gesture (taken from an initial or a repeated reference, produced with or without a screen), and they were asked which of two Greeble objects was the one the speaker was talking about. The results showed that gestures which were produced when the speaker knew that these would not be seen (in the no-visibility condition) were, as expected, less 'intelligible' than gestures taken from contexts of mutual visibility. However, participants did not perform better on this task when viewing gestures from initial descriptions.

In general, when looking at repetition and gesture form a clear and consistent picture emerges. Gestures produced during initial descriptions are judged to be more precise than those produced during repeated descriptions, even though the manual coding does not reveal reliable differences. This suggests that the reduction is gradient, and that the form of the gesture (e.g., whether it is produced with one or two hands) generally does not change between initial and repeated references. Moreover, even though they are reduced in precision, we found that gestures in repeated references are still effective at communicating information; when participants are asked to decide which target object is being referred to based on just one gesture (a hard task!), they can do this roughly equally well when the gesture was produced during an initial or a repeated reference. The resulting picture is conceptually very similar to the way words are articulated when referring to initial or new compared to repeated or given information (e.g., Bard et al., 2000). However, visibility is an important factor in all these analyses.

On the effects of visibility

In general, we found clear effects of visibility. A reduction due to lack of mutual visibility was found for the overall number of gestures, as well as for both measures of gesture rate. Lack of mutual visibility also had an effect on general gesture form, with speakers in that case producing smaller gestures that were also shorter in duration (Experiment I). We also found that gestures produced when there was no mutual visibility were less intelligible (Experiment III). It is interesting to observe that while gesture and speech in our data seem to go hand in hand when considering the effects of repetition (at least when considering gesture rate per word), this does not appear to be the case when considering the effects of visibility. Lack of visibility impacts gesture but not speech; participants produce substantially fewer gestures when separated by a screen, but the same amount of speech with and without visibility.

Earlier gesture studies using a visibility design have led to sometimes conflicting results (see e.g., Alibali et al., 2001; Bavelas & Healing, 2013, for discussion). Interestingly, Alibali et al. (2001, p. 184) when discussing conflicting effects of visibility on gesture rate per word, observe that "[a]mong visibility studies, those that have demonstrated effects of visibility on gesture production (e.g., Cohen, 1977; Cohen & Harrison, 1973; Krauss, Dushay, Chen, & Rauscher, 1995) used tasks with high spatial content (giving directions, describing abstract figures), which may have elicited primarily representational gestures". This suggestion nicely ties in with our findings obtained with the Greeble objects, which are both highly spatial and abstract. Bavelas and Healing (2013) argue that in a number of earlier visibility studies – including Alibali et al. (2001) and Mol et al. (2009) – the visibility manipulation may have confounded with addressee responsiveness. Since we based this part of our design on the aforementioned studies, this criticism may be applied to our study as well (although it is interesting to observe that Alibali et al., 2001, p. 182, discuss and discard this possible alternative explanation of their results). In any case, this issue certainly warrants further study.

Importantly, Bavelas and Healing (2013, p. 79) stress that gesture rate is not the best way to assess visibility effects, and write: "A closer look at how speakers use their gestures reveals that visibility affects many aspects of gestures including the kinds of gestures, their size, location, and relationship to words. All of these differences seem to be done for the addressee's benefit." Our results on gesture form are perfectly in line with this. This suggests that many of the gestures produced by speakers in the mutual visibility condition were indeed designed with the addressee in mind, which has implications for models of speech and gesture production.

Implications for models of speech and gesture production

Over the years, various models of speech and gesture production have been proposed, including Krauss, Chen and Gottesman's (2000) Process model, Kita and Özyürek's (2003) Interface model, de Ruiter's (2000) Sketch model, and McNeill and Duncan's (2000) Growth Point theory (see e.g., Chu & Hagoort, 2014; Hostetter & Alibali, 2008; Wagner, Malisz, & Kopp, 2014, for recent comparisons and discussion). These models all seek to describe how speakers produce multimodal utterances and are concerned with issues such as the timing and integration of gesture and speech, and the role that gestures play in communication. Our present findings are relevant for both of these issues.

Many of the aforementioned models take Levelt's (1989) 'blueprint for the speaker' as their starting point. In this blueprint, speech production is assumed to be a modular process involving three main, consecutive stages. A speaker first has to decide what she wants to say, a decision made in the conceptualizer stage, and resulting in a semantic "preverbal message". Notice, importantly, that this is the stage in which speakers in our Experiment I decide which attributes of the target Greeble to include in their referring expression, based on how helpful they are in distinguishing the target from the other Greebles (cf., Gatt, Krahmer, van Deemter, & van Gompel, 2014; Olson, 1970). In a second stage, known as the formulator and involving lexical retrieval and grammatical encoding, the words of the actual utterance are planned, based on the preverbal message. Finally, in the third stage, the utterance plan is phonologically encoded and articulated, resulting in overt, auditory speech. Models of gesture production typically involve two stages: a Motor Planning stage, sometimes referred to as the Gesture Planner or the Action Generator, during which the motor instructions are produced, and a Motor Execution stage, during which these programs are executed, resulting in overt, visible gestures (Chu & Hagoort, 2014; Wagner et al., 2014).

The main difference between the various extensions of Levelt's (1989) model concerns the exact points of interaction between the speech and gesture production processes. All agree that there is early interaction, with a joint origin for speech and gesture, either in working memory or in the conceptualizer stage. However, some models assume that after this initial interaction, the two processes develop independently—or "ballistically", in terms of Levelt, Richardson, and La Heij (1985). This is true, for instance, for the Sketch model (de Ruiter, 2000), while both the Process model (Krauss et al., 2000) and the Interface model (Kita & Özyürek, 2003) assume that there is further interaction during later stages of the production process. Krauss and colleagues, for example, argue for interaction between the Motor system and the formulator, to account for their observation that the production of gestures may facilitate lexical retrieval. McNeill's Growth Point theory makes the strongest claim concerning interaction, by arguing that speech and gesture are two inseparable parts of a single process (rather than two interacting processes), jointly arising from a single idea (growth point).

Our data do not allow us to draw conclusions about the underlying representations from which gestures arise, but it seems plausible that visual inspection of the Greebles allows speakers to select distinguishing visual features of the target to be expressed (say the "Dunth" protrusion in Fig. 1), comparable to how the Sketch Generator in de Ruiter's (2000) model accounts for this. At this early stage, the "Dunth" attribute becomes part of the pre-verbal message, and since our participants do not have words for Greeble "Dunths", they may express the spatial properties of this shape in gesture, combined with, say, a phrase such as "a protrusion shaped like this".

Interestingly, the two different gesture rates we reported in this study can be seen to operate at two different levels in models of speech and gesture production - the gesture rate per attribute relates to the early interaction of speech and gesture at the pre-verbal level of conceptualisation, while the gesture rate per word is more directly related to later interactions, at the level of the formulator, where words arise. Given that all models assume early interactions between speech and gesture, our gesture per attribute findings do not clearly differentiate between the models. However, the gesture rate per word findings, generally suggesting that speech and gesture go "hand-in-hand", are arguably more difficult to explain for a "ballistic" model, than for an interactive model assuming that the production of speech and gesture also interact at the later stages of speech production, such as McNeill and Duncan's (2000) Growth Point theory. Also our suggestion that with repetition the qualitative reduction in gesture production is comparable to the acoustic reduction in speech production is consistent with this perspective.

A second, partly related issue concerns the question whether gestures communicate information, and whether they were intended as such by the speaker. Models of gesture and speech production are usually not explicit about whether gestures communicate information to an addressee, which is perhaps not surprising given that they are models of the *speaker* (a limitation also discussed by, among others, Mol, Krahmer, Maes, & Swerts, 2012). Still, our findings clearly show that addressees may obtain information from gesture, since in Experiment III we found that participants could determine which of two Greebles was being described, at least based on certain, single gestures. Presumably, this is because these gestures tended to be not redundant with the accompanying speech, but really added information to it, for instance, about the precise form of the target Greeble described by the speaker (see e.g., Singer & Goldin-Meadow, 2005 for comparable

observations in a very different setting, namely child learners). Moreover, the finding that in Experiment III participants performed better when seeing gestures produced in the mutual visibility condition strongly suggests that these gestures were intended by the speaker to be communicative, as we observed above. This is in contrast with the Process model (Krauss et al., 2000), which assumes that gestures are not part of the speaker's communicative intention, but rather have a facilitative function for the speaker herself (since they may help with lexical retrieval).

Acknowledgments

This research came out of various discussions we had in our research group in 2009. The data for Experiment I was collected by the first three authors and analyzed by the first two authors. Later on the first author collected the datasets for Experiments II and III, conducted the analyses on these data sets and took the lead in writing this paper. The last two authors supervised and contributed to all stages of the research. We would like to thank Elsa Jonkers, Kristel Bartels, Joost Driessen, and Manon Yassa for practical help in collecting the data, Nick Wood and Bas Roset for technical support, Judith Holler and Jette Viethen for helpful comments and Martin Pickering, Jan de Ruiter and two anonymous reviewers for comments on a previous version of this paper. Earlier versions of this study were presented at the Optimal communication colloquium at Radboud University Nijmegen (2011), and at the 33rd Annual Conference of the Cognitive Science Society, Boston (2011). We received financial support from The Netherlands Organization for Scientific Research, via a Vici Grant (NWO Grant 27770007), which is gratefully acknowledged.

References

Alibali, M., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44*, 169–188.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., et al. (1991). The HCRC map task corpus. *Language and Speech, 34*, 351–366.

Anderson, A. H., Bard, E. G., Sotillo, C., Newlands, A., & Doherty-Sneddon, G. (1997). Limited visual control of the intelligibility of speech in face-to-face dialogue. *Perception and Psychophysics, 59*, 580–592.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31–56.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science, 15*, 415–419.

Bard, E. G., Anderson, A. H., Sotillo, C., Aylett, M., Doherty-Sneddon, G., & Newlands, A. (2000). Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language, 42*, 1–22.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495–520.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture, 13*, 63–92.

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology, 22*, 1482–1493.

Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67–78). New York: Springer-Verlag.

Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General, 143*, 1726–1741.

Clark, H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

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

Clark, H., & Brennan, S. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & J. S. D. Teasley (Eds.), Perspectives on socially shared cognition (pp. 127–149). American Psychological Association.

Clark, H., & Krych, M. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*, 62–81.

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1–39.

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

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

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.

de Ruiter, J. P. (2006). Can gesticulation help aphasic people speak, or rather, communicate? *Advances in Speech-Language Pathology, 8*, 124–127.

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the trade-off hypothesis. *Topics in Cognitive Science, 4*(2), 232–248.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31*(4), 307–319.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language, 26*(5), 489–504.

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

Galati, A., & Brennan, S. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*, 35–51.

Galati, A., & Brennan, S. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience, 29*(4), 435–451.

Gatt, A., Krahmer, E., van Deemter, K., & van Gompel, R. P. G. (2014). Models and empirical data for the production of referring expressions. *Language, Cognition and Neuroscience, 29*(8), 899–911.

Gauthier, I., & Tarr, M. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*, 1673–1682.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture, 4*, 157–195.

Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering and referential over specification. *Topics in Cognitive Science, 4*, 269–289.

Gullberg, M. (2006). Handling discourse: Gestures, reference tracking, and communication strategies in early L2. *Language Learning, 56*(1), 155–196.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology, 26*(1), 4–27.

Holler, J., Tutton, M., & Wilkin, K. (2011). Co-speech gestures in the process of meaning coordination. In *Paper presented at the 2nd GESPIN – Gesture and Speech in Interaction Conference*, Bielefeld.

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267–289.

Hostetter, A. B., & Alibali, M. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review, 15*, 495–514.

Jacobs, N., & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language, 56*, 291–303.

Kaland, C., Krahmer, E., & Swerts, M. (in press). White bear effects in language production: Evidence from the prosodic realisation of adjectives. Language and Speech 58(1), http://las.sagepub.com/content/early/2013/12/16/0023830913513710.

Kendon, A. (1972). Some relationships between body motion and speech. In A. W. Seigman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–216). New York: Pergamon Press.

Kendon, A. (1980). Gesture and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *Nonverbal communication and language* (pp. 207–227). The Hague: Mouton.
Kendon, A. (2000). Language and gesture: Unity or duality? In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). Cambridge: Cambridge University Press.
Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 162–185). 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.
Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*, 396–414.
Krauss, R. M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science, 7*, 54–60.
Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). New York: Cambridge University Press.
Krauss, R. M., Dushay, R., Chen, Y., & Rauscher, F. (1995). The communicative value of conversational hand gestures. *Journal of Experimental Social Psychology, 31*, 533–552.
Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology, 4*, 343–346.
Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38*, 1137–1146.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24*, 133–164.
Levy, E., & McNeill, D. (1992). Speech, gesture and discourse. *Discourse Processes, 15*, 277–301.
Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6*(3), 172–187.
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., & Duncan, S. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.
Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes, 22*(4), 473–500.
Melinger, A., & Levelt, W. J. M. (2004). Gesture and the communicative intention of the speaker. *Gesture, 4*, 119–141.
Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures. Evidence from a comparative analysis of human–human and human–machine interactions. *Gesture, 9*(1), 97–126.
Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language, 66*, 249–264.
Olson, D. R. (1970). Language and thought: Aspects of a cognitive theory of semantics. *Psychological Review, 77*, 257–273.
Singer, M. A., & Goldin-Meadow, S. (2005). Children learn when their teachers' gestures and speech differ. *Psychological Science, 16*, 85–89.
So, W. C., Kita, S., & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand. *Cognitive Science, 33*, 115–125.
Van Deemter, K., Gatt, A., van der Sluis, I., & Power, R. (2012). Generation of referring expressions: Assessing the incremental algorithm. *Cognitive Science, 36*, 799–836.
Van der Sluis, I., & Krahmer, E. (2007). Generating multimodal referring expressions. *Discourse Processes, 44*, 145–174.
Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209–232.
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In Paper presented at the LREC 2006, Fifth international conference on language resources and evaluation, Genoa, Italy.